Semiempirical Implementations of Molecular Orbital Theory

How one can make Hartree-Fock theory less computationally intensive without much sacrificing its accuracy? The most demanding step – calculation of two-electron (four-index) integrals (J and K integrals) appearing in the Fock matrix elements (N^4 where N is the number of basis functions). One way to save time – to estimate their value accurately in an *a priori* fashion and thus to avoid numerical integration.

Coulomb integrals measure the repulsion between electrons in regions of space defined by the basis functions. When the basis functions in the integral for one electron are very far from the basis functions for the other, the value of that integral will approach zero. In a large molecule, one might be able to avoid the calculation of a very large number of integrals simply by assuming them to be zero.

HF theory is intrinsically inaccurate as it does not include correlation energy. Therefore, modifications of the theory introduced in order to simplify its formalism may actually improve it, provided the new approximations somehow introduce an accounting for correlation energy. Most typically, such approximations involve the adoption of a parametric form for some aspect of the calculation where the parameters involved are chosen so as best reproduce experimental data – ‘semiempirical’.

Another motivation for introducing semiempirical approximation into HF theory was to facilitate the computation of derivatives (gradients, Hessians) so that geometries could be more efficiently optimized.

Extended Hückel Theory

Before considering semiempirical methods we revisit Hückel theory:

$$\begin{vmatrix} H_{11}-ES_{11} & H_{12}-ES_{12} & \cdots & H_{1N}-ES_{1N} \\ H_{21}-ES_{21} & H_{22}-ES_{22} & \cdots & H_{2N}-ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1}-ES_{N1} & H_{N2}-ES_{N2} & \cdots & H_{NN}-ES_{NN} \end{vmatrix} = 0$$

The dimension of the secular determinant depends on the choice of the basis set. EHT has two critical conventions: 1) all core electrons are ignored (all modern semiempirical methods adopt this approximation); 2) if an atom has occupied d orbitals, the highest occupied level of d orbitals is considered to contribute to the set of valence orbitals.

Each remaining valence orbital is represented by a Slater-type orbital (STO):

$$\varphi(r,\theta,\phi;\varsigma,n,l,m) = \frac{(2\varsigma)^{n+1/2}}{\left[(2n)!\right]^{1/2}} r^{n-1} e^{-\varsigma r} Y_l^m(\theta,\phi)$$

ζ - an exponent, depends on the atomic number

n – the principal quantum number for the valence orbital

$Y_l^m(\theta,\phi)$ - spherical harmonic functions, depend on the angular momentum quantum numbers l and m.

The size of the secular determinant is dictated by the total number of valence orbitals in the molecule. Example – MnO_4^- anion – 25 STO basis functions: one 2s and three 2p functions for each oxygen (16), one 4s, three 4p, and five 3d functions for manganese.

Attractive features of STOs – the correct exponential decay with increasing *r*, the angular component is hydrogenic, and the 1s orbital has, as it should a cusp at the nucleus (it is not smooth). From a practical point of view, overlap integrals between two STOs as a function of interatomic distance are readily computed – overlap matrix elements in EHT are not assumed to be 0 or 1 but computed.

Resonance integrals *H*: diagonal $H_{\mu\mu}$ is taken as the negative of the average ionization potential for an electron in the appropriate valence orbital: $H_{\mu\mu} = -13.6$ eV for a hydrogen 1s function. In many-electron atoms, the valence-shell ionization potential (VSIP) for the ground state atomic term may not be the best choice for the atom in a molecule, so this term is best regarded as an adjustable parameter. VSIPs have been tabulated for most of the atoms in the periodic table. Atoms in molecular environments may develop fairly large partial charges – schemes for adjusting the neutral atomic VSIP as a function of partial atomic charge have been proposed – Fenske-Hall effective Hamiltonian calculations, still find considerable use for inorganic and organometallic systems composed of atoms having widely different electronegativities.

Off-diagonal resonance integrals:

$$H_{\mu\nu} = \frac{1}{2} C_{\mu\nu} \left(H_{\mu\mu} + H_{\nu\nu} \right) S_{\mu\nu}$$

C – an empirical constant, *S* – an overlap integral. The energy associated with the matrix element is thus proportional to the average of the VSIPs for the two orbitals μ and ν multiplied by the extent to which two orbitals overlap in space (by symmetry, the overlap between two different STOs on the same atom is zero).

Using standard linear algebra methods, the MO energies and wave functions can be found from solution of the secular equation. The matrix elements do not depend on MOs (unlike HF theory), so the calculation is not iterative and is very fast even for very large molecules (the process becomes iterative if VSIPs are adjusted as a function of partial atomic charge, since the partial atomic charge depends on occupied orbitals). EHT is insufficiently accurate to predict geometries – the use is best restricted to systems with available experimental geometries. EHT predicts MO qualitatively correct – now is used for very large systems, solids, band structure.

CNDO formalism

Returning to the SCF formalism of HF theory, one can proceed in the spirit of an effective Hamiltonian method described above by developing a recipe for replacement of matrix elements in the HF secular equation. CNDO – complete neglect of differential overlap, **John Pople** and co-workers (1965):

1. The basis set is formed from valence STOs, one STO per valence orbital. In the original version, only atoms having s and p valence orbitals were addressed.
2. In the secular determinant, overlap matrix elements are defined by $S_{\mu\nu} = \delta_{\mu\nu}$

3. All two-electron integrals are parameterized using the following scheme:
   $(\mu\nu \mid \lambda\sigma) = \delta_{\mu\nu}\delta_{\lambda\sigma}(\mu\mu \mid \lambda\lambda)$ The only non-zero integrals have μ and ν as identical orbitals on the same atom, and λ and δ as identical orbitals on the same atom; the second atom might be different than the first (integrals involving overlap of different basis functions are completely neglected – CNDO).
4. For the surviving two-electron integrals: $(\mu\mu \mid \lambda\lambda) = \gamma_{AB}$
   A and B are the atoms on which basis functions μ and λ reside, respectively. The term γ can be computed explicitly from s-type STOs or can be treated as a parameter.
   Pariser-Parr approximation: $\gamma_{AA} = IP_A - EA_A$
   IP and EA are the atomic ionization potential and electron affinity, respectively. For the two-center term, Mataga-Nishimoto formalism: $\gamma_{AB} = \dfrac{\gamma_{AA} + \gamma_{BB}}{2 + r_{AB}(\gamma_{AA} + \gamma_{BB})}$
   At large distances, it goes to $1/r_{AB}$, at short distances it approaches the average of the two one-center parameters.
5. One-electron integrals for diagonal matrix elements:
   $$\left\langle \mu \middle| -\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{r_k} \middle| \mu \right\rangle = -IP_\mu - \sum_k \left(Z_k - \delta_{Z_A Z_k}\right)\gamma_{Ak}$$

Remember that the full Fock matrix element $F_{\mu\mu}$ is the sum of the one-electron integral and a series of two-electron integrals. If the number of valence electrons on each atom is exactly equal to the valence nuclear charge, then the repulsive two-electron terms will exactly cancel the attractive nuclear terms at the end of the equation and we will obtain the desired result: the energy associated with the diagonal matrix element is the ionization potential of the orbital.

6. One-electron integrals for off-diagonal matrix elements:

$$\left\langle \mu \middle| -\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{r_k} \middle| \nu \right\rangle = \frac{(\beta_A + \beta_B)S_{\mu\nu}}{2}$$

μ and ν are centered on atom A and B, respectively; β – semi-empirical parameters; $S_{\mu\nu}$ - the overlap matrix element computed using the STO basis set. The β parameters provide a measure of the strength of through space interactions between atoms – they were originally adjusted to reproduce certain experimental quantities.

The CNDO method – a vast simplification of HF theory. The number of two-electron integrals – N^2, they are computed by trivial algebraic formulas, evaluation of one-electron integrals is also entirely avoided. The simplifications of course have chemical cost. CNDO is quite uncapable of accurately predicting molecular structures, it cannot distinguish between different types of atomic orbitals (example – singlet and triplet states of methylene, CH_2) and between the orientation of those orbitals. Another example – if we consider the rotational coordinate for hydrazine, N_2H_4, one factor influencing the energetics will be the repulsion of the lone pairs, one on each nitrogen. However, this repulsion in CNDO is γ_{NN}, which does not depend of the orientation, only on the N-N distance.

INDO formalism

Of the two deficiencies of CNDO, the methylene problem is atomic in nature – it involves electronic interactions on a single center, while the hydrazine problem is molecular, as it involves two centers. Many ultraviolet/visible (UV/Vis) spectroscopic transitions in molecules are reasonably highly localized to a single center (transitions in mononuclear inorganic complexes). Pople et al. – modifications to the CNDO formalism to permit a more flexible handling of electron-electron interactions on the same center in order to model such spectroscopic transitions: "intermediate neglect of differential overlap" (INDO). The key change – to use different values for the unique one-center two-electron integrals:

$(ss|ss) = G_{ss}$ $(ss|pp) = G_{sp}$ $(pp|pp) = G_{pp}$

$(pp|p'p') = G_{pp'}$ $(sp|sp) = L_{sp}$

The G and L are free parameters, in practice they can be estimated from spectroscopic data. When the atomic valence orbitals include d and f functions, the number of unique integrals considerably increases and the estimation of appropriate vales from spectroscopy becomes considerably more complicated.

Valence bond angles in INDO are predicted with much greater accuracy then in CNDO, but nevertheless, overall molecular geometries predicted from INDO tend to be rather poor. If a good molecular geometry is available, the INDO method has considerable potential for modeling the UV/Vis spectroscopy. Ridley and Zerner (1973): INDO/S – careful parameterization of INDO for spectroscopy. The INDO/S model is very successful for d → d transitions within transition metal complexes (within 2000 cm^{-1}), but less robust for transitions that are not well localized to a single center (metal-to-ligand , etc.) and is not good for Rydberg states. INDO/S has a good accuracy for ionization potentials and oscillator strengths for weak transitions.

MINDO/3 and SINDO1

Michael Dewar established, as a goal, the development of a parameter set that would as robust as possible across the widest possible set of molecules (at the time, he has chosen organic chemistry, with a few inorganic compounds comprised of second- and third-row elements). A truly thorough test of a general model – efforts of more than one research group are needed – a computer code available to a general chemistry public. Dewar's code included geometry optimization routines, which made it particularly attractive to non-developers interested in using the code for chemical purposes.

The first general parameterization reported by Dewar and co-workers – MINDO/3 (1975). Modifications of the INDO framework: the use of different ζ exponents in s and p type STOs on the same atom, the definition of pair parameters β_{AB} between two atoms A and B were not averages of atomic parameters (four such parameters were introduced per pair of atoms: s_A, s_B; s_A, p_B; p_A, s_B; and p_A, p_B), adoption of a slightly different form for γ_{AB}, and some empirical modifications of the nuclear repulsion energy.

Every parameter was treated as a free variable subject to 'chemical common sense' restraints. To optimize parameter values, Dewar initially took a set of 138 small molecules containing C, H, N, and O, and constructed a penalty function depending on bond distances, valence angles, torsional angles, dipole moments, ionization potentials, and heats of formation. The performance of the MINDO/3 model was impressive overall. The mean absolute error in predicted heats of formation was 11 kcal/mol (all molecules), the corresponding error for IPs was 0.7 eV (46 molecules), for heavy-atom bond lengths 0.022 Å (81 molecules), and for dipole moments 0.45 D (31 molecules). Such mean errors would not be tolerated today but are very small for 1975.

Dewar's subsequent work on other semiempirical models rendered MINDO/3 effectively obsolete, but its historical importance remains unchanged. A modified INDO model that is not entirely obsolete – symmetric orthonogalized INDO (SINDO1) by Jug and co-workers (1980). SINDO1 includes d functions for atoms of the second row. This inclusion proves critical for handling hypervalent molecules; SINDO1 performs considerably better for phosphorus-containing compounds than do other semiempirical models that lack d functions.

Basic NDDO formalism

The INDO model extends the CNDO model by adding flexibility to the description of the one-center two-electron integrals. However, in INDO there continues to be only a single two-center two-electron integral γ_{AB} irrespective of which orbitals on atoms A and B are considered. The neglect of diatomic differential overlap (NDDO) method relaxes the constraints on two-center two-electron integrals: all integrals $(\mu\nu|\lambda\sigma)$ are retained provided μ and ν are on the same atomic center and λ and σ are on the same atomic center, but not necessarily the center of μ and ν. The order of μ and ν does not affect the value of the integral, we need to worry only about combinations – 10 unique combinations of s, p_x, p_y, and p_z. With 10 unique combinations on each atom, there 100 possible combinations of combinations for the integral. If we include d functions, the number of unique integrals increases to 2025. Although these numbers seem large, this is still a considerable improvement over evaluating every possible integral. Most modern semiempirical models are NDDO models.

MNDO – a modified neglect of differential overlap (Dewar and Thiel, 1977). Initially developed for C, H, O, and N.

A diagonal element of the Fock matrix:

$$F_{\mu\mu} = U_{\mu} - \sum_{B \neq A} Z_B(\mu\mu \mid s_B s_B) + \sum_{\nu \in A} P_{\nu\nu}\left[(\mu\mu \mid \nu\nu) - \frac{1}{2}(\mu\nu \mid \mu\nu)\right] +$$

$$\sum_{B} \sum_{\lambda \in B} \sum_{\sigma \in B} P_{\lambda\sigma}(\mu\mu \mid \lambda\sigma)$$

μ is located on atom A. The first term is the atomic orbital ionization potential, the second term the attraction to the other nuclei where each term is proportional to the repulsion with the valence s electron on the nucleus, the third term reflects the Coulomb and exchange interactions with the other electrons, and the fourth term reflects Coulomb repulsion with electrons on other atoms B.

An off-diagonal matrix element for two basis functions μ and ν on the same atom A:

$$F_{\mu\nu} = -\sum_{B \neq A} Z_B(\mu\mu \mid s_B s_B) + P_{\mu\nu}\left[\frac{3}{2}(\mu\nu \mid \mu\nu) - \frac{1}{2}(\mu\mu \mid \nu\nu)\right] +$$

$$\sum_{B} \sum_{\lambda \in B} \sum_{\sigma \in B} P_{\lambda\sigma}(\mu\nu \mid \lambda\sigma)$$

When μ is on atom A and ν on atom B: $F_{\mu\nu} = \frac{1}{2}\left(\beta_{\mu} + \beta_{\nu}\right)S_{\mu\nu} - \frac{1}{2}\sum_{\lambda \in A} \sum_{\sigma \in B} P_{\lambda\sigma}(\mu\lambda \mid \nu\sigma)$

The first term is the resonance integral that describes the one-electron kinetic energy and nuclear attraction terms; it is an average of atomic resonance integrals β times the overlap of the orbitals involved. The second term includes favorable exchange interactions. How to evaluate all the necessary two-electron integrals? Unlike one-center two-electron integrals, it is not easy to analyze spectroscopic data to determine universal values. Dewar and co-workers – replaced continuous clouds with classical multipoles.

An ss product was replaced with a point charge, and sp product was replaced with a classical dipole, and a pp product was replaced with a classical quadrupole – the evaluation is simple and evaluation of their analytical derivatives with respect to nuclear motion is simple.

The nuclear repulsion energy is added to the SCF energy:

$$V_N = \sum_{k<l}^{nuclei} Z_k Z_l \left(s_k s_l \mid s_k s_l\right)\left(1 + \frac{1}{\tau} e^{-\alpha_{Z_l} r_{kl}} + e^{-\alpha_{Z_l} r_{kl}}\right)$$

Z – the valence atomic number, α - a parameter specific for each atom type, r – interatomic distance, τ - 1 unless the two nuclei k and l are an O/H or N/H pair, in which case it is rXH. Internuclear repulsion is proportional to the repulsion between s electrons on the same center and is empirically increased slightly at short bond lengths to make up for imbalances in the electronic part.

As with MINDO/3, Dewar and Thiel optimized the parameters of the MNDO model against a large test set of molecular properties. Within the assumption of a valence orbital set comprised only of s and p orbitals, MNDO parameters are now available for H, He, Li, Be, B, C, N, O, F, Al, Si, P, S, Cl, Zn, Ge, Br, Sn, I, Hg, and Pb. MNDO now is mostly replaced by the later NDDO models (AM1 and PM3), but MNDO calculations still appear in the literature. MNDO forms the foundation for MNDO/d. A modified MNDO model explicitly adding electron correlation effects (MNDOC) was described by Thiel in 1981.

One critical flaw in MNDO is that it does very poorly in the prediction of hydrogen bonding geometries and energies. Modification – Austin Model 1 (AM1, Dewar, 1985). The nuclear repulsion energy between any two nuclei A and B is computed as

$$V_N(A,B) = Z_A Z_B (s_k s_l \mid s_k s_l) + \frac{Z_A Z_B}{r_{AB}} \sum_{i=1}^{4} \left(a_{A,i} e^{-b_{A,i}(r_{AB}-c_{A,i})^2} + a_{B,i} e^{-b_{B,i}(r_{AB}-c_{B,i})^2} \right)$$

Every atom has up to 4 each parameters *a*, *b*, and *c* describing Gaussian functions centered at various distances *c* that modify the potential of mean force between the two atoms. Simultaneous optimization of the original MNDO parameters with the Gaussian parameters markedly improved performance.

Now AM1 parameters are available for H, C, N, and O, as well as B, F, Al, Si, P, S, Cl, Zn, Ge, Br, I, and Hg. AM1 calculations are fast and the model is reasonably robust over a large range of chemical functionality – AM1 is included in many molecular modeling packages and AM1 results continue to be reported in the chemical literature for a wide variety of applications.

PM3

J.J.P. Stewart: the development of AM1 had been potentially non-optimal, from a statistical point of view because (i) the optimization of parameters was carried out stepwisely; (ii) the search of parameter space had not been completely exhaustive; (iii) human intervention based on perceived 'reasonableness' of parameters occurred in many instances.

Stewart had a somewhat more mathematical philosophy – a sophisticated search of parameter space using complex optimization algorithms might be more successful in producing a best possible parameter set within Dewar's NDDO framework. Stewart optimized simultaneously parameters for H, C, N, O, F, Al, Si, P, S, Cl, Br, and I. He employed a significantly larger data set in evaluating his penalty function – Parametrization Model 3 (PM3; Stewart 1989). PM3 will not necessarily be better than MNDO or AM1 for any particular problem but is likely to be optimal for systems closely resembling molecules found in the training set. PM3 parameters are also available for Li, Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, Bi, Po, and At. PM3 is a very robust NDDO model and continues to be used as widely as AM1.

General Performance Overview of Basic NDDO Methods

Energetics

Mean unsigned errors (kcal/mol) in predicted heats of formation

Elements Subset	MNDO	AM1	PM3	MNDO/d
Lighter (181)	7.35	5.80	4.71	
CH (58)	5.81	4.89	3.79	
CHN (32)	6.24	4.65	5.02	
CHNO (48)	7.12	6.79	4.04	
CHNOF (43)	10.50	6.76	6.45	

Radicals (14)	9.3	8.0	7.4	
Heavier (488)	29.2	15.3	10.0	4.9
Al (29)	22.1	10.4	16.4	4.9
Si (84)	12.0	8.5	6.0	6.3
P (43)	38.7	14.5	17.1	7.6
S (99)	48.4	10.3	7.5	5.6
Cl (85)	39.4	29.1	10.4	3.9
Br (51)	16.2	15.2	8.1	3.4
I (42)	25.4	21.7	13.4	4.0
Hg (37)	13.7	9.0	7.7	2.2
Normal (421)	11.0	8.0	8.4	4.8
Hypervalent (67)	143.2	61.3	19.9	5.4
Cations (34)	9.55	7.62	9.46	
Anions (13)	11.36	7.11	8.81	

AM1 and PM3 are always more accurate than MNDO. PM3 offers a slight advantage over AM1 for molecules composed of lighter elements (C, H, N, O, F) and a clear advantage for heavier elements. In the latter case, the difference is essentially entirely for hypervalent molecules. Errors for ions and radicals are larger than for neutral closed-shell molecules.

Another interesting energetic property – IP. In HF theory, the eigenvalue associated with each MO is the energy of an electron in that MO. Thus, a good estimate of the negative of the IP is the energy of the highest occupied MO – Koopmans' theorem. Employing this approximation, all of the semiempirical methods do reasonably well in predicting IPs for organic molecules. On a test set of 207 molecules containing H, C, N, O, F, Al, S, P, Cl, Br, and I, the average errors in predicted IP are 0.7 eV (MNDO), 0.6 eV (AM1), and 0.5 eV (PM3). For purely inorganic compounds PM3 gives as good results, the errors of MNDO and AM1 increase by a few tenths of eV.

Energetics of conformational changes and reactions – MNDO: steric crowding tends to be strongly disfavored and small ring compounds are predicted to be too stable. For the most part, these problems are corrected in AM1 and PM3. Activation enthalpies are still more likely to be overestimated because electron correlation energy tends to be more important in TS structures than in minima. Since correlation energy is introduced in only an average way by parametrization of the semiempirical HF equations, it cannot distinguish well between the two kinds of structures.

For weak intermolecular interactions (van der Waals or dispersion forces) or hydrogen bonding, semiempirical methods are in general unreliable. Hydrogen bonding is better described by AM1 and PM3 than by MNDO, still in most instances the interaction energies are systematically underestimated by up to 50% - basic NNDO methods are not well suited to the characterization of hydrogen bonded systems.

Energetic barriers to rotation about bonds with partial double bond character – significantly underestimated (by ~15 kcal/mol for the C-N bond). Conformational analysis – the NDDO methods are not quantitatively very accurate. For 8 different sets of conformer pairs, predictions from MNDO, AM1, and PM3 gave mean unsigned errors of 1.4, 1.3, and 1.8 kcal/mol, respectively.

Geometries

Correct molecular structures are dependent on the proper location of wells in the PES – they are intimately related to the energetics of conformational analysis. For organic molecules, most gross structural details are modeled with a reasonable degree of accuracy. Dewar et al. evaluated AM1 and PM3 for 344 bond lengtrhs and 146 valence angles in primarily organic molecules composed of H, C, N, O, F, Cl, Br, and I; the average unsigned errors were 0.027 and 0.022 Å, respectively, for the bond lengths, and 2.3 and 2.8°, respectively, for the angles. For PM3, Stewart performed a similar analysis for a larger set of molecules, some of them including Al, Si, P, and S. For 460 bond lengths, the mean unsigned errors wer 0.054, 0.050, and 0.036 Å for MNDO, AM1, and PM3, respectively. For 196 valence angles, the errors were 4.3, 3.3, and 3.9°.

The geometries for the molecules containing second-row elements are more difficult to predict than for simpler organics. MNDO and AM1 are less successful than PM3. For dihedral angles the errors were 21.6, 12.5 and 14.9°, respectively, for MNDO, AM1, and PM3. Heteroatom-heteroatom linkages are often problematic (for example, O-O, N-N, P-P, and Si-X). The energetics of hydrogen bonding is not handled well and geometries are simila problematic.

Charge Distributions

One of the most useful features of a QM model is its ability to provide information about the molecular charge distribution. Normally, even very low quality QM methods tend to give reasonable charge distribution. For a 125-molecule test set including H, C, N, O, F, Al, Si, P, S, Cl, Br, and I, Stewart found mean unsigned errors in dipole moments of 0.45, 0.35, and 0.38 D, respectively, for MNDO, AM1, and PM3. PM3 works better for compounds containing phosphorus. For partitioning into partial atomic charges, all three methods give reasonable results, except PM3 tends to predict nitrogen atoms to be too weakly electronegative.

Ongoing developments in semiempirical MO theory

d orbitals in NDDO methods

Thiel and Voityuk (1992, 1996) described the first NDDO model with d orbitals included, MNDO/d. It has high utility for thermochemical applications. MNDO/d parameters have been determined for Na, Mg, Zn, Zr, and Cd. However, since the model is based on MNDO and is identical to MNDO for light elements, it still performs rather poorly for intermolecular interactions and for hydrogen bonding.

Hehre and co-workers extended the PM3 Hamiltonian to include d orbitals. PM3(tm) is available in SPARTAN. Only geometrical data (from X-ray crystallography) were included in the penalty function for parameterization. The model may be regarded as an efficient way to generate reasonable molecular geometries whose energies may then be evaluated using m
sophisticated levels of theory.

Dewar and co-workers (1993, 1994) – semi-ab initio model 1 (SAM1 or SAM1D with d orbitals). Somewhat different from the NDDO formalism, some two-electron integrals are calculated analytically. Parameters exist for H, Li, C, N, O, F, Si, P, S, Cl, Fe, Cu, Br, and I. For molecules of light elements, SAM1 performs better than AM1 and very slightly better than PM3. The same is true for non-hypervalent molecules of heavier elements, but large improvements are observed for molecules containing hypervalent heavy atoms – across 404 compounds containing Si, P, S, Cl, Br, and I, the mean unsigned errors in heats of formation for AM1, PM3, SAM1, and MNDO/d are 16.2, 9.5, 0.3, and 5.1 kcal/mol.

Linear scaling

The developments of methods that scale linearly with respect to system size opens the door to the modeling of very large systems with maximal computational efficiency. NDDO approximation is already rather efficient in forming the Fock matrix, so it serves as an excellent basis on which to build a linear scaling QM model and such models have been reported. The motivation for developing linear scaling methods – to permit QM calculations of biomolecules. However, because of the relative poor performance of semiempirical models for non-bonded interactions, one can expect that such models would be bad at predicting biomolecular geometries – inferior to the far more efficient force fields developed and optimized for this purpose. But QM calculations provide rather accurate charge distribution, so they are more sensitive to charge-charge interactions and functional group interactions. In addition, one may be also interested in the modeling of a bond-making/bond breaking reactions within a very large molecular framework, in which case the availability of appropriate force-field models is extremely limited. QM and MM models have complementary strengths – QM/MM models are being developed.