Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures

Jens Sadowski and Johann Gasteiger*

Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

Gerhard Klebe

BASF-AG, Hauptlabor, Carl-Bosch-Strasse, D-67056 Ludwigshafen, Germany

Received October 29, 1993®

Several criteria were defined to select a dataset of high-quality X-ray structures from the Cambridge file resulting in 639 molecules. Six currently available programs for automatic 3D structure generation were compared by converting the connectivity tables including appropriate stereodescriptors from this dataset of 639 molecular structures into 3D geometries: CONCORD, ALCOGEN, Chem-X, MOLGEO, COBRA, and CORINA. The geometries produced by the different programs were evaluated in terms of several quality criteria and are discussed in detail. These criteria measure how well the different programs reproduce the X-ray geometries of the 639 input structures. Accordingly, the major strengths and weaknesses of the programs are indicated.

INTRODUCTION

Biological activity is clearly a function of the 3D structure adopted by an organic molecule. In the pharmaceutical industry, 3D database searching for a given pharmacophoric pattern is an important strategy to find new lead compounds targeted toward a specific activity. This task requires large 3D databases containing high-quality structures from a wide variety of organic compounds.

The largest source of experimentally determined molecular structures is the Cambridge Crystallographic Database,[1] which contains at present about 100 000 X-ray structures. Nevertheless, this amount of data is often not sufficient for the purpose of pharmacophoric pattern searching. Furthermore, the data in the Cambridge file are rather heterogeneous, and some classes of compounds are only represented by a few examples, if not completely missing. Especially for the generation of new lead compounds, structural data of known pharmaceuticals or in-house data of compounds synthesized within the pharmaceutical company would be highly desirable. As a consequence, in recent years a number of 3D structural databases with some hundreds of thousands of computer-generated structures have become available.[2] Most of these data have been obtained from the popular 3D structure generator CONCORD.[3] However, a number of additional programs for automatic 2D-to-3D conversion have been reported (for a recent review see ref 4).

Although some of these programs are widely used in constructing large 3D databases, only a few studies were communicated on experiences with CONCORD[5] and the Chem-X 2D-to-3D builder.[6] In two recent studies, Milne et al. compared CONCORD-generated structures[5d] and 3D structures obtained from the Chem-X builder[6b] with X-ray data using a small dataset of 90 molecules. No further attempt to compare a larger dataset of computer-generated geometries with those determined experimentally has yet been reported. In any case, the decision to use a specific conversion program should only be made after a very careful evaluation process. Once a particular 2D-to-3D conversion program has been selected, switching to another program or even program version might involve a number of difficulties and incompatibilities. First, the required computer resources for the conversion of hundreds of thousands of structures are rather large. Second, the reproducibility of scientific work based on the generated database cannot be guaranteed. Accordingly, any change might lead to questionable or obsolete results. Therefore, the quality of the generated 3D coordinates is of great importance.

In the course of the development of the 3D structure generator CORINA,[7] we performed a comparison of several automatic 3D model builders using a dataset of 639 X-ray structures. We tried to perform a comprehensive evaluation of all those programs that were known and made available to us. The methods and results of this study are presented in this paper.

EXTRACTION OF THE DATASET FROM THE CAMBRIDGE STRUCTURAL DATABASE

An automatic retrieval of structures from the Cambridge Structural Database[1] (CSD version of February 1992 with 90 496 entries) was performed using the Crystal Structure Search and Retrieval program system (CSSR).[8]

Extraction Criteria. The following criteria for extracting the dataset have been used:

- The dataset should be of the size of 500–1000 structures.
- The experimental structures should be of high quality.
- The dataset should represent a wide variety of organic compounds.
- Since rings represent a special challenge in 3D structure generation, the originally retrieved structures should contain at least one ring.

Extraction Steps. The above criteria have been applied in the following sequence. First, 19 chemical classes according to the definition in the CSD were selected. These classes represent a wide variety of organic compounds and include only structures with at least one ring. In particular, the

® Abstract published in *Advance ACS Abstracts*, March 15, 1994.

0095-2338/94/1634-1000$04.50/0

following chemical classes were selected (the class numbers of the CSD file are given in parentheses):

- benzene nitro compounds (15)
- anilines (16)
- phenols and ethers (17)
- monocyclic hydrocarbons (3,4,5-membered rings) (20)
- monocyclic hydrocarbons (6-membered rings) (21)
- monocyclic hydrocarbons (7,8-membered rings) (22)
- naphthalene compounds (24)
- anthracene compounds (26)
- hydrocarbons (two fused rings) (27)
- hydrocarbons (three fused rings) (28)
- bridged ring hydrocarbons (31)
- hetero-nitrogen (3,4,5-membered monocyclic) (32)
- hetero-nitrogen (6-membered monocyclic) (33)
- hetero-nitrogen (two fused rings) (35)
- hetero-nitrogen (more than two fused rings) (36)
- hetero-nitrogen (bridged ring system) (37)
- hetero-oxygen (38)
- hetero-(nitrogen and oxygen) (40)
- hetero-(nitrogen and sulfur) (41)

Note that this selection also comprises most of the other ring structures contained in the CSD classes 42–59. For example, pyrimidines and purines (class 44) are covered by classes 33 and 35.

Thus, a subset of 24 980 structures (28%) of the entire CSD (90 496 records) was obtained. In the next step, this set was reduced to structures with connectivity **and** coordinate information. Only those structures that had been refined to an R factor of less than 3% were considered. This resulted in 568 structures. Structures were removed that had been flagged to be erroneous.

Conversion to MDL MOLFILEs. The structures were converted into MDL MOLFILE format.[9] For this purpose, the connectivity information given in the *.ct files has to be projected onto the coordinates given in the *.xr files. This is not a simple one-to-one mapping, since the atom numberings in the *.ct and *.xr files are not necessarily the same. First, the atoms in both representations were brought into a canonical order using only the information on atom types and neighboring atoms given in both file types. Then, in an iterative procedure all possible projections of the atoms in the *.xr files onto the equivalent atoms in the *.ct files were checked. The procedure stops if an error-free projection is achieved. Structures were removed for which inconsistencies between the *.ct and *.xr files were obtained. Aromatic bonds and nitro groups were converted into the required valence bond notation with alternating single and double bonds and localized charges, where necessary. Stereodescriptors were calculated for asymmetric atoms. Missing hydrogen atoms were added. Records containing more than one molecule in the asymmetric unit were split into single molecule entries. Molecules with less than five atoms (also containing hydrogen atoms) were removed. The crystal coordinates were transformed into Cartesian coordinates. Finally, a total of 639 molecular structures were obtained. The CSD Refcodes of these structure are given in the Appendix. It is assumed that the described procedure revealed a widely and well spread dataset.

Characteristics of the Dataset. Some simple statistics to characterize the selected dataset of 639 X-ray structures were calculated: (1) The number of structures containing elements other than H, C, N, O, F, Cl, Br, I, P, and S was determined to be 43 molecules or 6.7%. (2) The number of non-hydrogen atoms, $N_{non\text{-}H}$, in each particular molecule was determined. Table 1 gives the number and percentage of structures in intervals of 10 non-hydrogen atoms. The majority of structures has less than 30 non-hydrogen atoms. (3) The number of rotatable bonds, N_{rotor}, per molecule was determined (Table 2). Rotatable bonds are defined as only those bonds that are not contained in a ring and bear at least one neighboring non-hydrogen atom at either one of the two terminal atoms of the considered bond. Accordingly, isobutane has no rotatable bonds, whereas n-butane has one. (4) To capture an indication on the complexity of the ring systems contained in the dataset, the complexity measure C_R was estimated by[10]

$$C_R = \sum S_i / N_{RA} \quad (1)$$

C_R is the sum of the ring sizes (S_i) of all fundamental rings i in a molecule divided by the total number of atoms (N_{RA}) composing the rings. For molecules having only isolated rings, the C_R value is 1. Fused or bridged ring systems obtain a $C_R > 1$; molecules with no rings have a C_R value of 0. For example, decaline obtains a C_R value of $(6 + 6)/10 = 1.2$ and cubane a C_R value of $(5 \times 4)/8 = 2.5$. Table 3 gives the number of structures that fall into intervals of 0.2 of the C_R value. Fifty-one molecules (8%) are acyclic ($C_R = 0$). The latter type of structures, originally not intended to be included in the dataset (retrieval of examples with at least one ring) were obtained by the procedure of splitting those entries of the CSD which possess more than one molecule in the asymmetric unit. Most of the structures (51%) contain only isolated rings ($C_R = 1$). (5) The number of molecules containing asymmetric centers was determined to be 213, and the number of molecules containing stereochemically relevant double bonds not included in rings with less than eight atoms (smaller ring sizes imply Z-configurations in the ring skeleton) was 35.

Table 1. Numbers of Non-Hydrogen Atoms in the Dataset of 639 X-ray Structures

$N_{non\text{-}H}$						
5–10	10–20	20–30	30–40	40–50	50–60	>60
141 (22%)	295 (46%)	168 (26%)	30 (5%)	4 (1%)	0	1

Table 2. Numbers of Rotatable Bonds in the Dataset of 639 X-ray Structures

N_{rotor}						
0	1	2	3	4	5	>5
248 (39%)	93 (15%)	83 (13%)	77 (12%)	48 (8%)	28 (4%)	62 (10%)

Table 3. Complexity of the Ring Systems in the Dataset of 639 X-ray Structures

C_R						
0.0	1.0	>1.0–1.2	>1.2–1.4	>1.4–1.6	>1.6–1.8	>1.8
51 (8%)	325 (51%)	54 (8%)	141 (22%)	38 (6%)	21 (3%)	9 (1%)

EVALUATION PROCEDURE

The dataset described above was submitted to the various 3D structure generators. Figure 1 illustrates the evaluation procedure. Only connectivity information as expressed by the connection tables and stereodescriptors obtained from the local geometries as found in the X-ray structures (set A) were used to generate 3D atomic coordinates (set B). Computation times and the number of program crashes are reported as criteria for the performance of the different programs.

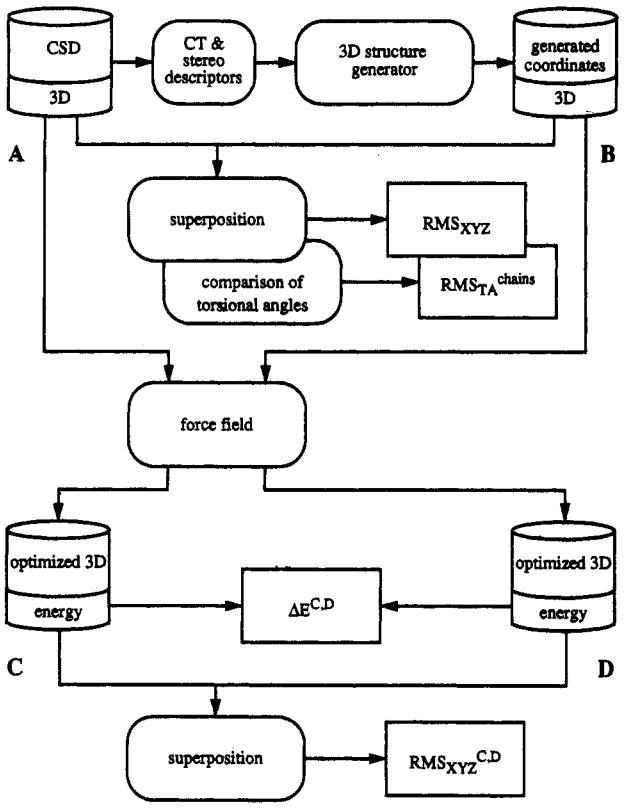

Figure 1. Flow diagram of the evaluation procedure performed independently for each 3D structure generator. CT, connection table; CSD, Cambridge Structural Database; RMS_{XYZ}, root mean square deviation of the Cartesian coordinates; RMS_{TA}, root mean square deviation of the torsional angles; ΔE, difference between the force field energies of the generated 3D models and their X-ray references after minimization by the force field (see text).

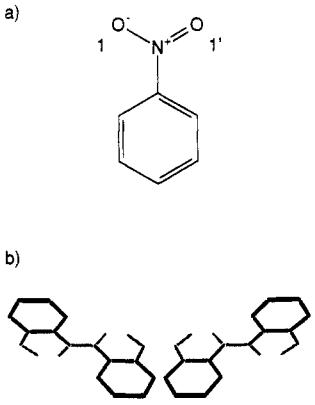

Figure 2. (a) Topologically equivalent oxygen atoms (1 and 1′) in nitrobenzene. (b) Two mirror symmetric conformations of 1,1′-binaphthyl.

Subsequently, the structures were checked for their quality in terms of close atom contacts and incorrect stereochemistry. In addition, the two sets of coordinates (A and B) were superimposed to determine the root mean square (RMS) deviation of the Cartesian coordinates of the atoms (RMS_{XYZ}) and of the torsional angles (RMS_{TA}).

The Euclidean distances between corresponding atoms were minimized using an approach described by Sippl and Stegbuchner.[11] Two additional problems had to be solved in this process. Figure 2a shows an example for the first type: nitrobenzene. The two oxygen atoms 1 and 1′ of the nitro group are topologically equivalent. Due to this C_S symmetry, both possible mappings, 1–1/1′–1′ and 1–1′/1′–1, have to be checked to find the lowest possible RMS value. Figure 2b illustrates a case of mirror symmetric conformers of 1,1′-binaphthyl. Matching image and mirror image will result in a large RMS deviation of energetically equivalent conformers. X-ray structures without asymmetric centers were accordingly superimposed either with the image or with the mirror image of the generated structure to obtain the lowest possible RMS deviation.

Molecular mechanics calculations provide energies as a measure of the quality of a structure. However, these energies are quite sensitive to small deviations in bond lengths and bond angles which are parametrized somehow differently in different force fields. To eliminate these effects inherent to a specific force field that would have mainly penalized structures that are already of reasonable quality, we relaxed both the generated and the experimentally determined structures into the next local minimum of a force field. The standard force field MM2 as obtained from QCPE could not be used because many parameters were missing. Our own implementation of the Chem-X force field[12] (excluding the electrostatic interactions) was used. This allowed us to control every single parameter and to use generalized parameters in cases where appropriate parameters were missing. The minimization of the experimental data by the Chem-X force field provided set C of conformers. The same minimization procedure was applied to the structures obtained from the 3D structure generators and gave set D. Comparison of these two sets C and D of optimized structures provided the energy differences $\Delta E^{C,D}$ and the root mean square deviations between the Cartesian coordinates of the atoms $RMS_{XYZ}^{C,D}$.

EVALUATION CRITERIA

The following criteria were defined to estimate the performance of the different programs and to assess the quality of the obtained results:

(1) The conversion rate is used as a measure of the general applicability of the programs.

(2) The number of program crashes is a measure of the robustness and reliability of the programs.

(3) Correctness of stereochemistry indicates the reliable conversion of chiral centers and the stereochemistry at double bonds.

(4) CPU time is used as a measure of the mean conversion time per structure. This time was approximately scaled to that on a VAX 6600 for comparison, since not all programs were available on a common hardware platform. For scaling, runs of the entire dataset of 639 structures with our program CORINA were performed on the various machine types.

(5) $RMS_{XYZ} < 0.3$ Å is a measure for the deviation of the generated structures from their X-ray reference structure. In our experience, for molecules of the size considered in this study a value of 0.3 Å represents an approximate limit between equivalent or different conformations.

(6) $RMS_{XYZ}^{rings} < 0.3$ Å is a measure of how often the programs reproduce the experimentally determined ring conformations. In cases with more than one ring system connected through acyclic chains, the RMS value was determined separately for each individual ring system and subsequently averaged by the number of ring atoms contained in the different systems.

(7) $RMS_{TA}^{chains} < 15°$ is a measure of the deviations of the torsional angles along the bonds in acyclic portions in the molecules. A value of 15° represents an empirical limit for different conformations. For example, for bonds between an

sp^2-type and an sp^3-type atom, the torsional potential shows approximately every 30° a local minimum.

(8) CCR is the close contact ratio, defined as the ratio between the actual interatomic distance and a minimal acceptable distance. For each molecule the smallest CCR of all nonbonded distances is reported. This criterion is similarly defined as the close ratio within CONCORD.[3c] An interatomic distance of less than 75% of the sum of the van der Waals radii of the atoms is attributed to be unacceptable. To get an idea of how this measure works, the distribution of the CCR values of the original X-ray geometries of the 639 test structures has been determined: 94% of the X-ray structures have a CCR of ≥1.0, 3% of 0.9–1.0, and 2% of 0.8–0.9. Thus, structures with a CCR > 0.8 seem to be acceptable. The criterion **CCR > 0.8** refers to structures with no problems caused by close contacts. On the contrary, the criterion **CCR < 0.5** refers to structures with rather serious close contacts caused by nearly identical positions of two or more atoms. These two criteria are measures of how reasonable the considered programs cope with nonbonded interactions.

(9) $N_{\text{insufficient}}$ is the number of structures with a geometry of insufficient quality. In these cases a difference between the force field energies of more than 50 kcal/mol ($\Delta E^{C,D}$ in Figure 1) has been obtained. Furthermore, structures with an RMS between the ring atoms of the optimized structures ($RMS_{XYZ}^{C,D}$) of more than 0.2 Å **and** an energy difference of more than 5 kcal/mol were considered to be insufficient. These are structures that do not reproduce the geometries of the ring systems and represent energetically less stable conformations. This criterion is used to detect rather poor 3D models.

EXPERIMENTAL SECTION

In this section the program versions used in this study are given together with options and parameters required to reproduce the results. Input to all programs was as MDL MOLFILEs.[9]

CONCORD.[3] Originally, program version 2.9.7 on a VAX 6600 under VMS was used. However, to keep this study as current as possible, the latest version 3.0.1, which was released in July 1993, has been used.

ALCOGEN.[13] Program version 1.02 on a Sun SPARC station 10/20 under SunOS 4.1.3 was used. The ALgorithmic COnformations GENerator ALCOGEN is a program for conformational analysis which generates for a given 2D connection table with stereodescriptors a set of 3D structures with low energy. The program was forced to generate only the lowest-energy conformation of the input structures by the following options:

INPUT (MOFILE) = *input file*
OUTPUT (MOFILE) = *output file*
CONFORM = 1
EXECUTE

Chem-X 2D-to-3D Builder.[6,14] The Chem-X program system of January 1993 on a VAX 3800 under VMS was used. The input structures were all read in batch mode into a temporary database. During the read-in procedure 3D coordinates were generated. This required the dimension flags of the input structures in the MOLFILEs to be set from "3D" to "2D". Finally, the temporary database with the newly generated 3D coordinates was written to disk. The following logfile recommended by Chemical Design Ltd. to create a 3D database was used:

write/dbs seg
/create temp
/finish
read/dbs maccs
/open *input file*
/dbs_write *
/finish
read/dbs maccs
/create *output file*
/write/answer_set all_segments
/finish
exit

MOLGEO.[15] Program version 2.4 on a VAX 3800 under VMS was used. MOLGEO is a combination of distance geometry[16] and internal coordinate conformation search. The program can only process single entry input structures. Thus, it was necessary to split the input file into single records and to convert each of them by consecutive calls of the program. The following options recommended by the program authors were used:

- I0, generation of an initial geometry for the distance geometry optimization by the internal coordinates conformation search procedure;
- /W3, output in the MOFILE format;
- /T5, break condition for the optimizer.

The following recommended parameters for the internal coordinate conformation search were written into an "options.dat" file: "60.0 1.2 0".

COBRA.[17] Program version of June 1993 on a Sun SPARC station 10/20 under SunOS 4.1.3 was used. COBRA uses many of the techniques that were developed as part of the WIZARD project.[18] The program is mainly targeted for conformational analysis. The default settings in this version are an attempt to tailor the program so that it generates a single conformation as rapidly as possible, though this structure will not necessarily be the lowest-energy structure. The distance geometry capabilities[17b] were switched off. The program is capable to read but not to write MOLFILEs. Thus, output in the PDB[19] format was produced. Subsequently, the generated coordinates were remapped onto the initial MOLFILE structures.

CORINA.[7] Program version 1.5 of May 1993 on a Sun SPARC station 10/20 under SunOS 4.1.3 was used.

RESULTS AND DISCUSSION

Table 4 summarizes technical details of the program jobs. The conversion rates, the numbers of program crashes, the CPU times for the jobs of the various 3D structure generators, and the machine types are given.

Conversion Rate. The number indicates the scope of a program. CORINA achieved a substantially larger conversion rate (100%) in comparison with the other programs (74–84%). Conversion rates could be increased by adding further atom types, rules, or fragments. Whereas for most of the programs this can only be done by the program developers, COBRA and Chem-X provide features for the user to update the fragment libraries.

Robustness. CONCORD and ALCOGEN encountered one and two program crashes, respectively, a rather high rate

Table 4. Technical Details of the Program Runs

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
generated 3D models	534	503	473	502	479	639
conversion rate (%)	84	79	74	79	75	100
program crashes	1	2	0	0	0	0
CPU time (s)	75	433	1431	41865	1830	401
machine type	VAX 6600	Sun SPARC	VAX 3800	VAX 3800	Sun SPARC	Sun SPARC
CPU time (s), VAX 6600	75	397	154	4508	1672	368
CPU time (s) per molecule, VAX 6600	0.14	0.79	0.33	8.98	3.49	0.58

Table 5. Incorrectly Assigned Stereochemistry in the Generated Models

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
asymm atoms	0	1	6	1	0	0
double bonds	0	0	17	0	0	0
total	0	1	23	1	0	0

Table 6. Numbers of Structures within 0.3-Å Intervals of the RMS Value of the Non-Hydrogen Atom Positions[a]

RMS_{XYZ} (Å)	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
<0.3	202 (38%)	202 (40%)	155 (33%)	94 (19%)	182 (38%)	267 (42%)
0.3–0.6	89 (17%)	94 (19%)	67 (14%)	84 (17%)	92 (19%)	103 (16%)
0.6–0.9	84 (16%)	79 (16%)	73 (15%)	77 (15%)	49 (10%)	95 (15%)
0.9–1.2	65 (12%)	43 (9%)	55 (12%)	63 (13%)	57 (12%)	59 (9%)
1.2–1.5	30 (6%)	33 (7%)	49 (10%)	67 (13%)	39 (8%)	43 (7%)
1.5–1.8	20 (4%)	23 (5%)	26 (5%)	38 (8%)	19 (4%)	27 (4%)
1.8–2.1	16 (3%)	8 (2%)	18 (4%)	33 (7%)	16 (3%)	19 (3%)
>2.1	28 (5%)	21 (4%)	30 (6%)	46 (9%)	25 (5%)	26 (4%)
total	534	503	473	502	479	639

[a] The percentages refer to the total number of structures converted by each of the different programs as given in the last row.

Table 7. Numbers of Structures within 0.3-Å Intervals of the RMS Value of the Ring Atom Positions[a]

RMS_{XYZ}^{rings} (Å)	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
<0.3	474 (89%)	445 (88%)	419 (89%)	346 (69%)	425 (89%)	566 (89%)
0.3–0.6	47 (9%)	39 (8%)	46 (10%)	91 (18%)	41 (9%)	43 (7%)
0.6–0.9	6 (1%)	13 (3%)	6 (1%)	28 (6%)	11 (2%)	22 (3%)
0.9–1.2	4 (1%)	5 (1%)	2	18 (4%)	2	3
1.2–1.5	3 (1%)	1	0	17 (3%)	0	4 (1%)
1.5–1.8	0	0	0	2	0	1
1.8–2.1	0	0	0	0	0	0
>2.1	0	0	0	0	0	0
total	534	503	473	502	479	639

[a] The percentages refer to the total number of structures converted by each of the different programs as given in the last row.

considering the rather limited size of the dataset of 639 molecules. The two program crashes in ALCOGEN were investigated further. They both occurred with the perchlorate anion. Separate input of the perchlorate anion, however, resulted in smooth generation of 3D coordinates. Thus, the reasons for the program crashes remain obscure.

CPU Time. The CPU times for runs of the different programs were scaled according to run times obtained with CORINA on the various computers (VAX 6600/SPARC 10/20/VAX 3800 = 1.00/1.09/9.30). The reported timings are for building and reading/writing the molecules. For comparison, the relative CPU times required on average per molecule were determined. CONCORD requires extremely short computation times. MOLGEO and COBRA need substantially larger times (factors of 60 and 20, respectively); the other programs require CPU times that are quite similar, being by a factor of 2–5 longer than those of CONCORD.

Correctness of Stereochemistry. The entire dataset contained 213 structures with asymmetric centers and 35 structures with stereochemically relevant double bonds. Table 5 gives the numbers of incorrectly generated 3D models, listed individually for asymmetric centers and double bonds. Only the Chem-X builder gives a nonacceptable failure rate. Focusing on double-bond configurations, the Chem-X builder failed to correctly reproduce more than 50% of the structures. This suggests some deficiencies in the MOLFILE interface of Chem-X.

RMS_{XYZ}. The achieved values were grouped into 0.3-Å intervals. Table 6 gives the number of structures in each of these intervals (absolute and relative percentage). The percentages were calculated using the total number of structures converted by the individual programs. CORINA gives the best results; MOLGEO performs rather unsatisfactorily.

RMS_{XYZ}^{rings}. The RMS deviations of the atomic positions of the ring atoms were calculated and grouped into 0.3-Å intervals (Table 7). Due to the conformational restraints in cyclic systems, the probability for obtaining similar geometries

Table 8. Numbers of Structures within 15° Intervals of the RMS Value of the Torsional Angles of Acyclic Bonds[a]

RMS_{TA}^{chains} (deg)	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
<15	243 (46%)	253 (50%)	186 (39%)	182 (36%)	217 (45%)	318 (50%)
15–30	47 (9%)	52 (10%)	32 (7%)	19 (4%)	49 (10%)	53 (8%)
30–45	36 (7%)	39 (8%)	24 (5%)	25 (5%)	32 (7%)	40 (6%)
45–60	41 (8%)	25 (5%)	36 (8%)	42 (8%)	24 (5%)	40 (6%)
60–75	44 (8%)	28 (6%)	31 (7%)	44 (9%)	38 (8%)	52 (8%)
75–90	45 (8%)	21 (4%)	43 (9%)	76 (15%)	40 (8%)	45 (7%)
90–105	30 (6%)	32 (6%)	31 (7%)	59 (12%)	27 (6%)	35 (5%)
>105	48 (9%)	53 (11%)	90 (19%)	55 (11%)	52 (11%)	56 (9%)
total	534	503	473	502	479	639

[a] The percentages refer to the total number of structures converted by each of the different programs as given in the last row.

Table 9. Numbers of Structures within Different Intervals of the Close Contact Ratio CCR[a]

CCR	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
>0.8	486 (91%)	474 (94%)	337 (71%)	431 (86%)	415 (87%)	593 (93%)
0.5–0.8	30 (6%)	29 (6%)	75 (16%)	61 (12%)	53 (11%)	41 (6%)
<0.5	18 (3%)	0	61 (13%)	10 (2%)	11 (2%)	5 (1%)
total	534	503	473	502	479	639

[a] The percentages refer to the total number of structures converted by each of the different programs as given in the last row.

Table 10. Numbers of 3D Models with an Unsatisfactory Geometry

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
$\Delta E^{C,D} > 50$ kcal	8 (1%)	1	21 (4%)	11 (2%)	3 (1%)	1
$\Delta E^{C,D} > 5$ kcal AND $RMS_{XYZ}^{rings,C,D} > 0.2$ Å	20 (4%)	8 (2%)	33 (7%)	63 (13%)	23 (5%)	11 (2%)
$N_{insufficient}$	21 (4%)	8 (2%)	37 (8%)	67 (13%)	23 (5%)	12 (2%)

Table 11. Results with the Subset of 317 Structures Being Processed by All of the Considered Programs

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
$RMS_{XYZ} < 0.3$ Å[a] (%)	42	40	31	24	40	42
$RMS_{XYZ}^{rings} < 0.3$ Å[b] (%)	90	93	89	79	91	94
$RMS_{TA}^{chains} < 15°$ [c] (%)	45	45	37	33	46	44
CCR > 0.8[d] (%)	97	95	69	91	90	95
CCR < 0.5[e] (%)	1	1	13	1	1	1
$N_{insufficient}$[f] (%)	2	1	6	6	2	0

[a] Structures with an RMS value of the positions of the non-hydrogen atoms of less than 0.3 Å. [b] Structures with an RMS value of the positions of the ring atoms of less than 0.3 Å. [c] Structures with an RMS value of the torsional angles at acyclic bonds of less than 15°. [d] Structures with a close contact ratio of greater than 0.8. [e] Structures with a close contact ratio of less than 0.5. [f] Insufficient structures.

Table 12. Results with the Subset of 43 Structures Containing Other Than H, C, N, O, Cl, F, Br, I, P, and S Elements

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
conversion rate (%)	5	26	5	19	0	100
$RMS_{XYZ} < 0.3$ Å[a] (%)		45		38		60

[a] Structures with an RMS value of the positions of the non-hydrogen atoms of less than 0.3 Å.

between generated and X-ray structures (RMS < 0.3 Å) is higher for the ring parts than for the complete structures. All programs apart from MOLGEO obtain satisfactory results.

RMS_{TA}^{chains}. The RMS deviations of the torsional angles in the acyclic portions were sorted into intervals of 15° (Table 8). On average, a larger portion of the ALCOGEN- and CORINA-generated structures comes closer to the X-ray references compared to the other programs; less satisfactory are the results obtained by MOLGEO and Chem-X.

CCR, the close contact ratio, is shown in Table 9. ALCOGEN, CORINA, and CONCORD obtain the best results, with more than 90% of structures with an acceptable CCR of less than 0.8. Problems are indicated by the large number of structures with a CCR of less than 0.5 generated by Chem-X (13%). Note that Chemical Design Ltd. recommends the use of Chem-X generated 3D structures only in combination with conformational analysis and/or flexible search strategies. Thus, the large databases of Chem-X generated 3D structures are to be understood only as starting geometries for further computations.

$N_{insufficient}$. Table 10 shows the number of structures with an unsatisfactory geometry as defined by the evaluation criteria. The first two rows show the number of structures matching the two conditions $\Delta E^{C,D} > 50.0$ kcal/mol and ($\Delta E^{C,D} > 5.0$ kcal/mol AND $RMS_{XYZ}^{ringsC,D} > 0.2$ Å). The third row shows the number of structures matching both conditions for unsatisfactory geometries. ALCOGEN and CORINA reveal the best results, whereas the high percentage

Table 13. Summary of Results (Percentages Refer to the Total Number of Structures Converted by Each of the Different Programs and Not to the Total Number of 639 Structures in the Original Dataset)

	CONCORD	ALCOGEN	Chem-X	MOLGEO	COBRA	CORINA
conversion rate (%)	84	79	74	79	75	100
program crashes	1	2	0	0	0	0
stereo errors	0	1	23	1	0	0
CPU times (s) per molecule[a]	0.14	0.79	0.33	8.98	3.49	0.58
$RMS_{XYZ} < 0.3$ Å[b] (%)	38	40	33	19	38	42
$RMS_{XYZ}^{rings} < 0.3$ Å[c] (%)	89	88	89	69	89	89
$RMS_{TA}^{chains} < 15°$[d] (%)	46	50	39	36	45	50
CCR > 0.8[e] (%)	91	94	71	86	87	93
CCR < 0.5[f] (%)	3	0	13	2	2	1
$N_{insufficient}$[g] (%)	4	2	8	13	5	2

[a] On a VAX 6600. [b] Structures with an RMS value of the positions of the non-hydrogen atoms of less than 0.3 Å. [c] Structures with an RMS value of the positions of the ring atoms of less than 0.3 Å. [d] Structures with an RMS value of the torsional angles at acyclic bonds of less than 15°. [e] Structures with a close contact ratio of greater than 0.8. [f] Structures with a close contact ratio of less than 0.5. [g] Insufficient structures.

Chart 1

AMXNON10	ANFFUR	ANFFUR	ANLINT20	ATDZSA	DOXSEO	DOXSEO	DOXSIS	DOZMIO	DOZXOF
BHNACN	BOPTZO	BPYPNO	BRTPCP	BTNPOC	DUBKOA	DUDDOV	DUDMOE	DUDMUK	DUDNAR
BZAMTZ	BZMTZO	BZOXZO	CATDAC10	CTMTNA	DUFXEH	DUFXIL	DUKMOL	DUKMOL	DUVZAV
CYCYPR	CYPHEP	CYURAC12	DCLPAB	DMAFBZ	DUWKEL	DUZTAT	DUZTIB	FADXIR	FAFSAG
DMIMZT01	DMIMZT10	EPNPHD10	ETCYPY	FORBHC	FAFSAG	FALWUK	FALWUK	FANPAL	FAPWOI
FUACAM	GRISFL	HHXPOZ	HXMTAM03	KECYBU06	IPATZL01	PYRZOM01	CITRUS10	CITRUS10	FAGXIU
KTNASO10	MELAMI01	METHIC	METHPZ10	MHPTEC	FAKKOR	FAVKOC	FAVMUK	FAXMEW	FECPAE
MMANSC	MMORPB	MPASEP	MPTVTP	MSULIN	FECZES	FECZES	ADAMAN08	DUTNIP10	FADFUL01
MTFPTZ	MXTRYP	PALDIM	PCLDTR	PHNDMC11	FADFUL01	FADFUL01	FAFTOV10	FAFTOV10	FEGLUY
PHNDMO	PMTSZD	PROPRA10	SSULAN02	TASANF	FEGNOU	FEGNOU	FEGZEW	FEHCUQ	FELRUJ
THAMCU10	THAMCU10	TIPMAN	TMCCDM	TMCCDM	FELRUJ	FEMTIA	FENDAD	FENWOK	FENWOK
TMHPTC	TOYOCM10	TPYDTS10	TUBERC01	TXBNON	FESPAU	FESPAU	FEVNID	FEZDAP	FIBMUY
VIMNOL10	VIMNOL10	XANOPT	DCLNAQ	DIPKGA	FIBMUY	FIBNAF	FIDTER	FIDTIV	FIDTIV
KTDZOX	MENPPO	ACBUOL	BRPSYD10	DMHXCA	FIJFUZ	FILBIL	FILHEN	FILHEN	FILHOX
PYPTHI	AXMQOL	DCHXMO	DCHXMO	HTBCDC	FILHOX	FILHUD	FIMPEW	FINSUQ	FISHIY
MBTZCU	PARBAC02	BPTHNB	BQUINI	DKHOOD	FIVPIJ	FIVXUD	FIWWIR	FIXPIL	FIXWEO
IPDGLF	IPDGLF	QUOLTS10	BZTZTL	HBYPZT	FIXWIS	FIXWUE	FIXXAL	FOBKEM	FOGZOQ
HBYPZT	POXHDO	PRSMCA	TBCBCH	THMBNP	FOHMEU	FOHPAT	FOKRUS	FOKSAZ	FOKSED
TOHBPA10	BPSMPY	MAEPTZ	AZSTBI	BRNPHL	FOKSUT	FOKVAC	FONWIO01	FONYUC	FORHUP
CIMGUA	IMCLUR	IMCLUR	MORTRS10	NFMLEB	FORVEN	FOSDIA	FUCNAS	FUFHUJ	FUFVIL
TSEMOR10	TSEPIP10	MEPYRI	MOXADM	PAHQUN	FUJSOS	FUMWIT	FUNVUF	ZZZTVO01	BCPROP02
PYRCYM01	CBZPOX	CEMIND	CHXBQU	MEPHCB01	CESVIF10	CIPYAB10	CIRPIC10	CIVPOM10	CIVPOM10
MOPHLD	NOPHTZ	TCOCBR	THETAC	THPDCB	CIVVEI10	CLANIC05	COPREE10	CUPNUW10	CUPNUW10
DMOPHS	EPOCMO	EPOCMO	HYDTZA	OMPIPL	DEHSIS10	FAXTUT01	FEHCOK01	FIYTOW10	FOGVIG01
THZOLA	TSABAN	BAKBAQ	AZTPHZ10	AZTPHZ10	FURNIP	FURNIP	FURNIP	FUVWOI	FUVWOI
BARGIK	BARVEV	BATCEE	BAWFEK	BAWVUQ	FUWVUO	FUZXAZ	FUZZAB	GADGUN	GADGUN
BOIMPH10	NOPHTE01	NOPHTE01	THMBNP10	XMSITZ10	GAFPEI	GAHWUH	GAKGUU	GALXIA	GAMJAF
AMPHOM02	BAWKEP10	BEDJUP	BEDJUP	BEDNON	GANJEK02	GAPVEY	GAPWUP	GASJUF	GASYOO
BEFROT	BELDAX	IMAZOL06	PYRCYM03	PYRCYM04	GEBYER	GEBYER	GECVAL	GEFCUP	GEFJOQ
PYRIDS02	PYRIDS11	BESNUI	BEWXUW	BEXTED	GEGGAA	GEGGAA	GEHLIO	GEKVIB	GEKVOH
BINWAW	BIRDIP	BIRDIP	RBFRPZ10	BISNOG	GEKZIF	GEKZIF	GELVAU	GEPCAF	GEPCAF
BISNOG	BIXFOD	BIXFOD	BIXFOD	BIXFOD	GESLIZ	GEVWIN	GEXGOF	GEYSEI	GEZPOQ
BIXGOE	BIZSAE	BOBHOP	BOCLAG01	BOCPIS	GEZSEJ	GICJUX	GICLIN	GICLIN	GICLIN
BOHNUH	BOHREV	BOHREV	BOHREV	BOMJIW	GICLUZ	GICTIV01	GIFNOY	GIFRAO	GIFRAO
BOSPEE	EXOBDO01	GUPICL10	GUPICL10	GUPICL10	GIGPOB	GIGYOK	GIHMEP	GIKNUJ	GINMUL
GUPICL10	HMNCXB01	SLFNMG01	BOWWOZ	BUBYIG	GINWAB	GINWAB	GIRTIK	GIRTIK	GIRTIK
BUCJIS	BUDTEZ	BUDTEZ	BUDTEZ	BUDTEZ	GIRTIK	GIRTOQ	GIRTOQ	GIRXOU	GOBSAR
BIXPUT10	BUMDOC	BUMDOC	BUSFUQ	BUSKUV	GOBSAR	GOJNIC	GOJNIC	JABZIV	JADVOZ
BUWMOV	CADCOZ	CADCOZ	CADCOZ	CADMUP	JAFFUR	JAFGAY	JAGLIM	JAHKUY	JAHTOB
CADNAW	CADPOM	CAGREH	CAGTEJ	RFMPPM10	JAHXOF	JAKTIY	JALSIY	JALSIY	JALSIY
TRAZOL02	CANFUS	CANFUS	CANRAK	CAXLES	JAMKEN	JAMLUE01	JANWOK	JAPTUP	JASSUR
CAXNEU	CAXYUV	BAWVUQ01	BIGTIU01	BUYPUG10	KACREL	KAGLAF	KAGLAF	KAHHUW	KANZOO
CECHIB	CECHIB	CEHZUK	CEJJAC	CEJPAI	KANZOO	KARGEP	KARHOA	KARHOA	KARSAX
CENGUX	CENGUX	CENHAE	CENHAE	CEPRUK	KATBOW	MCHTEP03	MCHTEP04	MCHTEP04	PARBAC04
CEPRUK	CERMUH	CERRAS	CESVIF	PYCLSB01	SACSEU	SACYOK	SAFFUA	SAFSAT	SAGTUP
PYCLSB01	ZZZKVU10	ZZZKVU10	BIVBOX10	BUBYIG10	SAGTUP	SAKYUY10	SAKYUY10	SAPKID	SAPKID
CEZFES	CEZFES	CEZFUI	CEZJEW	CEZKOH	SARBOC	SASNOP	SASNOP	SASNOP	SAXCID
CEZXEK	CIDJOO	CIDJOO	CIGFAZ	CIHWUL	SAXYIZ	SAYCIE	SUTHAZ03	VAGVAA	VAGXIK
CIKSAQ	CIMBAB	CINVUQ	CINZAA	BUYTIY10	VAKXIO	VALSEG	VAMZOY	VAMZOY	VAPZOB10
CIXGOF	COBMUB	COCYOI	COCYOI	COCYOI	VARDOH	VARKOO	VARZOD	VAWMEL	VAWMOV
COGHIP	COHNES	COHNOC	COHNOC	COJMET	VUXRIP	VUYPOU	VUZGOM	VUZGUS	VUZGUS
COKROJ	COMDIR	COPHEU	COPREE	COPROO	ZZZWOU01	CIRPOI10	DAXTAX01	JALSIY10	JALSIY10
COPROO	COPRUU	COTXAK	COWNAD	QUDSUL20	JALSIY10	JAVFUH	JAVRED	JAWVOS	JAXZUD
ALOXAN11	CEBTEI10	COFNUG10	CONYAF10	COXMIL01	JAZDOD	JEBYEU	JEBYEU	JEDTAN	KAXXAI
COZYUL	CUGBAH	CUGBAH	CUGBAH	CUGCEM	KECMEK	KECMEK	KECMEK	KECMEK	KEDZIC
CUNNUU	COFJIQ10	CUXTOE	DACRAA	DADPAZ	KEDZIC	KEGJEL	SECSAU	SEFBUA	SEFPIC
DAHROT	DAKFUQ	DAKYIX	DARMOY	DARNIT	SEFRIE	SEGROL	SEGWEG	SEGWIK	SEGZIN
ETTHUR02	PYRIDO04	COFNAM10	DASWAV	DAZVIJ	SEGZOT	SEJTAC	SEJTAC	VEDWUW	VEHNIF
DECZEQ	DEDBET	DEHKEG	DELNEN	DESKER	VEHNIF	VEHREF	ZZZBQV01	ZZZUEU01	ZZZVCQ01
DESKER01	DESMOD	DESMOD	DEWYIN	DEYYIP	DUZEB10	JAZVEL10	JAZVIP10	JAZVOV10	JEJZAZ
DEZPUT	DIBREL	DIVCIU	DIWNIG	DIWRAC	JEKGEL	JELWIG	KELKOB	KENWAB	KENYEH
DIZPUX	COFGOT10	DIWDES	DIXKAW	DIXREH	KENYEH	KEPJIY	KETFEU	SEMTUZ	SESMUY
DIXREH	DODWAU	DOMNOI	DONBAJ	DONBAJ	SEVHUW	SEWGEG	SEWGIK	SIDJUK	TACGIN
CIVTUW10	COPDOA10	DOVVIT	DOWTUE	DOWVAM	TAGCEJ	VEJGUM	VEPWES	VEVCAA	

of insufficient structures generated by MOLGEO (13%) is rather unsatisfactory.

Subset of Structures Converted by All of the Programs. Since only one of the programs converted the entire dataset, the bias introduced into the evaluation criteria as a consequence of such reduced conversion rates was estimated by an additional test. An analysis was performed based on the subset of molecules that can be handled by **all** of the considered 3D structure generators. This subset consisted of 317 structures, less than 50% of the entire dataset. Table 11 gives the obtained quality criteria for this subset. It can be concluded that this dataset is composed of structures less complicated for 3D conversion. Particularly the ring systems seem to be less complex, as indicated by the RMS values of the ring atom positions, RMS_{XYZ}^{rings}. Thus, it has to be kept in mind that the failure of some of the programs to convert the entire dataset

leads to a distortion of the resulting numbers. Quite often the quality measures come out favorably as the more complicated structures are not converted.

Subset of Structures Containing Unusual Elements. To investigate how the different 3D structure generators handle molecules containing unusual elements, the subset of 43 structures with atoms other than H, C, N, O, F, Cl, Br, I, P, and S was used. It is to be understood that some of the programs tested explicitly specify that they are **not** applicable to molecules containing atoms other than those just specified. However, as not all programs are very clear about this point, we thought it interesting to investigate the scope of atoms allowed by the various 3D structure generators. Table 12 gives the conversion rates and the percentage of generated structures which reproduce the X-ray geometries (RMS_{XYZ} < 0.3 Å). Only CORINA, ALCOGEN, and MOLGEO processed a significant portion of this subset.

SUMMARY

Table 13 summarizes the important results of the comparison of the six 3D structure generators. The achieved values of the quality criteria (see Evaluation Criteria) are given. The percentages given refer to the number of 3D models generated by the particular programs and not to the number of 639 molecules contained in the original dataset. The same holds for the computation times per molecule. They can only be given for the subset of the data that were converted by the programs. The figures giving the percentage of structures with an RMS_{XYZ} < 0.3 Å indicate how well the overall X-ray geometries of the input structures are reproduced by the different programs. In the following two rows of the table, the percentages of structures with perfectly reproduced ring geometries (RMS_{XYZ}^{rings} < 0.3 Å) and open-chain portions (RMS_{TA}^{chains} < 15°) are listed. The following two rows give the percentages of structures without problems (CCR > 0.8) and with serious problems caused by close contacts (CCR < 0.5), measures of how reasonable the structure generators cope with nonbonded interactions. The last row gives the number of structures with unsatisfactory geometries, $N_{insufficient}$, thus indicating how well the different programs assess their results and avoid an output of rather poor geometries.

Since two of the authors have developed one of the 3D-structure generators discussed in the present paper (CORINA), some bias might be anticipated. However, we hope that we performed an objective, fair, and scientifically sound analysis.

Nevertheless, we refrain from a more detailed analysis of the results. Users of 3D structure generators have to decide for themselves which of the criteria are the most important ones with respect to their application: conversion rate, quality of the 3D models, and/or computation time. Through this comparative study giving details about the performance of different 3D structure generators, we hope to stimulate further investigations to improve these highly important tools.

ACKNOWLEDGMENT

We gratefully acknowledge the assistance and advice of the following colleagues and institutions in providing copies of the 3D structure generators and giving instructions on their usage: Chemical Concepts, Weinheim, Germany (ALCOGEN); Chemical Design Ltd., Oxford, U.K. (Chem-X); A. R. Leach, University of Southampton, U.K. (COBRA); A. R. Katritzky, University of Florida, and V. V. Shcherbukhin, Zelinsky Institute, Moscow, Russia (MOLGEO).

APPENDIX

The CSD Refcodes of the 639 X-ray structures used in this study are given in Chart 1. Some codes appear two or more times. In these cases there is more than one molecule per asymmetric unit which have been split into single molecular entities. The Refcodes (shown in Chart 1) appear in the chronological order as retrieved from the CSD. The 639 X-ray structures can be obtained from the authors in MDL MOLFILE format if due acknowledgment of the origin of these data from the Cambridge Crystallographic Data Centre is given.

Supplementary Material Available: Tables listing the CSD Refcodes, the various RMS deviations of the Cartesian coordinates and the torsional angles, the force field energy differences between the generated structures and the X-ray references, the ring complexity values, the numbers of rotatable bonds, and the close contact ratio for each of the considered programs (67 pages). Ordering information is given on any current masthead page.

REFERENCES AND NOTES

(1) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187–204.

(2) (a) CAST-3D: CAS 3D Structure Templates File, available from Chemical Abstracts Service, Columbus, OH 43210 (370 000 CONCORD-generated structures). (b) MDDR-3D: MACCS-II Drug Data Report-3D, FCD-3D: Fine Chemicals Directory-3D, available from Molecular Design Ltd., San Leandro, CA, (12 000 and 57 000 CONCORD-generated structures. (c) Chapman & Hall Chemical Databases, available from Chemical Design Ltd., Oxford, England (216 000 structures from the Chem-X 2D-to-3D builder).

(3) (a) Rusinko, A., III. Tools for Computer-Assisted Drug Design. Ph.D. Thesis, University of Texas at Austin, Austin, TX, 1988. (b) Pearlman, R. S. Rapid Generation of High Quality Approximate 3D Molecular Structures. *Chem. Des. Auto. News* **1987**, *2*, 1/5–6. (c) *CONCORD User's Manual*; TRIPOS Associates: St. Louis, MO, 1988. (d) CONCORD is available from TRIPOS Associates.

(4) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567–2581.

(5) (a) Rusinko, A., III; Sheridan, R. P.; Nilakantan, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. Using CONCORD To Construct a Large Database of Three-Dimensional Coordinates from Connection Tables. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251–255. (b) Fisanick, W.; Cross, K. P.; Rusinko, A., III. Characteristics of Computer-Generated 3D and Related Molecular Property Data for CAS Registry Substances. *Tetrahedron Comput. Methodol.* **1990**, *3*, 635–652. (c) Henry, D. R.; McHale, P. J.; Christie, B. D.; Hillman, D. Building 3D Structural Databases: Experiences with MDDR-3D and FCD-3D. *Tetrahedron Comput. Methodol.* **1990**, *3*, 531–536. (d) Hendrickson, M. A.; Nicklaus, M. C.; Milne, G. W. A.; Zaharevitz, D. CONCORD and CAMBRIDGE: Comparison of Computer-Generated Chemical Structures with X-ray Crystallographic Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 155–163.

(6) (a) Davies, K.; Upton, R. Experiences Building and Searching the Chapman & Hall Dictionary of Drugs. *Tetrahedron Comput. Methodol.* **1990**, *3*, 665–671. (b) Nicklaus, M. C.; Milne, G. W. A.; Zaharevitz, D. Chem-X and CAMBRIDGE: Comparison of Computer-Generated Chemical Structures with X-Ray Crystallographic Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 639–646.

(7) (a) Hiller, C.; Gasteiger, J. In *Software-Entwicklung in der Chemie*; Gasteiger, J., Ed.; Springer: Berlin, 1987; Vol. 1, pp 53–66. (b) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547. (c) Sadowski, J.; Rudolph, C.; Gasteiger, J. The Generation of 3D Models of Host-Guest Complexes. *Anal. Chim. Acta* **1992**, *265*, 233–241. (d) Sadowski, J.; Gasteiger, J. In *Software Development in Chemistry*; Ziessow, D., Ed.; Gesellschaft Deutscher Chemiker: Frankfurt am Main, 1993; Vol. 7, pp 65–76.

(8) CSSR Program Manual, SERC Daresbury Laboratory, Warrington, U.K.

(9) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure

File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.
(10) Gasteiger, J.; Jochum, C. An Algorithm for the Perception of Synthetically Important Rings. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 43–48.
(11) Sippl, M. J.; Stegbuchner, H. Superposition of Three-dimensional Objects: A Fast and Numerically Stable Algorithm for the Calculation of the Matrix of Optimal Rotation. *Comput. Chem.* **1991**, *15*, 73–78 (correction of two errors in the mathematical equations can be obtained from the authors).
(12) Davies, E. K.; Murall, N. W. How Accurate Does a Force Field Need to Be? *Comput. Chem.* **1989**, *13*, 149–156.
(13) ALCOGEN is available from Chemical Concepts, Weinheim, Germany.
(14) (a) Davies, K.; Dunn, D.; Upton, R. "An Algorithm to Generate 3D Structures from 2D Connection Tables"; Poster of the 5th Molecular Modeling Workshop; Darmstadt, 1991. (b) The Chem-X 2D-to-3D builder is available from Chemical Design Ltd., Oxford, England.
(15) Gordeeva, E. V.; Katritzky, A. R.; Shcherbukhin, V. V.; Zefirov, N. S. Rapid Conversion of Molecular Graphs to Three-dimensional Representation Using the MOLGEO Program. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 102–111.
(16) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformations, Chemometrics Research Studies 15*; Wiley: New York, 1988.
(17) (a) Leach, A. R.; Prout, K. Automated Conformational Analysis: Directed Conformational Search Using the A* Algorithm. *J. Comput. Chem.* **1990**, *11*, 1193–1205. (b) Leach, A. R.; Smellie, A. S. A Combined Model-Building and Distance Geometry Approach to Automated Conformational Analysis and Search. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 379–385. (c) COBRA is available from Oxford Molecular Ltd., Oxford, England.
(18) (a) 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*, 36–47. (b) Dolata, D. P.; Leach, A. R.; Prout, K. WIZARD: AI in Conformational Analysis. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 73–85. (c) Leach, A. R.; Prout, K.; Dolata, D. P. An Investigation into the Construction of Molecular Models by the Template Joining Method. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 107–123. (d) Dolata, D. P.; Leach, A. R.; Prout, K. In *Computer Aided Molecular Design*; Richards, W. G., Ed.; IBC Technical Services: London, 1989; pp 67–82. (e) Leach, A. R.; Prout, K.; Dolata, D. P. Automated Conformational Analysis: Algorithms for the Efficient Construction of Low-energy Conformations. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 271–282. (f) Leach, A. R.; Prout, K.; Dolata, D. P. The Application of Artificial Intelligence to the Conformational Analysis of Strained Molecules. *J. Comput. Chem.* **1990**, *11*, 680–693. (g) Leach, A. R.; Dolata, D. P.; Prout, K. Automated Conformational Analysis and Structure Generation: Algorithms for Molecular Perception. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 316–324. (h) Dolata, D. P.; Walters, W. P. MOUSE: A Teachable Program for Learning in Conformational Analysis. *J. Mol. Graphics* **1993**, *11*, 106–111. (i) Dolata, D. P.; Walters, W. P. Short-term Learning in Conformational Analysis. *J. Mol. Graphics* **1993**, *11*, 112–117.
(19) Bernstein, F. C.; Koetzle, T.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *12*, 535–542.