Pure & Appl. Chem., Vol. 69, No. 5, pp. 1137–1152, 1997.
Printed in Great Britain.

INTERNATIONAL UNION OF
PURE AND APPLIED CHEMISTRY

CHEMISTRY AND HUMAN HEALTH DIVISION
MEDICINAL CHEMISTRY SECTION

GLOSSARY OF TERMS USED IN COMPUTATIONAL DRUG DESIGN

(IUPAC recommendations 1997)

Prepared for publication by

H. VAN DE WATERBEEMD[1] (CHAIRMAN)
R. E. CARTER[2], G. GRASSY[3], H. KUBINYI[4], Y. C. MARTIN[5],
M. S. TUTE[6], P. WILLETT[7]

[1]F. Hoffmann-La Roche Ltd., Pharma Research New Technologies, CH-4070 Basel, Switzerland
[2]Astra Hässle AB, Computational Chemistry, S-43183 Mölndal, Sweden
[3]Centre de Biochimie Structurale, Faculté de Pharmacie, F-34060 Montpellier, France
[4]BASF AG, ZHB/W A30, D-67056 Ludwigshafen, Germany
[5]Abbott Laboratories, Computer-Assisted Molecular Design, Abbott Park IL 60064-3500, USA
[6]University of Canterbury, Kent CT2 7NH, UK
[7]University of Sheffield, Department of Information Studies, Sheffield S10 2TN, UK

Membership of the Section during the period (1992–1995) when this report was prepared was as follows:

President: J. G. Topliss (USA); *Vice-president:* N. Koga (Japan); *Past-president:* C. G. Wermuth (France); *Secretary:* W. D. Busse (FRG); *Titular members:* C. R. Ganellin (UK); L. A. Mitscher (USA); *Co-opted members:* P. Anderson (USA); P. R. Andrews (Australia); W. A. Denny (New Zealand); W. Granik (Russia); Y. Guindon (Canada); C. A. G. Haasnoot (The Netherlands); J. Ide (Japan); R. Imhof (Switzerland); P. Lindberg (Sweden); G. Tarzia (Italy); R. S. Xu (China); *National representatives:* O. A. M. Stoppani (Argentina); E. J. Barreiro (Brazil); A. Again (Bulgaria); J. Krepelka (Czechoslovakia); E. K. Pohjala (Finland); A. Monge Vega (Spain).

CONTRIBUTORS

C.A.G. Haasnoot[8], L.B. Kier[9], K. Müller[1], S.V. Rose[10], J. Weber[11], K.S. Wibley[12], S. Wold[13]

[8] *Diosynth BV, PO box 20, NL-5340 BH Oss, The Netherlands*

[9] *Virginia Commonwealth University, Department of Medicinal Chemistry, Richmond, USA*

[10] *BioFocus Molecular, Central Avenue, Chatham Maritime, Chatham ME4 4TB, UK*

[11] *Université de Genève, Departement de Chimie Physique, CH-1211 Genève 4, Switzerland*

[12] *University College London, Department of Chemistry, London WC1H 0AJ, UK*

[13] *University of Umea, Department of Organic Chemistry, Research Group for Chemometrics, S-90187 Umea, Sweden*

REVIEWERS

D.B. Boyd[14], D.E. Clark[15], Chr. de Haën[16], N.D. Heindel[17], P. Kratochvíl[18], B. Kutscher[19], R.A. Lewis[20], M. Mabilia[21], W.V. Metanomski[22], E.E. Polymeropoulos[19], J.P. Tollenaere[23], M.D. Turnbull[24], W.E. van der Linden[25], E.J. Van Lenten[26]

[14] *Indiana University-Purdue University at Indianapolis, Department of Chemistry, 402 North Blackford Street, Indianapolis, Indiana 46202-3274, USA.*

[15] *Proteus Molecular Design Ltd., Proteus House, Lyme Green Business Park, Macclesfield, Cheshire SK11 0JL, UK.*

[16] *Bracco spa, Via Egidio Folli 50, I-20134 Milano, Italy.*

[17] *Lehigh University, Department of Chemistry, Bethlehem, Pennsylvania 18015-3172, USA.*

[18] *Academy of Sciences of the Czech Republic, Institute of Macromolecular Chemistry, CZ-16206 Prague 6, Czech Republic.*

[19] *Asta Medica AG, Weismüllerstrasse 45, D-60314 Frankfurt am Main, Germany*

[20] *Rhône-Poulenc Rorer, Dagenham Research Centre, Rainham Road South, Dagenham, Essex RM10 7XS, UK*

[21] *Via Salvemini 9, I-36100 Vicenza, Italy*

[22] *CAS, PO Box 3012, Columbus, Ohio 4310-0012, USA*

[23] *Janssen Research Foundation, Turnhoutseweg 30, B-2340 Beerse, Belgium*

[24] *Zeneca Agrochemicals, Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EZ, UK*

[25] *University of Twente, The Netherlands*

[26] *American Chemical Society Committe on Nomenclature, National Library of Medicine, Index Section Biobliographic Services, 1105 Cedrus Way, Rockville, MD 20854, USA*

Glossary of terms in computational drug design (IUPAC Recommendations 1997)

Abstract: Computational drug design is a rapidly growing field which is now a very importar component in the discipline of medicinal chemistry. At the same time many medicinal chemist lack significant formal training in this field and may not have a clear understanding of some of the terminology used but need to grasp concepts, follow research results, define problems for, anc utilize findings of, computational drug design.

In this context the IUPAC Medicinal Chemistry Section Committee felt it would be useful to develop a glossary of terms used in computational drug design for easy reference purposes. Also there is the possibility that in different countries certain terms may not have the same meaning and in such a case there would be value in trying to establish an international definition standard. Accordingly a Working Party of seven experts in the field was assembled who constructed a glossary of some 100 terms. Concise but sufficiently explanatory definitions have been formulated based on a variety of literature sources and selected key references provided.

ALPHABETICAL ENTRIES

Some of the definitions also appear in the Glossary of Terms used in Medicinal Chemistry (IUPAC recommendations 1996; © *1996 IUPAC*). These are marked with an asterisk.

For some definitions the more extended form taken from the Glossary of Terms in Theoretical Organic Chemistry (IUPAC recommendations 199*; © *199* IUPAC*) is included in smaller font.

Ab initio calculations

Ab initio calculations are quantum chemical calculations using exact equations with no approximations which involve the whole electronic population of the molecule.

> ***Ab initio* quantum mechanical methods** (Synonymous with non-empirical quantum mechanical methods) - Methods of quantum mechanical calculations independent of any experiment other than the determination of fundamental constants. The methods are based on the use of the full Schrödinger equation to treat all the electrons of a chemical system. In practice, approximations are necessary to restrict the complexity of the electronic wavefunction and to make its calculation possible.

AM1 calculations

AM1 calculations are semi-empirical molecular orbital calculations developed at the University of Austin in Texas (*AM1* = Austin Model 1). These calculations involve the valence electrons of the atoms of the molecule. They are a further development of MNDO calculations (Wylie, 1994).

(see MNDO calculations)

AMBER

AMBER is a well-known molecular mechanics program for calculations on proteins and nucleic acids.

(see Molecular mechanics)

Artificial neural networks

Artificial neural networks (*ANN*) are algorithms simulating the functioning of human neurons and may be used for pattern recognition problems, e.g., to establish quantitative structure-activity relationships.

Atomic orbitals (AO)

Atomic orbitals are mathematical functions (e.g., Gaussian, or Slater functions) used in quantum chemical calculations. A set of atomic orbitals described by a defined function is the basis set of atomic orbitals.

(see Slater-type orbitals)

Orbital (Atomic or Molecular) - A wavefunction which depends explicitly on the spatial coordinates of only one electron.

Basis set

A basis set is a set of mathematical functions used in molecular orbital (*MO*) calculations, e.g., the *6-31G** basis set used in *ab initio* calculations. 6-31G* and similar expression refers to the type of mathematical function used.

(see Molecular orbital (MO) calculations)

Basis set - A set of basis functions employed for the representation of molecular orbitals. One may distinguish the minimal basis set (includes one basis function for each SCF (*SCF = Self- Consistent Field*) occupied atomic orbital with distinct principal and angular momentum quantum numbers); split valence basis set (includes two or more sizes of basis function for each valence orbital); double zeta (DZ) basis set (a split valence basis set that includes exactly twice as many functions as the minimal basis set; extended basis set (the set larger than the double zeta basis set); polarized basis set (incorporates basis functions of higher angular quantum number beyond what is required by the atom in its electron ground state; allows orbitals to change not only size, but also shape); basis set with diffuse functions and others.

Chemometrics

Chemometrics is the application of statistics to the analysis of chemical data (from organic, analytical or medicinal chemistry) and design of chemical experiments and simulations.

CLOGP values

CLOGP values are calculated 1-octanol/water partition coefficients, frequently used in structure-property correlation or quantitative structure-activity relationship (*SPC/QSAR*) studies (Leo, 1993)

(see Structure-property correlations (SPC) and Quantitative structure-activity relationships (QSAR))

Cluster analysis

Cluster analysis is the clustering, or grouping, of large data sets (e.g., chemical and/or pharmacological data sets) on the basis of similarity criteria for appropriately scaled variables that represent the data of interest. Similarity criteria (distance based, associative, correlative, probabilistic) among the several clusters facilitate the recognition of patterns and reveal otherwise hidden structures (Rouvray, 1990; Willett, 1987, 1991).

CNDO/2 calculations

CNDO/2 calculations are semi-empirical molecular orbital (MO) calculations using complete neglect of differential overlap.

(*see Molecular orbital (MO) calculations*)

*Comparative molecular field analysis (CoMFA)**

Comparative Molecular Field Analysis (CoMFA) is a 3D-QSAR method that uses statistical correlation techniques for the analysis of the quantitative relationship between the biological activity of a set of compounds with a specified alignment, and their three-dimensional electronic and steric properties. Other

properties, such as hydrophobicity and H-bonding can also be incorporated into the analysis (Cramer *et al.*, 1988; Kubinyi, 1993b).

(see 3D-QSAR, Hydrophobicity)

*Computational chemistry**

Computational chemistry is a discipline using mathematical methods for the calculation of molecular properties or for the simulation of molecular behaviour. It also includes, e.g., synthesis planning, database searching, combinatorial library manipulation (Hopfinger, 1981; Ugi *et al.*, 1990).

*Computer-assisted drug design (CADD)**

Computer-assisted drug design involves all computer-assisted techniques used to discover, design and optimize biologically active compounds with a putative use as drugs.

(*see Drug design*)

Computer-assisted molecular design (CAMD)

Computer-assisted molecular design involves all computer-assisted techniques used to discover, design and optimize compounds with desired structure and properties.

Computer-assisted molecular modeling (CAMM)

Computer-assisted molecular modeling is the investigation of molecular structures and properties using computational chemistry and graphical visualization techniques.

Computer chemistry

Computer chemistry is often used as equivalent to computational chemistry, and can also refer to the use of computers in synthesis planning (Ugi *et al.*, 1990; Boyd, 1990).

Conformational analysis

Conformational analysis consists of the exploration of energetically favorable spatial arrangements (shapes) of a molecule (conformations) using molecular mechanics, molecular dynamics, quantum chemical calculations or analysis of experimentally-determined structural data, e.g., NMR or crystal structures.

Molecular mechanics and quantum chemical methods are employed to compute conformational energies, whereas systematic and random searches, Monte Carlo, molecular dynamics, and distance geometry are methods (often combined with energy minimization procedures) used to explore the conformational space.

(see Distance geometry, Molecular dynamics, Molecular Mechanics, Monte Carlo technique, Quantum chemical methods)

Conformationally flexible searching (CFS)

Conformationally flexible searching is a three dimensional-structure database search taking into account the flexibility of molecules.

Connolly surface

The Connolly surface is the envelope traced out by the point of contact of a defined probe (e.g., a sphere) and a molecule of interest where they touch once, plus the van der Waals surface of the probe where it touches twice or more (the re-entrant surface), It is used to visualize the molecular surface.

Craig plot

A Craig plot is a plot of two substituent parameters (e.g., Hansch-Fujita π and Hammett σ values) used in analog design.

CSSR

The CSSR (*Crystal Structure Search Retrieval*) file format is one of several used by the Cambridge Crystal Structure Database (*CSD*) to store molecular structures. This format is used in many molecular modeling software packages.

De novo design *

De novo design is the design of bioactive compounds by the incremental construction of a ligand model within the receptor or enzyme active site, the structure of which is known from X-ray or nuclear magnetic resonance (*NMR*) data.

Discriminant analysis

Discriminant analysis is a statistical technique to find a set of descriptors which can be used to detect and rationalize separation between activity classes.

Distance geometry

Distance geometry is a mathematical method used to build three-dimensional (*3D*) molecular models from a set of approximate interatomic distances (e.g., nuclear Overhauser effect (*NOE*) experiments in nuclear magnetic resonance (*NMR*) suggest only ranges of distances). Distance geometry can be used to define a 3D pharmacophore starting from a set of molecules with the same mechanism of action, or for the generation of likely geometries for drug-receptor complexes using intermolecular distance constraints. (Crippen, 1988).

Docking studies

Docking studies are computational techniques for the exploration of the possible binding modes of a substrate to a given receptor, enzyme or other binding site.

D-optimal design

D-optimal design is an experimental design technique based on the optimization of the determinant calculated from the variance-covariance matrix of the descriptors. It is used to maximize the efficiency of fractional (uncomplete) factorial design.

(see Factorial design, Fractional factorial design)

3D-QSAR (three-dimensional quantitative structure-activity relationships) *

Three-dimensional quantitative structure-activity relationships (*3D-QSAR*) involves the analysis of the quantitative relationship between the biological activity of a set of compounds and their three-dimensional properties using statistical correlation methods.

Drug design

Drug design includes not only ligand design, but also pharmacokinetics and toxicity, which are mostly beyond the possibilities of structure- and/or computer-aided design. Nevertheless, appropriate chemometric tools, including experimental design and multivariate statistics, can be of value in the planning and evaluation of pharmacokinetic and toxicological experiments and results. Drug design is most often used instead of the correct term "Ligand Design".

Electrostatic field and potential

The electrostatic field and potential are properties of a molecule arising from the interaction between a charged probe, such as a positive unit point charge reflecting a proton, and a target molecule. These fields and potential are being used in three-dimensional quantitative structure-activity relationship (*3D-QSAR*) studies and to compare or assess the similarity of a set of molecules.

Electrostatic potential - A physical property equal in magnitude to the electrostatic energy between the static charge distribution, r(r), of an atomic or molecular system and a positive unit point charge located at r. The electrostatic potential V(r) that is produced at any point r by the electrons and nuclei (A) of the system is given by i.e. $V(r) = \Sigma Z_A/|R_A-r| - \int r(r')dr'/|r'-r|$.

Energy minimization

Energy minimization is a mathematical procedure to locate the stable conformations of a molecule (energy minima), as determined by molecular mechanics or quantum mechanical calculations.

(see Molecular mechanics, Quantum chemical calculations)

Experimental design

Experimental design is the use of mathematical and statistical methods to select the minimum number of experiments or compounds for optimal coverage of descriptor or variable space.

Extended Hückel (EH) calculations

Extended Hückel calculations are low-level semi-empirical molecular orbital (*MO*) calculations.

Extended Hückel method - A semi-empirical all-valence electron quantum mechanical method which uses the same approximations, apart from p-approximation and neglect of overlap integrals, as those of Hückel molecular orbital theory. The method reproduces relatively well the shapes and the order of energy levels of molecular orbitals. The account for overlap makes it possible to describe the net destabilization caused by interaction of two double occupied orbitals.

Extrathermodynamic approach

The extrathermodynamic approach involves the correlation between variables which, from a strictly thermodynamic standpoint, are not related. It is the basis of Hansch analysis used in traditional QSAR (Kubinyi, 1993a)

Factorial design (FD)

Factorial design is an experimental design technique in which each variable (factor or descriptor) is investigated at fixed levels. In a two-level *FD*, each variable can take two values, e.g., high and low lipophilicity.

File format

The (molecular) file format describes the layout of a computer data file. It is a set of instructions on how a molecule is encoded with respect to its connectivity, atom types, coordinates, and may also contain bibliographic data.

Force field

The force field is a set of functions and parametrization used in molecular mechanics calculations.

Force field - Within the molecular mechanics approach, a set of potential functions defining bond stretch, bond angle (both valence and dihedral) distortion energy of a molecule as compared with its nonstrained conformation (that characterized by standard values of bond lengths and angles). A set of transferable empirical force constants is preassigned and the harmonic approximation is usually employed. Some force fields may contain terms for interactions between non-bonded atoms, electrostatic, hydrogen bond and other structural effects as well as account for anharmonicity effects.

In vibrational spectroscopy, the inverse problem is solved of determining a set of force constants and other parameters of a choosen potential energy functions which would match with experimentally observed vibrational frequencies of a given series of congeneric molecules.

Fractional factorial design (FFD)

Fractional factorial design is an experimental design technique, using a reduction factor in order to limit the number of experiments to a lower number than obtained by factorial design.

Free energy perturbation calculations

Free energy perturbation calculations are mathematical procedures used in molecular dynamics studies to gradually convert one chemical species to another in a thermodynamic cycle.

Free-Wilson (FW) analysis

Free-Wilson analysis is a regression technique using the presence or absence of substituents or groups as the only molecular descriptors in correlations with biological activity (Kubinyi, 1993a).

Gaussian-type orbitals (GTO)

Gaussian-type orbitals are mathematical functions used in *ab initio* calculations. They have superceded Slater-type orbitals because of the greater computational efficiency that results.

(see Slater-type orbitals)

Genetic algorithm

A genetic algorithm is an optimization algorithm based on the mechanisms of Darwinian evolution which uses random mutation, crossover and selection procedures to breed better models or solutions from an originally random starting population or sample (Rogers and Hopfinger, 1994).

GOLPE

Generating optimal linear *PLS* estimations. It is an advanced variable selection technique in partial least squares (*PLS*) used in three-dimensional quantitative structure-activity relationships (*3D QSAR*) studies to handle very large data sets.

(see Partial least squares (PLS))

GRID

GRID is a program for receptor/ligand mapping. It calculates interaction energies between probes and target molecules at interaction points on a 3D grid (Goodford, 1985).

Hamiltonian

The Hamiltonian is a mathematical operator function used in molecular orbital calculations (Wylie, 1994).

Hammett constant σ

The Hammett constant is an electronic substituent descriptor reflecting the electron-donating or -accepting properties of a substituent (Hansch *et al.*, 1995).

*Hansch analysis**

Hansch analysis is the investigation of the quantitative relationship between the biological activity of a series of compounds and their physicochemical substituent or global parameters representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology (Hansch and Fujita, 1964; Kubinyi, 1993a).

Hansch–Fujita π constant

The Hansch–Fujita π constant describes the contribution of a substituent to the lipophilicity of a compound (Hansch and Fujita, 1964).

Highest occupied molecular orbital (HOMO) energy

The highest occupied molecular orbital (HOMO) energy is obtained by molecular orbital calculations and relates to the ionization potential of a molecule and its reactivity as a nucleophile.

(see *Lowest unoccupied molecular orbital (LUMO) energy)*

Frontier orbital - The molecular orbitals that involve the highest occupied molecular orbital (*HOMO*) and the lowest unoccupied molecular orbital (*LUMO*) of a given molecular entity. In the case of an odd-electron molecular entity, when its HOMO is occupied by a single electron such a molecular orbital is termed a singly occupied molecular orbital (*SOMO*). Depending on the properties of the reactive partner, the SOMO of a given species may function as either HOMO or LUMO. The special importance of the frontier orbitals is due to the fact that a broad variety of chemical reactions

takes place at a position and in a direction where the overlap of HOMO and LUMO of the respective reactants is maximal.

Homology model

A homology model is a model of a protein, whose three-dimensional structure is unknown, built from, e.g., the X-ray coordinate data of similar proteins or using alignment techniques and homology arguments.

*Hydrophilicity**

Hydrophilicity is the tendency of a molecule to be solvated by water.

Hydrophobic fragmental constant (f or f')

The hydrophobic fragmental constant of a substituent or molecular fragment represents the lipophilicity contribution of that molecular fragment (Rekker and De Kort, 1979; Hansch and Leo, 1979; Rekker and Mannhold, 1992).

*Hydrophobicity**

Hydrophobicity is the association of non-polar groups or molecules in an aqueous environment which arises from the tendency of water to exclude non-polar molecules (Martin, 1978; Martin *et al.*, 1989; Dean, 1990).

Indicator variable

An indicator variable is a descriptor that can assume only two values indicating the presence (=1) or absence (=0) of a given condition. It is often used to indicate the absence or presence of a substituent or substructure. More broadly, it is a variable which can encode anything that the investigator chooses.

Ligand design

Ligand design is the design of ligands using structural information about the target to which they should bind, often by attempting to maximize the energy of the interaction.

(see Docking studies)

Linear combination of atomic orbitals (LCAO)

The linear combination of atomic orbitals (LCAO) is a mathematical method used in quantum chemical calculations. It expresses the approximation of the molecular orbital function as a linear combination of atomic orbitals chosen as the basis functions.

*Lipophilicity**

Lipophilicity represents the affinity of a molecule or a moiety for a lipophilic environment. It is commonly measured by its distribution behaviour in a biphasic system, either liquid-liquid (e.g. partition coefficient in 1-octanol/water) or solid-liquid (retention on reversed-phase high-performance liquid chromatography (*RP-HPLC*) or thin-layer chromatography (*TLC*) system).

Lowest unoccupied molecular orbital (LUMO) energy

The lowest unoccupied molecular orbital (LUMO) energy is obtained from molecular orbital calculations and represents the electron affinity of a molecule or its reactivity as an electrophile.

(see *Highest occupied molecular orbital (HOMO) energy*)

MINDO/3 calculations

MINDO/3 (*Modified Intermediate Neglect of Differential Overlap*) calculations are semi-empirical *MO* calculations (Bingham *et al.* 1975).

MM2 calculations

MM2 calculations involve molecular mechanical calculations using version 2 of the widely-distributed force field program *MM2* (Allinger, 1977).

MNDO calculations

MNDO calculations are semi-empirical molecular orbital (*MO*) calculations, using a modified neglect of diatomic (differential) overlap approximation.

MOL file format

The MOL file format is used to encode chemical structures, substructures and conformations as text-based connection tables. It is used by MDL Information Systems Inc. (e.g., in their MACCS or ISIS programs) (Dalby *et al.*, 1992).

Molar refractivity (MR)

The molar refractivity is the molar volume corrected by the refractive index. It represents size and polarizability of a fragment or molecule.

Molecular connectivity index

A molecular connectivity index is a numeric descriptor derived from molecular topology (Kier and Hall, 1976).

Molecular descriptors

Molecular descriptors are terms that characterize a specific aspect of a molecule (Van de Waterbeemd and Testa, 1987).

Molecular design

Molecular design is the application of all techniques leading to the discovery of new chemical entities with specific properties required for the intended application.

Molecular dynamics

Molecular dynamics is a simulation procedure consisting of the computation of the motion of atoms in a molecule or of individual atoms or molecules in solids, liquids and gases, according to Newton's laws of motion. The forces acting on the atoms, required to simulate their motions, are generally calculated using molecular mechanics force fields.

(see Molecular mechanics)

Molecular electrostatic potentials (MEP)

Molecular electrostatic potentials (MEP) are electrostatic properties of a molecule based on the charge density as calculated directly from the molecular wavefunction. The electrostatic potential (scalar with dimensions of energy) is calculated at a point in the vicinity of a molecule. The spatial derivative is the electric force (vector) acting on a unit positive charge at that point caused by the nuclei and the electrons of the molecule (Williams, 1991).

*Molecular graphics**

Molecular graphics is a technique for the visualization and manipulation of molecules on a graphical display device.

Molecular interaction potentials (MIP)

Molecular interaction potentials (MIP) are field properties arising from the interaction of a probe (e.g., methyl, proton or water) with a molecule. These are calculated in a space around the molecule.

Molecular lipophilic potentials (MLP)

Molecular lipophilic potentials are properties on the Van der Waals or solvent accessible molecular surface or any other point in space (e.g., in a 3D grid for CoMFA studies) calculated from atomic lipophilicity contributions. It can be used for log P calculations, CoMFA and docking studies (Gaillard *et al.*, 1994).

Molecular mechanics

Molecular mechanics is the calculation of molecular conformational geometries and energies using a combination of empirical force fields (Burkert and Allinger, 1982).

> **Molecular mechanics** - (synonymous with force field method) - Method of calculation of geometrical and energy characteristics of molecular entities on the basis of empirical potential functions (see force field) the form of which is taken from classical mechanics. The method implies transferability of the potential functions within a network of similar molecules. An assumption is made on "natural" bond lengths and angles, deviations from which result in bond and angle strain respectively. Repulsive or attractive van der Waals and electrostatic forces between nonbonded atoms are also taken into account.

*Molecular modeling**

Molecular modeling is the investigation of molecular structures and properties using computational chemistry and graphical visualization techniques in order to provide a plausible three-dimensional representation under a given set of circumstances.

Molecular orbital (MO) calculations

Molecular orbital (MO) calculations are quantum chemical calculations based on the Schrödinger equation, which can be subdivided into semi-empirical and *ab initio* methods.

(see Ab initio calculations)

> **Molecular orbital theory** - An approach to molecular quantum mechanics which uses one- electron functions (orbitals) to approximate the full wavefunction.

Molecular shape

The molecular shape is an attribute of a molecule dealing with spatial extension, form, framework, or geometry. It is often described by, e.g., principal axes, ovality, or connectivity indices.

Molecular (dis-)similarity

Molecular (dis-)similarity is a number to express structural relatedness between pairs of molecules, e.g., the so-called Carbo, Hodgkin or Tanimoto coefficient (Good, 1992; Willett and Winterman, 1986).

Molecular topology

Molecular topology is the description of the way in which the atoms in a molecule are bonded together.

(*see Molecular connectivity, Topological index*)

Molfile

A molfile is a table containing atom type, connectivity and a more or less arbitrary *2D* or *3D* information about a molecule. Well-known file formats include the MOLfile used by MDL Information Systems Inc. (e.g., in the database MACCS), the MOL2 file used by Tripos Associates (e.g., in the modeling package SYBYL), or the CSSR format.

Monte Carlo technique

The Monte Carlo technique is a simulation procedure consisting of randomly sampling the conformational space of a molecule.

Mulliken population analysis

Mulliken population analysis is a method for allocating electrons to atoms in order to generate partial atomic charges. The results are strongly dependent on the basis set used.

Mulliken population analysis - A partitioning scheme based on the use of density and overlap matrices of allocating the electrons of a molecular entity in some fractional manner among its various parts (atoms, bonds, orbitals). As with other schemes of partitioning the electron density in molecules, Mulliken population analysis is arbitrary and strongly dependent on the particular basis set employed. However, comparison of population analyses for a series of molecules is useful for a quantitative description of intramolecular interactions, chemical reactivity and structural regularities.

Multivariate statistics

Multivariate statistics is a set of statistical tools to analyze data (e.g., chemical and biological) matrices using regression and/or pattern recognition techniques.

Neural networks

(see Artificial neural networks)

Non-bonded energy terms

Non-bonded energy terms are potential energy functions describing van der Waals, electrostatic and hydrogen bonding interactions in a force field.

Parameter space

The parameter space is a multidimensional space spanned by the descriptors in a data set.

Partial least squares (PLS)

Partial least squares projection to latent structures (*PLS*) is a robust multivariate generalized regression method using projections to summarize multitudes of potentially collinear variables (Wold *et al.*, 1993).

*Pattern recognition**

Pattern recognition is the identification of patterns in large data sets, using appropriate mathematical methodology. Examples are principal component analysis (*PCA*), *SIMCA*, partial least squares (*PLS*) and artificial neural networks (*ANN*) (Rouvray, 1990; Van de Waterbeemd, 1995ab).

PCILO calculations

PCILO (Perturbative Configuration Interaction using Localized Orbitals) calculations are semi-empirical molecular orbital calculations related to *CNDO/2* and *MNDO* calculations.

PDB

The Protein Data Bank (*PDB*) maintained at Brookhaven National Library, Upton, New York, which contains X-ray structures of several hundreds of proteins.

(see PDB file)

PDB file

A *PDB* (Protein Data Bank) file is an *ASCII* (*American Symbolic Code for Information Interexchange* = text) file used to store the atomic coordinates of a molecule, usually a protein or nucleic acid.

(see PDB)

Pharmacophore generation

Pharmacophore generation is a procedure to extract the most important common structural features relevant for a given biological activity from a series of molecules with a similar mechanism of action.

PM3

PM3 is a widely used semi-empirical molecular mechanics program.

(see Molecular mechanics)

Principal components analysis (PCA)

Principal components analysis is a data reduction method using mathematical techniques to identify patterns in a data matrix. The main element of this approach consists of the construction of a small set of new orthogonal, i.e., non-correlated, variables derived from a linear combination of the original variables.

Principal properties

Principal properties are scales of substituent or amino acid values derived by principal components analysis from a large matrix of structure descriptor variables, and useful in series design and data analysis.

*Quantitative structure-activity relationships (QSAR)**

Quantitative structure-activity relationships (QSAR) are mathematical relationships linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. Methods which can be used in QSAR include various regression and pattern recognition techniques.

QSAR is often taken to be equivalent to chemometrics or multivariate statistical data analysis. It is sometimes used in a more limited sense as equivalent to Hansch analysis. QSAR is a subset of the more general term SPC (Kubinyi, 1993a).

Quantum chemical calculations

Quantum chemical calculations are molecular property calculations based on the Schrödinger equation, which take into account the interactions between electrons in the molecule.

*Receptor**

A receptor is a protein or a protein complex in or on a cell that specifically recognizes and binds to a compound acting as a molecular messenger (neurotransmitter, hormone, lymphokine, lectin, drug, etc). In a broader sense, the term receptor is often used as a synonym for any specific (as opposed to non-specific such as binding to plasma proteins) drug binding site, also including nucleic acids such as DNA.

*Receptor mapping**

Receptor mapping is the technique used to describe the geometric and/or electronic features of a binding site when insufficient structural data for this receptor or enzyme are available. Generally the active site cavity is defined by comparing the superposition of active to that of inactive molecules.

Regression analysis

Regression analysis is the use of statistical methods for modeling a set of dependent variables, Y, in terms of combinations of predictors, X. It includes methods such as multiple linear regression (MLR) and partial least squares (PLS).

Semi-empirical methods

Semi-empirical methods are molecular orbital calculations using various degrees of approximation and using only valence electrons.

Semi-empirical quantum mechanical methods - The methods which use parameters derived from experimental data to simplify computations. The simplification may occur at various levels: simplification of the Hamiltonian (e.g. as in the Extended Hückel method), approximate evaluation of certain molecular integrals (see, for example, zero differential overlap), simplification of the wave function (for example, use of π electron approximation as in Pariser-Parr-Pople).

Sequential simplex method

The sequential simplex method is an experimental design method used for the rapid optimization of properties.

SIMCA

The SIMCA (SIMple Classification Analysis or Soft Independent Modeling of Class Analogy) method is a pattern recognition and classification technique (Dunn and Wold, 1995).

Simulated annealing

Simulated annealing is a procedure used in molecular dynamics simulations, in which the system is allowed to equilibrate at high temperatures, and then cooled down slowly to remove kinetic energy and to permit trajectories to settle into local minimum energy conformations.

Slater-type orbitals (STO)

Slater-type orbitals are mathematical functions involving exponential functions, used in *ab initio* quantum chemical calculations. These functions mimic the electronic distribution in atoms and were used in *ab initio* calculations, but have now been superceded by Gaussian-type orbitals.

(see Gaussian-type orbitals)

> **Slater type atomic orbital (STO)** - The exponential function on an atom; its radial dependence is given by Nr^{n-1} $\exp(-\zeta r)$, where n is the effective principal quantum number and ζ is the orbital exponent (screening constant) derived from empirical considerations. The angular dependence is usually introduced by multiplying the radial one by a spherical harmonic $Y_{lm}(\theta,\Phi)$.

SMILES

SMILES (Simplified Molecular Input Line Entry System) is a string notation used to describe the nature and topology of molecular structures.

Solvent-accessible surface

The solvent-accessible surface is described as the surface traced out by of a probe molecule, e.g., water, rolling over the van der Waals surface of a molecule. There are two types: a) the surface formed by the locii of the centre of a spherical probe rolled around a molecule in the van der Waals contact and b) the contact surface (or Connolly/Richards surface).

(see Connolly surface)

STO-3G basis set

A STO-3G basis set is a set of Gaussian-type orbitals (*GTO*), each of which uses three Gaussian functions to approximate a Slater-type orbital (*STO*). More extended modern basis sets include STO-3-21G or STO-KG.

*Structure-based design**

Structure-based design is a design strategy for new chemical entities based on the three-dimensional (*3D*) structure of the target obtained by X-ray or nuclear magnetic resonance (*NMR*) studies, or from protein homology models.

*Structure-property correlations (SPC) **

Structure-property correlations (*SPC*) refers to all statistical mathematical methods used to correlate any molecular property (intrinsic, chemical or biological) to any other property, using statistical regression or pattern recognition techniques (Van de Waterbeemd, 1992).

Swain-Lupton parameters ($\mathcal{F}$ and $\mathcal{R}$)

The Swain and Lupton parameters ($\mathcal{F}$ and $\mathcal{R}$) are electronic field and resonance descriptors derived from Hammett constants (Hansch and Leo, 1979).

Taft steric parameter (Es)

The Taft steric parameter is a relative reaction parameter encoding the reaction rate retardation due to the size of a substituent group.

Three-dimensional database searching

Three-dimensional database searching is a lead finding technique using three-dimensional structures of compounds stored in a database.

*Topliss tree**

A Topliss tree is an operational scheme for analog design (Topliss, 1972).

Topological index

A topological index is a numerical value associated with chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity or biological activity.

(see *Molecular connectivity*)

Topological index - The numerical basis for topological indices is provided (depending on how a molecular graph is converted into a numerical value) by either the adjacency matrix or the topological distance matrix. In the latter the topological distance between two vertices is the number of edges in the shortest path between these.

United atom approach

The united atom approach is a simplification used by molecular mechanics programs such as AMBER and CHARMM which approximates the influence of groups of atoms or molecular fragments by treating them as single atoms.

Verloop STERIMOL parameters

The STERIMOL parameters defined by Verloop are a set of substituent length and width parameters (Verloop, 1987).

Complementary and additional information may be found in the following related documents:

- Glossary of Terms in Theoretical Organic Chemistry (V.I. Minkin)
- Guidelines for the Publication of Research Results from Empirical Force Field Calculations (D.J. Raber)
- Best Values of Substituent Constants (J. Shorter)
- Acronyms used in Theoretical Chemistry (R.D. Brown)

REFERENCES

Allinger, N.L., *J.Amer.Chem.Soc.* **99**, 8127 (1977)

Bingham, R.C. *et al.*, *J.Amer.Chem.Soc.* **97**, 1285 (1975)

Boyd, D.B., Ed., Reviews in Computational Chemistry, Vol. 1 (1990)

Burkert, U. and Allinger, N.L., Molecular Mechanics, ACS Monograph 177 (1982)

Cammarata, A. and Menon, J.K., *J.Med.Chem.* **19**, 739 (1976)

Cramer III, R.D., Patterson, D.E. and Bunce, J.D., *J.Amer.Chem.Soc.* **110**, 5959 (1988)

Crippen, G.M. and Havel, T.F., Distance Geometry and Molecular Conformation, Wiley, New York (1988)

Dalby, A., Nourse, J.G., Hounsell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A., and Laufer, J. *J.Chem.Inf.Comput.Sci.* **32**, 244-255 (1992)

Dean, P.M., In: Concepts and Applications of Molecular Similarity, Johnson, A.M. and Maggiora, G.M., Eds., Wiley, New York (1990), pp 211-238

Dunn, W.J. and Wold, S. In: Chemometric Methods in Molecular Design, Van de Waterbeemd, H., Ed., VCH, Weinheim (1995), pp. 179-193.

Gaillard, P., Carrupt, P.A., Testa, B. and Boudon, A., *J.Comput.Aided Mol.Des.* **8**, 83-96 (1994)

Rogers, D. and Hopfinger, A.J., J.Chem.Inf.Comp.Sci. **34**, 854-866 (1994)

Good, A.C., *J.Mol.Graph.* **10**, 144-151 (1992)

Goodford, P.J., *J.Med.Chem.* **28**, 849 (1985)

Hansch, C. and Fujita, T., *J.Amer.Chem.Soc.* **86**, 1616-1626 (1964)

Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York (1979)

Hansch, C., Leo, A. and Hoekman, D., Exploring QSAR, American Chemical Society, Washington (1995)

Hopfinger, A.J., *J.Med.Chem.* **24**, 229 (1981).

Kier, L.B. and Hall, L.H., Molecular Connectivity in Chemistry and Drug Research, Academic Press, London (1976)

Kubinyi, H., QSAR: Hansch Analysis and Related Approaches (1993a), Vol. 1 of Methods and Principles in Medicinal Chemistry, Mannhold, R. *et al.*, Eds., VCH, Weinheim

Kubinyi, H., 3D-QSAR in Drug Design. Theory, Methods and Applications (1993b). Escom, Leiden.

Leo, A.J., *Chem.Revs.* **93**, 1281-1306 (1993)

Martin, Y.C., Quantitative Drug Design, Marcel Dekker, New York (1978)

Martin, Y.C. *et al.*, Modern Drug Research, Marcel Dekker, New York (1989)

Rekker, R.F. and De Kort, H.M., *Eur.J.Med.Chem.* **14**, 479-488 (1979)

Rekker, R.F. and Mannhold, R., Calculation of Drug Lipophilicity, VCH, Weinheim (1992)

Rouvray, D.H., In: Concepts and Applications of Molecular Similarity, Johnson, A.M. and Maggiora, G.M., Eds., Wiley, New York (1990), pp 15-42.

Tollenaere, J.P., In: Guidebook on Molecular Modeling in Drug Design (1995), Academic Press, London, pp. 337-356.

Topliss, J.G., *J.Med.Chem.* **15**, 1006-1011 (1972)

Ugi, I., Wochner, M., Fontain, E., Bauer, J., Gruber, B. and Karl, R., In: Concepts and Applications of Molecular Similarity, Johnson, A.M. and Maggiora, G.M., Eds., Wiley, New York (1990), pp 239-288.

Van de Waterbeemd, H., *Quant.Struct.-Act.Relat.* **11**, 200-204 (1992)

Van de Waterbeemd, H. and Testa, B., *Adv. Drug Res.* **16**, 85-225 (1987)

Van de Waterbeemd, H. (Ed.), Chemometric Methods in Molecular Design (1995a), Vol. 2 of Methods and Principles in Medicinal Chemistry, Mannhold, R. *et al.*, Eds., VCH, Weinheim

Van de Waterbeemd, H. (Ed.), Advanced Computer-Assisted Techniques in Drug Discovery (1995b), Vol.3 of Methods and Principles in Medicinal Chemistry, Mannhold, R. *et al.*, Eds., VCH, Weinheim

Verloop, A., The STERIMOL Approach to Drug Design, Marcel Dekker, New York (1987).

Willett, P., Similarity and Clustering in Chemical Information Systems, John Wiley, New York (1987).

Willett, P. and Winterman, V., *Quant.Struct.-Act.Relat.* **5**, 18-25 (1986)

Willett, P., Three-dimensional Chemical Structure Handling, John Wiley, New York (1991).

Williams, D.E., *Rev.Comp.Chem.* **2**, 226 (1991).

Wold, S., Johansson, E. and Cocchi, M., In: 3D-QSAR in Drug Design. Theory, Methods and Applications, Kubinyi, H., Ed., Escom, Leiden (1993), pp. 523-550.

Wylie, W.A., In: Molecular Modeling and Drug Design, Vinter, J.G. and Gardner, M., Eds., Macmillan, London (1994).