National Cancer Institute Drug Information System 3D Database

G. W. A. Milne, Marc C. Nicklaus, J. S. Driscoll, and Shaomeng Wang

Laboratory of Medicinal Chemistry, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4255

D. Zaharevitz*

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

Received April 28, 1994®

A searchable database of three-dimensional structures has been developed from the chemistry database of the NCI Drug Information System (DIS), a file of about 450 000 primarily organic compounds which have been tested by NCI for anticancer activity. The DIS database is very similar in size and content to the proprietary databases used in the pharmaceutical industry; its development began in the 1950s; and this history led to a number of problems in the generation of 3D structures.

BACKGROUND

Since 1955, the National Cancer Institute (NCI), one of the U.S. National Institutes of Health (NIH), has conducted extensive testing of materials for possible activity against different forms of cancer.[1] Most of the substances tested have been pure organic compounds. Some inorganic compounds have been examined, and, particularly since 1990, extracts of terrestrial plants and marine organisms such as algae have been emphasized.

Few assumptions were made as to promising types of structures, and the program examined an extraordinarily eclectic assembly of organic structures. Currently, almost 4% of all known organic structures have been tested,[2] a proportion that arguably is representative of the whole *corpus* of organic chemistry. Almost every natural element in the periodic table is represented in the database, and, in general, such elements are present either as the anions or cations of salts or covalently bound to an organic structure.[3] Some elements, *e.g.*, platinum, have often been associated with antitumor activity,[4,5] and this has supported the decision to include all elements in the antitumor screening process.

Because samples of compounds were needed for testing, it was necessary to develop a large acquisition effort and a repository, both of which are still functioning today. Continuous scanning of the chemistry literature allows identification of compounds that are novel[6] and of interest to the program, and the authors are approached for a sample, typically under 100 mg. Academic and government laboratories often provide samples freely. Commercial laboratories usually request confidentiality before releasing materials, and the NCI has responded with a "Discreet Agreement" under which the government can acquire and test a compound while keeping all the data concerning it confidential with results transmitted to only the NCI and the compound's supplier. Compounds that are so classified are termed "discreet" and currently make up approximately 48% of the NCI database. Access to discreet compounds and information relating to them is controlled. The balance of the database is "open", and NCI supports public access to the open data.[7]

Efforts to maintain computer control of the enormous quantities of data associated with this program have often preceded development of the necessary computer hardware and software. In the 1950s, when computer development was still in its early stages, elaborate electromechanical devices were built[8] to provide machine support of the voluminous cardfiles. In the 1960s, in collaboration with Chemical Abstracts Services, one of the very first substructurally searchable databases was created from the NCI file. This search was a batch operation, the epitome of user-unfriendliness, with a turnround time measured in days. It nevertheless was a great improvement over the earlier manual methods, and as faster hardware became available the search capability improved steadily. In 1980, an entirely new data management system, the Drug Information System (DIS), was built.[9] This permitted real time interactive searching of all the database records, including the chemical structure file, with search times of seconds. There have been changes in hardware since then, but the overall design of the system is intact.

During the last decade, structural biology has become established as a powerful paradigm which increasingly is driving drug development efforts such as the NCI program. The empirical acquisition and screening efforts still continue but are receiving more and more competition from structure-based approaches to drug design. One of these depends upon the certain fact that drug-receptor interactions cannot be accounted for, far less predicted, if one persists in a two-dimensional view of the process. Receptor sites and the drugs that would lodge in them are three-dimensional, and in order to study processes such as enzyme inhibition, a means must be found to develop three-dimensional models of both of them. X-ray crystallography is a method of obtaining three-dimensional data directly[10] and has been heavily exploited, there being few other sources of such data. X-ray methods however have serious shortcomings. They require crystalline material, which is not always available, and the early methods were slow and expensive, although with improving technology, this is ceasing to be a major problem. Perhaps most serious is the fact, long suggested[11] but only recently subjected to systematic study,[12] that structures in the crystal state, at least of nonrigid compounds, bear only a limited resemblance to those in the normal biological medium. For these reasons, the three-dimensional structures of small molecules are often developed by modeling programs today. The structures of enzymes and other macromolecules have generally been measured by X-ray methods, but multidimensional NMR

* Abstract published in *Advance ACS Abstracts*, August 1, 1994.

0095-2338/94/1634-1219$04.50/0

H																	
Li	Be											B	C	N	O	F	Ne
113	16											1590	45851	1366145	392981	22956	0
Na	Mg											Al	Si	P	S	Cl	Ar
5960	165											93	2241	14554	120638	128555	0
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
1224	197	9	235	103	274	321	991	1027	841	1503	900	83	288	1298	984	24173	0
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
13	19	14	74	38	236	0	267	305	487	154	125	26	2336	385	66	8394	0
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
42	145	38	23	19	214	62	45	92	1820	169	835	44	306	108	0	0	0
Fr	Ra	Ac															
0	1	0															
		Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu		
		39	26	43	0	27	83	24	42	28	13	20	13	17	12		
		Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Md	Lo	Fm	Lr		
		20	0	38	0	0	0	0	0	0	0	0	0	0	0		

Figure 1. Occurrence of different elements in compounds in the NCI DIS database.

spectroscopy[13] is challenging crystallography and seems certain ultimately to surpass it.

Very few of the compounds in the NCI database have been analyzed by X-ray diffraction and in view of the need to work with three-dimensional structures, it was decided to generate a three-dimensional database from the NCI 2D chemical file. The problems encountered during this process and the results that were finally obtained form the subject of this paper.

METHODS

The task to be addressed was to develop a procedure which could take 500 000 structures in the form of 2D connection tables and produce the corresponding 3D records, consisting of connection tables plus the 3D coordinates for every atom. Further, it must be possible to search the 3D database for specific geometric arrangements of atoms or in the medicinal chemistry context, *pharmacophores*. For a single structure, generation of a 3D structure is not particularly difficult; numerous computer programs exist for this purpose. It is necessary however that the 3D structure that is produced be reliable, and it is important that the conversion be completed very quickly, in 1 s for example, or the whole database cannot be processed. There are after all, only 604 800 s in 1 week and a conversion that requires 1 min per structure will run for about 1 year if applied to a database of this size.

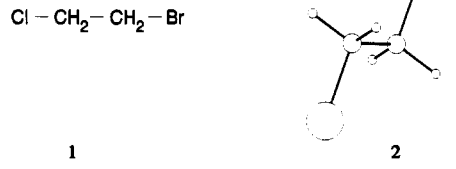
When a 3D structure is modeled, using the 2D structure as a starting point, various problems can arise. These concern stereochemistry, which is sometimes incompletely defined in the 2D structure, and elements beyond the first two rows of the periodic table, for which geometric and electronic parameters are not well established. Further problems surround the choice of conformation when conformational flexibility is present and the question of the potential energy of any particular 3D structure. Each of these issues is discussed separately below.

1. Stereochemistry. Although our contemporary notions of stereochemistry were well established in the 1950s, the means, or even the need to incorporate these into chemical structure handling programs was not always recognized. As a result, stereochemistry was often ignored when chiral structures were entered into the database during this period. Worse, the precise stereochemistry of compounds that were tested in those days was frequently not even known. The attitude adopted, which persisted well into the 1970s, was that it was not worth determining stereochemistry unless the compound showed some antitumor activity. From the Institute's perspective, this is defensible in a way, but it led to the presence in the file of as many as 150 000 inactive compounds whose structures, because they lack chiral definition, present special problems to the 2D → 3D conversion program. For many of these structures, the chirality was simply unknown. For others, it was known but not recorded, and for still others it was known and recorded but was either incomplete and/or incorrect. Against this background, it was decided to continue with the earlier policy; if chirality at a specific stereocenter was known and recorded, it should be incorporated into the 3D model. In all other cases, the program would arbitrarily assign a chirality to the stereocenter. Approximately half of these assignments will be incorrect, and this will lead ultimately to low levels of both false positives and false negatives in 3D searches. The former can be detected in post-search examinations; the latter cannot be recovered.

2. Different Elements. The remarkable inclusion of heavier elements in this database stems from a deliberate NCI decision to examine as many elements as possible. This was driven further by the discovery that platinum-containing compounds often showed desirable antitumor activity[4,5] and by the ready availability of compounds containing, for example, copper or tin. In 1994, an analysis of the database, summarized in Figure 1, showed that of a grand total of 468 678 compounds, over 16 000 contain elements other than H, C, O, N, S, P, F, Cl, Br, I, and Si. Over a period of nearly 40 years, only 12 of the elements ($Z \leq 92$) have been overlooked as can be seen from Figure 1. While this comprehensiveness may be commendable, it leads to a problem for the 2D → 3D conversion because when this conversion was carried out (1992), few of the 3D modeling programs were able to handle elements other than H, C, O, N, S, P, F, Cl, Br, I, and Si. This stemmed from the lack of the necessary parameters—bond lengths, bond angles, and so on—for all but these well-studied atoms as well

as a paucity of information concerning the coordination state of a metal in any specific compound—deficiencies which have since been addressed by the various workers such as Zerner.[14] At the time of the 2D → 3D conversion, the options again were limited. For platinum, for example, determination of the appropriate parameters by examination of the literature permitted inclusion of platinum-containing compounds in the 3D database, but this task, for one element, required weeks of a skilled person's time. Even if such resources could be committed to the task for other elements, many of the required parameters are simply unknown, and experimental measurement would be necessary. Faced with such a boundless problem, the decision was taken initially to drop compounds containing covalently bound elements not in the above list of 11. This had the effect of removing 16 634 compounds (3.7%) from the database. At a later stage, subsequent to performing the searches described in this paper, platinum was parameterized and the 1820 Pt-containing compounds were merged into the 3D database.

3. Conformational Flexibility. Conformational flexibility has emerged as one of the most serious problems associated with 2D → 3D conversion. The problem can be illustrated with reference to 1-chloro-2-bromoethane (**1**). This structure contains one carbon–carbon bond, whose rotation is significant. The 2D structure (**1**) is mute as to the torsion angle of this rotatable bond or "rotor" [A rotatable bond, or rotor, is defined as a bond, usually acyclic, which can rotate freely and which has attached at either end at least one non-hydrogen. Bonds to methyl, primary amino, and hydroxy groups are thus excluded from this definition.] In the 3D structure (**2**) a value, 180° in the conformation shown, must be assigned to this angle, but the modeling program does not know what the appropriate angle is; it must be determined. Two options are available; the energies can be computed for all the possibilities and the lowest energy conformation selected,



1

2

or the program can generate and store more than one 3D structure, making no *a priori* choice between them. Energy calculation, while easily feasible for a single structure, is unrealistic for such a large database and one is forced into the second option, generation of multiple structures. This leads to significant problems associated with data storage and search times, but management of these problems is possible, as will be discussed below.

4. Potential Energies of 3D Structures. Having built several alternative 3D structures from a single 2D structure, it is possible in principle to identify the energetically most stable of these and reject the others. Numerous commercially available computer programs accomplish this reliably and reproducibly, but the difficulty, as mentioned above, lies in the computation time. Calculation of the potential energy of a structure typically requires less than a second on computers such a Silicon Graphics IRIS[15] using programs such as CHARMm,[16] the molecular mechanics program built into the molecular modeling package Quanta.[17] Other hardware/software combinations give comparable results. Unfortunately, this potential energy value requires refinement, or "minimization", in which all the atomic positions are adjusted, seeking a "most stable" geometry for the structure as a whole. Sophisticated minimization algorithms exist, but none can overcome the fact that this is an n^2 problem—a set of 10 atoms implies on the order of 100 nonbonded atom–atom interactions, all of which must be computed repeatedly. The outcome is that although the energy calculation may take less than a second, the minimization takes several seconds, even minutes, putting the whole method out of bounds for a large database.

A serious criticism of this "global minimum" approach is that there is an increasing body of evidence[11,12,18] to support the idea that the global energy minimum structure is not necessarily favored or even particularly relevant to enzyme binding. Consequently, after some examination of these points, it was decided that because potential energy calculation during 3D building is prohibitively slow and the structures developed by the method are not clearly valid in the enzyme binding context,[12] this approach would not be pursued.

A less time-consuming means of avoiding high-energy structures was more successful, and required examination of each newly generated 3D structure for bad contacts, defined as atoms which approach one another too closely in terms of their van der Waals radii. This is also an n^2 computation but needs however to be done only once. Detection of bad contacts can be accomplished very rapidly and can be used as grounds for rejection of that conformation in favor of a better one. If no conformation free of bad contacts can be found, the entire structure can be rejected. This approach can be used in either the building step or the searching; in the present work it was not used in the structure building but was used in searches, as is described in the next section.

3D STRUCTURE BUILDING

Three different approaches were explored to the generation of the 3D database. Two of these failed to meet the criteria laid out above; one was successful. The approaches are described in this section.

1. Use of Energy-Minimized Structures. A generalized molecular modeling program such as Quanta[17] can accept a 2D structure in a number of formats and produce from it a 3D structure or, if the structure contains rotors, several possible 3D conformations. These structures approximate energetically plausible structures but they must be energy-minimized. It was felt to be common wisdom that the minimum energy structure of an isolated molecule, calculated *in vacuo*, would closely approximate the structure actually present in a binding site and that such energy-based approaches thus allow a choice between alternative 3D structures. This assumption however probably is not supportable in the enzyme-binding context,[12] and, in any case, the energy minimization is time-consuming—noticeably so in comparison to energy calculation and extravagantly so in comparison to structure building.

Measurement of the CPU time required for these operations and subsequent extrapolation to the large database indicated that, even on a supercomputer, CPU times on the order of years would be called for, and this approach was therefore abandoned.

2. Building Single Conformers. The program CONCORD[19,20] is a popular means of creating a 3D database from the corresponding 2D file. Exhaustive testing on CONCORD was conducted to assess its suitability for generation of the 3D database from the NCI 2D database. The connection tables in the DIS were converted to the MDL SD format[21] and submitted to CONCORD which then builds a 3D

Table 1. 2D → 3D Conversion of the NCI DIS Database

	open	discreet	total	percentage
2D database; 10 June, 1992	233 985	218 322	452 307	100.0
no connection table	4686	4430	9116	2.0
incomplete connection table	1669	2243	3912	0.9
excluded atom	11 963	4671	16 634	3.7
miscellaneous error	60	55	115	<0.1
valid connection table	215 607	206 923	422 530	93.4
3D structure built successfully	206 876	201 036	407 912	90.2

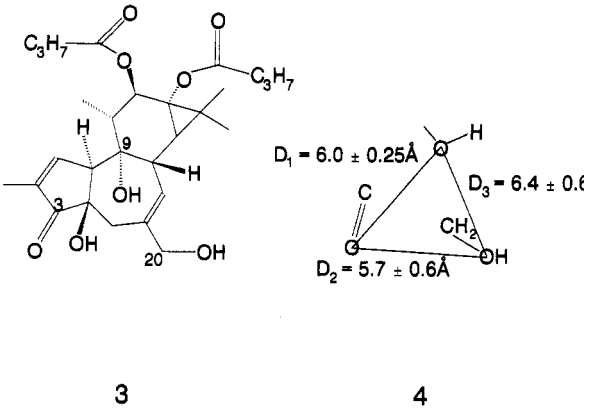
structure. This program is rule-based and uses standard bond lengths and angles but does not attempt energy calculations although bad contacts (*v. supra*) can be detected. Experiments with this program revealed that while it performs very well with rigid molecules it is less successful with structures which contain rotors because it has only limited means to examine different possible torsion angles.[22] In the version of CONCORD that was examined,[23] conformational flexibility was not explicitly supported, and the generation of multiple conformations was not carried out. A large subset consisting of some 396 000 structures from the NCI database was converted to 3D structures by CONCORD.[22] Examination of these resulting 3D structures revealed that this restriction relating to torsion angles is crippling for the applications envisaged in this work, and accordingly Chem-X[24] was used for the building of the large 3D database.

3. Building Multiple Conformers. Chem-X is a molecular modeling package[24] which contains a module for generation of 3D structures. This program has benefitted from the pioneering work of Dolata and Leach[25,26] who reported their research in conformational analysis as long ago as 1987. An important feature of the Chem-X model builder is its ability[27] to build multiple conformations for structures containing rotors. The program can conduct a superficial examination of the 3D structures produced, rejecting those that are seriously flawed by steric crowding, for example, and all the conformations that survive are organized into a database which is searchable with substructural fragments and 3D coordinates.

The two-step process by which multiple conformations are considered is as follows. The connection tables in MDL SD format[21] are read into the model builder which uses a standard set of molecular fragments to build the 3D structures. The maximum allowable deviation from standard bond angles was increased to 60° in order to allow structures with three-membered rings (primarily cyclopropanes and epoxides) to pass the quality check that is made after the structure is built. With these conditions, 96.5% (407912/422530) of the valid connection tables were successfully converted into 3D structures (Table 1). The next step in the process is to perform the conformational analysis, calculate interatomic distances, and pack them into a series of bit screens, which ultimately form the basis of a 3D coordinate search system.[27] When the program encounters a rotor in the 2D structure, it checks the nature of the atoms at either end of the rotor bond. The allowable values for the torsion angle depends upon the atoms involved and upon parameters set in the program. We used the Chem-X default values,[27] so if the bond is between two sp^3 atoms, the program allows the rotor to adopt one of three torsion angles, 60°, 180°, and 300°, and thus generates three different conformations. The number of torsions used is adjustable; a C_{sp^2}–C_{sp^2} bond, such as that joining two aryl rings, will be assigned one of two values, 0° or 180°. Thus amongst sp^3 carbons, after the first C–C rotor, three conformations will be built. If a second C–C rotor is subsequently encountered, three conformations will be built for each of these, making a total of nine conformations. As each conformation is generated, it is checked to ensure that it meets various criteria such as conformational rules[25] and the absence of close contacts. If it passes this check, all the interatomic distances in the conformation are calculated and packed into the set of bit screens. The number of conformations generated by the model builder increases exponentially, and even the number of accepted conformations often becomes very large. As an example, a structure with n C–C bonds will produce 3^n conformations, *i.e.*, 59 049 conformations for 10 bonds, and over 14 million for 15 bonds. There are 15 such bonds in a common *n*-pentadecyl side chain, and so this unrestrained generation of conformers must be restricted in some way. Accordingly, as a practical matter, the maximum number of rotors to be examined in any single compound was set to 15 (excluded 2.5% of open database), the maximum number of conformations was set to 1.6 million, and a 5-min limit was placed upon the CPU time to be used in the build step for any compound. This is far in excess of the average CPU build time per structure, established below, of approximately 20 s. In preliminary work with these choices for the conformational analysis, a significant number (>10%) of the 3D structures produced no conformations that passed all the checks. Investigation of this problem indicated that it was due in large part to a parameter that set to 2 the allowable settings for the torsion angle of bonds joining two aromatic rings. This led to many structures with unacceptably close contacts between different rings. This parameter was dynamically reset to 6 when necessary, and in the build of the entire database approximately 4% of the structures produced no acceptable conformations.

Under these circumstances, a total of 90.2% of the 2D database structures were successfully converted to 3D structures. The resulting 3D database contained 206 876 "open" (nondiscreet) structures and 201 036 discreet (proprietary) structures, for a total of 407 912 3D structures. The two subfiles are kept separately, and the data reported in this paper are derived exclusively from the 206 876 open structures. The building of this 3D database was carried out on a VAX 9000 and an SG Indigo with a 33 MHz IP12 processor and required about 4 months of CPU time or approximately 20 CPU-s per compound. The major losses are recorded in Table 1, and the reasons for them include the absence of a complete connection table in the 2D database (2.9%) and the presence of an excluded atom—*i.e.*, any atom other than H, C, O, N, S, P, F, Cl, Br, I, and Si—(3.7%). A factor which eliminated relatively few structures, but which is nevertheless serious, is the program's inability to process large rings. Structures with rings of more than seven atoms cannot be built by Chem-X with its default set of build fragments, and this excluded 8983 compounds, accounting for 61% of the build failures. Significantly, a number of biologically important natural products contain large rings and their omission on these grounds is a cause for concern. One possible method for attacking this problem would be to use another program to build the initial 3D structure and then use Chem-X to perform the conformational analysis. Preliminary studies indicate that CONCORD 3.0 can successfully build 75% of the Chem-X failures, including compounds with up to 25-membered rings.

Searching in the 3D Database 3D database searching is an exciting new approach to the solution of a variety of problems, and its potential in the area of drug design is of particular interest to the NCI. The search algorithms were pioneered by groups such as that led by Willett[28] at the University of Sheffield and are now installed in numerous commercially available software packages such as Chem-X,[24,27] ISIS/3D,[29]



3 4

Figure 2. Phorbol dibutyrate and the phorbol pharmacophore.

or UNITY.[30] In this paper, we present one example of a pharmacophore search conducted in the NCI 3D database using the standard Chem-X search software. Further examples are described in other publications from this laboratory.[31]

Protein kinase C (PK-C) is an enzyme which is important in intracellular signaling[32] and has long been regarded[33] as an attractive target for chemotherapy of some cancers. Diesters of the diterpene phorbol such as its 12,13-dibutyrate (**3**) bind very effectively to PK-C and much evidence[34] has been accumulated suggesting that the operative pharmacophore in phorbol esters consists of three oxygens, at C_3, C_9, and C_{20}. The 3D structure of phorbol has been established by X-ray crystallography,[35] and the pharmacophore derived from the X-ray data has the dimensions shown in **4** (Figure 2).

A search of the open compounds in the NCI 3D database was carried out with the pharmacophore **4**, and 535 compounds were identified as having structures with an arrangement of oxygen atoms approximating this pharmacophore. The first step of the search, a very fast screen on interatomic distances, passed some 4000 structures to the second search step, which examines substructure, geometry, and conformation. This step processes 100 structures per CPU-hour on a Silicon Graphics IRIS 3000, and 40 h were thus required to complete this search. This results list was manually reviewed, and compounds such as sugars, which contained no clearly hydrophobic moiety were discarded, as were compounds for which no physical sample was available. The 125 compounds which were retained were tested experimentally for binding to PK-C. Of these, five compounds were found to bind strongly to the enzyme. Among the most potent binders was the bis-(2-hydroxyethyl)-2-aminosteroid (**5**), with a binding constant of 16.1 ± 3.3 μM. A superpositioning of a low energy conformation of **5** upon phorbol is depicted in Figure 3 which shows the excellent match (RMS = 0.28 Å) of the three pharmacophore oxygen atoms. Compound **5** is representative of a previously unknown class of PK-C binders and that it was found in a search of the NCI 3D database constitutes a validation of this approach to drug lead discovery.

SUMMARY

The various laboratories that have seriously investigated the utility of 3D searching have generally reported positive results, and the structure building program that has been described in this paper leads to a 3D database which produces valid searches. Many of the details require more work. The inability to accommodate all elements is troublesome, as is the failure of these (and most other) structure building

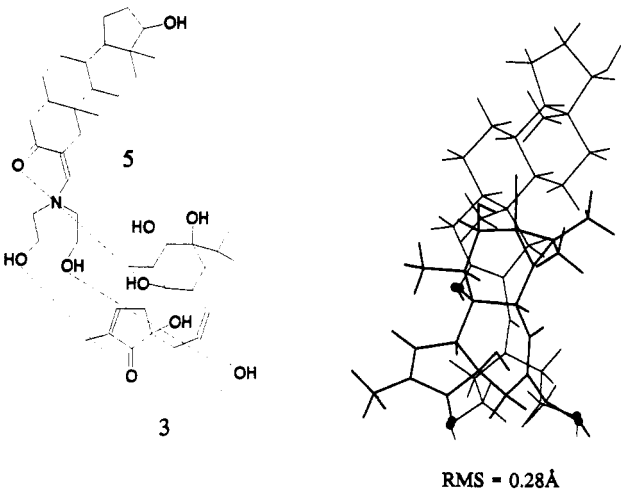


Figure 3. Superposition of the phorbol pharmacophore on the hydroxyamine **5**.

programs to build 3D models of the large flexible rings found in many biologically important compounds. The results presented here however indicate that searching of the NCI 3D database is worthwhile and makes it clear that work should be undertaken to resolve the problems described here. How many positives were missed because of the deficiencies in the database remains unknown, but, even in its presently incomplete form, our 3D database was found immediately to be useful in identifying PK-C inhibitors.

This laboratory's experience in this area parallels that of others and may be indicative of a new direction in drug discovery. The NCI database contains some 460 000 compounds, and NCI holds physical samples of over 65% of these. The acquisition and testing of new compounds is slow and uncontrolled in the sense that it is limited to those compounds whose availability becomes known to the Institute. 3D searching of such a database can serve to identify lead compounds rapidly because of the immediate availability of samples for bioassay. The database is probably large enough to include an adequate selection of substructures which are commonly used and useful in drug development. For pharmaceutical companies, for whom patent considerations are important, the use in this sense of an in-house, proprietary database is not only convenient but also simplifies the patent questions that can arise when a new use is discovered, because a composition-of-matter patent may already be held.

For these reasons, 3D searching is thus not only technically promising as a tool in medicinal chemistry research, it also has potential advantages for the pharmaceutical industry. Consequently, we expect to see considerable activity in this area in the future.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Peter Blumberg of the Division of Cancer Etiology; NCI for providing a PK-C bioassay capability, Dr. Kevin Haraki of Lederle Laboratories, Pearl River, NY for providing a copy of his program for conversion of connection tables in MDL SD format; and Dr. Kenneth Paull of the Division of Cancer Treatment, NCI for providing project input and the financial resources used to obtain the copies of software used in this work.

REFERENCES AND NOTES

(1) The starting point for this effort is generally regarded as the Congressional authorization, in 1955 of $5 million for the establishment by the NCI of a Drug Development Program. See: DeVita, V. T.; Oliverio, V. T.;

Muggia, F. M.; Wiernik, P. W.; Ziegler, J.; Goldin, A.; Rubin, D.; Henney, J.; Schepartz, S. The Drug Development and Clinical Trials Programs of the Division of Cancer Treatment, National Cancer Institute. *Cancer Clin. Trials* **1979**, *2*, 195–216.

(2) At the end of 1993, the CAS Registry contained over 12 million substances so the NCI database, which contained about 465 000 substances at that time, represented some 3.875% of the CAS Registry.

(3) In general, ionically bound elements such as sodium salts or hydrochlorides represent a trivial occurrence of the element in question. The majority of heavy elements (other than Na, K, Ca, and Mg) in the NCI database are covalently bound. Thus of the 274 chromium-containing compounds, for example, 229, or 84%, contain covalently bound chromium.

(4) Rosenberg, B.; Van Camp, L.; Krigas, T. Inhibition of Cell Division in *Escherichia coli* by Electrolysis Products from a Platinum Electrode. *Nature* **1965**, *205*, 698–699.

(5) Rosenberg, B.; Van Camp, L.; Trosko, J. E. Platinum Compounds: A New Class of Potential Antitumor Agents. *Nature* **1969**, *222*, 385–386.

(6) The structures of compounds whose acquisition is under consideration are examined by computer programs which determine their novelty with respect to the current database. See: Hodes, L. J. Computer-Aided Selection of Compounds for Antitumor Screening: Validation of a Statistical Heuristic Method. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 128–132. Selection of Molecular Fragment Features for Structure-Activity Studies in Antitumor Screening. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 132–136. Computer-Aided Selection of Novel Antitumor Drugs for Animal Screening. *ACS Symp. Ser.* **1979**, *112*, 583–603.

(7) The (2D) structurally searchable file can be accessed directly *via* Telnet, but 3D searches, which are much more time-consuming, are best conducted in conjunction with the NCI. For further details readers should contact K. Paull, Information Technology Branch, EPN-843, NCI, NIH, Bethesda, MD 20892.

(8) Richman, S.; Hazard, G. F.; Kalikow, A. K. The Drug Research and Developmental Chemical Information System of NCI's Developmental Therapeutics Program. In *Retrieval of Medicinal Chemical Information*, Howe, W. J., Milne, M. M., Pennell, A. F., Eds.; ACS Symposum Series; American Chemical Society: Washington, DC, 1978; Vol. 84; pp 200–221.

(9) Milne, G. W. A.; Miller, J. A. The NCI Drug Information System. 1. System Overview. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 154–159. Milne, G. W. A.; Feldman, A.; Miller, J. A.; Daly, G. P.; Hammel, M. J. The NCI Drug Information System. 2. DIS Pre-Registry. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 159–168. Milne, G. W. A.; Feldman, A.; Miller, J. A.; Daly, G. P. The NCI Drug Information System. 3. The DIS Chemistry Module. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 168–179. Milne, G. W. A.; Miller, J. A.; Hoover, J. R. The NCI Drug Information System. 4. Inventory and Shipping Modules. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 179–185. Zehnacker, M. T.; Brennan, R. H.; Milne, G. W. A.; Miller, J. A. The NCI Drug Information System. 5. DIS Biology Module. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 186–193. Zehnacker, M. T.; Brennan, R. H.; Milne, G. W. A.; Miller, J. A.; Hammel, M. J. The NCI Drug Information System. 6. System Maintenace. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 193–197.

(10) X-ray crystallographically derived structures of small molecules are all collected together in the Cambridge Structural Database (CSD), while similarly derived macromolecule structures are kept in the Protein Data Bank (PDB). See: Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Mothewell, 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. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; 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.

(11) See, for example: Jorgensen, W. L. Rusting of the Lock and Key Model for Protein-Ligand Binding. *Science* **1991**, *254*, 854–855.

(12) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational Changes of Small Molecules Binding to Proteins. 1. Geometric Changes. *J. Med. Chem.*, submitted for publication. Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational Changes of Small Molecules Binding to Proteins. 2. Energy Changes. *J. Med. Chem.*, submitted for publication.

(13) Brünger, A. T.; Nilges, M. Computational Challenges for Macromolecular Structure Determination by X-Ray Crystallography and Solution NMR Spectroscopy. *Quart. Rev. Biophys.* **1993**, *26*, 49–125, and references cited therein. Freund, C.; Ross, A.; Plückthun, A.; Holak, T. A. Structural and Dynamic Properties of the Fv Fragment and the Single-Chain Fv Fragment of an Antibody in Solution Investigated by Heteronuclear Three-Dimensional NMR Spectroscopy. *Biochemistry* **1994**, *33*, 3296–3303.

(14) Zerner, M. C. Semiempirical Molecular Orbital Methods. *Rev. Comp. Chem.* **1991**, *2*, 313–365. Anderson, W. P.; Cundari, T. R.; Zerner, M. C. An intermediate neglect of differential overlap model for second-row transition metal species. *Int. J. Quantum Chem.* **1991**, *39*, 31–45. Culberson, J. C.; Knappe, P.; Roesch, N.; Zerner, M. C. An intermediate neglect of differential overlap (INDO) technique for lanthanide complexes: studies on lanthanide halides. *Theor. Chim. Acta* **1987**, *71*, 21–39.

(15) Silicon Graphics Computer Systems, Mountain View, CA 94039-7311.

(16) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(17) QUANTA is produced by Molecular Simulations, Inc., 16 New England Executive Park, Burlington, MA 01803.

(18) Wiley, A. B.; Rich, D. L. Peptidomimetics Derived from Natural Products. *Med. Res. Rev.* **1993**, *13*, 327–384.

(19) CONCORD (copyrighted 1987, 1988, University of Texas, Austin, TX). *CONCORD User's Manual*; TRIPOS Associates, St. Louis, MO, 1988.

(20) Rusinko III, A. 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 3D Molecular Structures. *Chem. Des. Auto. News* **1987**, *2*, 1–6.

(21) 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 Ltd. *J. Chem. Inf. Comput Sci.* **1992**, *32*, 244–255.

(22) 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.

(23) This work was carried out in 1992 with versions 2.9.1 and 2.9.3 of CONCORD.

(24) Chem-X is a product of Chemical Design Ltd., 7 West Way, Oxford, OX2 0JB, England.

(25) Dolata, D. P.; Leach, A. R.; Prout, K. WIZARD: AI in Conformational Analysis. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 73–85.

(26) 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*, 271–283.

(27) Murrall, N. W.; Davies, K. R. Conformational Freedom in 3D Databases. 1. Techniques. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 312–316.

(28) Brint, A. T.; Willett, P. Pharmacophoric Pattern Matching in Files of 3D Chemical Structures: Comparison of Geometric Searching Algorithms. *J. Mol. Graph* **1987**, *5*, 49–56. Jakes, S. E.; Watts, N.; Willett, P.; Bawden, D.; Fisher, J. D. Pharmacophoric Pattern Matching in Files of 3D Chemical Structures: Evaluation of Search Performance. *J. Mol. Graph.* **1987**, *5*, 41–48.

(29) Moock, T. E.; Henry, D. R.; Ozkabak, A. G.; Alamgir, M. Conformational Searching in ISIS/3D Databases. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 184–189.

(30) Hurst, T. Flexible 3D Searching: The Directed Tweak Technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190–196.

(31) See: for example, Wang, S.; Zaharevitz, D.; Sharma, R.; Marquez, V. E.; Lewin, N.; Du, L.; Blumberg, P. M.; Milne, G. W. A. The Discovery of Novel, Structurally Diverse PK-C Agonists through Computer 3D-Database Pharmacophore Search. Molecular Modeling Studies. *J. Med. Chem.*, submitted for publication.

(32) Nishizuka, Y. The Role of Protein Kinase C in Cell Surface Signal Transduction and Tumour Promotion. *Nature* **1984**, *308*, 693–698.

(33) Blumberg, P. M. Protein Kinase C as the Receptor for the Phorbol Ester Tumor Promoters: Sixth Rhoads Memorial Award Lecture. *Cancer Res.* **1988**, *48*, 1–8.

(34) Rando, R. R.; Kishi, Y. Structural Basis of Protein Kinase C Activation by Diacylglycerols and Tumor Promoters. *Biochemistry* **1992**, *31*, 2211–2218. Wender, P. A.; Koehler, K. F.; Sharkey, N. A.; Dell'Aquila, M. L.; Blumberg, P. M. Analysis of the Phorbol Ester Pharmacophore on Protein Kinase C as a Guide to the Rational Design of New Classes of Analogs. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4214–4218. Jeffrey, A. M.; Liskamp, R. M. Computer-Assisted Molecular modeling of Tumor Promoters: Rationale for the Activity of Phorbol Esters, Teleocidin B, and Debromoaplysiatoxin. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 241–245.

(35) Brandl, B. F.; Röhrl, M.; Zechmeister, K.; Hoppe, W. X-ray Structure Analysis of Neophorbol, $C_{31}H_{35}O_9Br$, and Phorbol, $C_{20}H_{28}O_6$. *Acta Crystallogr.* **1971**, *B27*, 1718–1730.