

Chapter 8: Semi-Empirical Methods

Key Notes:

Fundamental Aspects:

Semi-empirical methods represent a “middle road” between the mostly qualitative results available from molecular mechanics and the computationally time-consuming quantitative results available from *ab initio* methods. Semi-empirical methods are a good choice for many users, especially those new to molecular modeling who are less interested in research-quality numerical results, and are more interested in developing their ability to use computing to understand structure, properties, and activities.

Methods:

Semi-empirical methods address the issue of limitations on calculations of large molecules and the length of computing time needed with *ab initio* methods. It does so by making several large assumptions, including ignoring core (non-valence) electrons and making major simplifications of the mathematics. Semi-empirical methods use many of the same mathematics as are found in the Hartree-Fock method, but look to reduce the computing time by replacing some of the mathematics with data (known as *parameters*) derived from experimental and computed data. The various types of semi-empirical methods all use different numbers and types of parameters, which affects the quality of the calculation. The term *semi-empirical* comes from the fact that *some* of the calculations come from *empirical* data. Modern semi-empirical methods include MNDO, AM1, and PM3.

Applications:

Semi-empirical methods are particularly useful in the study of organic chemistry and the structure and reactions of organic molecules. Semi-empirical methods were developed specifically for this area of chemistry, and organic continues to be the primary target for this method. Semi-empirical methods also provide researchers and students with a relatively quick way of studying the structure and behavior of molecules, especially as compared with *ab initio* methods.

Software Tools:

Semi-empirical methods are embedded in most of the modern computational chemistry software packages, including GAMESS and Gaussian, available on the North Carolina High School Computational Chemistry server. MOPAC, also found on the server, is only a semi-empirical tool, and does not perform *ab initio* calculations. MOPAC is the tool of choice for the user solely interested in semi-empirical calculations.

Advantages:

Semi-empirical methods allow the user to obtain qualitative and quantitative results on larger molecules than are possible using *ab initio* methods. It is also an excellent method for use in the study of organic molecules and related reactions. Semi-empirical methods give relatively good results in calculating and visualizing molecular orbitals.

Disadvantages:

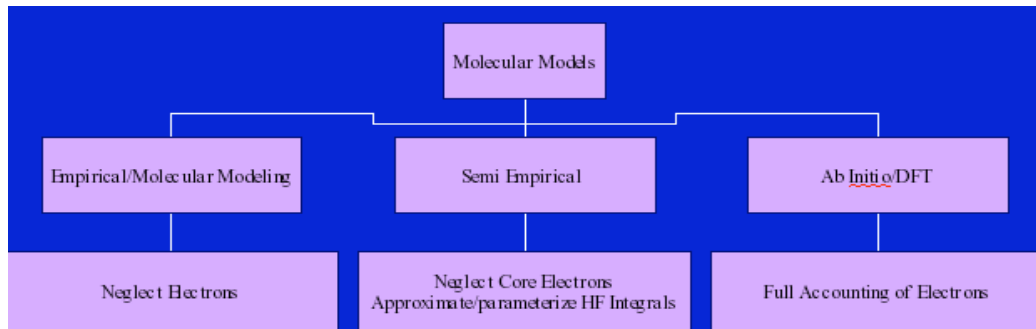
It is not possible to compute all molecules using this method. There are many atoms for which there are not suitable parameters, thus disallowing molecules that contain those atoms. For some parameterized molecules, such as nitrogen, there are well-known inaccuracies in the calculations. Semi-empirical methods do not behave well with hydrogen bonding, transition states, or molecules with non-parameterized atoms.

Fundamental Aspects:

As described in Chapter 7, *ab initio* methods are 100% mathematical, meaning that all of the information generated about an atom, molecule, or reaction comes from the fundamental quantum mechanical calculations (specifically, the Schrödinger equation). This requires significant computing resources; as such, most users are limited to small molecules, typically those consisting of less than 100 atoms.

Semi-empirical methods provide users with a way to study larger molecules. As the name suggests, semi-empirical methods are a combination of *ab initio* methods coupled with the use of data from empirical studies. Through the use of some pre-calculated data, semi-empirical methods allow the user to generate relatively standard information about a molecule. The results obtained are typically less accurate than that of *ab initio* methods, but they are generated more quickly and are possible for larger molecules. A common rule of thumb is that *ab initio* methods scale at a rate of N^4 in terms of computing time (where N is the number of basis functions), whereas semi-empirical methods scale at a rate of N^2 .

The graphic below shows the relationship between the three basic molecular modeling methodologies: molecular mechanics (empirical modeling), semi-empirical, and *ab initio*/density functional theory (DFT). In previous chapters, this *Guide* has described molecular mechanics, which uses classical physics to describe the motion of nuclei, as if they are attached to springs. With *ab initio* methods, we assume that the nuclei do not move (Born-Oppenheimer approximation), and we focus solely on the behavior of electrons. As is suggested by this graphic, semi-empirical methods try to “split the difference” between these two methodologies.



For scientists, educators, and students, the use of semi-empirical methods provides an efficient way to get meaningful chemical information in a relatively short amount of time that does not put heavy demands on computational resources, especially in a shared/distributed system such as the North Carolina High School Computational Chemistry Server.

Methods:

The graphic above suggests how semi-empirical methods differ from those of molecular mechanics and *ab initio* methods. Molecular mechanics methods do not take electrons into account at all, while *ab initio* methods work to fully account for electron behavior (while fundamentally ignoring nuclear motion and influence). The graphic above suggests that semi-empirical methods do *two* things to improve upon the accuracy of the calculations generated by MM methods, while reducing the significant computing time required for performing *ab initio* calculations:

1. Ignore core electrons
2. Approximate/parameterize the HF (Hartree-Fock) integrals

The first major approximation – ignoring core electrons – contains some chemical common sense. As a general rule, it is the outer electrons – the *valence* electrons – that are of particular chemical interest. Most of the characteristics of the molecule, including its reactivity, come from the specific characteristics and behavior of the valence electrons. In semi-empirical methods, core electrons are not really ignored. Rather, the computations are included with some rather simple (and simplistic) calculations of nuclear behavior. The technique of simplifying calculations on core electrons has the value of significantly reducing computing time, without a tremendous sacrifice on accuracy. For example, in a calculation of methanol (CH₃OH), the semi-empirical calculation only deals with 14 of the 18 total electrons, ignoring the *1s* electrons of carbon and oxygen. The user should appreciate that simply lumping core electrons with nuclei needs its own set of approximations and treatments. These details are beyond the scope of this *Guide*.

It is important and useful for readers to understand that semi-empirical methods are a *type of Hartree-Fock method*, but with some of the Hartree-Fock calculations replaced by the use of empirical (experimental) data. As such, some of the approximations used in HF methods, such as the *Born-Oppenheimer* approximation and others discussed in the previous chapter, also apply to semi-empirical methods. One of the major approximations in semi-empirical methods, however, is the relatively complete neglect of what are known as *two-electron integrals*. These integrals are mathematical representations of the fact that two electrons typically occupy a molecular orbital (for example, a *1s² orbital*), and that they have a repulsive effect on each other.

It turns out that by eliminating these two-electron integrals, the size of the calculation to be done decreases by a substantial amount. To do this, semi-empirical methods use a type of approximation known as *zero differential overlap*, or *ZDO*. These methods *parameterize* some of the calculations. Specifically, semi-empirical methods will replace the calculation of the two-electron integrals with data from spectroscopic experimental data. Parameterization means that empirical data is used to create equations or datasets that are stored in the computer code, and accessed at the appropriate point of the semi-empirical calculation.

ZDO actually refers to a family of approximations, grouped according to older and newer methods:

1. Older methods: these methods were developed by John Pople (who also developed the Gaussian software). In this group of methods, data generated by *ab initio* calculations are analyzed using various data fitting algorithms. The results of these data fittings are stored in the software for use during the calculation:
 - a) **CNDO: Complete Neglect of Differential Overlap**. Fundamentally, this method does not understand or care that there are bonds between atoms. It simply calculates a wavefunction (molecular orbital) based on the type of atom and its location. Methods such as CNDO/2 are a variant of CNDO.
 - b) **INDO: Intermediate Neglect of Differential Overlap**. The “intermediate” part of the method lies in the fact that *some* electron-electron repulsions are ignored, but not those that are centered over the same atom. INDO methods do not have any data (parameters) for atoms with atomic numbers greater than 9, so it cannot be used for molecules containing those atoms. MNDO/3 is a version of an INDO method, not to be confused with MNDO (described below).
2. New methods: all three of these methods use a variant of ZDO known as *Neglect of Differential Diatomic Overlap (NDDO)*. In these methods, the major result of the calculation is that the final energy is reported as a *heat of formation* (ΔH_f), rather than as total energies in units of Hartrees. These methods are attributed primarily to Michael Dewar:

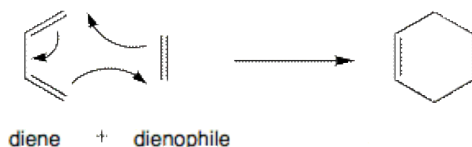
ELEMENT	AM1	PMS	MNDO
H	X	X	X
Li		X	X
Be		X	X
B	X		X
C	X	X	X
N	X	X	X
O	X	X	X
F	X	X	X
Na			
Mg		X	
Al	X	X	X
Si	X	X	X
P	X	X	X
S	X	X	X
Cl	X	X	X
Ca			
Fe			
Cu			
Zn	X	X	X
Ga			
Ge	X	X	X
As	X		
Se	X		
Br	X	X	X
Cd			X
In			
Sn	X	X	X
Sb	X		
Te	X		
I	X	X	X
Hg	X	X	X
Tl		X	
Pb			
Bi			

- a) MNDO: **M**odified **N**eglect of **D**ifferential **O**verlap. Parameters for this method come from a statistical analysis (a linear least squares regression fit) of enthalpies of formation and well-known molecular geometries. MNDO methods tend to overestimate repulsive forces between atoms.
- b) AM1: Austin Method 1 (so named after the University of Texas – Austin). The AM1 method looks to address the overestimation of repulsive forces by recalculation of the atom-to-atom forces. It does so by multiplying these forces by a sum of Gaussian STO (Slater-type orbital) calculations. In the AM1 method, there are somewhere between 10 and 19 parameters for an individual atom (refer to the list to see which atoms have parameters for a given method)
- c) PM3: Parameterization Method 3, developed by J.P. (Jimmy) Stewart in the late 1980s. The “3” comes from the fact that this is the third NDDO method (following MNDO and AM1). PM3 contains many of the same parameters as does AM1, but the parameters were derived more systematically. As compared with MNDO, the parameters are quite different, but the accuracy of the calculation is close to the same. The PM3 method has approximately 18 different parameters for each of the parameterized atoms (11 for hydrogen).

Of the three methods, PM3 is the most robust and the most commonly cited in the computational chemistry literature.

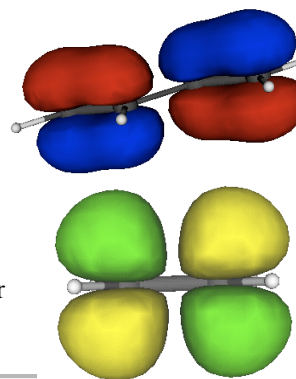
Applications:

Semi-empirical methods are designed specifically for organic systems, and, as such, are particularly well parameterized for those systems. For example, one might use semi-empirical methods for studying the reactivity of organic systems. One such example is the well known (at least to organic chemists!) *Diels-Alder* reaction. This reaction is described as the reaction of an alkene (a C-H compound with one double bond) and a diene (a C-H compound with two double bonds). The alkene is called a *dienophile*, in that it is attracted to the diene. The two compounds react to form a cycloalkene, a ring structure with a double bond. Cycloaddition reactions are an important class of reactions in organic chemistry. Graphically, the reaction is shown as follows:



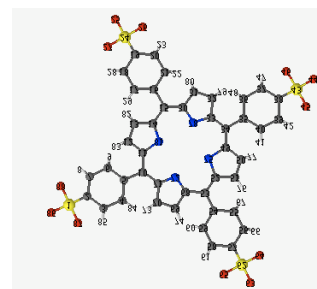
In this reaction, the diene is butadiene (C_4H_6) and the dienophile is ethene (C_2H_4). The two compounds react to form cyclohexene (C_6H_{10}), a six-carbon ring structure.

We can perform a semi-empirical calculation on both butadiene and ethene, running a molecular orbital calculation using PM3. The graphical results are shown. The top image shows the HOMO (highest occupied molecular orbital) for butadiene, and the bottom image shows the LUMO (lowest unoccupied molecular orbital) for the alkene. The molecular orbitals show lobes above and below the double bonds for the butadiene, with the blue representing a positive form of the wavefunction and the red representing negative. Likewise, for the ethene we see positive (green) and negative (yellow) wavefunctions above and below the plane of symmetry for the molecule. The significance of these calculations is as follows. Molecular orbital theory tells us that a reaction occurs *if* there is an overlap between a



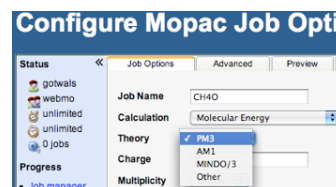
positive lobe of the HOMO of a molecule and the positive lobe of the LUMO of a molecule, and also between the negative HOMO-negative LUMO. The graphic shows that there is the potential for overlap between blue-green and red-yellow. As such, one can predict that there should be the potential for a reaction between these two compounds, which indeed happens.

Semi-empirical methods can also be used for larger molecules, one of its advantages (as discussed in the appropriate section). The molecule shown is tetraphenylporphine, a complicated molecule composed of a number of carbon and nitrogen ring structures. Typically this compound has a copper (II) atom in the center of the ring structure, and is of particular interest in the study of how metals form complexes with metal ions. Nitrogen compounds, by the way, sometimes give quantitatively questionable results in semi-empirical methods, particularly in the determination of the geometry of the molecule. AM1 methods sometime determine that non-planar molecules are flat, while planar molecules have a pyramidal structure. For all methods, not only semi-empirical, these mathematical “quirks” are part of the business of computational chemistry.



Software Tools:

Semi-empirical methods are found in virtually all major software codes. On the North Carolina High School Computational Chemistry server, semi-empirical methods are a part of GAMESS and Gaussian, and are the central (and only) method available through MOPAC. The three primary Hamiltonians are all from the NDDO family, and include PM3, and AM1. MNDO/3, an earlier semi-empirical technique and part of the INDO family, is also available in MOPAC. For GAMESS there are options for PM3, AM1, and MNDO, while Gaussian only includes PM3 and AM1 (although MNDO and MNDO/3 are available under the keyword system).



Advantages:

The use of semi-empirical methods offers a number of advantages to the user, especially one new to molecular modeling. As is the case with all of the methods – molecular mechanics, *ab initio*, semi-empirical, and density functional theory – there are always trade-offs to be made between accuracy of the answer and the computational resources/time needed to perform the calculation.

The advantages of the semi-empirical method are as follows:

1. Molecule size: Semi-empirical methods came into existence at a time when *ab initio* methods could only be reasonably performed on the smallest of molecules. Even with the increased computing power available now, molecule size is still a major consideration. Semi-empirical methods can produce reasonable results on relatively large molecules (100s of atoms) in a reasonable amount of time. This is especially of importance to the student who might not only have limited computing time (as is the case for classroom accounts on the North Carolina High School Computational Chemistry server) but who also is mostly satisfied with a balance between qualitative and quantitative results.
2. Application to organic molecules: Dewar and his group at the University of Texas – Austin developed semi-empirical methods primarily for the study of organic molecules. Indeed, it was Dewar’s goal to develop a type of “molecular orbital spectrometer”, specifically as a way to predict whether or not various types of organic reactions would occur. As such, semi-empirical methods are particularly well parameterized for organic systems. Semi-empirical methods can be demonstrated effectively in the study of organic mechanisms such as the

Woodward-Hoffman rules. These rules are of particular importance in the study of pericyclic reactions, which include organic rearrangement reactions.

3. Qualitative and quantitative results: semi-empirical methods give good results in calculating and visualizing molecular orbitals for use in describing the molecular system, in particular in use to predict whether or not a reaction will take place. For molecules for which the atoms are well parameterized, semi-empirical methods give relatively good results, especially as compared with the results one would obtain with *ab initio* methods in the same amount of time.

Disadvantages:

The disadvantages of semi-empirical methods are relatively few, especially for the new user. The major disadvantages are as follows:

1. Accuracy trade-offs: the advantages gained by a quicker calculation are offset by the decrease in accuracy obtained by semi-empirical methods.
2. Problem type limitations: semi-empirical methods tend not to work very well for these types of chemical systems:
 - a. Molecules that include hydrogen bonding
 - b. Transition structures
 - c. Molecules for which the atoms are not parameterized, or for which the atoms are poorly parameterized.

In the balance between advantages and disadvantages, the choice of semi-empirical methods for the computational chemistry student is *highly favored*. Given the relative efficiency of the calculations, coupled with the reasonableness of the results, semi-empirical methods are as close to a “best of both worlds” method as are currently available to the molecular modeler. This is especially true if the goal of the computation is to help the student develop “chemical intuition” about molecular structure, properties, and activities.