Glossary of Computer and Molecular Modeling Terms (Ga – Ho)

outtake from Molecular Modeling on the PC

gauche - a stereochemical term for the relationship between two **vicinal** substituents in a conformation characterized by a dihedral angle of approximately 60°. It is generally used to refer to the two largest or most important substituents attached to the two central atoms, and is generally a local energy minimum in **conformation space**. Gauche is a special case of **staggered**.

The term comes from French, literally meaning 'left' or 'left-hand' and figuratively meaning 'clumsy, awkward'.

The **gauche effect** is the observation in the area of conformational analysis that in a molecule possessing two electronegative groups in a vicinal relationship, the **gauche** rotamer will be more stable that the **anti** rotamer, despite the increase in non-bonded repulsions. A more specific example of the gauche effect is the **anomeric effect**.

Early explanations held this effect to be a result of the balance of attractive and repulsive forces between the substituents. A more recent explanation for this effect is that the C-C bond is subject to bending by each of the electronegative substituents, and this bending causes more of a net weakening of the C-C bond when the two electronegative substituents are **anti** than when they are gauche.

Gaussian - denotes an *ab initio* quantum chemical molecular orbital calculation package which is available in many forms and versions for various hardware platforms. It is the most widely used program of its kind. It utilizes a Gaussian-type orbital (GTO) basis set, and interfaces to many other software applications are available. A good description of Gaussian is available in Clark's monograph on computational chemistry.

GB/SA an acronym for generalized Born equation/surface area, *GB/SA* is a continuum treatment of solvation effects for use ion empirical force field modeling, developed by Still and coworkers. The solvation free energy is calculated as the sum of a solvent-solvent cavity term and a solute-solvent van der Waals term, which are calculated from the molecule's surface area, and a solute-solvent electrostatic polarization term, which is calculated from the Born solvation equation and a Coulomb electrostatic component. The *GB/SA* treatment is discussed in Section 5.6.

GEGOP - an acronym for GEometry of GlycOProteins, a program which carries out a conformational analysis of a glycoprotein, within a specified force field, by simultaneous variance of the torsional angles.

geometric isomer - one of a group of two or more stereoisomers cannot be interconverted because of hindered bond rotation. For example, this type of isomerism produces the **E** or **Z** (***trans*** or ***cis***) isomers about a double bond.

GERM - an acronym for Genetically Evolved Receptor Models, a program for generating a 3D description of a hypothetical receptor derived from structure activity relationships, developed by Walters and Hinds. The models are evolved by application of a **genetic algorithm**, and may be used to predict bioactivity for structures not used in the formulation of the receptor hypothesis.

GIF - an acronym for Graphics Interchange Format. A standard graphics format defined in 1987 by CompuServe Inc. for digitized images, converted for transmission on the Internet with the Unisys, Inc., LZW compression algorithm. GIF is currently an open (i.e. not proprietary) format.

global minimum - that **conformation** of a molecule which has the lowest relative potential energy. In some cases of conformational analysis, a local minimum may be confused with the global minimum. There is no method to determine whether a discovered conformational energy minimum is the global minimum except a rigorous, exhaustive search of **conformation space**. A number of searching techniques can be used to arrive at the global minimum with varying degrees of confidence, and the level of confidence decreases with increased complexity of the molecule.

GLYCAM is the name of a additional parameter set for the **AMBER** force field, developed by Woods et al. at Oxford University, to enable the treatment of oligosaccharides and glycoproteins. The **GLYCAM** parameter set is discussed in Section 5.4.3, and the parameters may be downloaded from the AMBER Web site.

glycosidic linkage - a term from carbohydrate chemistry, also referred to as the glycosidic bond. When two sugar residues are joined by a bridging oxygen between the anomeric carbon of one residue and a non-anomeric carbon of the other, the **C-O-C** unit is called a glycosidic linkage.

glycosyl acceptor - a term from synthetic carbohydrate chemistry; in the case of a reaction in which two sugar residues are being joined through the formation of a glycosidic bond, the glycosyl acceptor is that sugar residue in which a carbon other than the anomeric center forms part of that linkage.

glycosyl donor - a term from synthetic carbohydrate chemistry; in the case of a reaction in which two sugar residues are being joined through the formation of a glycosidic bond, the glycosyl donor is that sugar residue in which the anomeric carbon forms part of the linkage.

GOPT - an acronym for Graphics Optimization Tool, this is a graphics front end for the Lyngby version of the **CFF** modeling program, developed by Rasmussen and coworkers.

graphic user interface - a term for a type of computer operating environment which is characterized by the use of pictorial or visual screen objects to represent the input and output of a program, rather than only text on a command line. It is also known by the abbreviation **GUI** .A discussion of the GUI may be found in a recent review article.

A program with a GUI generally runs under a windowing system such as the Apple Macintosh™ operating system, or Microsoft Windows™. In this environment, the user controls the program by using a pointing device (such as a mouse, trackball, keyboard-driven cursor, etc.) to indicate objects (such as icons, buttons, dialog boxes, etc.) in the screen windows by directing the pointing device cursor at them and pressing a button to select it. Apple Computers™ popularized the GUI with their Macintosh™ operating system, but the concept originated in the early 1970's at Xerox's PARC laboratory.

graphics - specifically **computer graphics**, images produced on a monitor on a computer equipped with the proper hardware and software. The hardware requirements are a monitor and a video display card, and the software must incorporate a driver which allows the program to address the graphics hardware.

Earlier generation graphics were of the **vector graphics** type, where the image was produced as a series of straight or curved lines, polygons, etc., and groups of such objects. The lined features are mapped from beginning point to end point within the x,y-coordinate system corresponding to the screen (whence the name), and the images are continuously refreshed. An advantage of vector graphics is that any element of the picture (the features mentioned above) may be changed at any time, since each element is stored as an independent object.

The modern standard is the **raster graphics** type, in which an image is composed of an array of pixels arranged in rows and columns. The term comes from the Latin *rastrum*, meaning rake, from the motion of the cathode ray scanning the screen elements. In this case, a continuous image is produced from a pattern of closely spaced small square or rectangular elements (monochrome), or small triads of red/blue/green elements (color). These elements are called **pixels**, and higher numbers of rows and columns of pixels leads to higher resolution graphics. Important features besides the resolution (usually quantified by the column and row dimensions such as 640 x 480, 800 x 600, 1024 x 768, etc.), are the refresh rate, and in color graphics the number of colors (e.g. 8, 16, 256, etc.). The area of the screen is scanned from side to side in lines from top to bottom, and certain of the elements are activated which forms the graphical display.

In **bitmap graphics**, a type of **raster graphics**, each pixel in the display is mapped literally bit for bit from data in the image file. Strictly speaking, a bit for bit file to image translation is limited to a black and white graphic, since a color image requires more that one bit for each pixel, but the term bitmap is generally used for all images. Such a display can be updated extremely rapidly since changing a pixel involves only a single processor write to memory compared with a terminal connected via a serial line where the speed of the serial line limits the speed at which the display can be changed.

Most modern personal computers and workstations have bitmap graphics, allowing the efficient use of graphical user interfaces, interactive graphics and a choice of on-screen fonts. Because of the large amount of information required in such a specification, bitmapped graphics files tend to be relatively large in size. Some more expensive systems still delegate graphics operations to dedicated hardware such as graphic accelerators.

Some graphics displays have elements of both vector and raster format.

Graphics Optimization Tool - see **GOPT**.

grid search - a search conducted systematically over all available coordinates or values, an antonym for **stochastic** and **Monte Carlo**.

The **grid search method** involves the systematic selection or generation of input data or coefficient values, or systematic sampling techniques to obtain solutions to problems which are exact or nearly so. In a conformational analysis context, the grid search method may be used to find the minimum energy conformation as a function of all rotatable bonds. For "n" rotatable bonds, and after selecting a desired precision, this results in an "n"-dimensional grid, each point of which will be sampled. The grid search method is also referred to as the **brute force method**, because it painstakingly examines each possible value before arriving at an answer.

GSTAT - a multi-functional geometry program with the **Cambridge Structural Database Software** (**CSDS**). It will perform geometric calculation for complete CSD entries, location of 3D substructure fragments within the crystallographic connectivity record of the CSD, geometric calculations for substructure fragments, statistical and numerical analysis of fragment geometry. Most of the functionality of GSTAT is now available within a combination of the other CSDS components **QUEST3D** and **VISTA**.

GUI - an abbreviation for **Graphic User Interface**.

Hammond postulate - the observation (by the American chemist G. S. Hammond) that the structure of a transition state can be approximated best by the structure of that ground state species which is nearest in energy. More specifically, the transition state for an exothermic reaction will look more like the reactant(s), and the transition state for an endothermic reaction will look more like the product(s).

Applying the Hammond postulate aids the modeling of transition states by perturbation theory by showing whether the reactant or product structure is the best starting point for applying perturbation, based on a knowledge of the energetics of the reaction.

haptophilic - describes a molecule, or a functional group within a molecule, which may act as a donor in forming a **coordination bond** with a metal atom partner. A haptophilic group usually consists or one or more atoms which bear lone electron pairs, or two or more atoms which have π-orbital system. The word comes from Greek 'απτω' (to fasten) and 'φιλος' (loving) and was coined by H. W. Thompson.

Haptophilicity is the quality of being haptophilic. **Monohapto**, **dihapto**, **trihapto**, etc., refer to ligands which will interact with the metal center via one, two, or three atoms, resp.

hard acid - an acid with small acceptor atoms, high positive charge, and do not contain unshared lone pairs of electrons in their valence shells. A hard acid has low polarizability and high electronegativity, and will react preferentially with a **hard base**. In hard acids, the acceptor atom is of high positive charge and small size and does not have easily excited outer electrons. See **HSAB principle**.

hard base - a base with donor atoms of high electronegativity and low polarizability. A hard base is difficult to oxidize, and holds its valence electrons tightly, and will react preferentially with a **hard acid**. In hard bases, the donor atom is of low polarizability and high electronegativity, is hard to oxidize, and is associated with empty orbitals of high energy and hence inaccessible. See **HSAB principle**.

hardness - in the molecular modeling context, the hardness is equivalent to a force constant for the van der Waals nonbonded interaction potential. This parameter is symbolized by $\boldsymbol{\varepsilon}$ in the nonbonding potential of many force fields, and its use is described in Section 5.1.3.1. The hardness characterizes how easily the corresponding atom can be deformed as it approaches other atoms. The hardness can be understood in mathematical terms as the steepness of the potential curve for the van der Waals repulsion. See **van der Waals**.

For hardness in the physical organic context, see **HSAB principle**.

Hard Sphere, *exo*-Anomeric method - see **HSEA method**.

Hartree-Fock - an approach to molecular orbital (MO) calculations in which the energy of a set of molecular orbitals can be derived from the basis set functions used to define each orbital, and a set of adjustable coefficients which are used to minimize the total energy of the system. Thus, in Hartree's original method the wave function for a number of electrons 'N' was found as the product of 'N' one-electron wave functions. Fock modified the method to use Slater determinants (or matrices) to combine the one-electron wave functions instead of using their products. This method took account of most of the effects of electron spin. Electron-electron repulsions are addressed by considering the effect of the averaged Coulombic repulsions of the other electrons on each electron in turn.

Essentially, the **Hartree-Fock method** employs antisymmetric Slater determinants for combining the' one-electron wave functions, and through an iterative process the coefficients of the orbitals are modified from cycle to cycle until the electronic energy achieves a constant minimum value and the orbitals no longer change, i.e. a **self-consistent field** (**SCF**) has been reached.

The **restricted Hartree-Fock** (**RHF**) calculation is used for a closed shell system. It carries the restriction that there be an even number of electrons, and that they all be spin-paired. The restriction is that every orbital is either doubly occupied or unoccupied. The RHF calculation is appropriate for neutral systems, anions or cations.

The **unrestricted Hartree-Fock** (**UHF**) calculation is for an open shell system, where at least one orbital will be singly occupied. It calculates two sets of molecular orbitals, one for each type of spin, named 'alpha' and 'beta.' It thus considers each doubly occupied MO to consist of two spin-paired singly occupied MO's. The energies of these latter are not necessarily equivalent, as would be required in the RHF formulation. The UHF calculation accounts for the perturbation effect of an unpaired spin upon the formally paired spins, and generates more realistic spin densities.

The UHF calculation is appropriate for radicals or diradicals. A radical (doublet) will have one more 'alpha-type' electron than 'beta-type,' and a diradical (triplet) will have an excess of two 'alpha-type' electrons. The UHF calculation causes the electronic energy of the system to be lower does the RHF calculation for the same system, so that the results from the two types of calculation cannot be compared to each other.

heavy atom - a non-hydrogen atom. Any atom from the periodic table except hydrogen and its isotopes (e.g. deuterium).

Heisenberg Uncertainty - the principle or relation that the position and the velocity of a subatomic particle cannot both be measured exactly, at the same time, because the act of measuring one alters the other. The statement of this principle was made in 1927 by the German physicist W. Heisenberg.

Hessian - is the matrix of second partial derivatives of a function (assumed to be twice differentiable). This is often denoted $\mathbf{H}_\mathbf{F}(\boldsymbol{x})$, where **F** is the function and **x** is a point in its domain. The Hessian determinant is named for the 19th cent. German mathematician L. O. Hesse, who introduced it in 1842.

In molecular modeling, for a molecule of N atoms, the N x N Hessian matrix will contain the force constants which govern the interaction of each atom with each other atom.

heteroaromatic - a molecule which is both **aromatic**, and in which one or more of the ring members is a **heteroatom**.

heteroatom - an atom other than carbon or hydrogen, or their isotopes.

heterocyclic - a cyclic molecule in which one or more of the ring members is a **heteroatom**.

heuristic - a logic term meaning a "rule of thumb," a simplification or an educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood. Unlike an **algorithm**, a heuristic does not guarantee a solution.

HEV - an abbreviation for Hydrogen-Extension Vector, a parameter used in the **YETI** force field to represent the linearity of the hydrogen bond. The HEV will originate at the hydrogen bond donor, and its end point marks the ideal position for a hydrogen bond acceptor atom relative to that hydrogen bond donor fragment. It is one of the descriptors used in building pseudoreceptors in the **YAK** program. See also **YAK** and **LPV**.

HMO - an abbreviation for Hückel Molecular Orbital, see **Hückel calculation**.

HOMO - an acronym for Highest Occupied Molecular Orbital.

Hondo - an *ab initio* HFMO (Hückel Frontier Molecular Orbital) calculations package developed by M. Dupuis, J. D. Watts, H. O. Villar, and G.J.B. Hurst of IBM Corp. The program uses Gaussian-type basis functions up to 'd' type, and can handle up to 50 atoms and 120 shells for a total of 440 unique Gaussian exponential functions. The code is available as QCPE #544. A brief account of some of its calculational capabilities includes:

Wavefunctions: single-configuration self-consistent-field (closed-shell RHF, spin-unrestricted UHF, restricted open-shell ROHF), generalized valence bond and general multiconfiguration self-consistent-field, and configuration interaction, with full use of molecular symmetry.

Calculation of the electron correlation correction to the energy of closed-shell RHF wavefunctions; use of the effective core potential approximation; optimization of molecular geometries using the gradient of the energy with respect to nuclear coordinates for all SCF wavefunctions; calculation of the force-constant matrix in the Cartesian or internal coordinates space and of the vibrational spectrum (including infrared and Raman intensities) for all SCF wavefunctions; calculation of the dipole moment and polarizability with respect to the nuclear coordinates; determination of transition-state structures by taking advantage of the energy gradients for all SCF wavefunctions.