3DGEN: A System for Exhaustive 3D Molecular Design Proceeding from Molecular Topology

Julian Ivanov,† St. Karabunarliev,† and O. Mekenyan*,†,‡

Higher School of Chemical Technology, 8010 Burgas, Bulgaria, and Lake Superior Research Institute, University of Wisconsin, Superior, Wisconsin 54880

Received May 18, 1993®

A new combinatorial approach to 3D molecular design is presented. Proceeding from atom–atom connectivity augmented by the type of atoms and multiplicity of chemical bonds, it provides an exhaustive generation of all stereoisomers, enantiomers, and conformers, corresponding to the given molecular topochemistry. The core of the system is the so called idea of "propagating 3D spanning tree". The latter is an acyclic 3D model of the molecular skeleton. During propagation, it loses its acyclic character by closure of rings and eventually transforms into a 3D molecular skeleton. A current bond of 3D-spanning tree is positioned in the space by using a recursive procedure based on the 3D information of its already constructed predecessors. Bond lengths and valence angles are adjusted in advance by using MMPI parametrization. All possible rotamers at allowed torsional angle resolutions are explored, retaining those which pass the tests for geometric constraints, including nonbonded cutoff, ring-closure parameters, and torsional resolution. The number of generated isomers can be controlled by managing the values of these geometric constraints. The system provides a hierarchical set of rules for screening the generated 3D models according to the chemical expertise. In case of strained cyclic structures, some of the conformers obtained could be distorted with respect to the reference values of bond lengths and torsional and valence angles. Here, an original strain minimization technique could be applied based on a simple energy-like function, where the electrostatic terms are omitted. 3DGEN algorithm is incorporated in the 3DMOL module of the OASIS system for computer assisted structure–property studies. It can be applied also for investigation of the molecular similarity on a 3D level.

1. INTRODUCTION

Three-dimensional molecular modeling is an indispensable approach in many areas of applied and theoretical chemistry. By having information on 3D molecular shape, one can elucidate the physical properties, biological activities, and reactivity of organic molecules. Much effort is being paid to development of software systems providing a 3D molecular design; however, all of these systems are in load modules, which restrict their incorporation into large computer packages for quantitative structure–property relationship (QSPR) studies. This fact prompted the development of a variety of algorithms for a rational 3D molecular modeling of the studied chemical compounds.

The existing algorithms for obtaining 3D structures proceeding from molecular topology can be conventionally divided into five groups. The first group[1-3] includes the so-called methods for 3D structure elucidation (3DSE methods) providing generation of all constitutional[1,2] and configurational[3] stereoisomers. The obtained representations convey in a tabular or other symbolic forms stereospecificity (parity) of the stereo and chiral centers or the cis or trans orientation of double bonds. This 3D information has no molecular metric information (actual geometry): bond lengths, bond angles, and other geometric properties of the structure. The algorithms of Wipke and Dyott,[4,5] Nourse et al.,[6,7] and Sasaki and his co-workers[8] can be related to that group of approaches. In the method due to Pople,[9] the geometrical input information is specified by the ordering of the entries within the connectivity matrix according to the ring and chain conformations. The method is useful when one is able to specify correctly the rotameric state of each bond. Usually this information is difficult to be determined "by hand". On the other hand, arbitrary (or incorrect) ordering of entries as well as the presence of heteroatoms usually results in failure during ring closure in cyclic structures. In the method of Wipke and Dyott[10] the information input to the model builder is taken directly from the structure drawing on the graphics terminal. Approximate 3D coordinates are generated, proceeding from the location of the atoms with respect to the plane of the screen by using the special bond symbols ("up" and "down"). These coordinates are used for a next force field geometry optimization. Weiner and Profeta[11] developed an analogous method proceeding from 2D screen coordinates together with distance geometry approach introduced by Crippen and Havel[12-14] which has recently been extended by Waldman and Havel.[15] The general behavior of the above procedures is the necessity to perceive the connection information from two-dimensional (crude three-dimensional) screen information from "connection information" and provide 3D molecular building of a single stereoisomer.

The second group of approaches to 3D molecular design utilizes models and/or symbolic knowledge from a library of templates termed conventionally template-based methods for 3D molecular building. A set of rules govern the assembly of molecular template fragments into complete molecules.[16-23] The most representative is the algorithm of Cohen[17] where two main principles are applied. The first one is the "conformational diagram" symbolism which is a description of the sign of torsion angles for each ring and chain fragments of the molecule. This symbolic representation of the conformational moieties of the molecules provides a logical basis for the assembly rules depending on hybridization and stereochemistry of the atoms and the type of ring junction. Once the molecule is designed on a symbolic level, then the

* To whom the correspondence should be addressed.
† Higher School of Chemical Technology.
‡ Lake Superior Research Institute.
® Abstract published in *Advance ACS Abstracts*, February 15, 1994.

0095-2338/94/1634-0234$04.50/0

3D structures of the conformational surface are directly constructed by using the 3D coordinate representations of symbolic models. Cohen's procedure and related ones[17-23] are limited within a set of rings usually of size 4–8 whose conformational patterns are coded in special template libraries. This approach is not general enough, being limited by the size and the nature of the sets of templates and assembly rules. The latter are not large enough to provide 3D building for complicated molecules such as bridged and polycyclic systems. For example, it is known that Cohen's SCRIPT system fails in generating 3D models of polycyclic molecules with more than three cycles.

Another group of template-based model builders tries to account for the intra- and intermolecular forces of a molecule, thus aiming to discover a set of three-dimensional structures which would satisfactorily represent its low-energy conformers. The basic principles of these methods are embodied into the computer systems WIZARD[20] and COBRA.[22] Here, the conformational units are automatically recognized as connected fragments about whose conformational behavior (energies, coordinates, flexibility, etc.) the systems have expert assessments. It is assumed that the conformational properties of each unit do not depend on the molecules in which they appear. Once conformational units are recognized, a search of the conformational space is performed by examining all possible combinations of low-energy-unit subconformations. First, these structural combinations are performed at an abstract level. The symbolic suggestions are then examined to check for possible violations of assembling rules. If no problems are detected, the 3D structure representations of conformational units are read in from the template library and joined in a sequential fashion. The possible suggestions are then criticized, depending on geometry violation of common atoms and bonds, van der Waals repulsion, and other intramolecular forces.

A template-free molecular design procedure is the distance geometry method,[12-15] as implemented in refs 11, 15, and 24. The method generates a 3D representation of the molecule satisfying (a) distance constraints imposed by the distance ranges for bonded and nonbonded atoms and (b) stereo constraints expressed by the configuration designation of stereocenters. The distance and configuration constraints, however, do not specify a unique 3D structure, and hence a set of conformers is generated by maintaining input stereochemistry. As specific limitations of the method one can mention (a) the random placement of atoms within the allowed distance and configuration ranges thus reaching only random sampling of the conformational space and (b) the lack of energetic information in the procedure which can yield to chemically erroneous models in the case of highly strained structures.

Historically, the first practical realization of the distance geometry method is the EMBED system. The problem of geometrical constraints is solved here by local minimization of a *penalty function*, which consists of a sum of terms, one for each constraint. The optimization, however, starts from an arbitrarily chosen conformation and, because of that, ends nearly always in a local minimum (greater than zero penalty function). Different variations on EMBED have been programmed,[25,26] the most popular of which is DGEOM.[27] Linearized embedding approach[28] is another development of the distance geometry method in which all the constraints are organized into a linearized representation of the molecule.

Pearlman's CONCORD system[29] is another elegant template-free method for rapidly generating good quality three-dimensional structures, proceeding from molecular topology. The building module of the CONCORD algorithm combines aspects of an "expert system" approach and a "pseudo molecular mechanics" (PMM) approach. A collection of rules based on very fast logical analysis, attempting to minimize all steric interactions, administers the decisions regarding acyclic bond and torsion angle values. The angles within cycles, however, can be determined only through an optimization procedure. The latter is similar to the energy minimization by the molecular mechanics calculations. Different than the conventional molecular mechanics minimization, where the bond and torsional angles vary independently, in CONCORD they are coupled in a novel composite strain function. Minimization over this single variable energy-like function is fast and provides chemically reasonable geometries. CONCORD identifies gross conformation of each ring within each cyclic system in the molecule by making use of a logical analysis. Then, PMM is performed for the individual rings in a way that retains those gross conformations while optimizing the bond- and torsion angles. In a subsequent phase of the building algorithm, CONCORD optimizes the values of some of the dihedral angles associated with the rotatable bonds by minimizing the force field energy. Only those rotatable bonds are optimized, involved in paths connecting close-contact (nonbonded) atoms.

CONCORD is able to produce good quality structures for most organic chemicals, including those with complex heteroatom functional groups and ring systems (bridges, cages, and fused ring systems). It is one of the most frequently used 3D model builders that is available publicly. CONCORD was used to build 3D models for Lederle (225 000 structures), a large industrial database,[30] and has been used at Chemical Abstracts Service to generate approximately 5 million models. CONCORD, however, generates only a single conformer and cannot be used for conformational sampling.

An important area of molecular design encompasses the application of numerical methods to refine the geometry of the reasonable 3D models of chemical structures generated by the above described procedures for 3D molecular design. Commonly, molecular mechanics or quantum-chemical calculations are used for such a purpose. To solve the problem of obtaining energy optimized 3D models of a molecule, one should define different criteria of achieving computational speed and/or structural accuracy. Thus, one can sacrifice speed of the numerical calculation at the cost of simple (for the user) input or nonperfect algorithms generating strained 3D models for a subsequent optimization. In the extreme case, crude numerical calculations are invoked directly to convert 2D information taken from graphic screen (x, y, 0 coordinates and configuration qualifiers) to an (x, y, z) set. For example, computer systems like MOLY,[31] SCRIPT,[17] MACROMODEL,[32] CAMSEQ,[33] MMMS,[34] and TUTORS[35] (see also ref 16) use such an approach. Others can sacrifice accuracy for enhanced speed by applying simple potential energy functions. Levit[36] has proposed such an empirical function as the sum of strain energies due to stretching, bending, interaction between nonbonded atoms, and torsional angles. Nonbonded interactions are calculated between pairs of atoms separated by at least three bonds and closer together than a cutoff distance. Within the TUTORS system of Sasaki et al.[35] another term has been introduced additionally for minimization of strain energy. This fifth term is related to steric configurations of atoms. The latter are determined by the 2D screen coordinates of the input structural diagram. Another force field scheme for energy calculation

omitting the electrostatic terms is that of Vinter et al.[37] Because of the simple potential energy functions used in the above algorithms, they should be considered as geometry strain relief procedures or pseudo molecular mechanics rather than molecular mechanics methods. The simplified character of energy-like function significantly accelerates the optimization procedure at the cost of the relatively crude 3D geometries of the obtained 3D models.

Another group of algorithms has been developed which provides an exhaustive search of the conformational space which can be considered as a component of the 3DSE. Most of these approaches first produce crude starting geometries, which are then optimized by any of the numerical methods for energy minimization. In any stage, the current minimum energy conformer is compared with previously found conformers to check for possible duplication. The search of the conformation space is terminated when all starting geometries have been treated. Since the numerical methods simply refine the starting geometries, the overall effectiveness of these algorithms depends upon the quality of starting geometry generators. The latter, in turn, can be divided into two broad classes:[38] deterministic searchers, systematically exploring all areas of conformational space, and the stochastic (Monte Carlo) class, which uses a random element in the searching procedure. The first category of methods generates all combinations of selected values for (usually) all rotatable torsion angles to produce starting geometries. Although they are "absolute" in character, the algorithms of this group generate an extraordinarily large number of starting structures when conformationally flexible molecules are analyzed. The Monte Carlo methods do not search conformational space in a completely random manner. Usually they start with stable conformers and limit their conformational exploration by using random or pseudorandom variation of molecular geometry. Another grouping of the various conformation search methods is based on the coordinate system in which they operate. For the potential advantages of the external (Cartesian) and internal (bond lengths, bond and torsion angles) coordinate systems see refs 38 and 39.

A different technique used for exploring conformational space is the molecular dynamics method.[40] While it is effective at searching local conformational space, according to some statements,[38] it requires remarkably more computer time than the other methods for global search problems.

One of the most flexible systems for conformational space analysis is the internal coordinate tree-searching procedure, recently introduced by Still and Lipton.[39] It generates the starting geometries, proceeding from a local minimum of the conformational space, by rotating torsion angles around all rotatable bonds, thus analyzing all possible combinations of rotamers at some torsion angle resolution. In an acyclic system one alters succeeding dihedral angles starting from one end and traveling down the chain. In cyclic systems, rings are temporarily opened forming a 3D spanning tree which is then treated as the acyclic case but with additional constraints permitting ring closure. Only those structures which pass geometrical tests introduced to reject high-energy molecular geometries are retained. In such a way, the number of the starting geometries is limited while retaining as completely as possible a set of low-energy final conformers.

The final group of approaches providing a 3D molecular design are in fact interactive 3D graphics interface systems. According to the basic idea developed in these procedures, 3D molecular models are designed interactively starting from a 3D fragment. As examples of such systems one can cite MAGIC,[41] MOLBUILD,[42] CHEMMOD,[43] ALCHEMY,[44] MACROMODEL,[32] 3DEDIT,[45,46] etc. Analyzing capabilities of these programs, one can distinguish the following more significant 3D modeling processes: substitution of a single (terminal) atom by a mono- or polyatomic functional group; editing of a 3D model with respect to atom types and geometric parameters; adding single atoms by specifying a bond length, bond angle, and torsion angle; breaking or creating bonds; joining two molecules or fragments into a single molecule by pointing out the terminal atoms and length of a junction bond; automatically building hydrogen atoms; cyclizating an acyclic fragment into a ring and vice versa, etc. One of the main features of these systems is the user's control of conformational and other geometric characteristics of the growing molecular structure. This allows one to construct any conformation or nontrivial (transition state) 3D models.

On the basis of the above introduction to algorithms for 3D-molecular design, one can conclude that each of the methods is specialized in a particular area of the manifold procedures of 3D structure elucidation. Thus, some of the methods generate all stereoisomers and conformational isomers proceeding only from molecular topology. However, they provide 3D information in a symbolic way without constructing the actual 3D molecular models. Other 3D builders are limited within a set of template and construction rules, because of which they fail to analyze the pattern of the conformational potential surface. Moreover, they cannot even reach a single 3D model in the case of relatively small polycyclic systems. In general, the template-free algorithms provide good quality models proceeding from molecular topology. They, however, reach a random sampling (or a point) of the conformational space. In turn, the algorithms for conformational searching generate all (or almost all) representatives of this space; however, they proceed from some starting geometry(ies) but not from molecular topology.

In the present work we are introducing a template-free algorithm which provides, in a single procedure, generation of all but the high-energy conformers, proceeding from molecular topology. In other words, the goal of the present algorithm is to provide in a single procedure the following main tasks: (i) generation of all possible stereoisomers and optical isomers; (ii) exhaustive search of conformational space generating all possible conformers under some constraints; (iii) design of 3D models of generated stereoisomers and conformational isomers; and (iv) screening of the generated 3D molecular models according to chemical expertise.

2. OUTLINE OF THE APPROACH

The basic principles of the approach are next presented (Figure 1). They can be specified as the initial analysis of molecular graph, unique topochemical naming, propagation of the 3D spanning tree, and stereochemical and conformational naming.

2.1. Initial Structural Analysis of Molecular Graph. It provides the following.

(i) Ranking of the atoms in the respective chemical graph. Here, each non-hydrogen atom is presented by a vertex, and each chemical bond between non-hydrogen atoms, by an edge. The ranking algorithm starts by an ordering of the vertices according to their distance code,[47] namely, the sequence of numbers of the first, second, third, etc., neighbors. The distance codes of the vertices of structure 1 are given in Figure 2a. Then vertices are ordered starting by those with the smallest distance codes. The initial partitioning is next improved iteratively. First the sorted lists of ranks of all

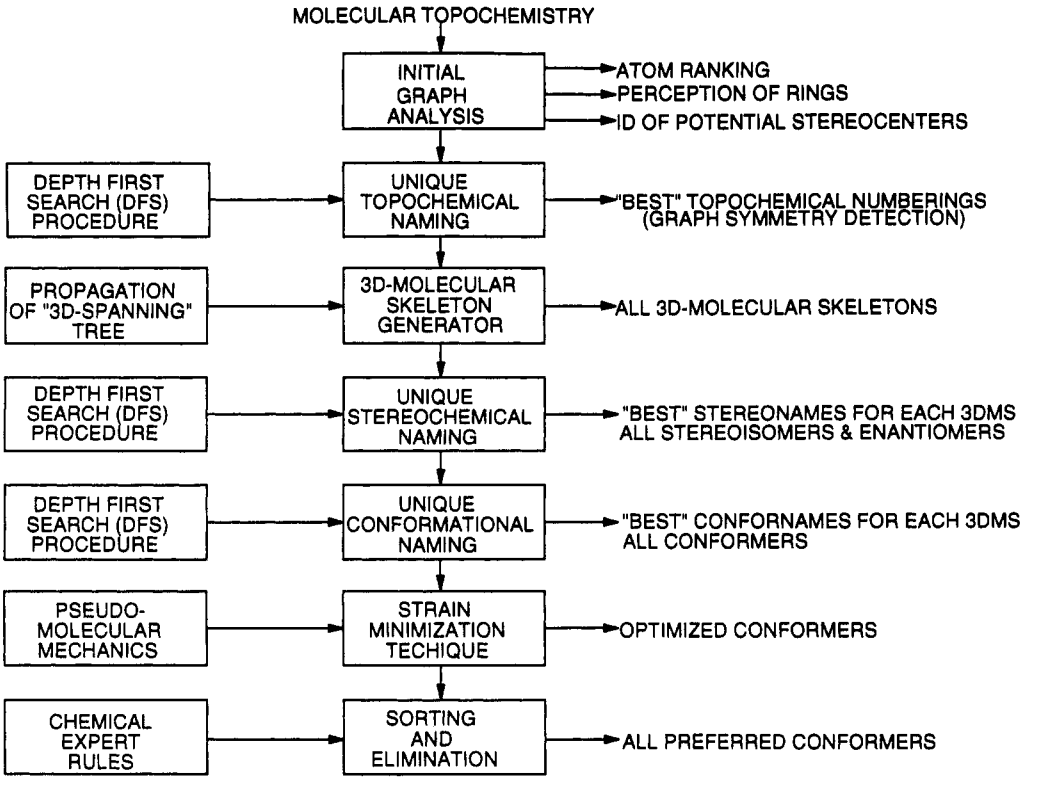


Figure 1. Flow chart of the 3DGEN system.

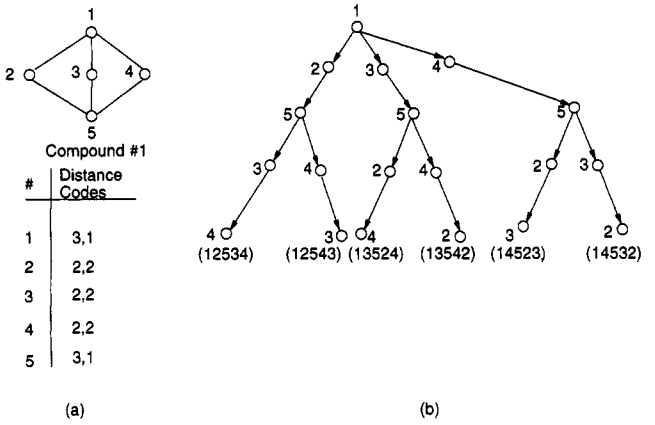


Figure 2. Vertex distance code (a) and oriented spanning tree of DFS procedure (b) for compound 1.

neighboring vertices are formed, and then vertices are partitioned according to the lexicographic ordering of those lists. The latter procedure is repeated until it does not produce a better partitioning. It was proved that this heuristic procedure for graph vertices numbering (hierarchically ordered extended connectivity, HOC,[48-50] illustrated for compound 2 in Figure 3) produces the graph orbits for graphs in which each vertex belongs to less than three fundamental rings.

(ii) Perception of the smallest rings incident to each pair of neighboring edges (in the case of cyclic structures). This information specifies the valence angles between these bonds. The size of the smallest rings are assumed to be crucial for the angle between two bonds, since the deformations of the small rings determine stability of the whole molecules.

(iii) Identification of the potential stereocenters. In this approach we deal with tertiary and quaternary carbons as well as carbon–carbon double bonds (with at least one non-hydrogen substituent at each end) as potential stereocenters but not heteroatoms and double bonds connecting heteroatoms.

2.2. Topochemical Naming. The unique topochemical naming of the molecule is obtained on the basis of an oriented spanning tree of the respective chemical graph and vertex ranking already defined. The spanning tree starts at a vertex with the lowest rank. Then one goes to an adjacent vertex, then to a third vertex adjacent to the second one, and so on. Vertices are visited only once. When all adjacent atoms have been visited, one returns to the last atom on the walk which still has neighbors which have not yet been visited and continues by one of them. Edges proceeding from the current vertex to a visited one but not to its ancestor are termed ring closure edges. Thus, an orientation of edges and ordering of vertices with respect to the sequence in which they are visited is induced. This well-known systematic method of exploring a graph is known as the depth first search (DFS).[51] The oriented spanning tree of the DFS procedure for structure 1 is illustrated in Figure 2b. The naming of the chemical graph is formed by the sequence of atom labels according to the ordering (numbering) they have in the DFS tree. If at some point during DFS procedure a choice must be made between two atoms, the preference is for the one with the lowest rank. In the case of a choice between n atoms ($1 < n < 4$) whose ranks are the same, the n names are generated, and the "better" numbering is taken. Better is defined to be lexicographically larger when the entire names are compared as a set of consecutive strings each presenting an atom defined by its atomic symbol, the type of the bonds connecting it with the neighboring atoms as well as their current numbering during DFS. Thus, in addition to molecular topology, the chemical nature of the atom and bonds (topochemistry) is taken into account. When two identical names are generated due to choice points, then the atoms interchanged by this choice are defined to be topochemically equivalent. In the case of m choice point between two equivalent neighbors and n choice points between three equivalent neighbors 2^m3^n, equivalent (but different) numberings of the graph will result. They correspond to the "best" topochemical name. The four equivalent numberings for compound 3 are presented in Figure 4. It is clear that the assignments of the vertices by their

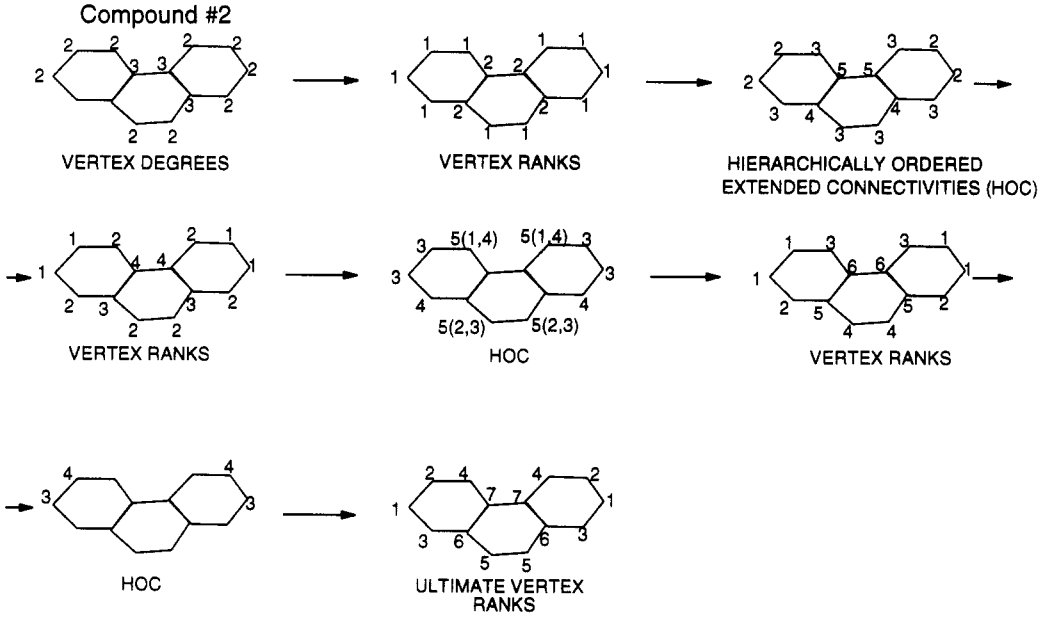


Figure 3. Successive stages in the application of the HOC algorithm to phenanthrene molecule (compound 2).

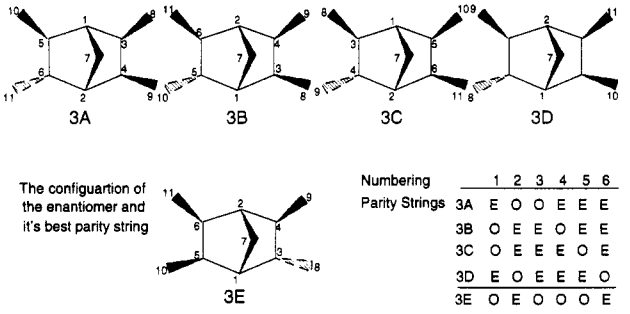


Figure 4. Equivalent topological numberings (3A–3D) and parity strings of the stereocenters for compound 3.

ranks reduce the walking in the graph during the DFS procedure.

2.3. Propagation of the 3D Spanning Tree. One of the principal novelties of the presented approach is the procedure of the "3D spanning tree propagation", proceeding from the molecular topochemistry information. The 3D spanning tree (3DST) is an acyclic 3D model of the molecular skeleton. Its construction starts from a particular atom, defined according to its topochemical ranking. During the propagation of the 3DST, the latter loses its acyclic character by rings' closure and eventually transforms itself into a 3D molecular skeleton (3DMS). A set of geometric requirements are imposed during the generation of 3DMS depending on the type of atoms, their hybridization, and incidence to the smallest rings (in the case of cyclic structures). For bond lengths and valence angles MMPI parametrization[52] is used. The possible combinations of torsion angles are explored by defining the position of a bond in the 3D space, depending on the hybridization of the initial atom: sp^3 (–60°, 0°, +60°, +180°), sp^2 (–120°, +120°), sp (+180°). All rotamers at allowed torsional angle resolutions are explored, retaining those 3DMSs which pass the tests for geometric constraints, including the nonbonded cutoff (NBC), ring-closure parameters (RCP), and torsional resolution (TR). Thus the number of generated isomers can be controlled by managing the values of geometric constraints. Still, the 3DGEN system provides an adapted set of constraint parameters for automated use.

As one can see later, using the "best" topochemical naming, a set of 3DMSs are generated by trying all possible combinations of the above torsion angles. After the detection of stereocenters (among the set of potential stereocenters) the best stereoname is specified for each 3DMS providing the best sequence of stereocenter parities. Because of the molecular symmetry, some of the best stereonames can be identical and the respective duplicating stereoisomers are eliminated. The screening of the identical stereoisomers is based on their conformational naming. The identification of each distinct stereoisomeric structure is performed by an original version of the SEMA unique naming algorithm.[48,50]

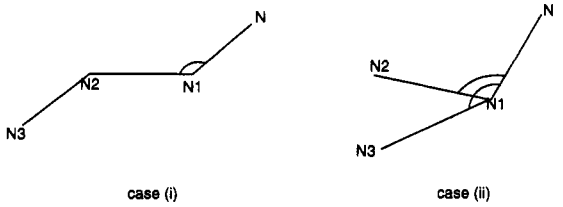
We present here a simplified description of the sequence of geometric operations for generation of 3DMSs. The process of generation starts by the vertex with the lowest topochemical numbering and follows the ordering of vertices according to their numbering. This ordered growing of 3DMS (ordered variation of torsion angles) permits a complete conformer generation process to be performed and presented as a treelike structure. The latter provides (i) an elimination of all 3D structures arising from a partially defined 3DMS having some undesirable geometric features (i.e., small nonbonded interatomic distances), both for hydrogen and non-hydrogen atoms (pruning of a branch from the generation tree) and (ii) continual propagation of the 3D spanning three through the ordered sequence of torsional rotations. Besides the constraints for atom crowding (defining the minimum allowable separations between nonbonded atoms) a second group of geometric tests is employed in the case of cyclic molecules. These constraints are applied to the ring-closure step when a cycle has to be constructed. They consist of the closure distant constraint, two bond angles and three dihedral angles. In the present realization of our approach we employ the closure distant constraint only. The above treelike generation procedure is similar to the Lipton and Still algorithm[39] applied for an exhaustive search of conformational space proceeding from a low-energy molecular conformation. Our algorithm, however, is basically different, proceeding from molecular topochemistry instead of a selected particular geometry.

A current bond of the 3D spanning tree is positioned in the space by using a sequential procedure based on 3D information of its already constructed neighboring bonds. The following two main cases hold here.

(A) The current NN_1 bond belongs to an acyclic part of the molecule. For its construction one needs 3D information for the following.

(i) A chain of two consecutive bonds (N_3N_2 and N_2N_1), if only one of them (N_1N_2) is a predecessor of the current bond. The bond (N_1N) length and valence (N_2N_1N) and torsional ($N_3N_2N_1N$) angles conditioning the position of the current bond are determined according to the type and hybridization of the atom N_1, which is a bond starting point.

(ii) The preceding bonds when both are incident to the current one. In this case the two valence angles (N_2N_1N and N_3N_1N) between the current bond and its neighbors uniquely determine the two possible positions of the current bond—up and down to the plane defined by the predecessors.

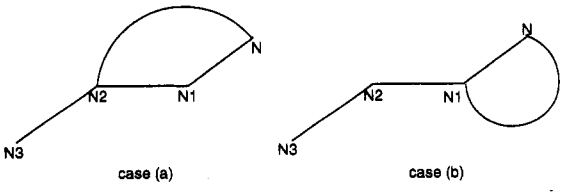


(B) The current bond belongs to a cyclic part of the molecule. Here, the sizes of the smallest rings incorporating the current bond and each of the neighboring (already constructed) bonds are determined during the algorithm. Three cases can be considered here, depending on the number of preceding incident bonds:

(i) In the case of one neighboring (and already constructed) bond, one of the following is considered.

(a) The preceding and the current bond belong to one and the same cycle. Here, the size of the smallest ring consisting of the two bonds defines the valence angles and the set of possible torsional angles determining the position of the currently constructed bond. For example, in the case of the cyclobutane ring all valence angles are equal to +90°, whereas the set of possible torsion angles are 0°, 35°, −35°, +60°, and −60°. Inclusion of one double bond in this ring (cyclobutene) yields valence angles +94° and +86° for sp^2 and sp^3 moieties, respectively, while inclusion of two double bonds (in cyclobutadiene) again yields +90° for all valence angles. Only a planar configuration is assumed for the last two cases. Such an "expert parametrization" is defined for rings up to size 5 on the basis of quantum-chemically optimized geometries.[53] The implementation of these specific (template-type) parametrizations is inevitable because of the deviations of small ring geometries from those corresponding to the standard hybridization of consisting atoms.

(b) The preceding bond and the current bond do not belong to one and the same cycle. Here, case A(i) holds.

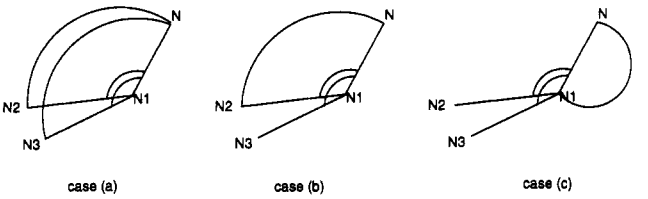


(ii) In the case of two already constructed bonds neighboring the current one, three cases can be considered.

(a) The current bond takes part in cycles with each of the neighboring bonds. In this case one determines the smallest rings formed by the current bond and its incident predecessors. The size of both cycles determines the possible pairs of valence angles between the current bonds and its preceding neighbors, as well as the up/down position of the current bond (the torsion angles with respect to the plane defined by the predecessors), similarly to case A(ii).

(b) The current bond takes part in cycles with one of the neighboring bonds. The pair of valence angles is determined by the size of the smallest ring between the current bond and its neighbor as well as the angle between the current and "acyclic" bond. The position of the current bond is determined analogously as in the above case.

(c) The current bond takes part in cycles with neither of the neighboring bonds. Here, case A(ii) holds.



(iii) In the case of three predecessors incident to the current bond, one determines the smallest two rings defined by the current bond and its already constructed neighbors. Then the position of the current bond is determined as in case B(ii).

In order to retain the cyclic nature of the ultimate structure during 3DST growth, rings are closed, thus transforming the 3D spanning tree into 3D molecular skeleton. Aiming at that, ring-closure constraints are defined. The most important of these constraints is the closure distance one—the range of acceptable distances between the two atoms forming the ring-closure bond. The results obtained show that this constraint is sufficient for an exhaustive generation of stereoisomers at a reasonably high resolution for conformational analysis.

2.4. Stereochemical and Conformational Naming. The above procedure for 3DST propagation eventually generates a set of 3DMSs by exploring every possible combination of rotamers at the allowed torsional angle resolution and retaining only those structures which pass the tests for geometrical constraints. After that, the DFS procedure is performed again providing n numberings of the atoms of *each* 3DMS, where n is the number of equivalent topochemical numberings corresponding to the best topochemical name ($n = 4$ for chemical 3 in Figure 4). Having the geometric configuration of each potential stereocenters and the numbering of their attachments, we determine the following:

(i) The parity of potential stereocenters, by using the Cahn and Ingold algorithm[54] (see also ref 3) specifying them as even and odd. The parities of the stereocenters of compound 3 are listed in Figure 4.

(ii) Stereonames, by using the stereochemical extension of the HOC algorithm.[48-50] Here, the sequence of the topochemical numbering is extended by the parity of stereocenters, assigning 0 and 1 for an even and odd stereocenter, respectively.

(iii) The real stereocenters. During the DFS procedure a set of names is generated due to the topological symmetry of the respective molecular graph. If these names are identical except for the parity of one potential stereocenter,

then the attachments bonded to this atom are equivalent and that center is removed from the set of real stereocenters.

(iv) The best stereoname, by comparing the list of stereonames corresponding to each of the topologically equivalent numbering and choosing the best one (with the best set of stereodescriptors). On examining compound 3 (Figure 4), it can be seen that 3B has the best parity string since the configuration of stereocenters 1–6 converted into a binary number gives 011011, which is minimal among all four topologically equivalent numberings (3A–3D).

(v) The existing enantiomeric relationship. The nature of the unique stereochemical naming[4,5,48,50] provides recognition of the enantiomeric relationship by simply comparing the stereonames of each distinct 3DMS. It was seen (Figure 4) that the best parity string corresponds to the minimal binary number. On looking among the parity strings for the other topological equivalent numberings, the maximum binary number will correspond to the best parity string of the antipode, if it exists. Analyzing the minimal and maximal binary numbers, whenever the parity is opposite, indicates that the vertex is not a chiral stereocenter. This is because the best parity strings of two enantiomers will have the chiral centers of opposite configuration. From Figure 4, it may be seen that 3D has the maximum binary number, 101110; therefore this corresponds to the best parity string of the enantiomer (3E), 010001. On comparing the latter with the best parity string (3B) of compound 3 for similarity, one can see that vertices 3 and 5 are chiral stereocenters because the only differences appear in the configuration of vertices 3 and 5.

The set of best stereonames, each corresponding to one of the 3DMSs, is further checked for uniqueness, and the redundant names (as well as the respective 3DMS) due to the molecular symmetry are eliminated from the list of stereonames. They are preserved, however, for the next conformational description and search for conformational uniqueness.

Another performance of the DFS procedure provides a complete conformational specification of the structure. By using 3D geometry of *each* 3DMS and the set of *n* numberings of its atoms, the values of all torsion angles are coded by specifying the two marginal edges of each path of length 3. For many reasons the dihedral angles need not be precisely coded. The system of 36 integers is used to describe the ranges of torsional angles, corresponding to the smallest dihedral angle resolution of +10°: 0° ÷ 10° = 1; 10° ÷ 20° = 2; ...; 340° ÷ 360° = 36. Again the best conformational names are chosen for the complete representation of each 3DMS. Next, they are checked for uniqueness, and the redundant (nonunique) conformers are rejected.

2.5. Strain Energy Minimization Procedure. The use of preliminary parametrized bond lengths and valence angles during 3DST propagation provides a generation of energetically favorable 3DMSs. Still, the obtained 3D models can be considered as crude suggestions which need to be refined mainly because of the possible deformations of cyclic structures due to the ring-closure step. In the case of strained cyclic structures or saturated cyclic systems with localized multiple bonds, the exhaustive generation of all stereoisomers usually requires a large closure distance range. Some of the conformers obtained by this "forced" algorithm, however, are distorted with respect to valence and torsional angles which requires a strain relief numerical procedure. In the present approach we have developed such a procedure based on a simple energy-like empirical function, where the electrostatic terms were omitted. Because of that, the strain releaf technique was named pseudo molecular mechanics. The energy minimization is estimated as a sum of four terms:

$$E(\text{strain}) = aE(\text{bond}) + bE(\text{valence angle}) + E(\text{torsion angle}) + E(\text{VW}) \quad (1)$$

The terms on the right side of eq 1 denote the strain energies due to stretching, bending, torsional angles, and (Van der Waals) interactions between nonbonded atoms. The equations used to calculate the above energy terms as well as the reference values for bond lengths, angles, and nonbonded interactions are taken from the original MMPI molecular mechanics program. A preference is given to stretching and bending terms by weighing parameters a and b ($a = 10^4$ whereas $b = 2.1914$ [52]). The energy function is minimized by correcting the atomic coordinates in the direction of steepest descent of the energy. The character of energy function significantly accelerates the optimization.

For example, if one performs the 3DGEN algorithm on compound 3 (Figure 4), applying the adapted set of geometric constraints (NBC = 1.5 Å, RC = 1.0–1.8 Å, and TR, depending on the hybridization of atoms), the systems fails to generate conformers. In this case, it requires more "loose" geometric constraints. The increase of RC tolerance to 1.0–3.0 Å provides the generation of two conformers, which are, however, distorted and need implementation of the strain releaf procedure. In a few seconds, the system reaches a minimum of the penalty function for each of the "optimized" conformers.

2.6. Expert System Rules and Conformational Analysis. The above "fast" energy minimization procedure is only one of the approaches utilized in our system to screen high-energy conformers. Having in mind the large number of possible conformers, especially in the case of flexible molecules, however, we have introduced a set of expert constraints for a preliminary avoidance of well-known high-energy structures. These hierarchical sets of rules criticize the generated portions of conformers according to experimental observations and theoretical results. For example, the high rotational barriers of conjugated fragments (e.g., esters, enones, enoles, ethers) provide a relatively secure reduction of associated torsional angles to 180° or 0°. If one detects the presence of atoms providing hydrogen bonding, then the conformers can be safely eliminated sterically, not allowing formation of such bonds. The maximization of the distances in saturated cyclic (and acyclic) fragments usually results in low-energy conformers. It is clear that the early avoidance of high-energy conformers during our treelike generation procedure improves its timing, interrupting construction of high-energy conformers. Keeping the general algorithm for an exhaustive conformational and stereo search, our system also provides a sorting of all generated isomers according to a chosen subset of the proposed expert rules.

2.7. Some Examples and Assessments of the 3DGEN Algorithm. The capabilities of 3DGEN as well as the relationship between the number of generated (stereo-, optical, conformational) isomers and the constraints parameters used, nonbonded cutoff (NBC), ring-closure (RC), and torsional resolution (TR)) can be analyzed by applying the system to a set of structurally diverse molecules. The purpose of this analysis is to specify the optimal parameter values which should be used with reasonable confidence for an exhaustive search of stereospace and conformational space. Apparently, these values will differ for different types of molecular structures.

Table I. Number of Generated Conformers for *n*-Hexane and *n*-Heptane, as Functions of the Geometric Constraints

torsional resolution (deg)	nonbonded cutoff (Å)	generated isomers			CPU time (s/structure)	
		n-hexane[a]	*n*-hexane[b]	*n*-heptane[b]	*n*-hexane	*n*-heptane
120	1.50	17	12	20	0.15	0.20
120	1.00		12	20	0.15	0.20
120	1.50		17	42	0.13	0.15
120	0.00	27	18	45	0.12	0.16
60	1.75	69	72	281	0.15	0.27
60	1.50	99	86	351	0.16	0.30
60	1.25		94	405	0.15	0.31
60	1.00		100	457	0.15	0.31
60	0.75		103	492	0.15	0.32
60	0.50		117	598	0.15	0.33
60	0.25		125	655	0.15	0.35
60	0.00	144	126	666	0.15	0.35
30	1.50	708	565	4728	0.31	1.71
30	1.00		754	7551	0.35	2.59
30	0.50		870	9371	0.40	2.51
30	0.00	1008	936	10440	0.42	2.62

[a] Reference 39. [b] This work.

Table II. Number of Generated Conformers for Cycloheptane, Cyclooctane, Cyclononane, and Cyclodecane, as Functions of the Geometric Constraints

torsional resolution (deg)	nonbonded cutoff (Å)	ring closure (Å)	generated conformers			
			cyclo-heptane	cyclo-octane	cyclo-nonane	cyclo-decane
120	0.00	1.0–3.0	7	15	38	102
120	0.50	1.0–5.0	19	42	84	184
60	0.50	1.4–1.6	6	10	78	158
60	0.50	1.0–2.0	8	59	172	813
60	0.50	1.0–3.0	57	201	818	3670
30	1.00	1.4–1.6	2	39	327	2990
30	1.00	1.0–2.0	84	349	2074	9961
30	1.00	0.5–2.5	198	938	5933	10315

Some preliminary examples and assessments of the 3DGEN algorithm will be presented here. The number of the generated conformers and the computer time necessary for that (on IBM/PC-486, 50 MHz) as functions of the torsional resolution and nonbonded cutoff are given in Table I for *n*-hexane and *n*-heptane. As shown, the number of generated isomers increases with the increase of TR. Thus, at a constant NBC (1.5 Å), the number of *n*-hexane isomers is 12, 86, and 565, for TR = 120°, 60°, and 30°, respectively. Similarly, at a constant torsional resolution (e.g., TR = 60°), the number of generated isomers for *n*-hexane increases from 72, for NBC = 1.75 Å, to 126, for NBC = 0.00 Å. On analyzing the generation time (CPU-s/structure), it can be seen that it increases with the increase of torsional resolution and reduction of nonbonded cutoff. For example, at TR = 60°, CPU increases from 0.27 s to 0.35 s/structure for NBC = 1.75 and 0.00 Å, respectively. The generation time depends also upon the size of the molecule; at TR = 30° and NBC = 1.5 Å, CPU time increases from 0.31 s/structure for *n*-hexane, to 1.71 s/structure, for *n*-heptane.

On comparing the numbers of generated isomers for *n*-hexane (Table I) by the 3DGEN algorithm and those of Still and Lipton (internal coordinate tree-search procedure, MULTIC program[39]), we can see some differences. With one exception (for TR = 60° and NBC = 1.75 Å), the number of generated conformers by Still and Lipton's algorithm is larger than ours. If the adapted set of torsional angles (–60°, 0°, +60°, +180° for sp^3 hybridized atoms) is applied, however, the number of generated conformers by both algorithms is equal (17 for *n*-hexane at NBC = 1.5 Å by both algorithms).

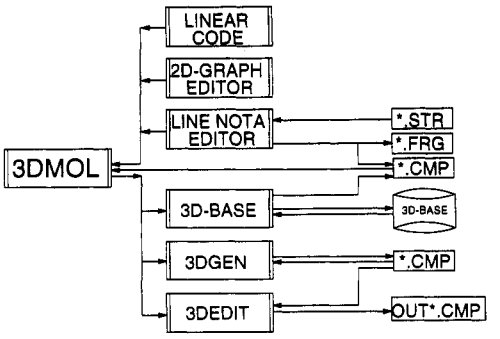


Figure 5. 3DGEN as a routine of the OASIS system.

This fact needs additional clarification.

Besides the torsional resolution and nonbonded cutoff controlling the generated conformers for the acyclic structures under investigation (Table I), another geometric constraint is included here—the ring-closure distance (Å). As one can expect, the number of generated isomers increases with the increase of ring-closure parameters; e.g., at TR = 60° and NBC = 0.5 Å, generated cyclodecane conformers are only 158 for RC = 1.4–1.6 Å but 3670 for RC = 1.0–3.0 Å. The number of conformers for four cyclic structures, cycloheptane, cyclooctane, cyclononane, and cyclodecane, obtained by 3DGEN algorithm at different geometric constraints are presented in Table II. Similarly to acyclic structures, the number of generated conformers increases with the system size: at TR = 60°, NBC = 0.5 Å, and RC = 1.0–2.0, generated conformers are 8, 59, 172, and 813, for ring sizes 7, 8, 9, and 10, respectively.

The detailed analysis of the capabilities and limitations of the algorithm will be presented in a subsequent publication.[55]

3. 3DGEN: AS A ROUTINE OF THE OASIS COMPUTER SYSTEM

The OASIS (optimized QSAR approach based on structural indices set) is a computer assisted approach aimed at helping in one of the major tasks in contemporary chemistry. Proceeding from quantitative structure–property or structure–activity relationships (QSPR/QSAR), it is a rational guideline in synthesizing new compounds with prescribed properties, e.g. drugs, agrochemicals, new materials. This sophisticated computer package has been successfully applied[56–63] to model biological activity of different classes of organic compounds as well as the acute toxicity of industrial chemicals.

The 3DGEN algorithm is incorporated in the 3DMOL module of the OASIS system. The general purpose of this module is the chemical structure input, 3D database management and generation of reasonable 3D models corresponding to the 2D input records. The scheme of information processing and file access of 3DMOL are presented in Figure 5.

OASIS provides three ways to enter the 2D structure of compounds from the studied series: by line notations, 2D graphic editor, and line codes from the console. In case the entered 2D structure (one of the items of input *CMP file) is not identified as a substructure of the records belonging to the 3D database, the user can implement the 3DGEN program for an automated generation of approximate 3D models of the input structure. The generation of 3DMSs is controlled optionally by the set of constraint parameters, NBC, RC parameters, and TR. An adapted set of such parameters is available for automated control of 3D structure generation. the 3DGEN program has a number of additional capabilities as options:

(i) Provides file management for three-dimensional models. Thus, it stores/retrieves a molecule on a file for later recall if necessary.
(ii) Provides graphic display of the 3D structures with options of rotation about X, Y, and Z axes, translation, zoom, and others.
(iii) Denotes the stereocenters (with their parity) and hiral centers.
(iv) Takes count of and indicates the number of generated conformers, stereoisomers, enantiomers, and structures providing hydrogen bonding.
(v) Creates the enantiomer of a query structure on the screen.
(vi) Deletes all hydrogens from the molecule as well as adds them in the proper positions.
(vii) Allows entering and subsequent treatment of charged species (cations, anions) and radicals (including radical cations, radical anions).
(viii) Answers user questions about particular interatomic distances, valence and torsional angles, and presence of crowded atoms (under some threshold distances between nonbonded atoms).
(ix) Displays all bond lengths and angles in geometry tables.
(x) Changes the conformation of a molecule by performing rotations about acyclic bonds trying to maximize the interatomic distances.
(xi) Sorts the generated conformers according to a set of hierarchically ordered expert rules which gives the priority of isomers, providing hydrogen bonding, sp^2 hybridization, maximum of interatomic distances (of acyclic and/or cyclic parts), specific interatomic distances, molecular "energies" after pseudo molecular mechanics, etc.

4. SUMMARY AND CONCLUSIONS

The 3DGEN algorithm provides, in a single procedure, generation of all but the high-energy conformers, including stereoisomers and optical isomers, proceeding from molecular topology. During the generation procedure, the system also constructs in parallel the 3D models of the generated conformers. One of the principal novelties of the presented algorithm is the procedure of the 3D spanning tree propagation, which is an acyclic 3D model of the molecular skeleton. Its construction starts from a particular atom, defined according to its topochemical ranking. A current bond of 3DST is positioned in the space by using a recursive procedure based on the 3D information of its already constructed bonds: the type and hybridization of consistent atoms; incidence to (a)-cyclic fragments and the size of the smallest ring(s) (if any) to which they belong. During the propagation, 3DST loses its acyclic character by closure of rings and eventually transforms into a 3D molecular skeleton. All possible rotamers at the allowed torsional angle resolutions are explored, retaining only those which pass the tests for geometric constraints: nonbonded cutoff, ring-closure parameters, and torsional resolution. Thus, the main difference from Lipton and Still's MULTIC system, which constructs the initial 3D spanning tree proceeding from a geometry optimized structure and which performs torsional rotations about all rotatable bonds (in cyclic systems, rings are temporarily opened) in analyzing conformational space, in the 3DGEN algorithm 3DST is generated proceeding from molecular topology only. A few differences appear between the number of conformers generated by Still and Lipton's MULTIC program and our algorithms, likely due to some differences in the values of rotational angles and geometric constraints which need further analysis.

The first screening of the highly unstable conformers is made by applying the set of the geometric constraints. Still, in the case of large and conformationally flexible molecules, the problem of finding all but the energetically unfavored conformers is intractable. For these reasons, during the generation of conformers one can apply the set of chemical expert rules, which can operationally screen the undesired structures according to the particular expert requirements. Thus for example, one can ask for the generation of all conformers providing intramolecular hydrogen bonding. In this case, 3DGEN will eliminate in advance (before the construction of the whole structure is completed) all conformers having geometric features not allowing formation of such bonds. Some of the expert rules (e.g., for geometric distance maximization), however, could be applied only after completion of the full-space search process, which is realistic for all but the simplest molecules.

Thus, the 3DGEN algorithm does not provide all conformational minima search, like the MULTIC system and related algorithms, which apply force field energy minimization consecutively to each of the generated crude geometries. Instead of using rigorous (e.g., energetic) criteria of screening, the 3DGEN approach aims at reducing the full-space number of conformers by a set of chemical expert rules. This feature of the algorithm makes it quite similar to other expert-based systems (e.g., CONCORD).

In the case of strained cyclic structures, the exhaustive generation of all stereoisomers usually requires looser geometric constraints. Due to this, some of the conformers obtained can be distorted with respect to valence and torsional angles. This imposed the development of a fast strain relief numerical procedure. The procedure used in the 3DGEN system is based on a simple energy-like empirical function, where the electrostatic terms were omitted. Because of that, the strain relief technique was named pseudo molecular mechanics (similarly to the strain relief algorithm used in CONCORD).

The present version of the 3DGEN algorithm does not take into account the ionization states of the 3D structure (not neglected by the Lederle group[23]), which is obviously important, because the pH of the system could affect the 3D structure generated. Another effect neglected by the 3DGEN system is the tautomeric one. For example, the different tautomeric forms of uracil have slightly different 3D molecular structures. Both effects, however, can be accounted for by suitable expert rules in 3DGEN. The algorithm was successfully applied to fullerene types of structures (C20). The generation process, however, was too time consuming. Apparently, the set of rules should be extended to deal with these more complicated polycyclic ring systems. Another development of 3DGEN should be directed to the recognition of stereocenters involving heteroatoms. Further investigations are necessary also to figure out the classes of structures for which the 3DGEN approach might fail. The system should be tested also on large 2D structural bases (hundreds or thousands of connection tables).

Two immediate implications follow based on the structural information generated by the 3DGEN algorithm. The first is related to QSAR studies. Usually, the computer-assisted methods for SAR analysis provide the low-energy conformer for each of the 2D items belonging to the series under investigation. In a complex biological environment with different polarity, however, this approach can lead to spurious results. Due to the expert systems for discrimination of conformers as well as the set of the geometric constraints used

to control the generation procedure, 3DGEN can eliminate only high-energy conformers, thus providing the complete list of potential 3D participants in the process. Taking into account that bond lengths and angles are parametrized in advance, within a first approximation one can avoid the geometry optimization by force field or quantum-chemical methods. Hence, one SCF should be performed for each conformer, assuming its geometry to be fixed, in order to assess the molecular stereoelectronic structure. Thus, the complete list of 3D isomers provides *bands* for the variation of stereoelectronic parameters of each 2D item of the studied series, which bands eventually can allow for the calibration of structure–property regression models.

The next implication is related to the area of structure similarity. The following two groups of observations need additional theoretical analysis, namely, the fact that many molecules with similar topology have different physicochemical properties and reactivities and many topologically dissimilar molecules behave similarly with respect to their properties. Apparently, for the basic explanation of these "irregularities" one should look at the 3D level. The set of 3D-dissimilar (3D-similar) conformers should be defined among all conformers of a series of topologically similar (dissimilar) molecules in an attempt to explain some differences (similarities) in their properties.

REFERENCES AND NOTES

(1) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. *J. Am. Chem. Soc.* **1975**, *97*, 5755.
(2) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. *J. Org. Chem.* **1981**, *46*, 1708.
(3) Nourse, J. G.; Smith, D. H.; Carhart, R. E.; Djerassi, C. *J. Am. Chem. Soc.* **1980**, *102*, 6289.
(4) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4825.
(5) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4834.
(6) Nourse, J. G. *J. Am. Chem. Soc.* **1979**, *101*, 1210.
(7) Nourse, J. G.; Carhart, R. E.; Smith, D. H.; Djerassi, C. *J. Am. Chem. Soc.* **1979**, *101*, 1216.
(8) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216.
(9) Gordon, M. S.; Pople, J. A. Program MBLD. *QCPE*, **1975**, *10*, 135.
(10) Wipke, W. T. In *Computer Representation and Manipulation of Chemical Information*; W. T. Wipke, S. R., Heller, R. J., Feldman, and F., Hyde, Eds.; Wiley: New York, 1974; p 153.
(11) Weiner, P. K.; Profeta, S., Jr. *Abstracts*, 181st National Meeting of the American Chemical Society, Atlanta, GA, Mar 29–Apr 3, 1981; American Chemical Society: Washington, D.C., 1981; COMP Division, Abstract No. 15.
(12) Crippen, G. M.; Havel, T. F. *Acta Crystallogr., Sect. A*, **1978**, *A34*, 282.
(13) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *Bull. Math. Biol.* **1983**, *45*, 665.
(14) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press: 1988.
(15) Waldman, M.; Havel, T. F. The Sketcher, available from BIOSYM Technologies Inc., San Diego, CA, 1992.
(16) Cory, E. J.; Feiner, N. F. *J. Org. Chem.* **1980**, *45*, 757.
(17) Cohen, N.; Colin, P.; Lemoin, G. *Tetrahedron* **1981**, *37*, 1711.
(18) DeClercq, P. J. *Tetrahedron* **1981**, *40*, 3717 (a program limited to joining ring templates).
(19) Wipke, W. T.; Hahn, M. *ACS Symp. Ser.* **1986**, *306*, 136 (a template assembly program which obtains its templates from an abbreviated X-ray database).
(20) Dolata, D. P.; Leach, A. R.; Prout, K. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 73.
(21) Bogel, H.; Sadowski, J. *Workshop "Computer in der Chemistry"*, 2nd, Magdeburg, Germany, Oct 2–4, 1990.
(22) Leach, A. R.; Prout, K. *J. Comput. Chem.* **1990**, *11*, 1193.
(23) Ruskino, A., III; Sheridan, R. P.; Nilakaten, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251.
(24) Wenger, J. C.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 29.
(25) Crippen, G. M. *J. Math. Chem.* **1991**, *6*, 307.
(26) Smellie, A. S. CONSTRICTOR, Oxford Molecular Ltd., Terrapin House, University Science Area, South Parks Road, Oxford, U.K., 1989.
(27) Blaney, J. M. DGEOM, Quantum Chemistry Program Exchange; Department of Chemistry, Indiana University: Bloomington, IN, 1989.
(28) Crippen, G. M.; Smellie, A. S.; Richardson, W. W. *J. Comput. Chem.* **1992**, *13*, 1262.
(29) Rusinko, A., III; Skell, J. M.; Balducci, R.; Pearlman, R. S. CONCORD, University of Texas at Austin; distributed by Tripos Associates, St. Louis, MO, 1987.
(30) Brint, A. T.; Willett, P. *J. Mol. Graphics* **1987**, *49*, 5.
(31) Dyott, T. M.; Stuper, A. J.; Zander, G. S. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 28.
(32) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
(33) Potenzone, R., Jr.; Cavicchi, E.; Weintraub, H. J. R.; Hopfinger, A. J. Comput. *Chem.* **1977**, *1*, 187.
(34) Gund, P.; Andose, J. D.; Rhodes, J. B.; Smith, G. M. *Science* **1981**, *208*, 1425.
(35) Sasaki, S.; Abe, H.; Hirota, Y.; Ishida, Y.; Kudo, Y.; Ochiai, S.; Saito, K.; Yamasaki, T. *J. Chem. Int. Comput. Sci.* **1978**, *18*, 211. Abe, H.; Fujiwara, I.; Nishimura, T.; Okuyama, T.; Kida, T.; Sasaki, S. *Comput. Enhanced Spectrosc.* **1983**, *1*, 55. Sasaki, S.; Abe, H.; Takahashi, Y. *TUTORS*-system.
(36) Levitt, M. *J. Mol. Biol.* **1974**, *82*, 393.
(37) Winter, J. G.; Davis, A.; Saunders, M. R. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31.
(38) Saunders, M.; Houk, K. N.; Wu, Y.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C.; *J. Am. Chem. Soc.* **1990**, *112*, 1419.
(39) Lipton, M.; Still, W. C. *J. Comput. Chem.* **1988**, *4*, 343.
(40) McCammon, J. A.; Karplus, M. *Acc. Chem. Res.* **1983**, *16*, 187.
(41) Bauer, J.; Schubert, W. *Comput. Chem.*, **1982**, (*2*), 61–65.
(42) Liljefors, T. *J. Mol. Graphics* **1983**, *1*, 111.
(43) White, D. N. J.; Tyler, J. K.; Lindley, M. R. *Comput. Chem.* **1986**, *3*, 193.
(44) Turk, D.; Zupan, J. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 116.
(45) Mekenyan, O.; Karabunarliev, S.; Bonchev, D. *Comput. Chem.* **1990**, *14*, 193.
(46) Mekenyan, O.; Karabunarliev, S.; Ivanov, J.; Dimitrov, D. *Comput. Chem.*, submitted for publication.
(47) Bonchev, D.; Balaban, A. T.; Mekenyan, O. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 106.
(48) Balaban, A. T.; Mekenyan, O.; Bonchev, D. *J. Comput. Chem.* **1985**, *6*, 538.
(49) Balaban, A. T.; Mekenyan, O.; Bonchev, D. *J. Comput. Chem.* **1985**, *6*, 552.
(50) Mekenyan, O.; Bonchev, D.; Balaban, A. T. *J. Comput. Chem.* **1985**, *6*, 562.
(51) Tarjan, R. E. *SIAM J. Comput.* **1972**, *1*, 146. Karabunarliev, S.; Mekenyan, O.; Ivanov, J. *Comput. Chem.*, submitted for publication.
(52) The standard MMPI parametrization.
(53) MNDO calculations (a private communication).
(54) Cahn, E. S.; Ingold, C. K. *J. Chem. Soc.* **1955**, 612. Cahn, E. S.; Ingold, C. K.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385.
(55) Ivanov, J.; Mekenyan, O.; Karabunarliev, St. Manuscript in preparation.
(56) Mekenyan, O.; Karabunarliev, S.; Bonchev, D. *J. Math. Chem.* **1990**, *4*, 207.
(57) Mekenyan, O.; Bonchev, D. *Acta Pharm. Yugosl.* **1986**, *36*, 225.
(58) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Bangov, I. *Arzneim. Forsch.* **1986**, *36*, 176.
(59) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Dimitrova, J. *Arzneim. Forsch.* **1986**, *36*, 629.
(60) Mekenyan, O.; Bonchev, D.; Rouvray, D. H.; Peitchev, D.; Bangov, I. *Eur. J. Med. Chem.* **1991**, *26*, 305.
(61) Mekenyan, O.; Bonchev, D.; Enchev, V. *Quant. Struct.-Act. Relat.* **1988**, *7*, 240.
(62) Mercier, C.; Mekenyan, O.; Dubois, J.-E.; Bonchev, D. *Eur. J. Med. Chem.* **1991**, *26*, 575.
(63) Bonchev, D.; Mountain, C. F.; Seitz, W. A.; Balaban, A. T. *J. Med. Chem.* **1993**, *36*, 1562.