Chem-X† and CAMBRIDGE. Comparison of Computer Generated Chemical Structures with X-Ray Crystallographic Data

Marc C. Nicklaus and George W. A. Milne*

Laboratory of Medicinal Chemistry, DTP, DCT, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

D. Zaharevitz

NCI/Frederick Cancer R&D Center, PRI-Dyncorp, Inc., Frederick, Maryland 21702

Received March 12, 1993

The structures of a number of molecules as determined by X-ray crystallography have been compared with the structures for the same molecules as calculated by the 3D structure generation program, Chem-X. In the group of molecules examined, ChemModel produced structures that were essentially identical to those based upon X-ray data in 57% of the cases. The corresponding figure for the widely used alternative model builder, CONCORD, was 38%. The superior performance of ChemModel was due entirely to that program's ability to generate multiple structures covering the entire conformational space.

INTRODUCTION

There is currently a great deal of activity aimed at determining the details of active sites in enzymes and designing molecules which can lodge in such a site and so interfere in some way with the normal biological action of the enzyme.[1,2] There are a variety of approaches employed, but, in general, they begin with some sort of characterization of the active site[3,4] and proceed to search for molecules which are commensurate with this characterization.

Geometry is commonly used in characterization; the dimensions of the enzyme cavity are defined and are then used to infer the shape and size of putative ligands. Once these required dimensions are established, various databases can be searched for molecules meeting the criteria and those compounds can then be tested.

Every step of this process is replete with difficulties, most of which are beyond the scope of this paper. The last step, however, searching for molecules with specific three-dimensional structures, depends upon the existence of the appropriate databases, and this is the issue that will be addressed here.

3D Databases. The establishment of the complete and precise 3D structure of a molecule can be regarded as a classical goal of analytical organic chemistry. X-ray crystallography has traditionally been the major source of such data, and a database of some 100 000 structures determined by X-ray and neutron diffraction techniques and reported in the literature has been built by the Cambridge Crystal Data Centre.[5] This database has been made available on a worldwide basis and has seen much use in this connection.

This database, however, possesses two deficiencies. First the content of the file is essentially uncontrolled. Crystallographers decide to solve structures for various good reasons, and most of those structures ultimately enter the database. The database is subsequently searched, in our present context, by medicinal chemists, who find that the compounds of interest to them typically are absent. A second systemic problem is that X-ray structure determination is carried out in the solid state, using crystalline material, while biological chemistry takes place in a medium which is usually aqueous. The "crystal structures" determined by X-ray methods do not reflect many of the important forces experienced by molecules in solution, but they do experience forces, such as packing forces, which are unknown in solution. In addition, as has been pointed out by Jorgenson,[6] the structure of a ligand bound to a macromolecule such as a protein is not necessarily the same as its structure in the crystal form, and work in this laboratory[7] has placed these discrepancies upon a quantitative footing. The relevance then of crystal structures to structures in biological systems can be ambiguous, particularly for flexible molecules. Nevertheless, crystal structures are widely believed to be frequently identical to the local, if not the global, minimum energy structures *in vacuo*, and since all model builders generate structures implicitly *in vacuo*, the crystal structure is still felt to be the most objective standard for comparison of different model builders. Discrepancy between a modeled structure and the crystal structure does not certify that the former is wrong, but rather that it is in disagreement with the crystal structure and may therefore merit further examination.

3D Structure Builders. In the late 1980s, several groups began to develop software which could develop a 3D structure from the corresponding 2D structure, usually represented by a connection table. The earliest and most successful of these was CONCORD[8,9] which relies upon a combination of expert system (rule-based) procedures and pseudo-molecular mechanics.[6] It uses standard bond lengths and angles but does not attempt energy calculations. In a previous paper,[10] we reported the results of a comparison of CONCORD-built structures with the same structures measured by X-ray diffraction. In this study, there was significant discrepancy between the two structures in over 50% of the cases examined. The major source of errors in the structures generated by CONCORD lay in the torsion angles that were assigned to bonds that were capable of rotation. The program avoids short-range nonbonded interactions but is less successful with longer-range interactions and also experiences difficulty accounting satisfactorily for electronic effects which often are important in the determination of the favored structure of the molecule.

Because of the large proportion of erroneous structures produced by CONCORD, it was felt that an alternative approach should be examined, and ChemModel, the model

† Chem-X is a suite of programs that includes the model builder ChemModel, which was used in this work.

0095-2338/93/1633-0639$04.00/0

building program within Chem-X,[11] was examined. ChemModel builds structures by relying upon a table of fragments which are selected and assembled as the program works its way through the 2D structure, which is usually provided in the form of a connection table. Like CONCORD, it does not attempt to detect long-range or electronic interactions; rather, it builds a "base conformation" and then attempts to finesse the problem by the systematic variation of the torsion angles of all rotatable bonds in the structure. In this paper, we define a rotatable bond, or rotor, as a bond, usually acyclic, which can rotate freely and which has at either end at least one non-hydrogen. Methyl, primary amino, and hydroxyl groups are excluded from this definition. Depending upon the atoms that are bonded, a rotor may be assigned two, three, or six torsion angles, one of which, presumably, is close to the correct structure. All the resulting structures are built and their coordinates stored.

All rotors are considered by ChemModel to be interdependent, and, as a result, the number of conformers generated increases exponentially as new rotors are encountered; generation of hundreds or thousands of conformers from one structure is not uncommon. The procedure can be justified theoretically on the grounds that all the conformers generated represent possible structures. As a practical matter, the large number of excess conformers do not present an insuperable data management problem, and, as a matter of experience, the presence of these excess conformers in the database does not appear to lead to false drops in subsequent coordinate searches. ChemModel begins by building the base conformation. Then it identifies the rotors in the structure and begins to generate different conformers of the base conformation. The conformers are generated by systematic rotation about the rotors in steps whose size is dependent on the type of bond and which can be controlled by the user. We have used the program defaults[12] as they represent the largest grid feasible with a database of more than 400 000 structures. Each new conformation is checked to see if it is consistent with a set of conditions, such as the absence of close contacts, and if so, it is accepted. As additional conformers of a structure are built and accepted, the different separations of every important atomic center are recorded and packed into bit screens.[12] If, for example, a separation of 2.5 Å is measured, then bit 9 will be set while a separation of 2.6 Å calls for a different bit, bit 10, to be set. The different bit screens for every atom pair are combined in a logical OR. An important result of this is that additional acceptable conformers lead to more bits being set, but not to additional storage. The bit screens can be used in 3D searches to eliminate structures that have no conformers that could satisfy the query, thus greatly reducing the number of structures for which conformers need to be regenerated.

For the purposes of this work we wanted to identify the conformer whose geometry was closest to that of the X-ray derived structure. This was accomplished by reading the X-ray derived structure into Chem-X and using the entire structure as a search query. Conformers that matched the query were written to a database. When more than one conformer satisfied the query, the one with the lowest root mean square (RMS) deviation from the X-ray derived structure was used. Several of the structures with the greatest deviation from the X-ray derived structure had no conformers satisfying the query using the default 0.5-Å tolerance, and in these cases the tolerance was increased until at least one conformer satisfying the query was found.

Database. Examination of the different model builders was undertaken in this laboratory in order to select one which can be used to effect a 2D → 3D conversion of the 450 000 structures in the NCI Drug Information System.[13] For approximately 44% of the compounds, the appropriate CAS Registry Number (RN) is provided and this provides a means of finding the same compounds in the Cambridge Structural Database (CSD),[5] in which 12% of the entries have CAS Registry Numbers. Using the CAS Registry Number, a total of 194 compounds were found to be common to both databases (the actual overlap is probably far higher but only seems to be low in both databases because of incomplete inclusion of Registry Numbers). Of these, several were not used in model building for various reasons. Some possessed undefined chiral centers, others contained heavy atoms whose parameters are not available or large rings, which ChemModel cannot process, and some, though present in the CSD, had no atomic coordinates. For one or another of these reasons, 104 of the original 194 compounds were dropped and 90 compounds remained for a direct comparison. Of the 90 compounds discussed here, 85 were in the set of 90 that were modeled by CONCORD.[9] Five compounds, however, contained large rings which, in the absence of a template, are not modeled by Chem-X, and these were therefore replaced. Because, in our earlier study,[11] CONCORD's failure to model these five structures correctly had been recorded, their removal from the set created a small bias in favor of Chem-X.

The 90 compounds[14] are listed in Table I, whose third column gives the number of rotors in the molecule. The fourth column of Table I gives the number of generated conformations that were accepted by Chem-X; the fifth column gives the RMS between the X-ray structure and the CONCORD model,[9] and the final column provides the corresponding RMS for the Chem-X model. As was shown in our previous paper,[9] the proportion of rigid compounds, those with zero rotors, in the 90 compound set (29%) is far higher than in the NCI database as a whole (6%) or the CSD (12%). The 90 compound set is therefore biased in favor of modeling programs, which may be expected to perform less effectively on a completely random set of structures.

RESULTS

Overall Data. Each of the compounds in the test set was processed by ChemModel; the results are displayed in Table I and summarised in Table II, which also shows the corresponding data derived from models built by CONCORD.[8] Structures were compared by computing the deviation between the position of an atom in one structure and that of the same atom in the other. These deviations were summed over all non-hydrogen atoms in the structure and the sum expressed as the root of the mean of the squares (RMS). For the sake of discussion, an RMS value of less than 0.5 Å is regarded here as an "exact" match.

The first result seen in Table II is that ChemModel delivered 51 exact matches (RMS < 0.5 Å) out of the 90 that were built (56.7%). This compares favorably with the corresponding figure from CONCORD (34/90 or 37.8%). ChemModel produces fewer "near-misses" with 0.5 Å < RMS < 1.0 Å (18/90 or 20.0%) than CONCORD (21/90 or 23.3%) and no "incorrect" structures with an RMS > 2, while CONCORD produces 5 of these (5.6%). Thus on an overall basis, ChemModel performs somewhat better than CONCORD, as can be seen from the bar plots in Figure 1. This is attributed to the only major difference between the programs; ChemModel's ability to consider conformational variation.

Table I. Set of 90 Compounds Modeled by Chem-X[a]

CAS RN	CSD no.	rotors	conformations	RMS(C)	RMS(X)	CAS RN	CSD no.	rotors	conformations	RMS(C)	RMS(X)
53-19-0	DCCPET	3	36	0.308	0.336	6067-31-8	FIPKAQ	2	2	1.610	1.006
54-36-4	BIHYEW10	3	72	1.930	0.184	6344-60-1	BESGEL	0	1	0.033	0.050
56-41-7	LALNIN03	1	9	0.148	0.122	6510-63-0	SEYJUB	2	36	1.040	0.610
57-85-2	ZZZRCG01	3	4	0.564	0.432	6829-31-8	CADZAI	2	12	1.280	0.693
62-55-5	THACEM01	0	1	0.035	0.073	N/A		1	6	0.049	0.092
70-25-7	NOGUNA02	4	4	1.800	0.336	6937-59-3	JEDAIC	5	64	0.678	1.305
75-52-5	NTROMA08	0	1	0.034	0.093	7544-65-2	IVLPIN	3	54	1.589	0.651
92-69-3	BOPSAA01	1	2	0.347	0.044	7597-43-5	JEBFEB01	1	2	0.264	0.179
100-15-2	FUXNAN	2	2	0.061	0.060	N/A		0	3	0.048	0.082
101-59-7	FIHBED	3	32	0.910	0.661	13652-13-6	DADLUP	5	324	1.680	1.004
101-81-5	ZZZMKS01	2	16	0.368	0.192	N/A		6	1296	0.104	1.405
109-99-9	BUNJAV01	0	1	0.052	0.037	N/A		4	7	1.100	1.113
139-91-3	FUPALP	5	16	3.979	1.424	15718-46-4	PYMSUL10	3	56	0.865	0.838
311-03-5	DOFSUM	0	1	1.060	0.270	N/A		0	1	0.351	0.398
431-03-8	CABBIQ01	1	1	0.035	0.038	21416-87-5	ICRFRA10	3	1	1.980	1.592
503-30-0	CIVXIO01	0	1	0.054	0.106	N/A		9	24040	1.880	1.440
504-02-9	FACRIK	0	1	0.288	0.363	24584-09-6	ICRFRB10	3	1	1.440	1.203
518-75-2	CITNIN02	1	2	0.178	0.306	27848-84-6	GICVUJ	5	18	2.110	1.406
526-99-8	BIVTUV	5	972	1.191	0.229	N/A		0	1	0.087	0.044
557-30-2	GLOXIM11	3	5	0.669	0.479	30868-30-5	PYRZOM01	3	18	1.510	1.615
579-43-1	VABVEZ	3	1	0.133	0.451	N/A		4	6	1.270	0.202
599-71-3	JEHWUO	2	14	0.722	0.088	N/A		2	2	0.293	0.296
640-19-7	FACETA01	1	1	0.817	0.169	N/A		3	2	1.330	0.567
N/A		0	1	0.200	0.204	64332-37-2	MEGONE	9	2916	2.600	1.780
N/A		4	1	0.821	0.481	66054-22-6	ICRFRD10	2	9	0.200	1.282
791-28-6	TPEPHO04	3	8	0.562	0.382	N/A		8	688	0.958	0.628
832-64-4	FUVGEI	0	1	0.118	0.131	N/A		1	1	0.308	0.132
838-41-5	DEMTIY	2	4	0.810	0.131	77097-65-5	BACLAS	5	32	1.310	0.325
948-44-7	POXTSO	0	1	0.434	0.290	76467-15-7	RUDMOL	1	2	0.736	0.672
N/A		0	1	0.104	0.070	77762-21-1	DUBPUL	0	1	0.136	0.357
1079-71-6	CEKWEU01	0	1	0.278	0.289	N/A		6	1	1.450	0.977
1135-32-6	AZSTBB	2	44	1.920	0.462	N/A		2	1	0.112	0.168
1138-48-3	VEMLUU	2	36	0.736	0.682	N/A		4	120	0.571	0.595
1226-42-2	CASGEI	5	2016	1.130	0.770	82891-67-6	BOKCEJ	8	1	0.791	1.185
1439-41-4	CETPOX01	9	1836	0.775	0.384	84472-85-5	GATHOY	3	6	1.490	0.996
1468-95-7	VAFMUK	1	6	0.081	0.092	85048-88-0	CAGLEB	0	1	0.222	0.230
2030-63-9	DAKXUI	4	16	1.270	1.027	85269-22-3	BURFID	3	2	0.975	1.014
2589-31-3	TABRIX02	2	16	0.212	0.075	N/A		4	12	1.023	1.084
2746-19-2	NBONAN	0	1	0.520	0.135	88430-84-6	CEVPIC	8	162	0.874	1.389
N/A		3	4	1.400	0.981	88946-46-7	CEWKUK	1	1	0.679	1.505
4023-53-4	CNETPD	9	1778	1.410	0.754	91190-12-4	COCZID	8	258	2.350	0.867
N/A		4	116	0.847	0.754	91296-23-0	BUYTOE10	2	12	1.420	0.655
4988-33-4	KAGMUA	0	1	0.239	0.117	91296-27-4	BUYTIY10	2	18	2.250	0.914
5441-02-1	DOPDAN	4	16	0.948	0.349	92900-65-7	COYDUP	7	1380	1.730	1.250
5693-87-8	GESTAZ	8	348	1.525	0.463	92952-33-5	GAPCUV	3	24	1.930	0.438

[a] CAS = Chemical Abstracts Service. CSD = Cambridge Structure Database. Rotors are defined as rotatable bonds with more than one non-hydrogen at either end. The number of conformations accepted by the program is given. RMS(C) and RMS(X) respectively are the RMS values obtained by comparing the CONCORD and Chem-X models with the X-ray structure. N/A indicates that the data are proprietary.

Table II. RMS Values for Chem-X and CONCORD Models

RMS (Å) all non-hydrogens	no. of rotatable bonds					
	0	1	2	3	>3	Σ
			Chem-X			
<0.5	19	9	8	7	8	51
0.5–1.0	0	1	5	6	6	18
1.0–1.5	0	0	2	3	13	18
1.5–2.0	0	1	0	1	1	3
>2.0	0	0	0	0	0	0
totals	19	11	15	17	28	90
			CONCORD			
<0.5	17	8	6	2	1	34
0.5–1.0	1	3	3	7	10	24
1.0–1.5	1	0	3	3	8	15
1.5–2.0	0	0	2	5	5	12
>2.0	0	0	1	0	4	5
totals	19	11	15	17	28	90

A closer look at the data in Table II allows some interesting observations. As the number of rotors in the structure is increased, CONCORD's ability to build correct structures is quickly impaired and it gives a correct structure in only one case, where more than three rotors are present. ChemModel, on the other hand, is still producing correct results (8/28) in structures with more than three rotors, and it also surpasses CONCORD in cases where there are two or three rotors. This is clearly a positive result of ChemModel's ability to examine rotors.

If the exact matches of the rigid structures with zero rotors (19 from ChemModel and 17 from CONCORD) are set aside, ChemModel builds 32 of the remaining 71 (flexible) structures (45.1%) correctly (RMS < 0.5 Å), while CONCORD does so with only 17/71, or 23.9%. This provides yet another argument in support of the need for conformational searching in the model building context.

Exact Matches. With structures which are rigid or which contain a small number of rotors, modeling programs have relatively little difficulty. Figure 2 shows two such structures for which ChemModel produced models that agree very closely with the X-ray derived structure. For nitromethane (**1**), the model is almost identical to the measured structure, the RMS for all non-hydrogen atoms being 0.093 Å. The RMS for the CONCORD model of this compound was 0.034 Å. Similarly, with *p*-nitro-*N*-methylaniline (**2**), ChemModel and CON-

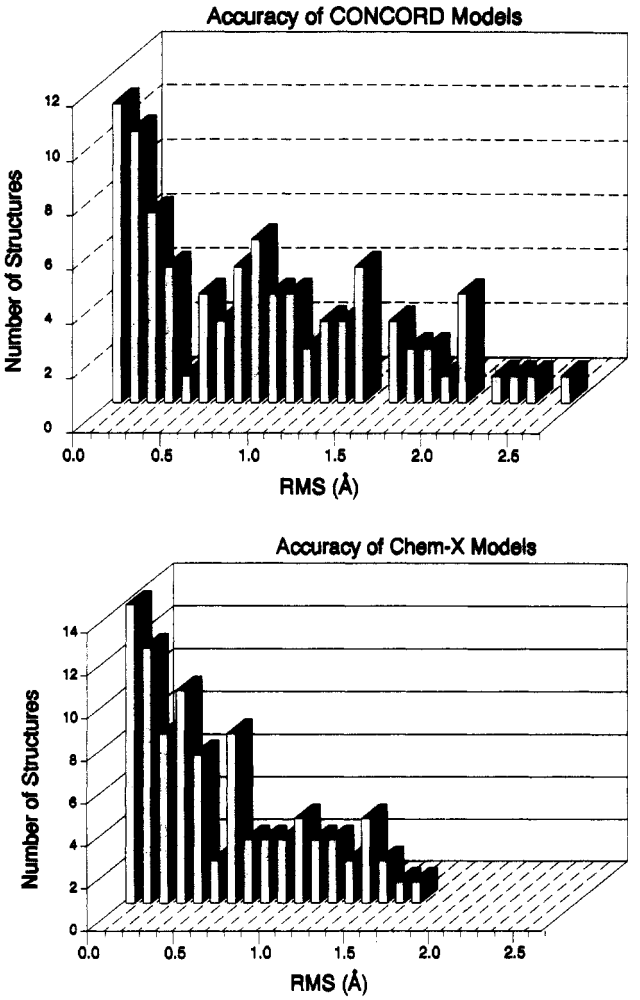

Figure 1. Accuracy of CONCORD and Chem-X models.

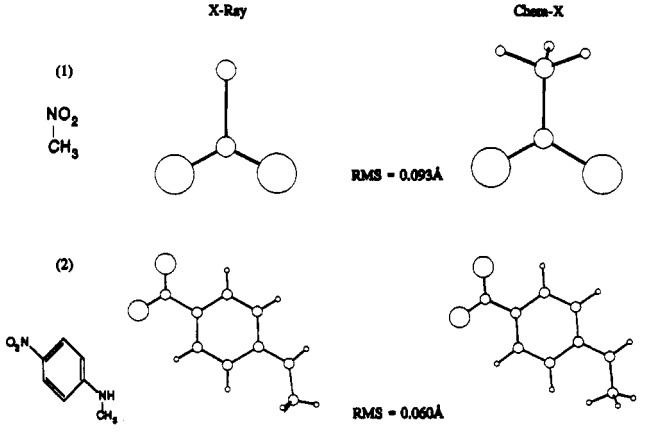

Figure 2. Exact matches built by ChemModel.

CORD both produce accurate models with RMS values of 0.060 and 0.061 Å, respectively.

Two more complicated cases are shown in Figure 3. With the *p*-hydroxybiphenyl structure (**3**), ChemModel correctly makes the two rings coplanar and the rest of the exercise is trivial; the resulting RMS is 0.044 Å. This stands in contrast to CONCORD which sees only steric crowding, not electronic overlap between the rings, and sets them at an angle of 40°. Six carbons are thus displaced from their correct position, and this is reflected in an RMS of 0.347 Å. The X-ray structure of 4-methylphenanthrene (**4**) shows how the B-ring is slightly deformed so as to alleviate the crowding caused by the methyl group. This relatively distant effect is overlooked by both ChemModel and CONCORD, but because the larger

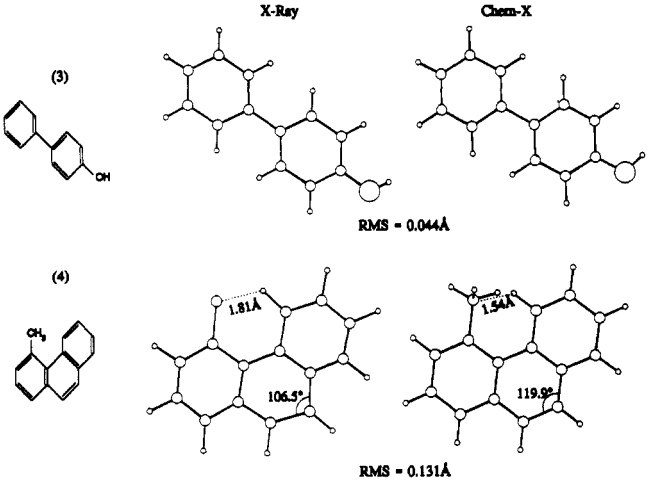

Figure 3. Good matches built by ChemModel.

part of the molecule is unaffected, the RMS values (0.118 for CONCORD and 0.131 for ChemModel) are still low. Thus in general, both programs do well with completely rigid structures. When rotors are present, as in the biphenyl (**3**), ChemModel may do better.

Near Matches. When more than one rotor is present, as in the examples shown in Figure 4, the difficulties facing modeling programs begin to be compounded and their performance tends to diminish. 2,6-(*N*,*N*′-Diacetyldiamino)-pyridine (**5**) is constrained by resonance energy to a planar conformation, but only at the expense of deformation of the side chains. ChemModel correctly produces a planar model but fails to detect the side chain deformation, and as a result, the RMS between its structure and the X-ray derived structure is 0.349 Å. In 4-methyl-4′-nitrodiphenyl sulfide (**6**), ChemModel misassigns the two crucial torsions. As a result, the plane of one ring in the model is quite wrong; four atoms (the *ortho* and *meta* carbons) are misplaced, and the RMS is 0.661 Å. CONCORD, however (RMS = 0.910 Å), does even worse.

Figure 5 shows the results with more flexible structures. The bis(*p*-methoxy)benzil structure (**7**) is not planar. ChemModel develops a nonplanar molecule but makes errors in several of the torsions, and the final RMS is 0.770 Å. In the 1,2-diphenylcyclopropane (**8**), ChemModel fails to assign the correct value to either of the two torsion angles. The errors (20° and 30°) are not large, but they displace several atoms and the RMS is 0.682 Å.

In the case of the bis(fluoranthenyl)methane (**9**), shown in Figure 6, ChemModel incorrectly assigns a value of 90° to one of the torsion angles; the correct value is 59°. This is not a huge error, but it moves an entire C_{13} fragment—half the molecule—out of place and the consequent RMS is 0.610 Å. Also in Figure 6, dipyrimidyl disulfide (**10**) has three important torsion angles whose values (0.30°, −83.92°, and 2.55°) can be measured from the X-ray data. The angles assigned by ChemModel (−120°, 60°, and −120°) seem suspiciously mechanical and, in any event, are quite wrong. As a result, half the atoms in the molecule are misplaced, and the RMS is 0.838 Å.

Mismatches. In 14 of the 90 test structures, ChemModel produced a structure which differed widely (RMS > 1.0 Å) from the X-ray derived structure, and it is instructive to examine some of these cases. Figure 7 shows the results obtained for 1-benzyl-2,3,4,5-tetraphenylcyclopentadiene (**11**). In this structure, there is steric *versus* electronic competition and the X-ray data show that the steric forces are supervening; all four phenyl rings are out of the plane of the

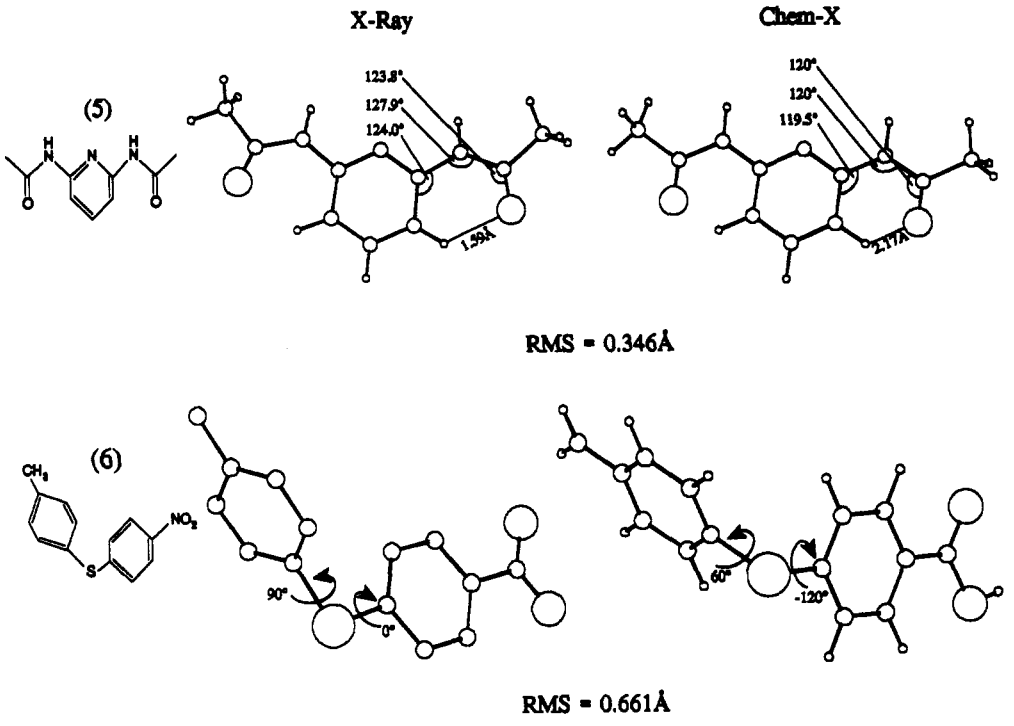

Figure 4. Approximate matches built by ChemModel.

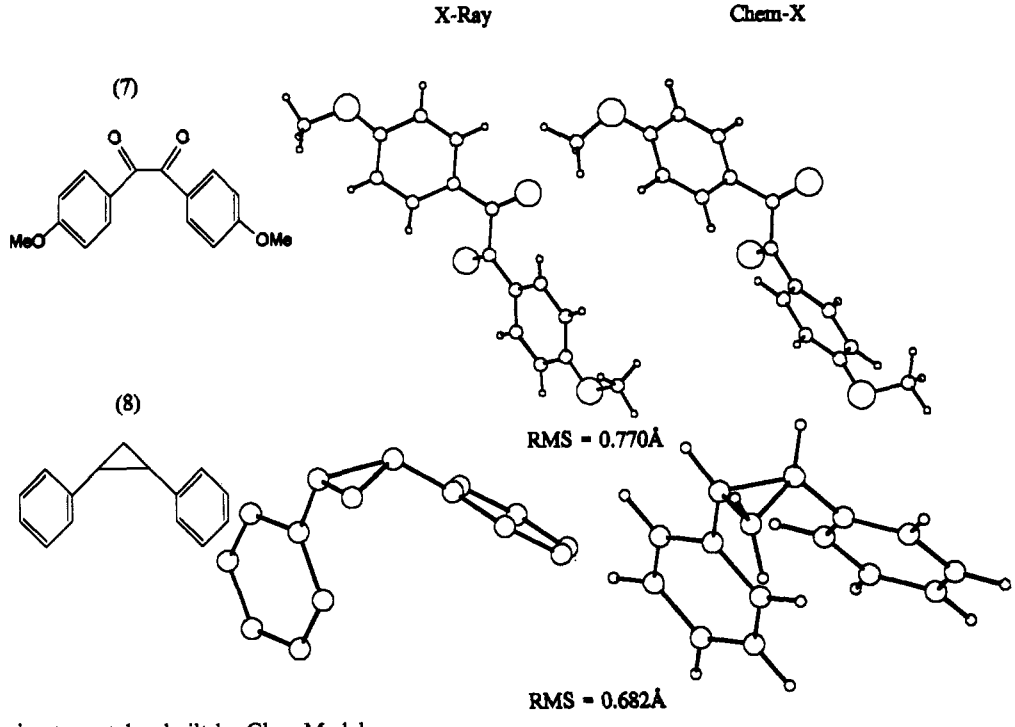

Figure 5. Approximate matches built by ChemModel.

cyclopentadiene ring by angles ranging from 43° to 67°. Accounting for this interplay of forces is far beyond ChemModel's capacity, and the model it delivers puts the benzyl ring out of plane but leaves the remainder of the molecule in a single plane, even though this leads to obvious and serious steric problems. A majority of the carbons are thus assigned erroneous coordinates, and a high RMS (1.305 Å) results. Interestingly, CONCORD does considerably better with this structure, because it places a higher priority upon relief of steric crowding and delivers an RMS of 0.678 Å. In the other example in Figure 7, the model of 1-phenyl-1-(trifluoromethyl)ethanol tosylate (**12**) is a poor match (RMS = 1.004 Å) to the X-ray structure. Closer inspection, however, shows this to stem almost entirely from the one torsion angle shown in the figure. In the X-ray structure, this angle has a value of −108°, but the model builder assigns it a value of 90° and in so doing misplaces nine atoms and contributes to the large RMS value.

When structures contain large numbers of rotors, both CONCORD and ChemModel have serious difficulties. Two examples are provided in Figure 8. In 1,1,1,3,3,3-hexaphenylpropane (**13**), every one of the six phenyl rings is misrotated relative to the propane backbone. In this way, 24 of the 33 atoms are out of place, compared to the X-ray structure, and the RMS is 1.185. There is a similar problem with the sulfur-containing compound (**14**). ChemModel makes the two ring

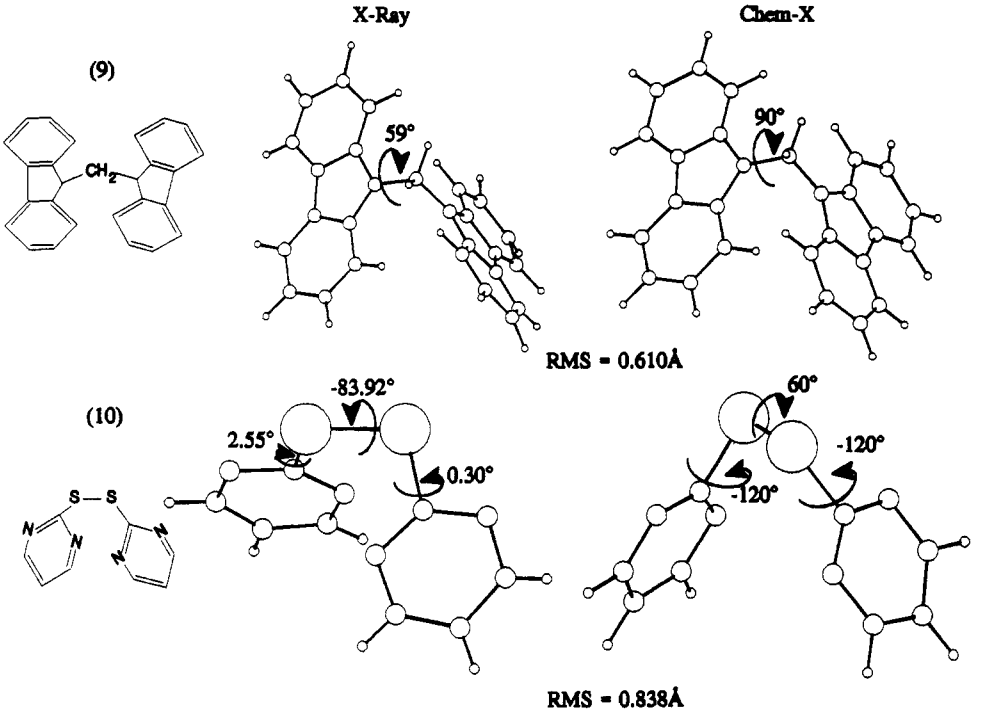

Figure 6. Poor matches built by ChemModel.

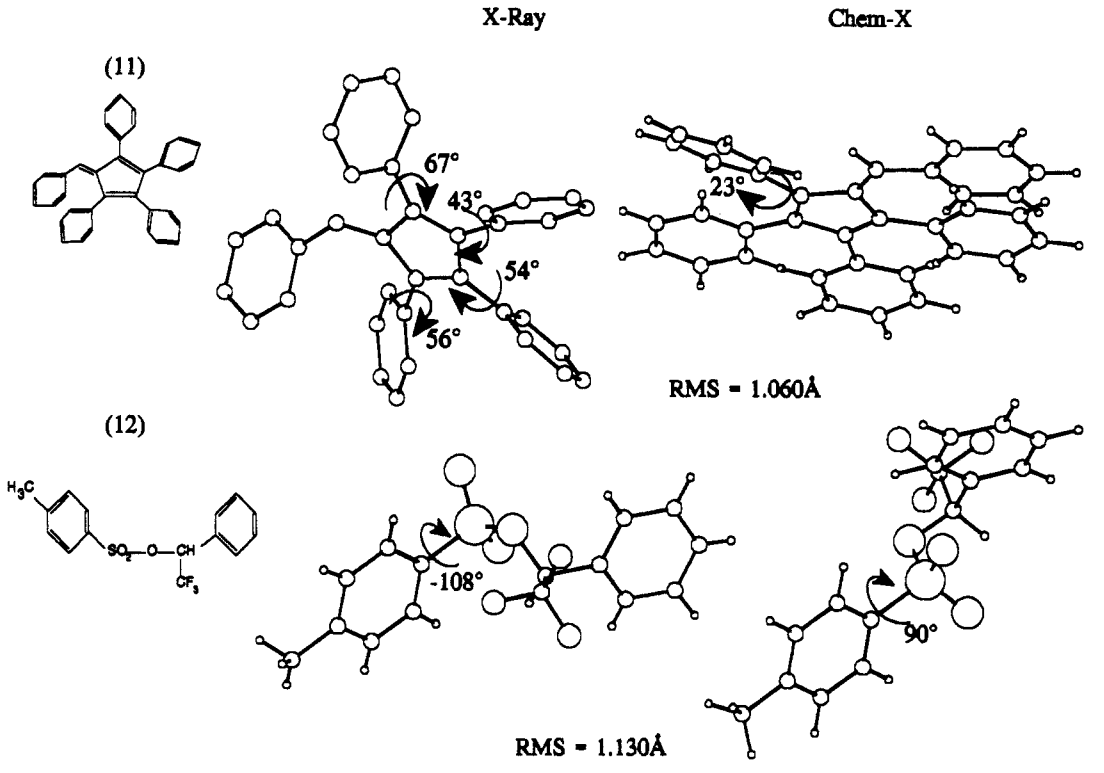

Figure 7. Mismatches built by ChemModel.

systems and every carbomethoxy group coplanar. The X-ray data show this compound to be quite noncoplanar, and a high RMS (1.389 Å) results.

DISCUSSION

CONCORD performs by using a set of *ad-hoc* rules which permit the assembly of atoms into the structure that was requested. The program delivers a single structure which is often close to the actual crystal structure but which more often is in error, sometimes in a minor way, sometimes more seriously.

Chem-X takes a philosophically different approach to the building of 3D structures. It begins with predefined fragments which can be assembled to produce, in principle, every possible structure from a 2D representation. In practice, however, it generates far fewer than this because its rotational space is explicitly limited. The bond between two sp^3 carbon atoms for example has a torsion angle which ChemModel sets at 60°, 180°, or 300°. Thus three models are built, and, unless some error condition supervenes, they are all retained. One of them will be the closest to the actual (X-ray) structure and will contribute, more or less, to the RMS values reported in

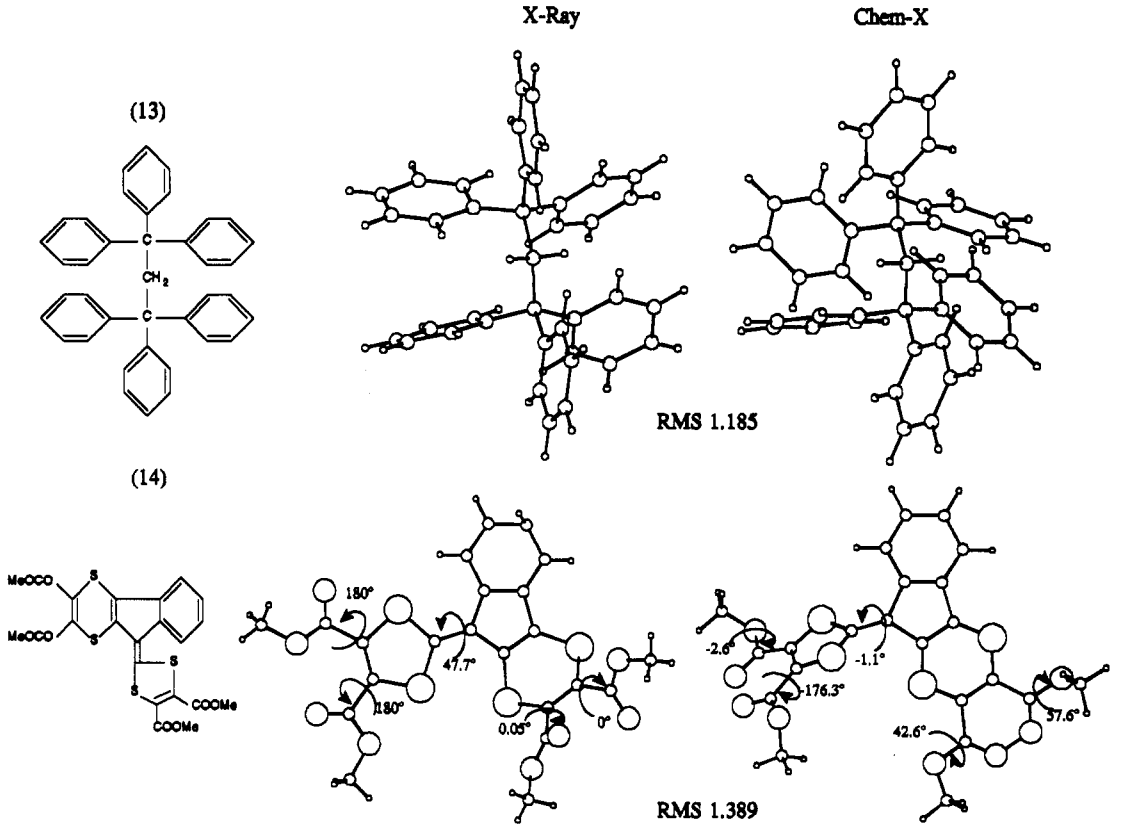

Figure 8. Highly flexible structures.

Table I. The search "grid" used by ChemModel is fairly sparse; most torsions are allowed to assume less than six values, and this accounts for the frequent minor misassignments of torsions and also for the "mechanical values", mentioned above, for torsion angles observed in Chem-X generated structures.

This situation can readily be improved by use of a less coarse grid in the ChemModel building, but a finer grid will increase the number of alternative structures that are generated, raising the demand for both storage and CPU time. For a large database such as NCI's DIS, this coarse grid is probably a reasonable compromise, but it should be reiterated that it is under the user's control. If more accuracy is required in the structures generated and the computer resources are available, then a finer grid should be used. Similar observations have been made by Haraki et al.[15] in a study comparing Chem-X and CONCORD with respect to their yield of active compounds from a 3D database of compounds with known biological activities. It was found that while the percentage of correct hits (i.e. compounds that are known to possess the activity associated with the pharmacophore used in the 3D search) was higher with CONCORD, the absolute number of hits as well as of active compounds retrieved was generally higher with Chem-X. It is also noted that use of a finer grid in Chem-X's conformational search would probably have increased the number of hits in those cases where Chem-X did not perform well. The intended application of the 3D database is relevant to a choice between CONCORD and Chem-X. If exhaustive retrieval from a database of all the active compounds is the goal (and the number of false positives is not a concern), then the Chem-X approach clearly seems to be more promising, especially since it can be fine-tuned by the user.

The use by Chem-X of predefined fragments is the reason for its inability, in our hands, to process structures containing large (greater than eight atoms) rings. All rings are regarded as fragments by Chem-X and while a six-membered ring can be easily defined, the geometry of larger rings is much more variable and easily affected by substituents. Such fragments therefore are not readily available, and structures containing large rings were discarded by the program because of this.

The data presented here suggest that examination of torsional space, as performed by ChemModel, greatly improves the overall accuracy of the model builder. It is clear that appropriate adjustment of the grid used by the program can further improve its accuracy, at the cost of computer resources. It is in fact surprising that ChemModel does not do better than was observed, because, in principle, it should generate near-correct conformers for any structure it builds. One possible explanation for ChemModel's behavior is that it generates the correct conformer but then discards it because the rules it uses for testing are flawed.

REFERENCES AND NOTES

(1) See, for example: Gubernator, K.; Ammann, H. J.; Broger, C.; Bur, D.; Doran, D. M.; Gerber, P. R.; Müller, K.; Schaumann, T. The Use of Computer-Assisted Molecular Modeling in Pharmaceutical Research. *Proc. Montreux Chem. Inf. Conf.* **1992**, *4*, 187–193.

(2) Müller, K.; Ammann, H. J.; Doran, D. M.; Gerber, P. R.; Gubernator, K.; Schredfer, G. Complex Heterocyclic Structures—a Challenge for Computer-Assisted Molecular Modeling. *Bull. Soc. Chim. Belge* **1988**, *97*, 655–667. Gerber, P. R.; Gubernator, K.; Müller, K. Generic Shapes for the Conformational Analysis of Macrocyclic Structures. *Helv. Chim. Acta* **1988**, *71*, 1429–1441.

(3) Lewis, R. A.; Roe, D. C.; Huang, C.; Ferrin, T. E.; Langridge, R.; Kuntz, I. D. Automated Site-Directed Drug Design Using Molecular Lattices. *J. Mol. Graph.* **1992**, *10*, 66–78.

(4) Leach, A. R.; Kuntz, I. D. Conformation Analysis of Flexible Ligands in Macromolecular Receptor Sites. *J. Comput. Chem.* **1992**, *13*, 730–748.

(5) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystal Data Centre: Computer-Based Search, Retrieval, Analysis, and Display of Information. *Acta Crystallogr.* **1979**, *B35*, 2331–2339.

(6) Jorgenson, W. L. Rusting of the Lock and Key Model for Protein-Ligand Binding. *Science* **1991**, *254*, 854–855.
(7) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Comparison of the Structures of Ligands in the Pure and Protein-Bound Forms. Unpublished.
(8) CONCORD (1987, 1988, University of Texas, Austin, TX). *CONCORD User's Manual*; TRIPOS Associates: St. Louis, MO, 1988.
(9) Rusinko, A., III. Tools for Computer-Assisted Drug Design. Ph.D. Thesis, University of Texas at Austin, Austin, TX, 1988. Pearlman, R. S. Rapid Generation of High Quality Approximate 3-D Molecular Structures. *Chem. Des. Auto. News* **1987**, *2*, 1–6. All CONCORD results cited here were obtained with Version 2.9.3.
(10) 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.
(11) Chem-X is a product of Chemical Design Ltd., 7 West Way, Oxford OX2 0JB, England. All results cited here were obtained with the July 1992 version of Chem-X.
(12) Murrall, N. W.; Davies, K. R. Conformational Freedom in 3-D Databases. 1. Techniques. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 312–316.
(13) The NCI, to maintain its database of ≈450 000 chemicals which have been examined for antitumor activity, uses the NCI Drug Information System, which was designed and built by NCI. See: Milne, G. W. A.; Miller, J. A.; *et al.* The NCI Drug Information System. Parts 1–6. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 154–197.
(14) Several compounds in the group of 90 modeled by CONCORD[8] were replaced to make up the current set of 90. Compounds corresponding to CAS RN 296-41-3, 1005-51-2, 25423-56-7, 38734-05-3, 65114-88-7, 71138-48-2, and 88589-00-8, contained large rings which are not processed by ChemModel. They were replaced by those with CAS RN 139-91-3, 504-02-9, 599-71-3, 786-78-7, 1135-32-6, 2030-63-9, and 7544-65-2.
(15) Haraki, K. S.; Sheridan, R. P.; Venkataraghavan, R.; Dunn, D. A.; McCulloch, R. Looking for Pharmacophores in 3-D Databases: Does Conformational Searching Improve the Yield of Actives? *Tetrahedron Comput. Methodol.* **1990**, *3*, 565–575.