Rapid Conversion of Molecular Graphs to Three-Dimensional Representation Using the MOLGEO Program

Ekaterina V. Gordeeva and Alan R. Katritzky*

Department of Chemistry, University of Florida, Gainesville, Florida 32611-2046

Vladimir V. Shcherbukhin and Nikolai S. Zefirov

N. D. Zelinsky Institute of Organic Chemistry, Leninsky prospect 47, 117913 Moscow, Russia

Received April 10, 1992

The MOLGEO program has been developed for the efficient conversion of molecular graphs into possible 3D molecular models. It rapidly generates acceptable 3D Cartesian coordinates from (i) the molecular connectivity and (ii) a table of bond lengths and angles which is user modifiable. Stereochemistry designated in the input structure is preserved throughout the optimization. The program accepts input files and provides output files in various formats as desired, including MOPAC, AMPAC, SMD, MMX, and MDL MolFiles. Our new depth-first search method for the rapid construction of a 3D model is described and compared with Crippen's embedding algorithm. Results of geometry calculation for various sets of structures are presented and discussed. Possible useful applications of the MOLGEO program for QSAR research are described.

1. INTRODUCTION

The rapid extension of computer applications in chemistry necessitates high accuracy and diverse requirements in the presentation of molecular structures. The vast majority of the computer programs operating in many different areas of organic and bioorganic chemistry represent structures as molecular graphs. This approach has deep historical roots and provides a simple and compact way to codify structural information. However, the graph-theoretical representation completely disregards the 3D structure of real molecules, although the spatial structure is crucial for physical and chemical properties of a molecule.

The problem of the rapid and reliable generation of molecular models solely from connectivity tables is presently highly significant in different areas of computational chemistry. The conventional methods such as molecular mechanics (MM)[1] or quantum chemistry packages,[2] while they produce quite precise geometry estimates, are time-consuming and, therefore, not convenient for handling large sets of structures. Moreover, all presently available packages of this type are very sensitive to the initial coordinates, and commencing with a 2D representation of the molecule could cause an optimization failure. Therefore, in applications of computational chemistry in which a large number of structures is treated or generated [as in, for example, quantitative structure–activity relationships (QSAR), computer-assisted synthesis, investigation of reaction mechanisms, etc.], the need to develop a rapid 2D → 3D convertor can hardly be overestimated.

Conventional methods for conformational calculations include numerous algorithms and programs designed especially for geometry calculations. The numerical result of the operation of a program is a set of coordinates which describes unambiguously the spatial structure of a molecule. The molecular mechanics and quantum chemistry methods are all based on the numerical minimization of the potential energy of a molecule, and the application of these methods requires that the initial geometry of a molecule be defined before starting the calculations. The choice of the starting coordinates is highly significant as it influences both the effectiveness of the energy minimization and whether convergency will be accomplished.

Some programs (e.g., SYMIN,[3] PRXBLD,[4] etc.) solve this problem by the application of 2D atomic coordinates taken from the structural drawings. However, this approach does not provide a completely satisfactory conversion since it depends on information not included in the graphical representation of structures.

Other methods are based on the designation of libraries of structural fragments and sets of heuristic rules for their combination. This approach is realized in a wide set of programs WIZARD,[5] AIMB,[6] ChemLabII,[7] MolBuild,[8] etc., and it appears to be very attractive at first sight due to its high efficiency. However, these programs are all extremely complex and require bulky databases with sophisticated rule sets. In addition, their applicability remains restricted by the available set of fragments and rules.

One can postulate two different template-free approaches to the problem of construction of a 3D molecular model for subsequent energy minimization. The first is a systematic search in conformational space, which is analogous to the way in which a chemist might go about the same task using a mechanical molecular model, namely, performing torsional rotations about all rotatable bonds at some convenient torsion angle resolution and retaining only those structures which pass geometrical tests designed to exclude high-energy molecular geometries.[9a] The search in conformational space is usually conducted as a back-track procedure or, in other words, is a tree-searching technique. To reduce computational time, stereochemical constraints (e.g., tests for interatomic contacts) must be used to prune the branches of the tree which apparently lead to the high-energy conformers. This approach was implemented in several computer programs designed for refined conformational analysis, such as MULTIC,[9a] RINGMAKER,[9b] SCA,[9c] CONCORD,[9d] and WIZARD[9e] (see also applications of this approach in refs 10a,b).

The second is an original template-free method for graph to 3D conversion based on the principles of distance geometry. This approach was suggested by Crippen[11a] and implemented in the DGEOM program.[11b] The distance geometry approach was successfully applied for modeling structural peculiarities in QSAR research.[11c]

0095-2338/93/1633-0102$04.00/0

Recently, three of us have discussed the ICAR program for the investigation of the mechanisms of reactions which pass through carbocationic intermediates.[12] Rearrangements of carbocations often proceed along multistep and branched pathways with huge numbers of intermediates involved. Combinatorial enumeration of the possible transformations requires that a strict system of the evaluation of solutions is used to prevent a combinatorial explosion. Here we encountered a problem when a rapid evaluation of geometry for structures was needed, and structures were represented as molecular graphs, with connectivity information only.

Thus, we realized the necessity to use a program for fast conversion of molecular graphs to 3D models. The program which could facilitate our investigation of carbocationic rearrangements, had to be able (i) to treat numerous intermediates for a reasonable time on a personal computer and (ii) to handle the complex caged structures which often appear in the rearrangement process. It was also very desirable that the program for molecular geometry optimization should be available in source code so that the procedure of 3D geometry optimization could be incorporated in the ICAR package in the most efficient way. The application of precise but time-consuming procedures such as quantum chemical methods (MOPAC) or advanced molecular mechanics (MODEL) seemed impractical because time and memory expenses are hardly affordable for personal computers. The necessity of treating unusual and strained molecules excluded the applications of template-oriented methods. These circumstances prompted us to create the MOLGEO program described below.

2. MOLGEO PROGRAM

The MOLGEO (MOLecular GEometry Optimization) program incorporates two different algorithms for converting a molecular graph to the corresponding 3D model: an embedding algorithm (distance geometry algorithm) based on Crippen's work[11] and our DFS algorithm, which is an implementation of a technique for conformational search. The DFS algorithm is especially effective for the evaluation of flexible molecules. It usually generates much better coordinates than those from the embedding algorithm, and it works faster since it does not require eigenvalues and eigenvectors to be evaluated (see also Table II). Moreover, the algorithm generates a model which fits the predefined stereochemical requirements. The embedding algorithm could be more effective in some rare cases where the approximation of fixed bond lengths and valence angles is strongly violated. To make the MOLGEO program more versatile, both algorithms are implemented in the current release. Comparisons of the performance of DFS vs the embedding algorithm will be presented in the Results and Discussion section.

The 3D model obtained in either way (DFS or embedding algorithm) is used as the initial approximation in the module of geometry optimization. Optimization is performed through three passes. Initially, the coordinates of the atoms which constitute the molecular skeleton are refined. Then the locations of the hydrogen atoms are evaluated. The conjugate gradients minimization procedure, suggested by Fletcher and Reeves,[13] is used to make these two preliminary steps faster. The third step includes the final refinement of the whole structure using the quadratic quasi-Newton[14] optimization procedure. When the starting point is not very far from the minimum, this algorithm is much more efficient than linear procedures.

Historically, the first version of the MOLGEO (MOLecular GEometry Optimization) program used only the embedding algorithm. However, we noticed some drawbacks of the distance geometry method:

(i) The quality of initial coordinates generated by the embedding algorithm depends strongly on the structural pecularities of a molecule under consideration. The program appears to generate very poor coordinates for molecules possessing large rings or long chains. Inconsistency in the choice of different but dependent elements of the distance matrix results in severe distortions of bond lengths and valence angles in the generated structure. The poor quality of the initial coordinates can highly complicate the subsequent optimization or even preclude it from reaching convergence.

(ii) Obviously the distance matrix is invariant in relation to the permutation of substituents at stereocenters. Hence, it is impossible to generate a model with a predefined geometry for any chiral structure. Stereocenter configurations are adjusted only during the geometry optimization. This requires additional amounts of CPU time and sometimes does not proceed smoothly.

(iii) The embedding algorithm is relatively inefficient and unstable since it requires the evaluation of eigenvalues and eigenvectors. Despite using the recommended[11] effective "exhaustion" algorithm,[15] the time required to accomplish convergence can be quite large.

Reasons i–iii prompted us to develop the second version of the MOLGEO program, which incorporates an alternative template-free method, namely, a systematic search in a conformational space. Thus, a depth-first search (DFS) algorithm for constructing the initial coordinates was created. The DFS algorithm is a back-track procedure for a search in conformational space with the application of stereochemical constraints at the earliest possible levels.

3. APPLICATION OF DEPTH-FIRST SEARCH FOR CONSTRUCTING MOLECULAR MODELS

The set of initial coordinates needs to satisfy three main requirements: (a) the molecular model must be embedded into 3D, (b) the model must not contain severe distortions of bond lengths or valence angles which can impede subsequent optimization, and (c) stereochemical constraints must be satisfied.

To derive a set of coordinates obeying criteria a–c, the MOLGEO program applies the method of "trial and error". This method is used extensively to solve various types of combinatorial and enumeration tasks, and it can be quite efficient, provided that (a) the search is reasonably restricted by suitable heuristics and (b) the search is performed systematically and nonredundantly.

The main reduction of the combinatorial search is achieved by the following way. The energy of a molecule can be approximated by the sum $E = E_{1,2} + E_{1,2,3} + E_{1,2,3,4} + E_{vdw}$. The first three terms are functions of the deviations of bond lengths, valence angles, and torsional angles from the standard values. The fourth term reflects interactions between nonbonded atoms (van der Waals interactions). The conformation of a molecule depends mainly on the first three terms, whereas the last is of smaller importance.

A molecule tends to adopt a conformation which conserves standard values of bond lengths and valence angles, and it is possible to exclude these parameters from the search. The task is, therefore, reduced to the generation of a self-consistent set of torsional angles which satisfies requirements a–c. The MOLGEO program solves this task by using the combinatorial enumeration technique. To evaluate a molecule which contains N_t independent torsional angles, which are scanned in steps

```
1  Logical procedure DepthFirstSearch
2      Assign coordinates to atoms 1,2,3
3      flag := DFS(3,0)
4      return flag
5  end of procedure DepthFirstSearch
6  Logical procedure DFS(current_atom,torsional_angle)
7  Evaluate Cartesian coordinates of the current_atom
8  Verify the compatibility of the position of the current_atom with
   structural and stereochemical constraints
9  If the constraints are satisfied
10     Declare current_atom to be fixed
11     Push the number of the current_atom into stack
12   else
13     return with flag := false
14 Evaluate the value of the initial_torsional_angle
15 flag := TRUE
16 For all the adjacent_atoms which are connected to the current_atom do
17   get the next adjacent_atom from the adjacency list
18   if the adjacent_atom is already fixed then
19     go to 33
20 for all the torsional_angles from the initial_torsional_angle to initial_torsional_angle +360° with step Δω
21   flag := DFS(adjacent_atom, torsional_angle)
22   if flag = true then
23     torsional_angle := torsional_angle + 180°/coordinative number of the current_atom- 1
24     go to 12
25   else
26     if torsional_angle is not fixed by the double bond then
27       torsional_angle := torsional_angle+ Δω
28       go to 21
29     else
30       flag := false
31       goto 34
32   end of loop for all torsional_angles
33 end of loop for all adjacent_atoms
34 if flag = false then
35     Pop numbers of atoms which were fixed after current_atom from stack and unfix the atoms
36 return flag
37 end of procedure DFS.
```

Figure 1. Flow chart of the DFS algorithm.

of $\Delta\omega$, it is necessary to investigate up to $(360°/\Delta\omega)^{N_t}$ points of conformational space, which is obviously possible only for small molecules. The efficiency can be significantly improved if the eligibility of intermediate conformations is verified during the process of generation, which enables the early elimination of most of the unreasonable solutions.

The program performs a depth-first search along the molecular graph with the assignment of assumed initial values of torsional angles to each successive group of four linearly linked atoms. At every step of the search, the program generates a new partial conformation which for the kth step includes $\omega_1, \omega_2, ..., \omega_k$ ($k \leq N_t$) torsional angles. Cartesian coordinates, which can be evaluated using k torsional angles, are then calculated using standard values of bond distances and valent angles. If the set of atomic coordinates fits criteria a–c, then the program recursively passes on to the $(k + 1)$th level. Otherwise the program starts scanning the conformational space by altering ω_k by increments of $\Delta\omega$ until an acceptable partial set of torsional angles is found. If the solution can not be established for the set of k torsional angles, then the program returns to the $(k - 1)$th step and ω_{k-1} is scanned. This procedure is repeated as described until a complete solution is generated or the conclusion is drawn that the task is unsolvable given the starting conditions. The method provides for the systematic and nonredundant investigation of the conformational space. The heuristics, which were used to reduce the search area, are described below.

We present now a more detailed description of the algorithm. The DFS procedure generates the 3D model which satisfies criteria a–c and is also acceptable for subsequent optimization (see Figure 1):

1–5: The DepthFirstSearch procedure initializes memory, assigns coordinates to the first three atoms, and then starts the DFS routine.

6: The DFS procedure constructs the whole molecule using the recursive depth-first search technique. The parameters are the sequential number of the atom under consideration and the value of the trial torsional angle which defines the orientation of the current atom in respect to the chain of the previous three atoms. The procedure returns "true" if the 3D model has been constructed successfully and "false" otherwise.

7: Evaluation of Cartesian coordinates of the current atom.

8: Obviously, not every combination of torsional angles can be used while constructing the 3D model of a molecule. Early recognition and elimination of ineligible partial solutions can strongly reduce combinatorial enumeration. Since conformations are described using the distance matrix, it seems to be rational to formulate the selection criteria in terms of interatomic distances.

To each pair of atoms there corresponds an interval of admissible interatomic distances $[l_{ij}, u_{ij}]$. However, it is generally impossible to fit this interval exactly due to the following reasons. First, the approximation of fixed bond lengths and valence angles is rarely fulfilled accurately. Second, the scanning of the conformational space is performed using fixed steps $\Delta\omega$. This is equivalent to the imposition of a grid with cells of size $\Delta\omega$ on the space of torsional angles. The precision of the determination of torsional angles thus depends on the value of $\Delta\omega$. The use of a small value for $\Delta\omega$ slows down the operation of the program, whereas the use of a large $\Delta\omega$ and, hence, large steps can result in the omission of potential solutions.

These two reasons necessitate the inclusion of a special parameter ϵ which defines the value of acceptable error and which is interpreted by the program as an extension of the interval to $[l_{ij} - \epsilon, u_{ij} + \epsilon]$. Testing the program revealed that ϵ = 1.0–1.1 Å is quite acceptable for evaluation of most structures and that it does not significantly worsen the quality of the 3D model obtained.

The algorithm of distance control just described allows the elimination of large distortions in the 3D structure. The stereochemical configuration is controlled by using the sign of the mixed product (**AB**,[**AC**,**AD**]) of three vectors **AB**, **AC**, and **AD** for the configuration of four substituents ABCD. The configuration of a stereocenter is considered fixed if the coordinates of the stereocenter and at least three of its substituents are fixed. The program evaluates the orientation of the fourth substituent and checks that the sign of the mixed product corresponds to the proper configuration.

9–13: If no violation of structural and stereochemical restrictions have been detected, then the atom under consideration is declared to be fixed, and its number is placed on the stack. If a violation is detected, the program returns to the previous vertex.

14: The program evaluates the value of the torsional angle ω_{init}, which defines the orientation of the subsequent atom in the chain. The angle is chosen within $-180° \leq \omega_{init} \leq 180°$, which gives the program an opportunity to generate a random conformer.

15: Initialization of the auxiliary **flag** variable.

16–19: Selection of the atom to be considered next from the set of atoms attached to the current atom and for which the coordinates are not so far established.

20: Beginning of the loop of scanning torsional angles. Scanning is performed for all angles from ω_{init} to $\omega_{init} + 360°$ in steps of $\Delta\omega$. Torsional angles can be sorted so that angles which are close to 180° will be taken in the first turn for the construction of molecular skeleton. Evidently, this cannot solve the challenging problem of the search for the global minimum, but it often results in the generation of more stable conformers.

21: Recursive call of DFS. The number of the subsequent atom and the trial value of torsional angle are passed on as parameters.

22–37: If DFS returns "true", it means that coordinates of all of the subsequent atoms in the chain attached to the atom

Table I. Comparison of 3D Optimization Results from MOLGEO Followed by MOPAC and MOPAC Only

structure	no.	MOLGEO + MOPAC, t_1, s	MOPAC only, t_2, s	t_2/t_1, 100%	MOLGEO + MOPAC, H_{f1}, kcal	MOPAC only, H_{f2}, kcal	$H_{f1} - H_{f2}$, kcal
trifluoromethanesulfonic anhydride	**1**	**86**	**460**	**535**	**-455.6**	**-388.5**	**-67.1**
perfluoropropionic anhydride	**2**	**302**	**490**	**162**	**-591.5**	**-417.2**	**-174.3**
perfluorobutyric anhydride	**3**	**473**	**1348**	**285**	**-771.4**	**-412.6**	**-358.8**
2,2-dimethyl-6,6,7,7,8,8,8-heptafluoro-3,5-octanedione	**4**	**378**	**1207**	**319**	**-419.8**	**-224.3**	**-195.5**
dodecyltrichlorosilane	5	305	417	136	-214.0	-217.1	+3.1
octadecyltrichlorosilane	6	1020	780	76	-255.0	-258.0	+3.0
dichlorodiphenylsilane	**7**	**61**	**58**	**95**	**-52.7**	**517.9**	**-570.6**
myristyl chloride	8	249	206	83	-128.8	-131.8	+3.0
cyclooctatetraene	9	28	184	657	63.5	63.5	0.0
norbornane	**10**	**19**	**94**	**376**	**-14.4**	**19.1**	**-33.5**
adamantane	**11**	**19**	**194**	**1021**	**-43.2**	**66.7**	**-109.9**
bromobenzene	12	7	17	242	26.75	26.75	0.0
bis(4-fluorophenyl)methane	13	78	491	629.5	-48.2	-47.98	-0.32
t-butylacetyl chloride	14	40	118	295	-69.8	-69.5	-0.3

under consideration have been successfully evaluated. In this case, the program passes to the next neighbor of this current atom and performs the depth-first search again. The standard difference of values of torsional angles for geminal substitutents is added to the trial torsional angle value. For instance, the difference for a sp^3 carbon atom equals 120°.

If DFS returns "false", then the procedure adjusts the current value of the torsional angle and calls itself again. If the torsional angle is fixed, e.g., by a double bond, and thus it cannot be altered, the partial solution is ineligible and must be discarded. In this case, all atoms which were fixed after the current vertex are declared as nonfixed, and their coordinates are zeroed.

4. RESULTS AND DISCUSSION

The MOLGEO program was incorporated in the GROUND[16a,b] (Generation and Recollection Of Updated Nonempirical Descriptors) package, which has been developed at the University of Florida. The package is especially designed for the calculation of a large set of molecular descriptors of various kinds including constitutional, topological, electronic, and combined descriptors, and to conduct subsequently QSAR/QSPR research using the descriptors.

The present version of the GROUND program includes many molecular descriptors which are mainly based on molecular geometry. Five of these, shadow indices,[17] shape parameter,[18] molecular volume,[19] solvent-accessible surface area,[20] and gravitation index,[21] are of pure geometrical nature. Charged partial surface area[22] is a combined descriptor which requires partial atomic charges to be evaluated in addition to the molecular geometry.

Structural sets to be optimized can include up to several thousand molecules that can be input in the form of MDL MolFiles[23] or in the GRD[16a] format. Application of molecular mechanics (MM2, MODEL) and quantum mechanics programs (MOPAC) for performing geometry calculations in such cases encounters three main obstacles. First, the files obtained from structural editors or retrieved from databases usually contain flat screen coordinates which provide a very poor initial approximation for geometry optimization procedure. This often results in the complete failure of the optimization, which is especially risky since the programs mentioned above, in general, do not alert the user to this situation. The only way to recognize whether the optimization performed is valid is to verify calculated structures visually or to compare the heats of formation with data obtained from an independent source, obviously a difficult and inefficient procedure for large sets of compounds. Second, molecular mechanics and particularly quantum calculations are enormously time-consuming for large arrays of molecules. Third, even if the direct application of screen coordinates is successful, it frequently results in the generation of the same final geometry, with no possibility to control violation of stereochemistry and perform conformational search.

All these reasons convinced us to perform a thorough analysis of the working ability of the MOLGEO program. The following trials have been performed.

(*a*) *Comparison of 3D models obtained through the MOLGEO optimization followed by the MOPAC optimization with 3D models optimized through the MOPAC only.* The MOLGEO program has been extensively tested on various structural sets, which included up to 3000 structures drawn in 2D representation. In all these trials, MOLGEO revealed itself to be a good and reliable source of 3D-optimized geometry. The high speed of the calculations enables us to optimize up to 1000 molecules of molecular weights 200–300 within 3 h of CPU time on a 25-MHz IBM PC AT 386/387. Moreover, the geometry optimized by the MOLGEO program can be further refined, if desired, using any one of the popular molecular mechanics or quantum mechanics programs MODEL, PCMODEL, MOPAC, or AMPAC. In such cases of further optimization, the total time of calculation was found to be diminished significantly in comparison with the optimization from scratch by any of these preexisting programs. The direct 2D → 3D conversion by either molecular mechanics or by a quantum chemical program frequently fails due to inappropriate initial coordinates. Prior treatment of molecular geometry by the MOLGEO program avoids such failures as well as accelerates the calculations. Advantages of the "MOLGEO + MOPAC optimization" in comparison with the "MOPAC only optimization" are shown in Table I. The table contains 14 structures treated (a) by the MOLGEO program followed by MOPAC and (b) by the MOPAC program only. The lines emphasized in boldface correspond to the high percentage of cases for which the MOPAC optimization failed to achieve the true energy minimization. Conformers of unreasonably high energy were generated during the MOPAC only sessions for structures **1–4**, **7**, **10**, and **11**, while the MOLGEO + MOPAC optimization was successful in all cases. All structures except **6–8** were optimized faster by MOLGEO + MOPAC than by MOPAC only (average acceleration of 300%). Structure **7** was not properly optimized by the MOPAC program alone, and the times of MOLGEO + MOPAC and MOPAC only sessions are, thus, in fact not comparable for skeleton **7** as well as for structures **1–4**, **10**, and **11**.

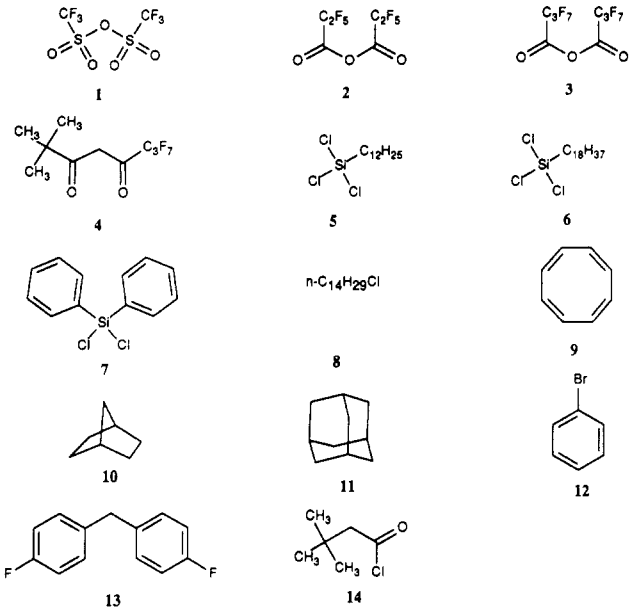

Figure 2. Structures from Table I.

(b) *Comparison of 3D models obtained from the MOLGEO program with data from Cambridge Structural Database.* A set of seven structures (Figure 2) was chosen randomly from CSD, and the corresponding Cartesian coordinates were converted to internal coordinates. The resulting values of bond lengths and valence angles were correlated with the lengths and angles obtained (i) from MOLGEO and (ii) from MOPAC. Torsional angles were excluded from the comparison to avoid problems of mismatching different conformations. Application of linear regression analysis revealed a good correlation between experimental and calculated data which confirms that MOLGEO provides a good reproduction of the experimental values of bond lengths and valence angles. Squared correlation coefficients are presented in the first two columns of Table II (see structures in Figure 3).

(c) *Comparison of 3D coordinates obtained by the depth-first search algorithm with 3D coordinates evaluated by the embedding algorithm.* Linear regression analysis was also used to compare the quality of 3D models obtained by the depth-first search algorithm and by Crippen's embedding algorithm without subsequent optimization. Correlation coefficients of the calculated geometry with experimental data are placed into the last two columns of Table II. The depth-first search method reveals significantly better results, which is particularly important if one takes into consideration that our method is significantly less time-consuming than that of Crippen.

(d) *Comparison of the energies for 3D structures derived from X-ray data vs 3D structures generated by MOLGEO from connectivity tables.* The comparison of energies for X-ray vs MOLGEO structures was conducted as follows:

(i) The 27 structures (Figure 4) were taken from randomly chosen sections of CSD and selected to provide a diverse set.

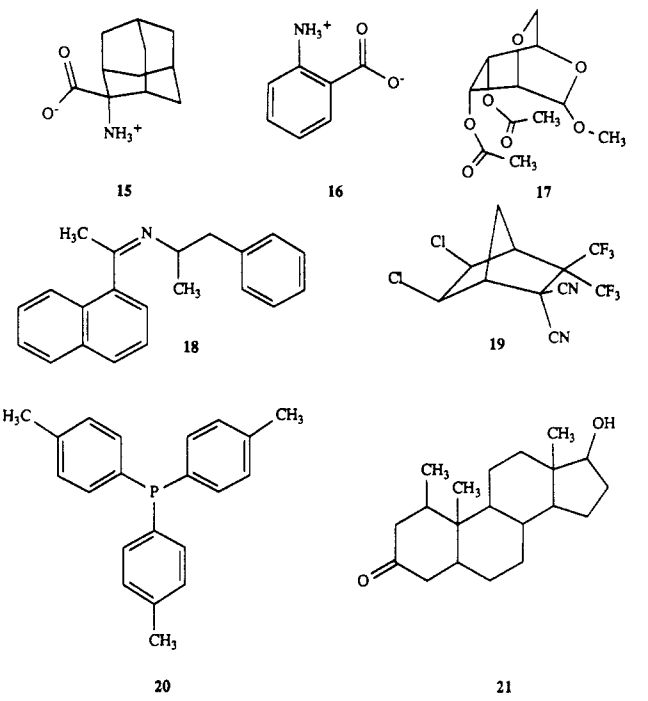

Figure 3. Structures from Table II.

As seen in Figure 4, the structures included in our study contain heterocycles, small rings, and bridged skeletons along with various substituents such as nitro, hydroxyl, carboxyl, carbonyl, halogens, and others. Also, the set includes *o*-, *m*-, and *p*-isomers (DOXKOQ, DOXKUW, and DOXLAD) and six steroid structures.

(ii) For each structure taken from the output file obtained during the CSD session, the X-ray coordinates were converted to the MOPAC format using the PCMODEL program (Serena Software, Inc.). This step was useful also for visualization of the structure, and thus, checking if connectivity has been correctly derived from the X-ray coordinates by PCMODEL. The connectivity information was stored in SYBYL format also available in PCMODEL. Then, the corresponding MOPAC file was loaded in the MOPAC program running with option 1SCF, i.e., no geometry optimization was performed. The heat of formation thus calculated is presented in second column of Table III.

(iii) The connectivity information from the SYBYL file obtained at the previous step was further used as an input for DFS algorithm with subsequent geometry optimization in the MOLGEO program. Then the 3D model generated by MOLGEO was converted to MOPAC format and loaded in MOPAC, again with option 1SCF. The heat of formation thus calculated is presented in the third column of Table III.

(iv) The heats of formation indicated in the fourth and fifth columns of Table III were calculated in a similar way, but 3D models obtained either from X-ray data or from MOLGEO were optimized in PCMODEL prior to MOPAC calculations.

Table II. R^2 Values for Correlations of Calculated and Experimental Bond Lengths and Valence Angles

		MOLGEO (opt)	MOPAC (opt)	DFS alg. (no opt)	emb. alg. (no opt)
compound	no.	CSD	CSD	CSD	CSD
2-aminoadamantane-2-carboxylic acid hydrobromide	**15**	0.985	0.986	0.920	0.668
anthranilic acid	**16**	0.997	0.997	0.997	0.787
methyl-3,4-di-*O*-acetyl-2,6-anhydro-α-D-altropyranoside	**17**	0.994	0.995	0.953	0.560
N-[1-(1′-naphthyl)ethylidene]-1-phenyl-2-propylamine	**18**	0.999	0.999	0.989	0.395
exo-cis-2,3-dichloro-5,5-dicyano-6,6-bis(trifluoromethyl)norbornane	**19**	0.996	0.999	0.979	0.715
tris(*p*-tolyl)phosphine	**20**	0.979	0.977	0.960	0.511
17β-hydroxy-1α-methyl-5α-androstan-3-one	**21**	0.998	0.998	0.977	0.640

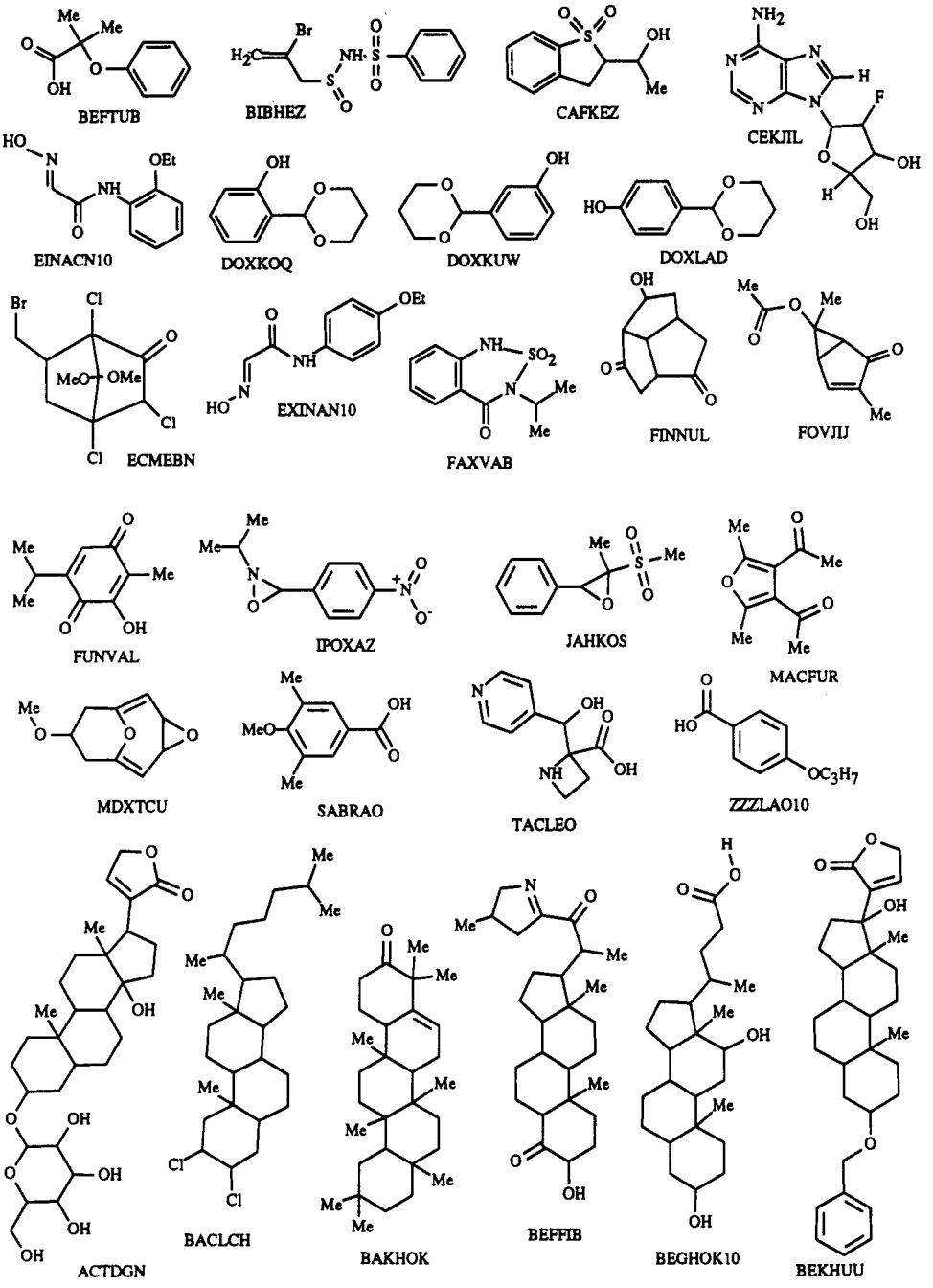

Figure 4. Structures from Table III.

(v) Finally, the sixth and seventh columns of Table III contain heats of formations calculated with the MOPAC program running in the mode of geometry optimization. The input MOPAC files were created as described above: see ii for conversion of X-ray data to MOPAC format and iii for conversion of connectivity information to MOPAC format via optimization in MOLGEO.

As seen from Table III, in the majority of cases the MOLGEO program generates a conformer with a lower energy than that corresponding to the X-ray structure. However, the comparison with X-ray conformers is not completely appropriate: due to packing effects, the energy of the X-ray structure is expected to be higher than for the completely relaxed structure. We found it more interesting to compare X-ray and MOLGEO conformers that were further optimized using MM and quantum semiempirical procedures. As shown in Table I, the MOPAC program is very sensitive to the initial conformation of structure under optimization, and heats of formation (see Table I) indicated the cases when different conformations resulted from optimization procedures. We expected that should MOLGEO generate a conformer very different from that corresponding to the X-ray data, then the heats of formation calculated via MOPAC would be very different. However, comparisons of the data in the sixth and seventh columns of Table III show that in the majority of cases the MOLGEO conformer is apparently very close to the X-ray conformer: after optimization in MOPAC the conformers became almost identical. The same conclusion can be derived after the comparison of data in the fifth and sixth columns of Table III: the optimization of both X-ray and MOLGEO structures in MM procedure results in almost the same conformer.

Although the MOLGEO program certainly cannot compete with well-developed and well-parametrized AM1 or MMX methods, the data from Tables I and III demonstrate that the MOLGEO program can be used efficiently for the rapid generation of a reasonable conformation from very poor initial coordinates or just from connectivity tables. For the most

Table III. Heats of Formation for CSD Structures Optimized in MOLGEO, PCMODEL, and MOPAC

CSD ref code	ΔH_f[a]		ΔH_f[b]		ΔH_f[c]	
	X-ray	MOLGEO	X-ray	MOLGEO	X-ray	MOLGEO
BEFTUB	+16.7	−50.6	−81.9	−79.4	−105.7	−104.8
BIBHEZ	+102.5	+21.5	−31.2	−31.5	−50.9	−66.2
CAFKEZ	+15.4	−51.2	−66.1	−66.3	−91.1	−91.7
CEKJIL	+123.6	+13.6	+11.3	+13.3	−92.3	−78.2
EINACN10	+85.3	+94.8	+15.6	+17.0	−46.4	−46.4
DOXKOQ	+18.9	+11.2	−88.7	−86.7	−111.2	−108.9
DOXKUW	−35.5	−51.5	−89.0	−86.4	−111.3	−110.4
DOXLAD	+18.6	−62.3	−89.7	−86.9	−111.9	−110.7
ECMEBN	+230.0	+152.2	−109.8	−97.2	−120.2	−114.3
EXINAN10	+291.4	+13.4	+14.6	+14.6	−46.4	−43.3
FAXVAB	+220.6	+129.9	−41.27	−41.4	−74.4	+32.1
FINNUL	−28.6	−97.4	−123.4	−121.7	−132.4	−132.2
FOVJIJ	+78.8	−100.2	−47.6	−59.7	−68.9	−67.6
FINVAL	−71.0	−68.0	−16.6	−48.4	−94.6	−94.8
IPOXAZ	+159.6	+138.1	+197.2	+197.3	+54.2	+54.3
JAHKOS	+53.9	+65.7	−11.2	−7.9	−39.8	−37.8
MACFUR	+282.1	−21.0	−53.5	−53.0	−84.1	−83.7
MDXTCU	+274.7	+2.0	−10.6	−10.6	−27.6	−27.7
SABRAO	+106.7	−46.5	−92.4	−92.5	−117.6	−51.7
TACLEO	+79.7	−25.4	−34.3	−27.9	−77.9	−77.9
ZZZLAO10	+30.5	−88.0	−95.0	−93.4	−118.6	−95.0
ACTDGN	+133.9	+145.1	−409.2	−365.6	−441.6	−406.9
BACLCH	−75.9	−80.2	−112.4	−111.7	−126.6	−151.1
BAKHOK	+407.4	+217.4	−69.4	−60.3	−66.1	−83.4
BEFFIB	−127.1	−72.3	−128.5	−117.8	−161.8	−152.4
BEGHOK10	+116.1	−144.5	−250.7	−240.1	−267.6	−257.8
BEKHUU	−75.7	−38.8	−143.7	−136.6	−165.9	−158.0

[a] No geometry optimization. [b] Geometry optimized in PCMODEL. [c] Geometry optimized in MOPAC.

effective conformational analysis, further optimization using the MM or quantum procedure is recommended. Thus, the MOLGEO program can be of most use as a fast preoptimization procedure.

(*e*) *Comparison of time requirements for DFS algorithm vs embedding algorithm.* Since the MOLGEO program incorporates both the DFS and the embedding method, it was interesting to compare the actual time consumption of the program while using either the DFS or the embedding algorithm. Although the embedding algorithm itself is, in general, slower than the DFS search, the time required for the calculation of the distance matrix and initial coordinates is still fairly small, even for a PC. However, the embedding algorithm frequently generates such poor initial coordinates that the subsequent optimization procedure is slowed down to a significant extent. Therefore, we conducted a comparison of the overall time required for obtaining the initial coordinates (in either DFS or embedding algorithm) followed by the optimization procedure. All times were measured with MOLGEO running on a PC 486/50.

Table IV shows the times of optimization using the MOLGEO program with the embedding algorithm and with the DFS method (emphasized in italics) for very flexible molecules (structures **22–39** in Figure 5), for less flexible structures (**40–44**), and for rigid molecules (**45–54** in Figure 5 and **10** and **11** in Figure 2). The program was run 11 times for each structure to provide a sufficient basis for statistical conclusions. In every run, the only input information was a connectivity table so that the result of the previous optimization would not affect the speed of the current calculations. The mean and median time of overall optimization was determined, along with the range for the fastest and slowest runs. As seen from Table IV, the DFS algorithm gives a better mean and median time in all cases, although in some individual runs the embedding algorithm can be faster than DFS. However, for several structures the intervals between slowest and fastest runs for DFS and embedding methods overlap only slightly, and for those structures the DFS algorithm was definitely more efficient than the embedding algorithm.

For bridged structures **40–44**, the DFS algorithm also showed higher speed (time in s, data for DFS are emphasized in italics):

	mean	median	range
40	29, *19*	31, *23*	19–45, *11–28*
41	25, *22*	25, *22*	21–29, *17–28*
42	24, *14*	24, *15*	20–37, *9–20*
43	19, *14*	18, *14*	12–26, *10–17*
44	14, *9*	14, *10*	10–17, *6–12*

For rigid structures **10** and **11** (Figure 2) and **45–54** (Figure 4), the speed of both algorithms was almost the same: 0 s (**45** and **46**), 1 s (**47** and **48**), 2–3 s (**49–52**), and 3–6 s (**53**, **54**, **10**, and **11**).

(*f*) *Comparison of structure–property correlations based on 3D models from the MOLGEO program, with correlations based on 3D models from MOPAC.* Since the geometry generated by MOLGEO has been proved to be reasonable (see sections a–d above), it was of great interest to compare structure–property correlations obtained using MOLGEO data vs correlations based on the molecular geometry optimized by the MOPAC program.

In one part of our QSPR (quantitative structure–property relationships) research, our goal was to derive dependencies of gas chromatography response factor (RF) on molecular structure. Regularity of this type is significant since it can be used for quantitative gas chromatographic analysis of compositions of complex mixtures.

The set of 121 structural 2D sketches[24] was initially converted to 3D models by the MOLGEO program, and then the structures were fit into MOPAC to produce a more precise geometry and to evaluate partial atomic charges. Then the complete set of 298 molecular descriptors was evaluated by the GROUND program for geometries optimized by MOPAC,

Table IV. Comparison of Time for MOLGEO Using Embedding and DFS Algorithms

structure no.	time, s[c]											mean value, s	median value, s	range, s
22	12[a]	10	10	9	9	9	9	9	9	8	8	9	9	8–12
	10[b]	*9*	*9*	*9*	*8*	*8*	*7*	*6*	*5*	*5*	*4*	*7*	*8*	*4–10*
23	20	18	15	13	12	12	12	11	11	11	11	13	12	11–20
	14	*14*	*10*	*10*	*8*	*8*	*8*	*6*	*6*	*6*	*5*	*9*	*8*	*5–14*
24	22	20	20	18	17	17	16	16	15	15	15	17	17	15–22
	14	*12*	*11*	*9*	*9*	*9*	*8*	*8*	*8*	*7*	*6*	*9*	*9*	*6–14*
25	29	28	26	23	21	20	19	18	18	17	17	21	20	17–29
	20	*19*	*19*	*19*	*19*	*13*	*12*	*12*	*9*	*8*	*7*	*14*	*13*	*7–20*
26	42	41	31	30	29	28	28	27	26	26	24	30	28	24–42
	24	*20*	*19*	*17*	*14*	*14*	*14*	*12*	*11*	*10*	*9*	*15*	*14*	*9–24*
27	46	44	40	40	39	37	36	33	31	30	29	37	37	29–46
	33	*31*	*31*	*29*	*29*	*25*	*24*	*21*	*20*	*14*	*12*	*24*	*25*	*12–33*
28	57	44	42	39	37	36	35	33	31	30	29	38	36	29–57
	41	*37*	*36*	*35*	*35*	*30*	*21*	*19*	*15*	*14*	*14*	*27*	*30*	*14–41*
29	57	56	54	54	50	47	42	42	39	38	37	47	47	37–57
	45	*30*	*28*	*27*	*27*	*26*	*24*	*21*	*19*	*19*	*15*	*26*	*26*	*15–45*
30	54	48	48	44	43	43	42	41	38	36	35	43	43	35–54
	31	*27*	*26*	*25*	*23*	*21*	*21*	*21*	*19*	*14*	*12*	*22*	*21*	*12–31*
31	152	125	120	113	108	106	105	102	100	97	95	111	106	95–152
	97	*88*	*88*	*83*	*77*	*70*	*65*	*53*	*52*	*50*	*39*	*69*	*70*	*39–97*
32	249	232	227	211	204	195	192	192	192	143	123	196	195	123–249
	192	*163*	*161*	*150*	*148*	*141*	*138*	*138*	*115*	*98*	*92*	*140*	*141*	*92–192*
33	92	91	88	75	75	69	69	68	67	63	60	74	69	60–92
	78	*68*	*56*	*53*	*50*	*48*	*45*	*44*	*38*	*37*	*32*	*50*	*48*	*32–78*
34	224	203	180	165	165	160	159	154	153	149	143	169	160	143–224
	203	*145*	*141*	*129*	*125*	*113*	*92*	*88*	*86*	*77*	*63*	*115*	*113*	*63–203*
35	126	108	99	92	91	73	69	67	66	63	61	83	73	61–126
	77	*68*	*66*	*66*	*63*	*57*	*43*	*43*	*42*	*40*	*38*	*55*	*57*	*38–77*
36	46	42	41	39	38	38	37	37	36	36	34	39	38	34–46
	39	*33*	*31*	*30*	*30*	*30*	*30*	*29*	*27*	*25*	*21*	*30*	*30*	*21–39*
37	103	91	90	83	82	80	75	75	73	72	70	81	80	70–103
	70	*69*	*67*	*61*	*57*	*53*	*51*	*50*	*49*	*38*	*37*	*55*	*53*	*37–70*
38	145	145	143	127	122	116	111	111	109	106	103	122	116	103–145
	138	*135*	*131*	*115*	*111*	*106*	*80*	*57*	*45*	*38*	*35*	*90*	*106*	*35–138*
39	280	235	235	228	202	201	194	191	188	183	182	211	201	182–280
	197	*188*	*174*	*160*	*150*	*135*	*135*	*133*	*132*	*86*	*40*	*139*	*135*	*40–197*

[a] Nonitalic numerals indicate runs of MOLGEO with embedding algorithm. [b] Italic numerals indicate runs of MOLGEO with DFS algorithm. [c] Times are sorted in decreasing values.

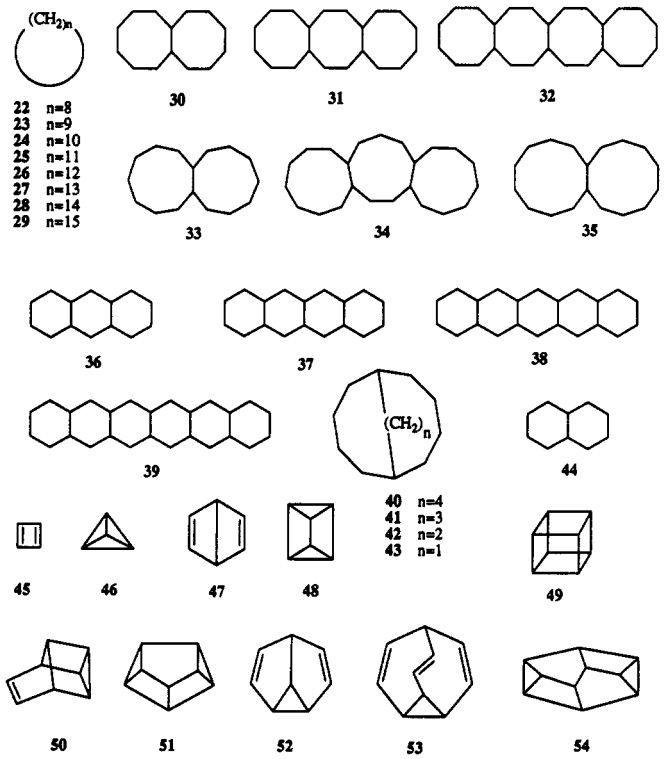

Figure 5. Structures used for comparison of performance of DFS and embedding algorithms.

as well as for the initial MOLGEO geometries, to enable subsequent comparison of correlations based on different sources of 3D structural data. Along with the partial atomic charges evaluated by the MOPAC program, the GROUND program performs two other independent procedures for partial charge calculation using Zefirov's[25] and Gasteiger's procedures,[26] both of which are fast and simple. Therefore, it was possible to compare correlations obtained using different sources of partial charges. The total set of 298 descriptors along with two subsets were generated using the different sources of 3D geometry and/or partial charges (Table V). The first subset of descriptors included the gravitation index[21] in combination with constitutional descriptors. The second subset consisted entirely of charged partial surface area descriptors.[22] The GROUNDSTAT[16b] package was then used to perform statistical analysis for the whole set and the two subsets, producing correlations which included at least one geometry-based parameter.

The set of 121 structures was subdivided into two subsets: a training set of 100 structures and a test set of 21 structures. Based on the regression modules obtained for the training set, predictions of the RF values were performed for the test set. The predicted RF values were correlated with the observed ones, and we also consider R^2 values for "observed–predicted" correlations in the test set of 21 structures. The R^2 values for multiparameter correlations are presented in Table V. As seen from Table V, the difference between the R^2 values obtained for regression models using 3D geometry from MOLGEO and from MOPAC is generally small. Therefore, in many QSAR/QSPR applications it seems possible to replace a time-consuming MOPAC optimization with the fast MOLGEO procedure.

Table V. R^2 Values for Correlations of Response Factor with Geometrical Descriptors Obtained from Different 3D Models and/or Different Partial Charge Distributions[a]

source of optimized molecular geometry	source of partial charge distribution	no. of parameters in regression model					
		1	2	3	4	5	6
MOLGEO[b]	GROUND	0.7860	0.8072	0.8327	0.8290	0.8201	0.8066
MOLGEO[b]	MOPAC	0.7860	0.8188	0.8321	0.8358	0.8280	0.8234
MOPAC[b]	GROUND	0.7982	0.7794	0.8230	0.8335	0.8401	0.8266
MOPAC[b]	MOPAC	0.7982	0.7794	0.7976	0.8221	0.8173	0.8102
MOLGEO[c]	GROUND	0.5050	0.5256	0.7261	0.7398	0.7437	0.7532
MOLGEO[c]	MOPAC	0.5050	0.5256	0.7261	0.7398	0.7437	0.7532
MOPAC[c]	GROUND	0.5100	0.5107	0.7180	0.7244	0.7368	0.7459
MOPAC[c]	MOPAC	0.5100	0.5107	0.7180	0.7244	0.7368	0.7459
MOLGEO[d]	GROUND	0.7860	0.7858	0.7790	0.7458	0.7460	0.7815
MOLGEO[d]	MOPAC	0.7860	0.7858	0.7840	0.7484	0.7453	0.7589
MOPAC[d]	GROUND	0.7982	0.7956	0.7744	0.7749	0.7730	0.7420
MOPAC[d]	MOPAC	0.7982	0.7973	0.8000	0.8001	0.7567	0.7830

[a] Only those correlations are compared which include at least one descriptor based on 3D representation of molecular structure.[7-22] [b] Total set of 298 molecular descriptors from the GROUND program. [c] Gravitation index[21] in combination with constitutional descriptors. [d] Charged partial surface areas.[22]

5. EXPERIMENTAL SECTION

The program is written in the C language following ANSI C standards. The first two versions of the program were designed especially for IBM PC; therefore, some modules have keywords which are specific for IBM PC. However, all these keywords are disabled when the program is compiled on another computer. Some auxiliary functions which are required to be supported by the compiler are also non-ANSI, but they are available in most modern C compilers' libraries. The MOLGEO 2.1 version has been ported to a VAX/VMS system and to an IBM RISC System/6000 under UNIX so the program should be ported easily to any other computer system which has a C compiler.

If the program is run on an IBM PC, then some restrictions should be taken into account: (i) the total amount of free memory must be not less than 400 K; (ii) a mathematical coprocessor, although not required, is very desirable to speed up the calculations; and (iii) the number of atoms including hydrogens must not exceed 130.

MOLGEO is used as a stand-alone executable file with all options pointed in the command line. This enables the creation of batch files to perform the calculation of large arrays of structures. MOLGEO can import and export molecular information using files written in MDL MolFile,[23] MMX,[2] MOPAC,[2] GRD,[16a] or SMD[27] format. Structural input can be also performed using a special mouse-driven graphics-oriented molecular editor.

The possibility to accept files written in different formats enables the use of MOLGEO as a format convertor. Geometry calculations are supressed in this case.

Empirical parameters (bond lengths and valent angles) which are necessary to perform calculations are stored in an external ASCII file, which enables modification and extension of the parameters set. The latest version of the MOLGEO program includes all these parameters for the following atoms: C, H, O, N, P, F, Cl, Br, I, S, Se, Si, B, Zn, Mg, and Li.[28]

6. CONCLUSION

The MOLGEO program has been developed to perform rapid conversion of a molecular graph to the corresponding 3D representation. Unlike most molecular/quantum mechanics programs, it is capable of generating spatial models merely from the connectivity table of the molecular structure. Since there is no one-to-one correspondence between a molecular graph and a spatial molecular structure, running the program requires stereochemical specifications, i.e., configurations at double bonds and stereocenters, and restrictions for the torsional angle values to be specified. Otherwise the program generates a random stereoisomer of the specified structural formula.

The MOLGEO program has the following advantages over the conventional molecular mechanics or quantum mechanics programs: (i) the program is capable of deriving Cartesian coordinates of atoms directly from the adjacency matrix of the molecular graph with no initial coordinates required; (ii) the program can construct 3D models which satisfy specified stereochemical requirements; (iii) the program works much faster than the conventional MM programs; and (iv) the program has a highly developed interface to packages with various formats.

The MOLGEO has been extensively tested on large sets of molecules. The generated 3D conformations revealed themselves to have reasonable energetic characteristics. If higher precision is required, then the coordinates can be refined using molecular mechanics or quantum mechanics programs. Convergence is achieved essentially faster, and the risk of optimization failure is significantly diminished in this case.

Comparative tests performed using the MODEL, MOPAC, and MOLGEO programs and data from Cambridge Structural Database demonstrated that the 3D geometry obtained from the MOLGEO program is reasonable and sufficiently precise to be used in many applications.

Therefore, the MOLGEO program can be considered as an effective and convenient tool for conversion of molecular graphs to the 3D representation, and it can be extensively used as a separate module in various chemical applications, especially where a large set of structures has to be rapidly optimized and/or evaluated from a stereochemical point of view.

REFERENCES AND NOTES

(1) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph 177; American Chemical Society: Washington, DC, 1982.
(2) Clark, T. *A Handbook of Computational Chemistry*; Wiley: New York, 1985.
(3) Wipke, W. T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W. SECS—Simulation and Evaluation of Chemical Synthesis: Strategy and Planning. *ACS Symp. Ser.* **1977**, *61*, 97–127.
(4) Wipke, W. T.; Verbalis, J.; Dyott, T. Three-Dimensional Interactive Model Building. Presented at the 162nd National Meeting of the American Chemical Society, Los Angeles, CA, 1972.

(5) Dolata, D. P.; Leach, A. R.; Prout, C. K. WIZARD: AI in Conformational Analysis. *J. Comput.-Aided Mol. Design.* **1987**, *1*, 73–85.
(6) Wipke, W. T.; Hahn, M. A. In *Artificial Intelligence. Applications in Chemistry*; Pierce, T. H., Hohne, B. A., Eds.; ACS Symposium Series 306; American Chemical Society: Washington, DC, 1986; pp 136–146.
(7) Potenzone, R., Jr.; Cavicchi, E.; Weintraub, H. J. R.; Hopfinger, A. J. Molecular Mechanics and the CAMSEQ Processor. *J. Comput. Chem.* **1977**, *1*, (3), 187–194.
(8) Liljefors, T. MOLBUILD: An Interactive Computer Graphics Interface to Molecular Mechanics. *J. Mol. Graphics* **1983**, *1*, (4), 111–117.
(9) (a) Lipton, M.; Still, W. C. The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space. *J. Comput. Chem.* **1988**, *9* (4), 343–355. (b) Still, W. C.; Galynker, I. Chemical Consequences of Conformation in Macrocyclic Compounds. *Tetrahedron* **1981**, *37* (23), 3981–3996. (c) De Clercq, P. J. Systematic Conformational Analysis. A Microcomputer Method for the Semiquantitative Evaluation of Polycyclic Systems Containing Five-, Six and Seven-membered Rings. 1. Program Characteristics. *Tetrahedron* **1984**, *40* (19), 3717–3727. (d) Pearlman, R. S. *Chem. Design Autom. News* **1984**, *2*, 4. (e) Dolata, D. P.; Carter, R. E. WIZARD: Applications of Expert System Techniques to Conformational Analysis. 1. The Basic Algorithms Exemplified on Simple Hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, (1), 36–47.
(10) (a) Fukazawa, Y.; Usui, S.; Uchio, Y. Conformational Study of the Cembranolide Diterpene Denticulatolide by Molecular Mechanics Method. *Tetrahedron Lett.* **1986**, *27*, 1825. (b) Bruccoleri, R. E.; Karplus, M. Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling. *Biopolymers* **1987**, *26*, 137–168.
(11) (a) Crippen, G. M. *Distance Geometry and Conformational Calculations, Research Studies*; New York, 1981. (b) DGEOM, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN. (c) Ghose, A. K.; Crippen, G. M. The Distance Geometry Approach to Modeling Receptor Sites. In *Comprehensive Medicinal Chemistry*; Hansch, C., Ed.; Pergamon Press: New York 1990.
(12) Gordeeva, E. V.; Shcherbukhin, V. V.; Zefirov, N. S. The ICAR Program: Computer-Assisted Investigation of Carbocationic Rearrangements. *Tetrahedron Comput. Methodol.* **1990**, *3* (6B), 429–443.
(13) Fletcher, R.; Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput. J.* **1964**, *7*, (2), 149–154.
(14) Fletcher, R. FORTRAN Subroutines for Minimization by quasi-Newton methods, Theoretical Physics Division, Atomic Energy Research Establishment, 1972.
(15) Faddeev, D. K.; Faddeeva, V. N. *Computational Methods of Linear Algebra*; Freeman: San Francisco, 1963.
(16) (a) Katritzky, A. R.; Gordeeva, E. V. The GROUND Program: Generation and Recollection of Updated Nonempirical Descriptors, in preparation. (b) Katritzky, A. G.; Gordeeva, E. V. The GROUNDSTAT Package for Statistical Treatment of Data in QSAR Research, in preparation.
(17) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of Molecular Shape Applied in Studies of Structure/Activity and Structure/Property Relationships. *Anal. Chim. Acta* **1987**, *199*, 99–109.
(18) Radecki, A.; Lamparczyk, H.; Kaliszan, R. A Relationship Between the Retension Indices on Nematic and Isotropic Phases and the Shape of Polycyclic Aromatic Hydrocarbons. *Chromatographia* **1979**, *12*, 595–599.
(19) Higo, J.; Go, N. Algorithm for Rapid Calculation of Extended Volume of Large Molecules. *J. Comput. Chem.* **1989**, *10*, 376–379.
(20) Jinno, K.; Kawasaki, K. Correlation between the Retention Data of Polycyclic Aromatic Hydrocarbons and Several Descriptors in Reversed-Phase HPLC. *Chromatographia* **1983**, *17*, 445–449.
(21) Osmialowski, K., private communications.
(22) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure–Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323–2329.
(23) Polton, D. J. Installation and Operational Experiences with MACCS (Molecular Access System). *Online Rev.* **1982**, *6* (3), 235–242.
(24) Musumarra, G.; Pisano, D.; Katritzky, A. R.; Lapucha, A. R.; Luxem, F. J.; Murugan, R.; Siskin, M.; Brons, G. Prediction of Gas Chromatographic Response Factors by the PLS Method. *Tetrahedron Comput. Methodol.* **1989**, *2* (1), 17–36.
(25) Zefirov, N. S.; Kirpichenok, M. A.; Ismailov, F. F.; Trofimov, M. I. Scheme for the Calculation of the Electronegativities of Atoms in a Molecule in the Framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR (Engl. Transl.)* **1987**, *296*, 440–443.
(26) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3288.
(27) Bebak, H.; Buse, C.; Donner, W. T.; Hoever, P.; Jacob, H.; Klaus, H.; Pesch, J.; Roemelt, J.; Schilling, P.; Woost, B.; Zirz, C. The Standard Molecular Data Format (SMD Format) as an Integration Tool in Computer Chemistry. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 1–5.
(28) *Interatomic Distances*, Supplement; Sutton, L. E., Ed.; Special Publication No. 18. The Chemical Society: London, 1965.