Scientific Research Publishing

Computational Chemistry, 2026, 14(4), 73-94
https://www.scirp.org/journal/cc
ISSN Online: 2332-5984
ISSN Print: 2332-5968

Semiempirical Molecular-Orbital Calculations of the Atomization Energies of Organic Molecules

Robert W. Smithwick[1], Sharani Roy[2]

[1]Independent Researcher, Knoxville, Tennessee, USA
[2]Chemistry Department, University of Tennessee, Knoxville, Tennessee, USA
Email: smithwickrw@gmail.com

How to cite this paper: Smithwick, R.W. and Roy, S. (2026) Semiempirical Molecular-Orbital Calculations of the Atomization Energies of Organic Molecules. *Computational Chemistry*, **14**, 73-94.
https://doi.org/10.4236/cc.2026.144005

Received: May 30, 2026
Accepted: August 14, 2026
Published: August 17, 2026

Abstract

A semiempirical molecular-orbital approach has been developed to calculate the atomization energies of 117 small molecules containing hydrogen, carbon, nitrogen, oxygen and fluorine. The molecules were mostly organic. Ten parameter sets were used to account for the hybridizations of the atoms of these five elements with single bonds (5 sets), with double bonds (3 sets) and with triple bonds (2 sets). One-electron energies with Slater orbitals were used with separate values of Z_{2S} and Z_{2P}. Calculations were made with a CNDO/INDO Fortran 77 program modified to a level of approximation that is more similar to an extended Hückel calculation rather than to a CNDO or INDO calculation. A repulsion work function was used in this work to calculate repulsion energies between pairs of atoms within each molecule. The correlation between calculated and experimental 117 atomization energies was favourable with a standard deviation of 0.39 eV. The standard deviation between calculated and experimental dipole moments of nine diatomic molecules was 0.04 debye. This semiempirical method gave favourable results for nicotine and caffeine molecules compared to the results of DFT calculations. The calculations in the present work for all 119 molecules were performed in double precision on a 64-bit personal computer within one second.

Keywords

Semiempirical, Molecular, Orbital, Organic, Molecule, Atomization, Energy

1. Introduction

The primary methods of computational chemistry [1] [2] are *ab initio* calculations, density functional theory (DFT) calculations, semiempirical molecular-

orbital calculations and molecular mechanics. These methods are used to calculate the energies of molecules and some of these methods can be used to determine the lowest-energy molecular geometry. *Ab initio* calculations are considered to be the most accurate but are also the most expensive in terms of computer time.

The present work concerns a semiempirical molecular-orbital calculation developed to calculate organic molecules containing H, C, N, O, and F. These calculations were performed on a 64-bit personal computer with a modification to a Fortran 77 program QCP281[1] [3]. That original program calculated energies, charges, dipole moments and bond distances at two levels of approximation known as Complete Neglect of Differential Overlap (CNDO) and Intermediate Neglect of Differential Overlap (INDO). That program utilized the Hartree-Fock method with Slater orbitals of valence electrons to determine orbital overlaps, electron-electron repulsion integrals and other interactions. The Hamiltonian matrix was solved with the Roothaan equations to calculate a minimum total energy for a set of molecular orbitals comprised of linear combinations of atomic orbitals.

In the present work, the nuclear-nuclear repulsions, the electron-electron repulsions and many other functions calculated by CNDO or INDO have been replaced or removed. The value of Z_{1S} equals one for the 1s electron of hydrogen. Distinct values of Z_{2S} and Z_{2P} are used for C, N, O, and F atoms of 2s and 2p electrons of the same atom. The orbital exponents are $Z_{1S}/1$, $Z_{2S}/2$ and $Z_{2P}/2$. The 1s orbital is used for hydrogen and 2s orbitals are used for second-row atoms, to calculate the repulsion energies between pairs of the atoms as they approach one another from a large internuclear distance to their bond distance.

In this work, a semiempirical MO method has been developed to calculate the atomization energies of 117 small molecules through the use of empirical parameters. Most of the molecules are organic, but also include diatomic molecules and many polyatomic molecules (such as water, nitric acid, hydrazine, ozone) that do not contain carbon. Dipole moments of some of the diatomic molecules were also calculated in this work. Ten sets of parameters were determined for hydrogen (single bond), carbon (single, double, or triple bond), nitrogen (single, double or triple bond), oxygen (single or double bond) and fluorine (single bond). These ten sets of parameters were determined by minimizing the sum-squared error between the calculated and experimental atomization energies for a set of 117 molecules. The sum-squared error of the dipole moments of nine of the diatomic molecules was multiplied by ten and added to the total sum-squared error of the atomization energies so that the atomization energy and the dipole moments of these nine diatomic molecules could both be calculated favorably by the same set of parameters.

The details of this semiempirical MO calculation of atomization energies, dipole moments, charges, and repulsion energies will be described in this paper. Input x, y and z coordinates for the atoms of each molecule in the data set were available from the NIST website [4]. A wide variety of organic groups (such as

[1]Pople, J.A. and Beveridge, D.L. Fortran 77 Computer Code QCPE 281 (Quantum Chemistry Program Exchange, Indiana University, Bloomington, Indiana). QCPE Is No Longer Operational.

alcohols, amines, amides, carbonyls, nitriles, alkenes, rings, double bonds, triple bonds and fluorocarbons) are represented within this data set. The molecules used to calibrate the parameters in this work contain up to six carbon atoms.

The atomization energies of nicotine and caffeine have also been calculated. The total electronic energy of each molecule was then computed by combining the total ground-state energies of the constituent unbound atoms with the calculated atomization energy. The total energies of each of these molecules will be compared to the corresponding energies calculated by density functional theory (DFT) in published work [5].

The parameters of the present method have been calibrated with the atomization data [4] [6] of 117 molecules. With input parameters and with input coordinates for the atoms in each molecule, the entire set of molecules, including nicotine and caffeine, could be computed within one second in double precision on a 64-bit personal computer with GNU Fortran running legacy Fortran 77 code that was modified. The dimensions of the legacy code were increased (from 35 atoms and 80 orbitals) to accommodate up to 200 atoms and up to 800 orbitals, but these dimensions could be increased further. This method or a variation of it may have useful applications.

2. Input to the Hamiltonian Matrix

The level of complexity represented by the modified computer program in this work is more similar to extended Hückel calculations [7] than to CNDO or INDO. The calculation uses valence basis sets: 1s orbitals only for hydrogen and 2s, $2p_X$, $2p_Y$, and $2p_Z$ orbitals only for carbon, nitrogen, oxygen and fluorine. The one-electron energies were calculated directly from values of Z_{1S}, Z_{2S} and Z_{2P} for each given element with zero charge. These values are calculated from the number of 1s, 2s and 2p electrons in the neutral atoms of each element by simple equations given in prior work [8]. The values of Z_{2S} and Z_{2P} are different from one another within the same atom and correspond to different energies and wavefunctions:

$$E(1s) = 13.604\ Z_{1S}^2 \tag{1}$$

$$E(2s) = 13.604\ Z_{2S}^2/4 \tag{2}$$

$$E(2p) = 13.604\ Z_{2P}^2/4 \tag{3}$$

For an atomic configuration, $1s^{N1S}\ 2s^{N2S}\ 2p^{N2P}$, for neutral atoms between Z = 1 and Z = 10, the following effective nuclear charges can be calculated [8]:

$$Z_{1S} = Z - [0.3125 - 0.030/Z - 0.006/Z^2]\ (N_{1S} - 1) - 0.0221\ N_{2S} - 0.0074\ N_{2P} \tag{4}$$

$$Z_{2S} = Z - 0.6989\ N_{1S} - 0.3008\ (N_{2S} - 1) - 0.2299\ N_{2P} \tag{5}$$

$$Z_{2P} = Z - 0.9270\ N_{1S} - 0.3668\ N_{2S} - [0.3526 - 0.1618/Z]\ (N_{2P} - 1) \tag{6}$$

For each of the ten sets of parameters, there is a value of K for each orbital type. Hydrogen has a value of K_{1S} (H_{1S}, single bond) for its 1s orbital. For carbon and nitrogen, each atom has a different value of K_{2P} and another value of K_{2S} for the single, double and triple bonds of 2p and 2s orbitals. Oxygen has different sets of values of K_{2S} and K_{2P} for single and double bonds. Fluorine has a set of these

parameters for single bonds. The values of K in each parameter set are used to calculate the off-diagonal elements of the Hamiltonian matrix as given by the following equation where S(i, j) is the overlap integral:

$$H(i, j) = S(i, j)\ [K(i)\ E(i) + K(j)\ E(j)/2 - (0.6 + 2.6\ G(i, j))]. \quad (7)$$

Orbitals i and j are associated with pairs of orbitals in different atoms. The first part of Equation (7) is similar to the commonly used Wolfsberg-Helmholtz [9] equation. The values of K_{1S}, for the 1s orbital (for H only) and the values of K_{2S} and K_{2P} for heavier atoms were determined from the experimental atomization energies for 117 neutral molecules and from the experimental dipole moments of nine diatomic molecules. An automated search was used to determine the parameters that gave the lowest sum-squared error for the total set of molecules.

The term, (0.6 +2.6 G(i, j)), in Equation (7) is used when the overlap S(i, j) is between the s orbitals [1s (for H only) and/or 2s (for C through F)] of a pair of different atoms. When the signs of the charges are the same, G(i, j) equals a value of $q_A \cdot q_B/R$ for Mulliken charges q_A and q_B of a pair of atoms separated by an interatomic distance R in bohr. When the charges are opposite in sign, G changes sign and equals $q_A \cdot q_B/(0.50\ R)$. The magnitude of H(i, j) is increased when the charges are opposite in sign. If any value of q exceeds one, the truncated value of one is used to calculate G(i, j). If any value of q is more negative than negative one, the truncated value of negative one is used to calculate G(i, j). The units of G(i, j) are eV. The factor of 0.50, which multiplies R in the denominator above, is related to the dipole moment of diatomic molecules and will be explained later. If the interatomic distance, R, between a pair of atoms is greater than 5. bohr (~2.7 Å), G(i, j) is set to zero.

Another type of parameter is V(a) for each of the ten cases. V(a) establishes the relative energy (in eV) between the energy terms on the diagonal of the Hamiltonian matrix for the valence electrons of the atoms of the different elements and hybridizations. For the 1s electron of hydrogen, V(H) was set to a value of zero and the diagonal value is H(1s). The values of V(a) for the nine other cases were allowed to vary relative to the constant value for hydrogen and to one another. For 2p electrons, the diagonal value of the Hamiltonian, H(2p) was set to its V(a) parameter for its corresponding element and bond order. For 2s electrons, the diagonal value of the Hamiltonian, H(2s) was set to E(2s) − E(2p) + V(a) in eV. With this notation, 2s and 2p electrons retain their electron-energy differences within the same atom. The choice of zero for V(H) was arbitrary because a constant value added to or subtracted from all of the values of V(a) does not change the calculation.

The calculation for each molecule was repeated until the sum of all of the positive charges in the molecule was constant within ±0.001 of charge between consecutive calculations. A weighted average, q(ave), is calculated [10] for new iterations to achieve self-consistent charges by the following equation:

$$q(ave) = [q(new) + q(old) \times (2 \times Iteration\# - 2)]/(2 \times Iteration\# - 1). \quad (8)$$

Typically, two to five iterations are required to achieve self-consistency within ±0.001 of the sum of the positive charges in the molecule. Some molecules require more iterations.

The input x, y and z coordinates for the atoms of each molecule in the data set were available from the NIST website [4]. For the simpler molecules, the coordinates in angstroms are given in a table labelled as Cartesians. For larger molecules, many sets of coordinates are available from different calculations within a section labelled Calculated Geometries. If available, the authors chose a geometry optimized by an *ab initio* method known as CCD (coupled cluster doubles) method. However, because the present work is semiempirical, the choice of the optimized geometry by these highly advanced methods was somewhat arbitrary and any of these geometries would be accurate enough for the present work.

3. Repulsion Work Function between Pairs of Atoms within a Molecule

In past work [10], a repulsion work function was calculated numerically as: $U(r) = \Sigma[Q_K \cdot Q_L / r^2]\,\Delta r$. This function represents the energy of the electrostatic repulsion between two neutral atoms. This summation was made numerically from a large distance with incremental reductions in r to give a repulsion energy near the experimental bond distance, R. The theoretical basis for this work function is that R defines the distance between two Gaussian surfaces (spherical shells). One is centered at atom A and the other centered at atom B. For spherical distributions of charges at these centers the net charge acts as if all of the charge is located at its center. When R is reduced, the distributions of charges can extend beyond the opposing nuclei to decrease the diameter of the shells and to increase the opposing net positive charges that repel one another.

In the present work, an approximation to an integral (from ∞ to the interatomic distance, R) is used. This function includes the 1s (for hydrogen) and the 2s Slater orbitals of carbon through fluorine to calculate the sum of repulsions among the pairs of the atoms within a molecule. For a given pair of atoms:

$$U(r) = W_{KL} \int \left[Q_K Q_L / r^2 \right] \partial r \,. \tag{9}$$

Q_K of a given atom equals a constant multiplied by $\int \phi_K^2 r^2 \partial r$ and Q_L equals a constant multiplied by $\int \phi_K^2 r^2 \partial r$. Each value of Q represents an amount of electronic charge that reaches beyond r, the internuclear distance with an adjacent atom. The ϕ_{1S} wave function of hydrogen has the form, $k_{1S}\, e^{-A \cdot r}$. The ϕ_{2S} Slater wave functions of carbon through fluorine have the form, $k_{2S}\, r \cdot e^{-A \cdot r}$. W_{KL} is a constant to be determined for pairs of different atoms. Because each wave function is squared, the value of A for these functions is twice the exponent for the respective exponent of each atom. The exponent for hydrogen is Z_{1S} is 1 and the exponents for the atoms of carbon through fluorine atoms are $Z_{2S}/2$, where Z_{2S} is calculated by Equation (5). The multiplication of the squares of the wave functions of the pairs of atoms represent approximations to definite integrals that are evaluated

from ∞ to a distance R, near the experimental bond distance. Three main forms for these products are obtained for H_2, for HX, and for XY forms products: 1) for a pair of hydrogen atoms the exponential term, $e^{-A \cdot r}$, is multiplied by r^2, 2) for a hydrogen atom and a second-row atom the exponential term, $e^{-A \cdot r}$, is multiplied by r^4, and 3) for a pair of second-row atoms the exponential term, $e^{-A \cdot r}$, is multiplied by r^6. These repulsion functions, which replace the numerical calculations in the past work [10], are all shown with the values of W_{KL} for different pair combinations in the Supplemental Information. If the interatomic distance between a given pair of atoms in the molecule is greater than 5. bohr (~2.7 Å), the value of U is set to zero.

In this work, the effect of the overreach of sigma-bonded 2p orbitals with adjacent atoms has not been directly accounted for. The manual scaling of W_{KL} described above may partially compensate for this deficiency through the calibration process.

Of the possible 100 different values of W_{KL} that could be used between the ten sets of parameters, only 17 different pair combinations of the atoms (among H, C, N, O, and F) have been used. Values of W_{KL} are shared between different elements except for carbon-carbon values of single, double and triple bonds. The 17 different values of K were adjusted by trial and error to position the maximum value of the calculated atomization energy at or near its experimental interatomic separation. The total energy of repulsion was calculated by the sum of the repulsions among all of the pairs of atoms in the molecule. The repulsion energy between atoms was subtracted from the electronic energy calculated from the electron populations of the energies derived from the Hamiltonian matrix. Of the 117 molecules used to calibrate parameters in this study, n-hexane had the largest calculated atomization energy of 76.48 eV after subtraction of a calculated repulsion energy of 18.58 eV. The experimental atomization energy of n-hexane is 76.92 eV.

It is counterintuitive that atomization energies are higher at 298 ˚K than they are at 0 ˚K. For example, the atomization energy of n-hexane is ~1 eV higher at 298 ˚K than at 0 ˚K. The reason for this is that the atomization energy is officially defined as an enthalpy, which is affected by the differences in heat capacities between the molecule and its constituent atoms at temperatures above 0 ˚K.

The experimental atomization energies [4] [6], which were used in this work, were corrected for vibrational energies down to a temperature of 0 ˚K. The zero-point energy (ZPE) is the vibrational energy of the molecule at absolute zero. For diatomic molecules, this energy is ½ hcν for the fundamental stretching vibration in wavenumbers. For polyatomic molecules, both stretching and bending vibrations are present. Although the ZPE is a component of the experimental atomization energy, the present work has not treated the ZPE explicitly.

In past work [10], the advantage of using a repulsion work function between atoms was illustrated by the example of diatomic oxygen. CNDO and INDO calculations use a repulsion energy for diatomic oxygen equal to the core charges of oxygen (which are six each since these exclude the 1s electrons of the oxygen atoms). For the bond distance of 1.208 Å for diatomic oxygen, this repulsion

between O^{+6} and O^{+6} was calculated to be ~430 eV. Electron-electron repulsions should increase this total repulsion further. (The total repulsion energy would also increase markedly for polyatomic molecules.) The experimental value of the atomization energy of diatomic oxygen is 5.12 eV. The INDO calculation did well to overcome more than 400 eV of repulsion energy to obtain a calculated atomization energy of 15.37 eV for diatomic oxygen [3]. In the present work, the atomization energy of diatomic oxygen was calculated to be 5.32 eV after subtraction of a calculated repulsion energy of 1.22 eV between the oxygen atoms. The example of diatomic oxygen will be revisited later.

4. Determination of Parameters

A total-error function was defined by the following equation where AE is the atomization energy (in eV) for 117 neutral molecules and where DM (in debye) is the dipole moment for each of nine diatomic molecules:

$$\text{Error} = \sum[\text{AE(exp)} - \text{AE(calc)}]^2 + 10\sum[\text{DM(exp)} - \text{DM(calc)}]^2. \quad (10)$$

These 117 molecules did not include the molecules, C_2, CO, nicotine and caffeine. These four molecules will be discussed later.

Multiplication by the sum-squared error of the dipole moments by ten was arbitrary but was found to be effective. The minimization of this error function yielded a set of parameters that allowed adequate calculation of both the atomization energies of 117 molecules and the dipole moments of nine diatomic molecules. Each of the ten parameter sets were varied one set at a time. For hydrogen, the 1s parameter for K(1s) was varied with three steps, starting with an increment that is −1, 0, and +1 steps added to starting value. For each of the other nine sets, K(2p), K(2s), and V(a) were varied in steps of −1, 0, and +1 each for a total of 27 combinations. When a given step size no longer gave a lower total error, the step size was reduced in half and the process was repeated for that parameter set until the increment size was below 0.001 for values of V and below 0.0001 for values of K(2p) and K(2s). All ten parameter sets were repeatedly searched one set at a time until the lowest error was established for all of the sets of parameters.

The parameters, K_{1S}, K_{2P}, K_{2S} and V(a), are shown in **Table 1**. The values of K_{2S} are much larger than those of K_{2P} for the atom of a given element and type of hybridization. The net effect of the overlap of filled 2s orbitals between adjacent atoms is that the bonding and antibonding molecular orbitals will both be filled (or mostly filled) and that a net amount of antibonding energy will occur. The significantly larger values of K_{2S} compared to the K_{2P} of the same atom and hybridization may compensate for the total electron-electron repulsion energies between the 1s, 2s and 2p electrons of different pairs of atoms in the molecule. Although the 1s electron orbitals of the second-row atoms are not included in the valence basis set, the higher value of K_{2S} may also compensate in part for repulsion energies between these 1s electrons with the electrons of other atoms.

Table 1. Parameters used in this work. The numbers correspond to atomic numbers of the atoms and the letters A, B and C correspond to the hybridization of carbon, nitrogen and oxygen atoms with single, double and triple bonds, respectively.

	1	6A	6B	6C	7A	7B	7C	8A	8B	9
K_{1S}	0.3126									
K_{2P}		0.0628	0.0631	0.0273	0.0453	0.0193	0.0212	0.0363	0.0560	0.0383
K_{2S}		0.3997	0.3967	0.4549	0.2928	0.3613	0.4593	0.3505	0.4052	0.4122
V(eV)	0	−0.436	−0.311	−0.230	0.304	−0.426	−0.172	0.362	−0.327	0.542

Sometimes there are multiple ways to assign single-, double- and triple-bond hybridization to the atoms within a molecule. One example is that an oxygen atom assigned to a double bond may have a small positive or negative charge; whereas, the assignment of a single-bond hybridization allows the oxygen atom to have a higher negative charge. Another example is the central carbon of the molecule allene, which has two double bonds due to two sets of pi bonds perpendicular to one another along the linear bond axis of its three carbons. In this case, the hybridization of the central carbon was assigned as a triple bond and the hybridization of the outer carbons were assigned as double bonds.

In contrast to other methods, the individual bonds are not assigned as being single, double or triple bonds. It is the atom that is assigned to have a hybridization with only single bonds (A), a hybridization with a double bond (B), or a hybridization with a triple bond (C). The specific assignments of hybridization are given in the input and output files on GITHUB (see Computer Programs section).

5. Alignment of the Maximum Atomization Energy to R_e

Seventeen factors of W of Equation (9) were adjusted manually by trial and error to align the maximum calculated atomization energy close to the experimental bond distance, R_e. Carbon-carbon single, double and triple bonds are represented by ethane, ethylene and acetylene. When one of the carbons of each type of carbon-carbon bond was moved closer or further away from the other carbon the directly bond hydrogen(s) were also moved with that carbon. These three carbon types were given individual, multiplicative carbon-carbon factors to align for the applicable single-, double- or triple-bonded carbon adjacent to other single-, double-, or triple-bonded carbons. The molecule CN was compared to HCN, FCN, and MeCN and only the nitrogen atom was moved closer or further away from the other atom(s) to determine the factor needed to align the maximum atomization energies to R_e. The CO molecule was replaced by formaldehyde, formyl fluoride, carbonyl fluoride and carbon dioxide and only the oxygen was moved closer or further away from the other atoms to determine the factor needed to align the maximum atomization energies near R_e.

The above-described alignment process, which was effective for the organic molecules with carbon-carbon double bonds and carbon-oxygen double bonds,

were not useful to describe the atomization energies for the singlet ground states of C_2 and CO molecules. After the removal of these two diatomic molecules from the data set, the remaining 23 molecules could be adjusted within approximately ±0.02 Å of R_e by the manual trial-and-error adjustment of the W factors. The set of molecules that were aligned in this fashion are shown in Table S1 of the Supplemental Information.

6. Calculated Atomization Energy of Diatomic Oxygen as a Function of Bond Distance

The previously described repulsions between pairs of atoms creates a minimum total energy (maximum atomization energy) at or near the observed bond distance(s). The example of diatomic oxygen is shown in Figure 1. The continuous red line represents calculations that were performed at different distances in the present work. The dashed blue line represents the *ab initio* calculations [11] performed at different interatomic distances. These curves are similar to the experimental potential curve [12] for diatomic oxygen and to the general shape of the well-known Morse function [13] for diatomic molecules.

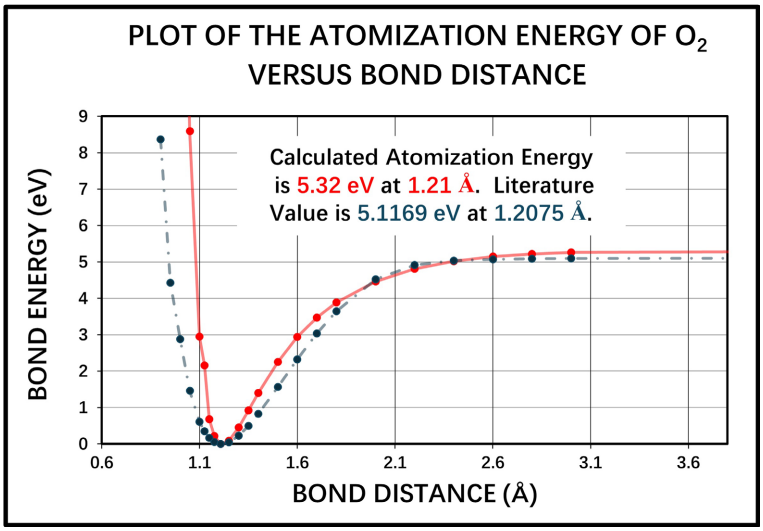


Figure 1. Plots of the energy of diatomic oxygen as a function of distance (eV vs Å). The continuous red curve was calculated in this work. The dashed blue curve was calculated by others [11].

7. Results for the Atomization Energies of the Molecules

The calculated atomization energy (AE) for each molecule is defined by the following equation:

AE = Absolute value of the calculated electron energy of the molecule
– Absolute value of the basis set electrons (H_{ii}) of constituent atoms
– Sum of pairs of repulsion energies in molecule. (11)

Figure 2 shows the plot of the 117 calculated and experimental atomization energies (in eV). The correlation between calculated and experimental energies is

strong with a slope of 1.000 and with an R^2 of 0.999. The standard deviation (sigma) was 0.39 eV and the relative standard deviation was 1.9%. In a standard normal distribution ~95% of the data should fall within ±2 sigma or ±0.78 eV.

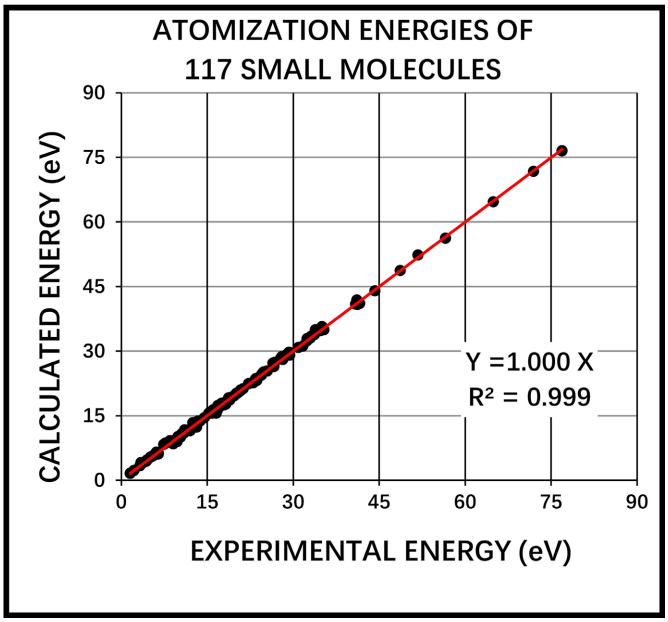


Figure 2. Plot of 117 experimental and calculated atomization energies (in eV).

The experimental energies (corrected to 0 ˚K) of these molecules were obtained primarily from coupled-cluster *ab initio* calculations with the total corrections in **Table 1** of [4], but energies from additional molecules were obtained from the NIST website [6]. The atomization energies were converted from units of kJ/mole to eV/molecule. Distances were converted from angstroms to bohr within the program. The calculated and experimental data is given in file FORT8 on Github.com (see Computer programs).

The values of the atomization energies for ten of these diatomic molecules had been calculated by INDO by past researchers [3]. These calculated values can be compared to the experimental values. The sum-squared error in the present work for these ten values is 2.0 eV^2; whereas, the sum-squared error for the INDO values was 670 eV^2. In the present work, the standard deviation for these ten molecules was 0.48 eV with a relative standard deviation of 9.1%; whereas, the standard deviation of the INDO values was 8.6 eV with a relative standard deviation of 160%.

8. Results for the Dipole Moments of Nine Diatomic Molecules

The plot of the calculated and experimental dipole moments for nine diatomic molecules is shown in **Figure 3**. The correlation is strong with a slope of 0.999 and an R^2 of 0.996. The sum-squared error for these nine values was 0.015 $debye^2$ with a standard deviation of 0.043 debye and with a relative standard deviation of 5.3%. The molecule CO was excluded from the data set.

In this work, the calculation of the dipole moment is based on the methods described in [3] and [14]. By the convention used in this work, the dipole moment is positive for a diatomic molecule if the charge on the first atom of the pair is negative and vice versa. The experimental dipole moments of all of these molecules were obtained from [4], except of that of NH, which was obtained from [15].

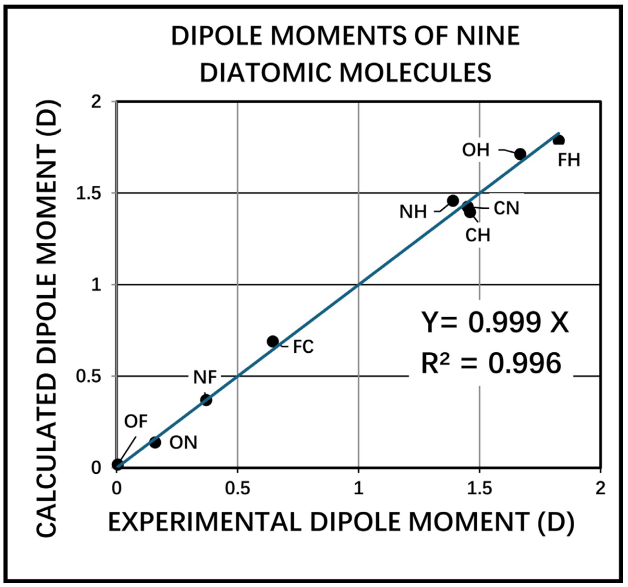


Figure 3. The plot of calculated and experimental values of the dipole moments (in debye) of nine diatomic molecules.

The calculation of dipole moments by the CNDO/INDO program uses the interatomic distance, R_e, to define the separation between the calculated charges. There is also a component added to the dipole moment due to electron polarization by 2p orbitals. In the present work, the charge separation in diatomic molecules was taken to be the distance 0.50 R_e. This factor gives the best agreement between calculated and experimental dipole moments for the nine diatomic molecules. The use of factors for calculating the dipole moments of diatomic molecules was described by [16] and the Supporting Information of [16]. The justification for this factor is that these charges could be closer together due to the mobility of electrons. In the present work, a single factor is proposed for the calculation of the dipole moment of diatomic molecules. Also, the positive charge of an atom must be centered at its nucleus because this charge results from the partial removal of electron charge from the atom. Therefore, for diatomic molecules, the distribution of negative charge is centered precisely in the middle of its bond axis. This is a simple result considering that the partial charges could be spread among different orbitals of the atoms in these diatomic molecules.

For polyatomic molecules, it might be expected that the positive charges are also located at the corresponding nuclei; but that the charge of each negatively charged atom might be spread among multiple bonds, orbitals and atoms. The component of the dipole moment due to electron polarization by 2p orbitals is

also present. The calculated dipole moments of the polyatomic molecules did not match the experimental values well but the calculated values (also multiplied by 0.5) are shown with experimental values in the output files on GITHUB (see Computer Programs section).

Often the charges calculated in this work within polyatomic molecules seem larger than those calculated by other methods. Based on the calculation of the dipole moments of diatomic molecules, calculated positive charges are located at the centers of the atoms deficient in electrons and calculated negative charges, due to excess electrons, are typically located within bonds between atoms where sigma overlap of electrons is calculated.

INDO data is available for six of the nine dipole moments that were plotted in Figure 3. In the present work, the sum-squared error for these six dipole moments (of CH, CN, NH, NO, OH, and FH) was 0.013 debye^2 with a standard deviation of 0.050 debye and with a relative standard deviation of 3.9%. For these same dipole moments calculated by INDO [3], the sum-squared error for these values was 0.85 debye^2 with a standard deviation of 0.41 debye and a relative standard deviation of 32.%.

9. Charges within Diatomic Molecules

The plot of Mulliken charges calculated in this work for seven neutral diatomic molecules are shown in Figure 4 with corresponding charges calculated by others [16]. The charges used for this comparison were calculated from Equation (5) of [16] by the use of fundamental constants, of the dipole moments, of the vibrational frequencies and of the reduced masses of each molecule. The charges in that referenced work had been shown to correlate well to a large set of *ab-initio* calculations. The method used in this work to calculate the dipole moments of diatomic molecules is different from that used in [16]. The correlation between the calculated charges in the present work and the referenced work is favorable with a slope of 0.96 and an R^2 of 0.80.

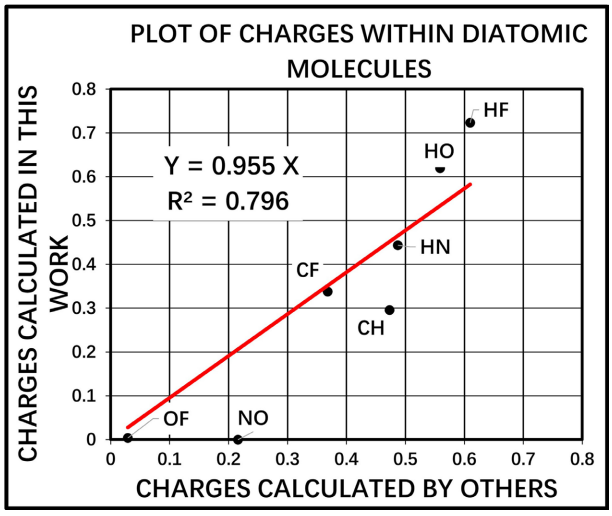


Figure 4. The plot of charges within seven diatomic molecules.

The results for the calculated charges complement the previously described calculation and interpretation of the dipole moments in diatomic molecules. The calculation of the dipole moment involves distances between the centers of the displaced charges and involves 2p-orbital polarization.

10. Calculation of Nicotine Compared to Calculation of Others

The previously described parameters were calibrated with the experimental values of the atomization energies (corrected to 0 ˚K) that were available [4] [6] for small molecules. To test the applicability of the present work to larger molecules, the atomization energy of nicotine ($C_{10}H_{14}N_2$) was calculated. The x, y and z coordinates for the atoms in nicotine were available [17].

In this work, the atomization energy of nicotine was calculated to be 113.33 eV. The total binding energies [18] of the electrons of the ground-state atoms of each carbon (−1030.1085 eV per C), of each hydrogen (−13.598435 eV per H), and of each nitrogen (−1486.058 eV per N) were summed. The total electron energy of the constituent atoms (−13463.58 eV) is lowered by the atomization energy to give −13576.91 eV for the total electron energy of the molecule. This value can be compared to the energy of −13581.43 eV for nicotine calculated by [5] (labelled gas phase) with density functional theory (DFT) with the Becke-Lee-Yang-Parr (B3LYP) functional. The DFT value is 4.5 eV more negative than the value calculated in the present work. This difference is ~0.033% compared to the total electron energy or ~4.0% compared to the calculated atomization energy.

The difference between total electron energies of the constituent atoms (-13463.58 eV) and the DFT value of −13581.43 eV equals 117.85 eV, which is the atomization energy calculated by DFT. This value can be compared directly to the calculated value of 113.33 eV for nicotine in the present work.

The charges calculated for the nitrogen atoms #1 and #2 of nicotine can be compared to those of previous calculations. In the present work, a charge of −0.96 was calculated on nitrogen atom #1 with a methyl group attached. This is consistent with a tertiary aliphatic amine being a strong base. The charge on the aromatic nitrogen atom #2 was calculated to be +0.11 in the present work. The charges on these nitrogen atoms (found with the atomic coordinates of nicotine) [17] were calculated by OpenEye Chem\MFF (molecular force field) to be −0.81 and −0.61, respectively. The charges on these nitrogen atoms were calculated by reference [5] to be near +0.2 for nitrogen atom #1 and to be near zero for nitrogen atom #2.

11. Calculation of Caffeine Compared to the Calculations of Others

In this work, the atomization energy of caffeine ($C_8H_{10}N_4O_2$) was calculated to be 107.98 eV. The x, y and z coordinates for the atoms of caffeine were available from [19]. The six-membered ring of caffeine was coded to be aromatic with carbon-carbon double bonds and carbon-oxygen single bonds. (The chemical structure of

caffeine is usually drawn with carbon-oxygen double bonds.) The total binding energies of the electrons to the ground-state atoms [18] listed in the previous section, plus that of the ground-state energy of each oxygen (−2043.8429 eV per O) were summed. The total energy of the constituent atoms (−18408.77 eV) is lowered by the calculated atomization energy to give −18516.75 eV for the total electron energy of the molecule. This value can be compared to the energy of −18518.8 eV for caffeine calculated [5] (labelled gas phase) with density functional theory (DFT) with the Becke-Lee-Yang-Parr (B3LYP) functional. The DFT value is 2.0 eV more negative than the value calculated in the present work. This difference is ~0.011% compared to the total electron energy or ~1.9% compared to the calculated atomization energy.

The difference between total electron energies of the constituent atoms (-18408.77 eV) and the DFT value of -18518.8 eV equals 110.0 eV, which is the atomization energy calculated by DFT. This value can be compared directly to the calculated value of 107.98 eV for caffeine in the present work.

The charges on oxygen atoms #1 and #2 of caffeine can be compared to those of previous calculations. In the present work, the oxygens were calculated to be −1.11 and −1.09, respectively. High negative charges are consistent with the assignment of C-O single bonds with a charge of −1 on the oxygen atom. The charges on these oxygen atoms, documented with the coordinates of caffeine [19], were calculated by OpenEye Chem\MFF (molecular force field) to be −0.57 each. The charges on these oxygen atoms were calculated by [5] to be ~−0.35 each. The Mulliken charges, Merz-Kollman charges, and NBO charges on the oxygen atoms of caffeine, calculated by these multiple methods [20], were all within the approximate range of −0.5 to −0.8.

12. Machine Learning

Machine learning is beyond the scope of the present work, but it is noteworthy that recent literature [21]-[23] has utilized machine learning with very large data bases to improve the calculated values of atomization energies, dipole moments and atomic charges. The improvements were achieved in part by modifying the Hamiltonians with functions that represent the repulsion energies between pairs of atoms.

13. Conclusions and Future Directions

The semiempirical method of the present work has the following features: The one-electron energies given by Equations 1 through 3 represent the energies of the individual atoms. The Slater orbitals have the exponents given by $Z_{1S}/1$, $Z_{2S}/2$ and $Z_{2P}/2$, where Z_{1S} equals one for hydrogen and where the values of Z_{2S} and Z_{2P} are given by Equation (5) and Equation (6). The electrons in the 2s and 2p orbitals of the same atom have different energies and exponents. At a large value of R, the electron energies correspond to those of the individual atoms and the calculated atomization energy is zero. Ideally, as r approaches R_e for each of the individual

bonds, the atomization energy is calculated at or near the minimum energy of the molecule. If desired, the calculated atomization energy can be combined with experimental ground-state electron energies of the constituent atoms to calculate the total electron energy of the molecule.

The present molecular-orbital method requires input coordinates for each atom of the molecule. Assignment of the hybridization (corresponding to single, double or triple bonds) to the atoms and running the program is mostly straightforward. A preliminary set of parameters has been determined with a data set of 117 small molecules. This data set included a variety of organic compounds with different bond types and functional groups. The calculation of each molecule requires several iterations to establish self-consistent charges. This method could be further developed.

Calculations of nicotine and caffeine each gave total electron energies for these molecules that were similar to the respective energies of advanced and accepted computational methods. This is a promising indicator that the present approach may have application to even larger molecules. Once the atomic coordinates were entered and the hybridization of atoms were assigned, the calculations for 117 small molecules plus the molecules of nicotine and caffeine were computed on a personal computer within one second.

For a molecule with unknown x, y, z coordinates for its atoms, molecular mechanics (or a Gaussian program) might be employed to determine the geometry of the molecule. With that calculated geometry, the molecular-orbital method described in the present work could then be used to calculate the atomization energy. These two computer programs run in sequence would be fast compared to many other computational chemistry programs.

The present work may have future potential in several areas. The authors imagine that variations of this method might be used: 1) to introduce students to computational chemistry, 2) to prescreen large molecules prior to the use of other computational methods, 3) to estimate the atomization energies (and total electron energies) of molecules, and 4) to provide a platform for improving molecular calculations through machine learning.

Computer Programs

Our Fortran 77 source code (see Supplemental Information) can be freely downloaded from the website, https://github.com/smithwickrw/organic. It can be compiled and run with GNU Fortran on a 64-bit personal computer.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Thiel, W. (2014) Semiempirical Quantum-Chemical Methods. *WIREs Computational Molecular Science*, **4**, 145-157. https://doi.org/10.1002/wcms.1161

[2] Lewars, E (2003) Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics. Kluwer Academic Publishers.

[3] Pople, J.A. and Beveridge, D.L. (1970) Approximate Molecular Orbital Theory. McGraw-Hill Book Company.

[4] Johnson III, R.D. (2022) NIST Computational Chemistry Comparison and Benchmark Database. NIST Standard Reference Database 101, Release 22. https://cccbdb.nist.gov/ea1x.asp

[5] Rijal, R., Sah, M., Lamichhane, H.P. and Mallik, H.S. (2022) Quantum Chemical Calculations of Nicotine and Caffeine Molecule in Gas Phase and Solvent Using DFT Methods. *Heliyon*, **8**, e12494. https://doi.org/10.1016/j.heliyon.2022.e12494

[6] Klopper, W., Ruscic, B., Tew, D.P., Bischoff, F.A. and Wolfsegger, S. (2009) Atomization Energies from Coupled-Cluster Calculations Augmented with Explicitly-Correlated Perturbation Theory. *Chemical Physics*, **356**, 14-24. https://doi.org/10.1016/j.chemphys.2008.11.013

[7] Hoffmann, R. (1963) An Extended Hückel Theory. I. Hydrocarbons. *The Journal of Chemical Physics*, **39**, 1397-1412. https://doi.org/10.1063/1.1734456

[8] Smithwick, R.W. (2021) Application of the Screened Hydrogenic Model to Light Atoms. *Computational Chemistry*, **9**, 131-143. https://doi.org/10.4236/cc.2021.93008

[9] Wolfsberg, M. and Helmholz, L. (1952) The Spectra and Electronic Structure of the Tetrahedral Ions MnO_4^-, CrO_4^-, and ClO_4^-. *The Journal of Chemical Physics*, **20**, 837-843. https://doi.org/10.1063/1.1700580

[10] Smithwick, R.W. and Roy, S. (2023) Semiempirical Molecular-Orbital Calculations of Dissociation Energies of Small Molecules Containing Light Elements. *Molecular Physics*, **122**, e2298242. https://doi.org/10.1080/00268976.2023.2298242

[11] Bytautas, L. and Ruedenberg, K. (2010) Accurate *Ab Initio* Potential Energy Curve of O_2. I. Nonrelativistic Full Configuration Interaction Valence Correlation by the Correlation Energy Extrapolation by Intrinsic Scaling Method. *The Journal of Chemical Physics*, **132**, Article ID: 074109. https://doi.org/10.1063/1.3298373

[12] Herzberg, G. (1950) Molecular Spectra and Molecular Structure. Volume I. Spectra of Diatomic Molecules. Van Nostrand Reinhold Co.

[13] Morse, P.M. (1929) Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Physical Review*, **34**, 57-64. https://doi.org/10.1103/physrev.34.57

[14] Pople, J.A. and Segal, G.A. (1965) Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *The Journal of Chemical Physics*, **43**, S136-S151. https://doi.org/10.1063/1.1701476

[15] Liu, X., Meijer, G. and Pérez-Ríos, J. (2020) A Data-Driven Approach to Determine Dipole Moments of Diatomic Molecules. *Physical Chemistry Chemical Physics*, **22**, 24191-24200. https://doi.org/10.1039/d0cp03810e

[16] Hou, S. and Bernath, P.F. (2015) Relationship between Dipole Moments and Harmonic Vibrational Frequencies in Diatomic Molecules. *The Journal of Physical Chemistry A*, **119**, 1435-1438. https://doi.org/10.1021/acs.jpca.5b00993

[17] PubChem (CID #89594). https://pubchem.ncbi.nlm.nih.gov/compound/89594

[18] Kramida, A., Ralchenko, Y., Reader, J. and NIST ASD Team (2024) NIST Atomic Spectra Database (Ver. 5.12). National Institute of Standards and Technology. https://physics.nist.gov/asd

[19] PubChem (CID #2519). https://pubchem.ncbi.nlm.nih.gov/compound/2519

[20] Almeida, A.L.d., Barbosa, L.P.G., Santos, R.L. and Martins, J.B.L. (2016) Chemical

Reactivity Indices of the Caffeine Molecule. *Revista Virtual de Química*, **8**, 483-492. https://doi.org/10.5935/1984-6835.20160035

[21] Zhou, G., Lubbers, N., Barros, K., Tretiak, S. and Nebgen, B. (2022) Deep Learning of Dynamically Responsive Chemical Hamiltonians with Semiempirical Quantum Mechanics. *Proceedings of the National Academy of Sciences*, **119**, e2120333119. https://doi.org/10.1073/pnas.2120333119

[22] Fedik, N., Nebgen, B., Lubbers, N., Barros, K., Kulichenko, M., Li, Y.W., *et al.* (2023) Synergy of Semiempirical Models and Machine Learning in Computational Chemistry. *The Journal of Chemical Physics*, **159**, Article ID: 110901. https://doi.org/10.1063/5.0151833

[23] Hu, F., He, F. and Yaron, D.J. (2023) Treating Semiempirical Hamiltonians as Flexible Machine Learning Models Yields Accurate and Interpretable Results. *Journal of Chemical Theory and Computation*, **19**, 6185-6196. https://doi.org/10.1021/acs.jctc.3c00491

Supplemental Information

In Table S1, the underlined numbers are calculated atomization energies (corresponding to the minimum total energies) at values near their equilibrium distances for the bond distances that were varied. The values under EXP on the right side of the table are the experimental values. Molecules #2 through #4 were calculated correctly for the C-C single, C=C double and C=C triple bond but the C=C double bond of molecule #5, $C_{2,}$ was not at its correct distance nor at its correct energy. Molecules #7 through #10 were calculated correctly for the C-N triple bond. #10 CN has a high error but is included in the calculation. Molecules #21 through #24 were calculated correctly for the C=O double bond (but #25 CO was not at its maximum atomization energy and also its calculated dipole moment did not match its experimental value). For the reasons stated above, C_2 and CO have been removed from the calculations. The parameters chosen to represent the hybridization of the C=C and C=O double bonds for the other organic molecules did not properly describe the bonding in the singlet ground-state energies of C_2 and CO.

Table S1. Atomization energies as a function of bond distances near the experimental bond distance.

		R-0.08	R-0.06	R-0.04	R-0.02	Re	R+0.02	R+0.04	R+0.06	R+0.08	EXP
1	H_2	4.622	4.650	4.671	4.684	4.689	4.688	4.680	4.667	4.648	4.485
2	C_2H_6	28.739	28.846	28.910	28.941	28.944	28.925	28.889	28.837	28.776	28.875
3	C_2H_4	22.241	22.429	22.553	22.622	22.648	22.637	22.596	22.532	22.449	23.074
4	C_2H_2	15.853	16.106	16.274	16.371	16.409	16.397	16.345	16.259	16.145	16.847
5	C_2-OUT-	5.212	5.835	6.306	6.649	6.884	7.031	7.104	7.116	7.079	6.219
6	CH	3.923	3.989	4.032	4.056	4.064	4.057	4.038	4.008	3.970	3.469
7	HCN	11.592	11.952	12.176	12.290	12.314	12.267	12.162	12.013	11.828	13.147
8	FCN	12.305	12.667	12.897	13.019	13.055	13.020	12.930	12.796	12.627	12.965
9	MeCN	24.717	25.058	25.266	25.365	25.376	25.315	25.198	25.038	24.843	25.488
10	CN	8.315	8.540	8.651	8.670	8.614	8.498	8.336	8.138	7.912	7.721
11	CF	4.435	4.999	5.352	5.549	5.628	5.621	5.552	5.438	5.294	5.565
12	NH	3.391	3.482	3.543	3.58	3.595	3.594	3.578	3.550	3.512	3.405
13	NO	5.381	5.759	5.974	6.067	6.067	5.998	5.879	5.725	5.547	6.499
14	NF	2.365	2.853	3.151	3.311	3.373	3.365	3.310	3.222	3.113	3.266
15	OH	4.09	4.237	4.329	4.376	4.385	4.365	4.320	4.256	4.178	4.413
16	OF	1.529	1.890	2.101	2.206	2.239	2.222	2.174	2.104	2.022	2.236
17	O_2	3.897	4.573	4.994	5.225	5.320	5.317	5.245	5.125	4.973	5.120
18	F_2	1.131	1.382	1.518	1.577	1.585	1.561	1.515	1.457	1.392	1.594
19	HF	5.413	5.676	5.835	5.911	5.921	5.879	5.797	5.684	5.548	5.865
20	N_2	8.104	8.581	8.862	8.987	8.991	8.898	8.733	8.512	8.250	9.751
21	COH_2	15.532	15.745	15.842	15.848	15.785	15.670	15.516	15.334	15.132	15.511
22	COHF	16.007	16.359	16.569	16.665	16.672	16.610	16.495	16.340	16.155	16.904
23	COF_2	16.861	17.314	17.596	17.745	17.788	17.752	17.653	17.508	17.328	17.797
24	FO_2	14.216	14.806	15.200	15.441	15.562	15.590	15.546	15.448	15.309	16.573
25	CO-OUT-	5.798	6.803	7.522	8.013	8.322	8.490	8.546	8.515	8.419	11.116

```
C       This is the repulsion work function used to calculate the repulsion (eV)
C        between a pair of atoms:
C FUNCTION UX6
      FUNCTION UX6(NA,STR1,NB,STR2,DD)
      IMPLICIT REAL (A-H,O-Z)
      IMPLICIT INTEGER (I-N)
      CHARACTER*1 STR1,STR2
C     ENERGIES, CALCULATED IN HARTREES (27.2114 eV/HARTREE), ARE CONVERTED TO EVs
C     NA MUST NOT BE LARGER THAN NB WHEN THIS FUNCTION IS CALLED.          EXP(H)=1.
C     THESE EXP VALUES ARE THE EMU VALUES (Zeff/2) FOR THE 2S ORBITALS FOR C, N, O, AND F.
      I=NA*NB
      EXP1=    1.0000
      EXPH=       DD
      EXP6S= 1.9208
      EXP6S5=1.9208*1.9208*1.9208*1.9208*1.9208*DD*DD*DD
      EXP7S= 2.3059
      EXP7S5=2.3059*2.3059*2.3059*2.3059*2.3059*DD*DD*DD
      EXP8S= 2.6909
      EXP8S5=2.6909*2.6909*2.6909*2.6909*2.6909*DD*DD*DD
      EXP9S= 3.0759
      EXP9S5=3.0759*3.0759*3.0759*3.0759*3.0759*DD*DD*DD
      IF(I.EQ.1) PROD=2.*DD*(EXP1 +EXP1)
      IF(I.EQ.1) UX=    (4.1)*EXPH*EXPH
      IF(I.EQ.6) PROD=2.*DD*(EXP1 +EXP6S)
      IF(I.EQ.6) UX= 14.*EXPH*EXP6S5
      IF(I.EQ.7) PROD=2.*DD*(EXP1 +EXP7S)
      IF(I.EQ.7) UX= 12.*EXPH*EXP7S5
      IF(I.EQ.8) PROD=2.*DD*(EXP1 +EXP8S)
      IF(I.EQ.8) UX= 18.*EXPH*EXP8S5
      IF(I.EQ.9) PROD=2.*DD*(EXP1 +EXP9S)
      IF(I.EQ.9) UX= 28.*EXPH*EXP9S5
      IF(I.EQ.36) PROD=2.*DD*(EXP6S +EXP6S)
C     CARBON-CARBON:
      IF(I.EQ.36) B= EXP6S5*EXP6S5
      IF(I.EQ.36) CA= 15.6
      IF(I.EQ.36) CB=    7.5
      IF(I.EQ.36) CC=    4.7
      IF(I.EQ.36) CAA= CA*CA
      IF(I.EQ.36) CAB= CA*CB
      IF(I.EQ.36) CAC= CA*CC
```

```
      IF(I.EQ.36) CBB= CB*CB
      IF(I.EQ.36) CBC= CB*CC
      IF(I.EQ.36) CCC= CC*CC
      IF(I.EQ.36.AND.STR1.EQ.'B'.AND.STR2.EQ.'B') UX=B*CBB
      IF(I.EQ.36.AND.STR1.EQ.'C'.AND.STR2.EQ.'C') UX=B*CCC
      IF(I.EQ.36.AND.STR1.EQ.'A'.AND.STR2.EQ.'A') UX=B*CAA
      IF(I.EQ.36.AND.STR1.EQ.'B'.AND.STR2.EQ.'C') UX=B*CBC
      IF(I.EQ.36.AND.STR1.EQ.'C'.AND.STR2.EQ.'B') UX=B*CBC
      IF(I.EQ.36.AND.STR1.EQ.'B'.AND.STR2.EQ.'A') UX=B*CAB
      IF(I.EQ.36.AND.STR1.EQ.'A'.AND.STR2.EQ.'B') UX=B*CAB
      IF(I.EQ.36.AND.STR1.EQ.'C'.AND.STR2.EQ.'A') UX=B*CAC
      IF(I.EQ.36.AND.STR1.EQ.'A'.AND.STR2.EQ.'C') UX=B*CAC
C     CARBON-NITROGEN:
      IF(I.EQ.42) PROD=2.*DD*(EXP6S +EXP7S)
      IF(I.EQ.42) UX= 36.*EXP6S5*EXP7S5
C     CARBON-OXYGEN
      IF(I.EQ.48) PROD=2.*DD*(EXP6S +EXP8S)
      IF(I.EQ.48) UX=    98.*EXP6S5*EXP8S5
C     CARBON-FLUORINE
      IF(I.EQ.54) PROD=2.*DD*(EXP6S +EXP9S)
      IF(I.EQ.54) UX= 900.*EXP6S5*EXP9S5
C      NITROGEN-NITROGEN
      IF(I.EQ.49) PROD=2.*DD*(EXP7S +EXP7S)
      IF(I.EQ.49) UX= 42.5*EXP7S5*EXP7S5
C     NITROGEN-OXYGEN
      IF(I.EQ.56) PROD=2.*DD*(EXP7S +EXP8S)
      IF(I.EQ.56) UX= 115.*EXP7S5*EXP8S5
C      NITROGEN-FLUORINE
      IF(I.EQ.63) PROD=2.*DD*(EXP7S +EXP9S)
      IF(I.EQ.63) UX= 3100.*EXP7S5*EXP9S5
C     OXYGEN-OXYGEN
      IF(I.EQ.64) PROD=2.*DD*(EXP8S +EXP8S)
      IF(I.EQ.64) UX= 740.*EXP8S5*EXP8S5
C      OXYGEN-FLUORINE
      IF(I.EQ.72) PROD=2.*DD*(EXP8S +EXP9S)
      IF(I.EQ.72) UX=11000.*EXP8S5*EXP9S5
C      FLUORINE-FLORINE
      IF(I.EQ.81) PROD=2.*DD*(EXP9S +EXP9S)
      IF(I.EQ.81) UX=69000.*EXP9S5*EXP9S5
C       DD IS IN ATOMIC UNITS
C       UX6 IS IN UNITS OF eV WHEN MULTIPLIED BY 27.2114 eV/har-
tree:
      UX6= 27.2114*UX*EXP(-PROD)
```

```
      RETURN
      END
```

The input file of acetonitrile is informative. There are six atoms with no net molecular charge and the multiplicity is 1. The coordinates of the nitrogen (atomic# 7) is on the next line. The hybridization is "C" corresponding to a triple bond. The next atom is carbon (atomic# 6) hybridized for a single bond "A". The next carbon is the nitrile carbon hybridized as "C" for a triple bond. The next three atoms are hydrogen atoms (atomic# 1). They are associated with single bonds (no letter is needed). The next line contains the experimental atomization energy in eV and the experimental dipole moment in Debye. These numbers are used to adjust or compare calculated values to experimental values. FINI in capital letters is used to stop reading input into the program.

```
#47     ACETONITRILE
CNDO.CLSD
    6    0    1
    7C      0.00000          0.00000         -2.61500
    6A      0.00000          0.00000          0.00000
    6C      0.00000          0.00000         -1.45800
    1       0.00000         -1.04100          0.36750
    1       0.90160          0.52050          0.36750
    1       0.90160         -0.52050          0.36750
25.4878     3.919
FINI
```

For molecules with double bonds, atoms hybridized for double bonds are denoted with the letter "B". For molecules containing fluorine, the fluorine atoms are associated with single bonds and no letter is needed.

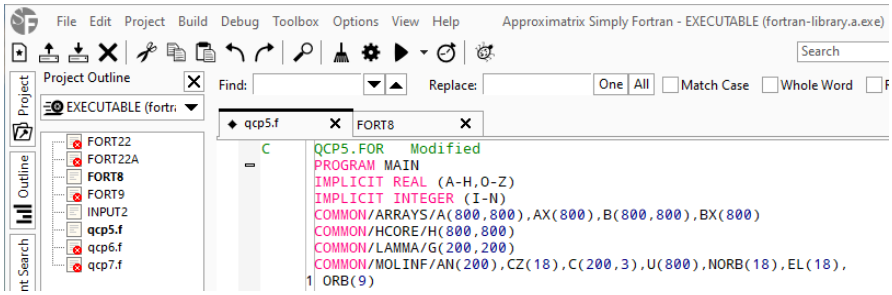
To calculate the 117 molecules (plus nicotine and caffeine and CO and C2), one needs only three files: 1) qcp5.f, which has built-in parameters, 2) input2, which has the molecular inputs followed by 'FINI' in capital letters to end the program, and 3) fort8, which is the pre-formed output file that is overwritten. After the program is compiled and program starts it takes only ~1 second total time to calculate all the molecules. Hydrogen has one orbital per atom; C, N, O, and F have four orbitals per atom.

To generate Table S1 above, we enabled QCP6.F with FORT22, INPUT2 and FORT9 (and disabled the other files). The parameters are read from FORT22. FORT9 is the pre-formed output file.

To modify the program and to subsequently determine alternative parameters, QCP7.F, FORT22A, FORT22, and FORT8 are enabled. This is a slower program and one has to set how many times (ILM = 1, 2, for example) one wants to cycle through the parameters for each of the data sets. Some sets were omitted in the past; therefore, sets ILL = 1, 12 are cycled with sets (#5 and #9) skipped to end up with 10 sets. "ILK = 1, 12" initializes several parameters prior to the calculation of each molecule. After the cycles are run, Fort22 may be copied into Fort22A to

upgrade the initial parameters. (Fort22 and Fort22A are different files because an error is generated if the same file is used for both reading and writing.)

The GNU Fortran (GFortran) is an open-source Fortran compiler. The program that we used is an inexpensive commercial version of GNU Fortran Compiler called Simply FORTRAN. We found this platform to be convenient for editing and running our legacy code. The files should all be located in the same computer folder. Under project options, we used fortran-library.a.exe and Executable mode and 64 bit (x64). We set Fortran options, to double precision and legacy code. We plan to put our programs and input/output files on GITHUB for free public access.



Used with permission from Approximatrix.