$$\Delta E = -\sum_{ab}(q_a + q_b)\beta_{ab}S_{ab} \quad + \quad \sum_{k<l}\frac{Q_k Q_l}{\varepsilon R_{kl}} \quad + \quad \sum_{r}^{occ.}\sum_{s}^{unocc.} - \sum_{s}^{occ.}\sum_{r}^{unocc.} \frac{2(\sum_{ab}c_{ra}c_{sb}\beta_{ab})^2}{E_r - E_s}$$

Closed-Shell Repulsion        Coulomb Interaction        Frontier Orbital Interaction

# Chem 1140; Molecular Modeling

- *Molecular Mechanics*
- *Semiempirical QM Modeling*
- *CaCHE*

---

## A. Modeling Software

Overview: http://cmm.info.nih.gov/software.html

- CaCHe: Structure building. Extended MM2(87) energy minimization, extended Hückel, MOPAC, ZINDO MO programs. Orbital electron density and electrostatic maps. Prediction of octanol/water partition coefficient and water solubility

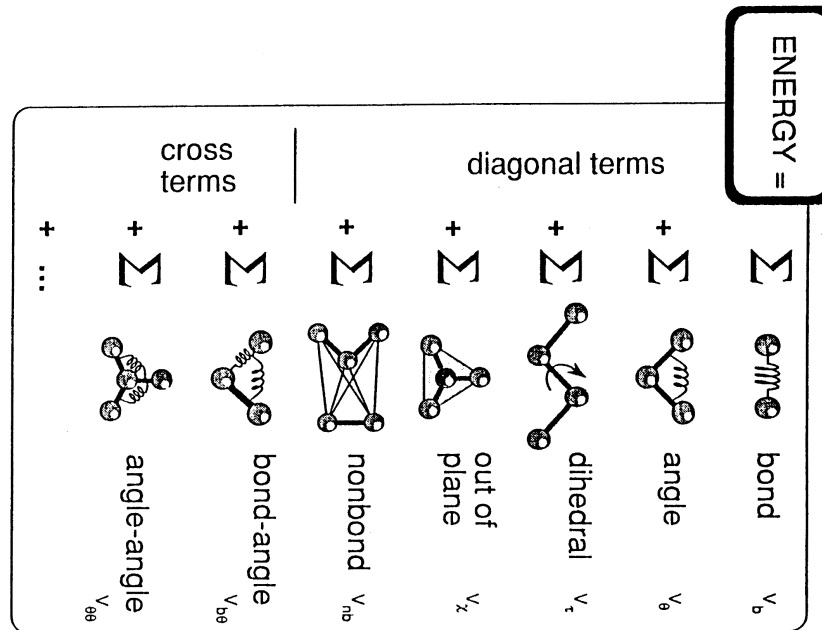# B. Introduction to Molecular Modeling

Objectives:
- Computer graphics visualization of molecules ("Dreiding models")
- Matching (overlays, docking) of molecules
- Use of empirical force-fields to determine molecular properties as well as interatomic distances.
- Correlate molecular properties with an electronic structure from *ab initio* quantum mechanics or semiempirical quantum mechanics.
- Gain information on dynamic molecular movements.
- Use computer-assisted design for molecular recognition in organic, bioorganic, and medicinal chemistry and material science.
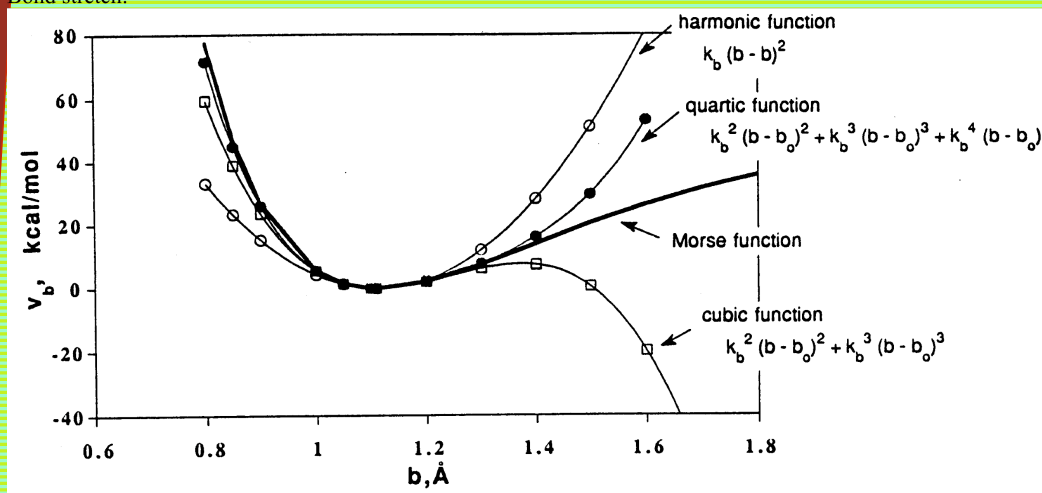
## Comparison of Quantum and Molecular Mechanics

| Advantage of using | Quantum mechanics | Molecular mechanics |
|---|---|---|
| Conceptual simplicity | | ● |
| Speed of computations | | ● |
| Capability to treat large molecules | | ● |
| Inclusion of solvation effects | | ● |
| Conformational analysis of "normal" molecules | | ● |
| Determination of geometries and relative energies | ● | ● |
| Generality of parameterization | ● | |
| Applicability to short-lived species | ● | |
| Treatment of intermediate/unknown hybridizations | ● | |
| Determination of electronic properties | ● | |
| Theoretical rigor | ● | |

Schematic of molecular force field expression. Diagonal terms refer to interactions that can be expressed as functions of a single internal coordinate, whereas cross terms introduce coupled interactions involving two or more coordinates.
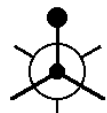


Bond stretch:



harmonic function
$$k_b (b - b)^2$$

quartic function
$$k_b^2 (b - b_o)^2 + k_b^3 (b - b_o)^3 + k_b^4 (b - b_o)^4$$

Morse function

cubic function
$$k_b^2 (b - b_o)^2 + k_b^3 (b - b_o)^3$$

Dihedral potentials

The traditional expression for the dihedral potential energy is a cosine series:

$$V_T(\tau) = \sum_{n=0}^{N} A_n \left[ 1 + s\cos(n\tau) \right]$$
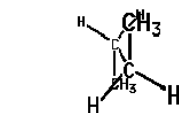
$s = 1$ for staggered minima
$s = -1$ for eclipsed maxima

$n$ = periodicity of rotation
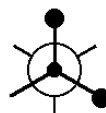$n = 3$ for ethane

TOC

**Anti conformations**

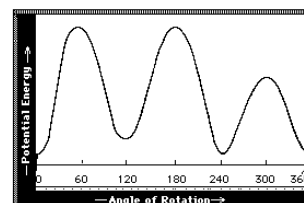anti

Consider the rotation of n-butane about the $C_2$-$C_3$ bond:
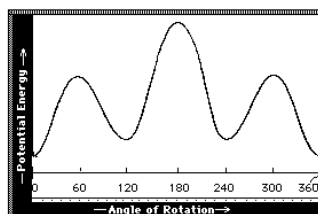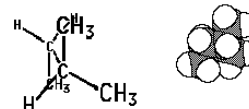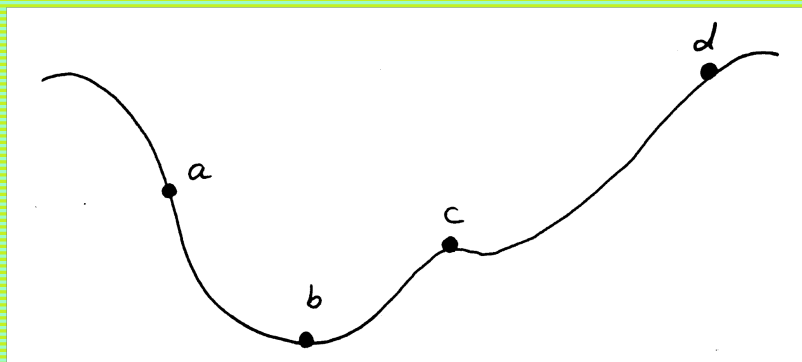
TOC

**Anti Conformation**

anti

Consider the rotation of methylbutane about the $C_2$-$C_3$ bond:

---

# Computational Options

**Single point calculation**. The energy of the system at a specified configuration.

In local methods for finding a minimum effort is directed toward a nearby local minimum, and steps that may increase the function are not permitted.

Descent structure of local minimization algorithms:
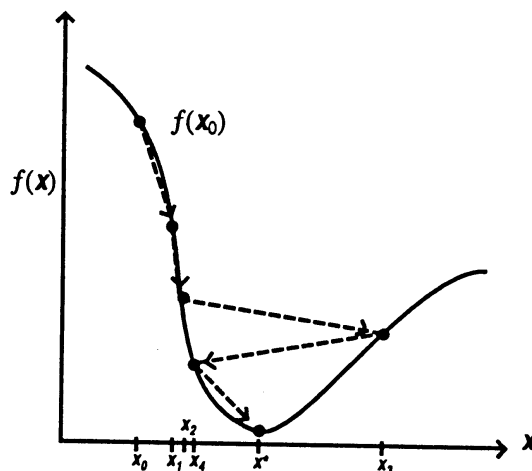


A comparison of the steepest descents and conjugate gradients methods for initial refinement and a stringent minimisation.

| Method | Initial refinement (Ave. gradient $< 1$ kcal $Å^{-2}$) | | Stringent minimisation (Ave. gradient $< 0.1$ kcal $Å^{-2}$) | |
|---|---|---|---|---|
| | Cpu time (s) | Number of iterations | Cpu time (s) | Number of iterations |
| Steepest descents | 67 | 98 | 1405 | 1893 |
| Conjugate gradients | 149 | 213 | 257 | 367 |

**The local minima problem**
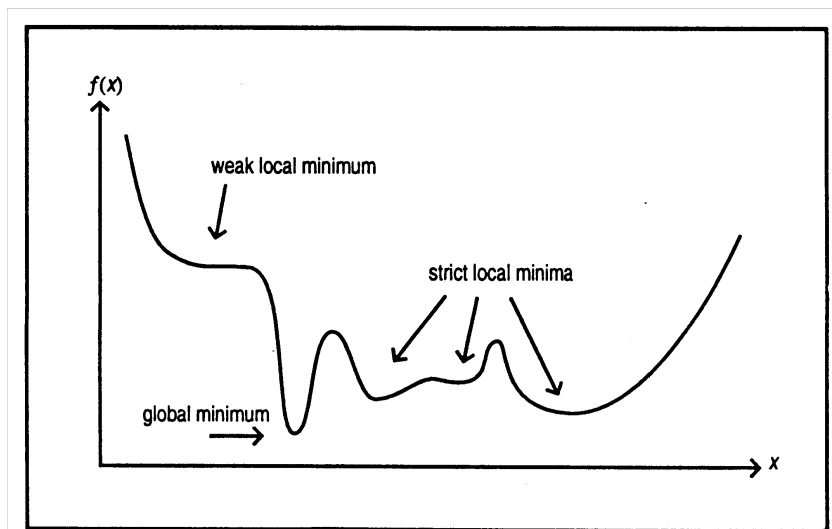


**The search for the global minimum**[1]

Generally, we are most interested in the preferred (most stable) conformation of our molecules and the conformers which are s ignificantly populated at r oom temperature (within approx. 3 kcal/mol of the global minimum). A number of conformational search techniques have been developed to answer these questions.

Structures of local and global minimization algorithms:



---
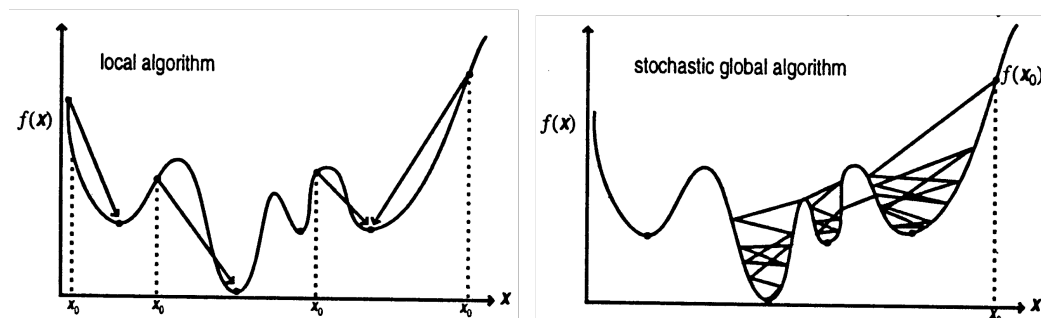
[1]    Saunders, M. et al. *J. Am. Chem. Soc.* **1990**, *112*, 1419

→ - **G r id search**; For very simple structures, this can be do ne manually. (MM2/3: <u>angle driver</u> to create reasonable starting geometries for conformational searching).
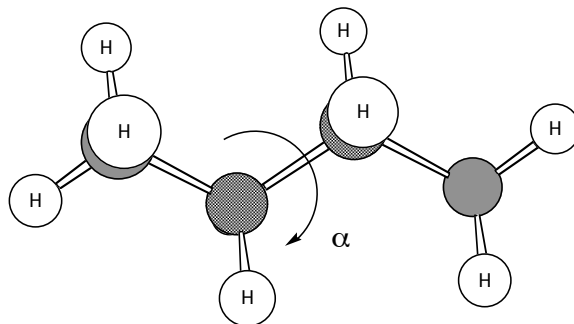


T h is <u>systematic</u> procedure provides guaranteed coverage of all regions of space if a sufficient number of starting points is selected (at least 3 per dihedral angle).

P r o b l ems: an extraordinary large number of starting structures is generated, which renders the problem very computer-time intensive, especially for flexible structures. For example, for polypeptides of $n$ residues, a rough range of $10^n$ to $25^n$ reasonable starting points by coarse subdomain partition of backbone and sidechains have to be investigated.

I f the number of values of for each torsion is reduced, the resolution with which space is covered may be inadequate to provide a starting geometry in the vicinity of each minimum.

Newer molecular modeling packages offer several force fields. Choose wisely!

**MM2**: Organic molecules. Last release: 1987.[1]
**MM3**: Organic molecules. Last release: 1994 (The MM3* on Macromodel is the MM3/91 release and is limited to fewer functionalities).
**AMBER**: Proteins and nucleic acids, carbohydrates.[2]
**OPLS**: Peptides. The united-atom simulation treats certain atom groups as one atom, with one atom type.[3]
**BIO**+: (CHARMM) Biological macromolecules.[4]

---

1    Allinger, N. L. et al. *J. Am. Chem. Soc.* **1992**, *114*, 1.
2    Kollman, P. A. et al. *J. Comp. Chem.* **1986**, *7*, 230.
3    Jorgensen, W. L. et al. *J. Am. Chem. Soc.* **1991**, *113*, 2810.
4    Karplus, M. et al. *J. Comp. Chem.* **1983**, *4*, 187.

(1 a.u. = 627.510 kcal/mol)

| Force Field | Eq. (kcal/mol) | Ax. (kcal/mol) | Δ (kcal/mol) |
|---|---|---|---|
| MM3/92 | 9.10 | 10.87 | 1.77 |
| Chem3D (MM2) | 6.90 | 8.67 | 1.77 |
| CaCHe (MM2?) | 6.89 | 8.68 | 1.79 |
| MM2* | 6.89 | 8.67 | 1.78 |
| MM3* | 9.10 | 10.87 | 1.77 |
| Amber* | 1.84 | 2.87 | 1.03 |
| OPLS* | 0.78 | 2.57 | 1.79 |
| Sybyl (tripos) | 1.30 | 2.70 | 1.40 |
| Sybyl (Spartan) | 1.30 | 2.70 | 1.40 |
| MMFF (Spartan) | 0.70 | 2.07 | 1.37 |
| **Exp.** | | | **1.74** |
| MNDO | -36.19 | -35.16 | 1.03 |
| AM1 | -43.70 | -42.28 | 1.42 |
| PM3 | -36.87 | -35.75 | 1.12 |
| STO-3G  E(HF) | -270.063379 a.u. | -270.060444 a.u. | 1.84 |
| 6-31G*  E(HF) | -273.2436640 a.u. | -273.2399991 a.u. | 2.30 |
| 6-31+G* E(HF) | -273.245892 a.u. | -273.242122 a.u. | 2.37 |
| 6-311+G** E(HF) | -273.308789 a.u. | -273.305906 a.u. | 1.81 |
| DFT/LSDA/VWN | -272.6182101 a.u. | -272.6149537 a.u. | 2.04 |

The *ab initio* calculation at the 6-311+G** E(HF) level required 1,800 x more time than  MM2.

## C. Semiempirical Methods

Molecular mechanics methods are based on classical concepts and require computer time roughly proportional to the square of the number of atoms. In contrast, semiempirical methods at the Hartree-Fock level use a combination of quantum chemical models and experimentally determined parameters to strive for accuracy and scale up as $(4N)^3$. Finally, *ab initio* quantum mechanics proceeds as $(10N)^4$ (for glucose: 1:1,500:6,000,000).

There are important differences between classical molecular mechanics and quantum mechanical methods. MM methods are extremely fast and able to handle very large systems. Particularly for hydrocarbons, they are also as accurate as the best *ab initio* methods. But they are only parametrized for ground state systems and common bonding situations (functional groups). MM methods are unable to anticipate the making and breaking of most bonds, the electronic properties of molecules, and the chemistry of electronically excited states.

Use molecular orbital methods to compute:

- bond orders
- dipole moments
- ionization potentials
- vibrational frequencies and IR, UV spectra
- MO energies
- partial charges
- potential energy maps
- reaction pathways
- transition states
- heats of formation
- mapping bond breaking and bond formation

The transition between semiempirical and *ab initio* quantum chemistry is not always obvious. Basis sets, for example, are empirical in nature, as are effective core potentials.
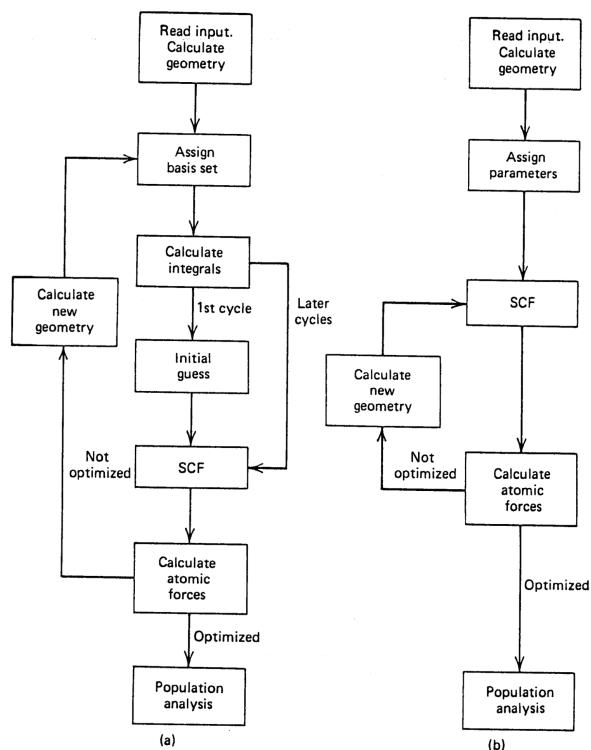


Typical flow charts for (a) an *ab initio* optimization, and (b) a corresponding semi-empirical calculation.

Approximate solutions of the Schrödinger equation:

$$\mathcal{H}\Psi = E\Psi_1$$

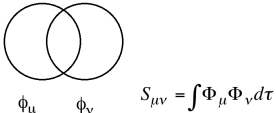The MO-LCAO approximation:

$$\Psi_i = \sum_v C_{vi}\Phi_v$$

The overlap between two atomic orbitals $\phi_\mu$ and $\phi_v$ defines the elements of a matrix S:

The Hartree-Fock method introduces a on e-electron Hamiltonian, $\mathcal{F}$, which is the sum of the kinetic energy operator for an electron, a potential that a single electron would feel coming from the fixed nuclei, and an average of the effects of all the other N-1 electrons: $\mathcal{F}\Psi_i = E\Psi_i$.

The Fock equations can be most easily be solved by solving the matrix equations: FC = SCE.

The following steps are then common to all molecular orbital programs:

1. Calculate the integrals needed to form the Fock matrix F. Guess $\Psi_i$'s.
2. Calculate the overlap matrix S.
3. Diagonalize S.
4. Diagonalize F for the MO eigenvalues E.
4. Solve the matrix equation to obtain the MO coefficients C.

$$\phi_\mu \qquad \phi_v \qquad S_{\mu v} = \int \Phi_\mu \Phi_v d\tau$$

---

[1]   "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble". Dirac, P. A. M. *Proc. Roy. Soc. (London)* **1929**, *123*, 714.

Semiempirical methods:

The neglect of differential overlap approximation says that $S_{\mu v} \ll 1$ in many situations and that the two electron integrals can be ignored or approximated in various ways (one-center terms). A second basic approximation is the neglect of diatomic differential overlap: $\Phi_\mu^A \Phi_v^B = \Phi_\mu^A \Phi_v^A \delta_{AB}$ (two-center terms).

The greatest simplification is achieved with complete neglect of differential overlap (**CNDO**), f. ex. CNDO/S.[1] Because INDO calculations take very little more time than CNDO calculations, they are re placing the CNDO models.

The intermediate neglect of differential overlap (**INDO**) schemes keep all of the one-center terms above: INDO/S[1] (spectroscopy of transition metal complexes; it was used on the reaction center of photosynthetic bacteria).
MINDO/3 was de veloped in 1975 to reproduce such diverse experimental properties as molecular geometries, heats of formation, dipole moments, and ionization potentials. All quantities in the Fock matrix and the energy expression were treated as free parameters, and the resonance integral was chosen to be a pair parameter rather than the average of two atomic parameters. H, B, C, N, O, F, Si, P, S, and Cl are parametrized.

---

[1]   Mataga, N.; Ishimoto, K. *Z. Physik. Chem. (Frankfurt)* **1957**, *13*, 140.

MNDO,[1] AM1,[2] and PM3[3] are all NDDO techniques developed by Dewar and his coworkers and use essentially all the same model functions. MNDO has been parametrized for H, Li, B, C, N, O, F, Al, Si, P, S, Cl, Zn, Ge, Br, Sn, I, Hg, and Pb. It uses monoatomic parameters and is computationally quite efficient but failed for biologically interesting molecules. AM1 has been parametrized for H, B, C, N, O, F, Al, Si, P, S, Cl, Zn, Ge, Br, I, and Hg, and PM3 for H, Be, C, N, O, F, Mg, Al, Si, P, S, Cl, Zn, Ga, Ge, As, Se, Br, Cd, In, Sn, Sb, Te, I, Hg, Tl, Pb, and Bi (the third- and fourth-row main-group elements have only recently been added). PM3 and AM1 use a different equation for the core-core repulsion term (this is a problem for AM1 with P, Al, and N). PM3 differs from AM1 in that the former treats the one-center, two-electron integrals as pure parameters, as opposed to being derived from atomic spectroscopy. Spartan's PM3(tm) model is parametrized for Ti, Zr, Hf, Ta, Cr, Mo, W, Co, Rh, Ni, and Pd, used in conjunction with PM3 for non-transition metals.

AM1:
- seems to be most accurate for calculations of dipole moments for elements that it has been parametrized for.
- reproduces hydrogen-bonds slightly better and
- gives energies that are still too large for activation barriers.
- reproduces ground-state geometries quite accurately.
- is poor in reproducing lon-pair lon-pair repulsion.
- has an average error of heats of formation of C, H, N, O compounds of ca. 40%.

---

1  Dewar, M. J. S. et al. *J. Comp. Chem.* **1981**, *2*, 433.
2  Austin Model 1; Dewar, M. J. S. et al. *J. Comp. Chem.* **1981**, *2*, 433.
3  Parametric Method Number 3. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221. This represents the third parametrization of MNDO (AM1 being the second).

Characterizations of the wave function

In a semi-empirical calculation, the energy, the first derivatives of the energy and the wave function for the molecule are computed. Other molecular properties are basically an average over the wave function of certain operators describing the properties.
Among the visualizable properties that are generally computed are:

Total electron density represents the probability of an electron being at a particular point in space.

Spin density defines the excess probability of finding spin-up over spin-down electrons at a point in space. The spin density at the position of a nucleus is the a prime determinant of ESR spectra.

Orbital plots are used to visualize molecular orbitals, especially frontier orbitals, for the analysis of organic reactions.

Electrostatic potentials can be used to predict initial attack positions of ions during a reaction. The potential of a positive probe charge is calculated as a function of the distance to the molecule, the nuclear charge, and the electronic density.

Atomic partial charges give a crude idea for the distribution of charges in a molecule.

IR, UV/VIS absorption spectra.

Kondru, R. K.; Wipf, P.; Beratan, D. N., "Theory assisted determination of absolute stereochemistry for complex natural products via computation of molar rotation angles." *J. Am. Chem. Soc.* **1998**, *120*, 2204-2205.

Wipf, P.; Kerekes, A. D., "Structure reassignment of the fungal metabolite TAEMC161 as the phytotoxin viridiol." *Journal of Natural Products* **2003**, *66*, 716-718.