Molecular Modeling in the Discovery of Drug Leads

George W. A. Milne,* Shaomeng Wang, and Marc C. Nicklaus

Laboratory of Medicinal Chemistry, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received November 21, 1995[⊗]

The National Cancer Institute of the U.S.A. maintains a repository of about 500 000 chemicals which it has tested at some time for anticancer activity. As new chemotherapeutic targets present themselves, methods have been developed by which this large database can be re-examined without resorting to expensive high-volume biological screening. Electronic screening, the method described in this paper, begins with the identification of a target enzyme. The pharmacophore used by the enzyme in binding to substrates is identified, and then all compounds in the database that contain the pharmacophore are found. These compounds are then further filtered, for example, by physical properties such as solubility, and the relatively small number of compounds that survive are submitted for biological testing. This use of a primary electronic screen in the search for ligands of protein kinase C is described. The screen is very fast, and the method is quite generally applicable to different enzymes.

INTRODUCTION

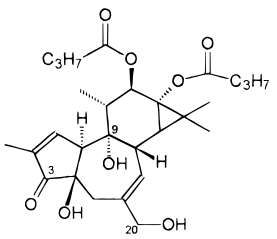
Beginning in 1955, the National Cancer Institute (NCI) began acquiring and testing chemicals for possible anticancer activity and chemotherapeutic potential.[1] The program still functions, currently acquiring about 10 000 new compounds per year, and, to date, about 500 000 compounds have been acquired and tested.[2] Each compound is logged into a database,[3] and samples of the compound are stored in a repository. Data from the biological testing that was carried out with these compounds have been stored in different databases but are not relevant to this paper. Here, the database is best regarded simply as a collection of approximately 4% of all known organic compounds.

When an enzyme performs its catalytic function, an early step is usually an association reaction between it and its normal substrate. Thus *cysteine transaminase* binds L-cysteine, *adenosine kinase* binds adenosine, and so on. This binding is more or less reversible and involves primarily hydrogen bonding between the enzyme and its substrate. The specificity of the enzyme for a particular substrate is predicated entirely upon the position and geometry of the related hydrogen bonding sites within the enzyme. The well-defined geometry of this "active site" suggests that substrates must have a well-defined complementary geometry, without which binding will not take place. This complementary geometry constitutes what is called a "pharmacophore".

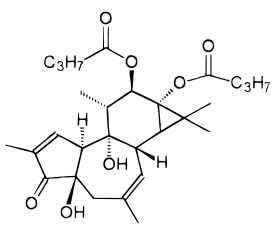
If a compound other than the natural substrate can bind in an active site, that molecule of enzyme will be deactivated, at least temporarily, and if sufficient unnatural substrate is available for binding, the enzyme will be "inhibited"—unable to perform its normal function. If that function is important to the organism in which the enzyme operates, this could be a threat to the organism. In this way, chemicals which are enzyme inhibitors can kill, for example, viruses or cancer cells. The goal of this work then is, by use of pharmacophore-based searches,[4,5] to locate compounds which can inhibit specific key enzymes and then to see if those compounds possess activity *in vitro* or *in vivo*, against the host organisms. Active compounds discovered by this method will need much further elaboration to provide clinically useful agents; for this reason, they are referred to as "lead compounds".

PHARMACOPHORE CHARACTERIZATION

If a compound is known to be a substrate for a specific enzyme, it may be assumed that it binds to the enzyme and the questions that arise are which substrate atoms are directly involved in this binding and which enzyme atoms are directly involved. That is, what new hydrogen bonds are formed during enzyme−substrate binding? Identification of the important atoms in the substrate is most often accomplished by traditional chemistry involving structure−activity relationships. If cyclohexane-1,2-diol binds much more effectively to an enzyme than cyclohexanol, this suggests that the second hydroxyl group is involved in the binding. In a less trivial case, the 12,13-dibutyl ester of phorbol (PDBU, **1**) binds to protein kinase C (PKC) about 100 000 times more efficiently than its 20-deoxy relative (**2**),[6,7] and this is powerful evidence for the importance in the PDBU binding of the oxygen atom in the side chain. During the 1980s many experiments of this sort were carried out with PKC,[8] and, as a result, it became increasingly clear that three oxygens in the phorbol skeleton, those at C_3, C_9, and C_{20}, were essential to its binding to PKC. If any of these three was missing, the binding activity was destroyed. With this piece of evidence in hand, it was fairly



1



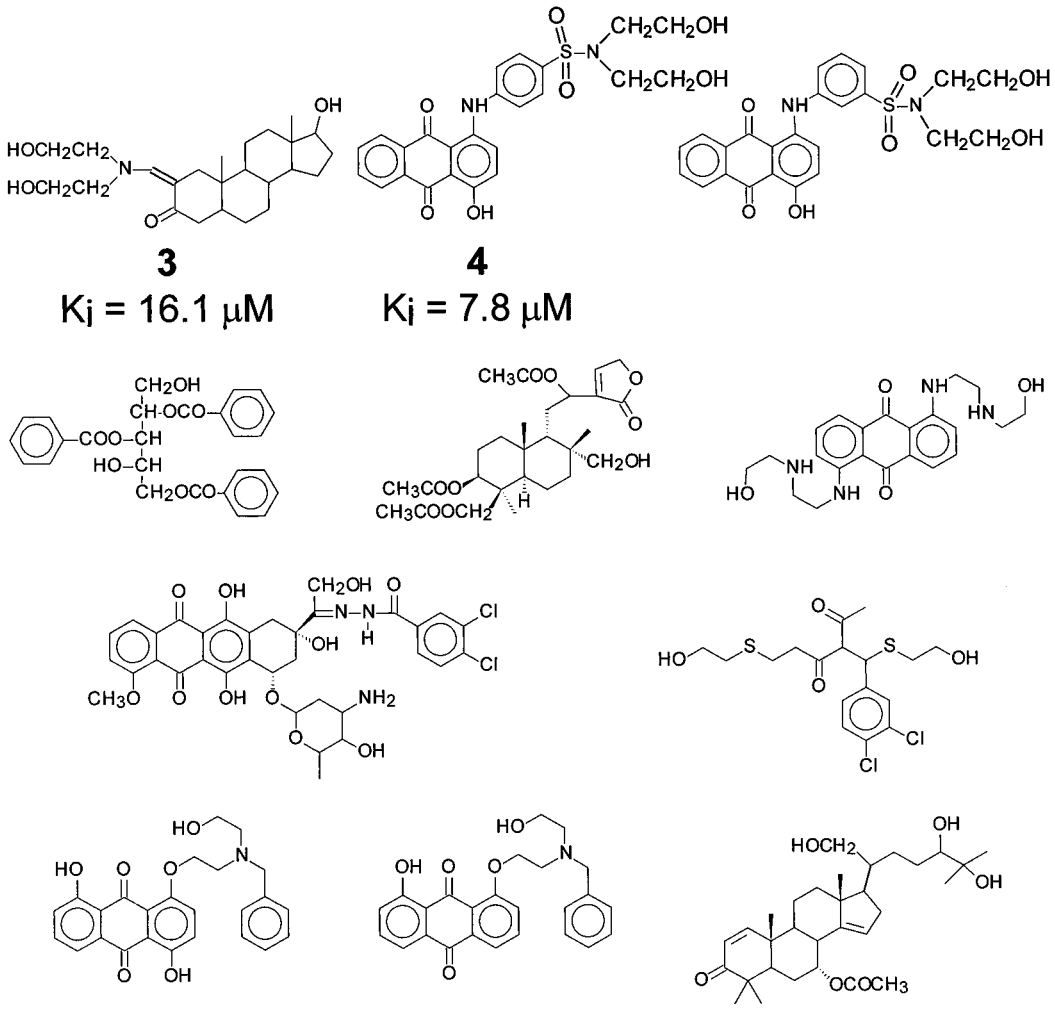
2

straightforward to determine the dimensions of the pharmacophore made up of these three atoms. The structure of

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1996.

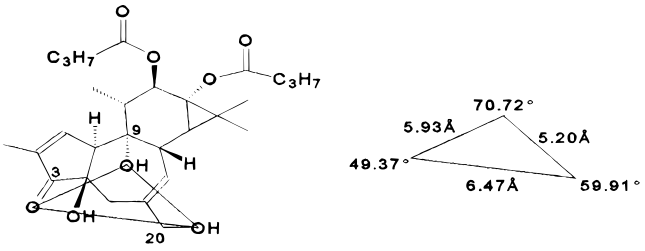
S0095-2338(95)00084-9

Chart 1



phorbol as determined by X-ray diffraction had been reported,[9] and this provided coordinates for the oxygens at C_3 and C_9 which are in a relatively rigid part of the structure. The side chain which carries the C_{20} hydroxyl group can rotate, and it was necessary to conduct a conformational search to find its most stable conformation and thus the best position for that oxygen.

Scheme 1



Once this was done, the pharmacophore in Scheme 1 was identified as important in the binding of PDBU to PKC. The dimensions of this pharmacophore were used in the pharmacophore searching that is described below.

In other cases, the characteristics of pharmacophores are measured in other ways. The operative pharmacophore in HIV protease, for example, was arrived at from a study of both the enzyme crystal structure and of different ligands. In general, however, the crucial information concerning pharmacophore dimensions is derived from the structures of different known substrates. Recently, computer programs which automate this deductive process have been developed[10] and are being used increasingly.

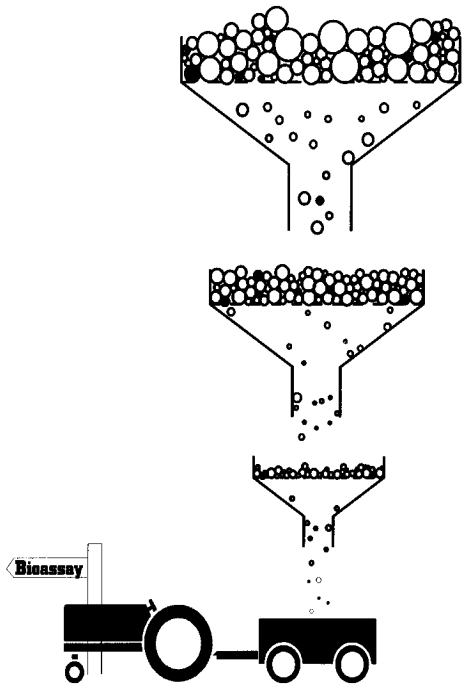
ACTIVE SITE CHARACTERIZATION

Information pertaining to the structure of the active site in the enzyme, which complements the pharmacophore, can be obtained in a number of ways. The most direct method is to measure the structure of an enzyme−substrate complex by X-ray diffraction.[11] If the solution to the structure of the native enzyme is in hand, analysis of the diffraction data is relatively easy; the difficult step is producing a crystal of the complex. In recent years, NMR spectroscopy has made tremendous progress in 3D structure analysis[12] and determination of the structure of enzyme−substrate complexes by NMR is now fairly common. A final technique that should be mentioned is site-directed mutation.[13] This is a biochemical method which allows one to replace a single amino acid residue in a protein with, say, glycine. If such a change eliminates enzymatic activity, then it is at least possible that the replaced residue was somehow involved in that activity. Analyses of this sort are very useful in identifying the residues involved in enzymatic activity.

PHARMACOPHORE SEARCHING

3D Database. Once the pharmacophore has been at least partially identified, the next step is to find compounds which

Chart 2



Full "open" database (250K)

Sample available (200K)

Possess pharmacophore (250)

Possess pharmacophore and appropriate solubility (25)

contain it embedded in their structure. To do this a database of three-dimensional structures is needed. The NCI database which was referred to earlier was assembled as a 2D database.[3] Structures were entered and organized so that the file could be searched for 2D structures or structure fragments. The main value of such substructure searching was not scientific, as had been hoped, but administrative. That is to say, substructure searches rarely yielded new active compounds, but the search system permitted easy control of the chemical content of the file and this proved to be one of its more useful aspects.

In 1990 we began the task of generating a 3D database from the 2D file.[14] At around this time, it became clear to all in the field that one major problem, that of conformational flexibility, was to dog this effort. When a 2D structure such as methane is used to produce 3D atomic coordinates, the process is a simple one of looking up bond lengths and bond angles and using them to build the only 3D structure which is consistent with all of these numbers. In the case of ethane, however, there is no information, *a priori*, about the torsion angle of the C−C bond, and unless it is known what this should be (eclipsed, 0° or skewed, 180° or some other value) a single 3D structure cannot be built. Early 2D → 3D programs understood this and made sensible attempts to arrive at the best torsion angle, but the successful approach was to allow many different torsions and so generate many different alternative 3D structures. The word "many" should be treated with care; the multiple alternatives at each rotatable bond combine multiplicatively, and it is not unusual to produce 10 000 3D structures from a single 2D structure.

After some exploratory work with programs which generated a single 3D structure from the 2D form,[15,16] attention was turned[15] to the use of Chem-X[17] which produces several conformationally isomeric 3D structures from a single 2D structure and assembles them into a searchable 3D database. It proceeds by assembling 3D molecular fragments and whenever it encounters a rotatable single bond (a "rotor") it assigns two (0°, 180°) or three (60°, 180°, 300°) specific

Table 1

	open[b]	discreet[b]	total	%
2D file (6/10/92)	233 985	218 322	452 307	100.0
no CT[a] available	4686	4430	9116	2.0
incomplete CT	1169	2243	3912	0.9
excluded atom	11 963	4671	16 634	3.7
other error	60	55	115	<0.1
valid CT	215 607	206 923	422 530	93.4
3D built	206 876	201 036	407 912	90.2

[a] Connection table. [b] See ref 2.

torsion angles, depending upon the nature and hybridization of the atoms at either end of the bond. Thus, one 2D structure can spawn many 3D structures. The generated structures are examined to remove those in which there are close contacts or which violate any of several conformational rules, and then all the interatomic distances are computed from every accepted 3D structure and these distances are packed into bit screens which form the basis of the 3D search system. With the NCI database, 2D → 3D conversion with Chem-X was largely complete; the statistics are given in Table 1. With another program, CORINA,[18] a much higher conversion rate was obtained because that program succeeded in converting many of the structures which were dropped as problems by Chem-X.

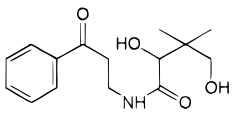
3D Searching. The search system is designed to search for all compounds which contain, for example, a carbonyl oxygen 3.50 ± 0.50 Å from the nitrogen of a tertiary amine. Multiple criteria can be combined with AND logic and so searching for a pharmacophore such as that of PDBU (Scheme 1) is quite straightforward. When this search was carried out through the 206 876 open compounds in the NCI 3D database, 535 compounds were found to contain the pharmacophore. Of these 410 were discarded because they contained no hydrophobic moiety—this being a known prerequisite for PKC activity. The remaining 125 compounds were tested experimentally for binding to PKC and 11 of these were found to bind strongly to the enzyme. The two best binders, **3** and **4** (Chart 1), had a measured binding

constant of 16.1 and 7.8 μM, respectively. All of these compounds are drug leads, and several of them have since been used as leads in our efforts to develop structures possessing superior inhibitory activity toward PKC.

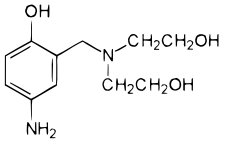
It is of interest to enquire why the 114 of the 125 compounds had no activity. They all contain the correct pharmacophore as well as the required hydrophobic residue, so why are they not active? The answer is that possession of the pharmacophore is a necessary-but-not-sufficient property, and the failure of the 114 to show activity could be attributed to either or both of two reasons.

The first of these can be traced back to the conformational flexibility issue. When generating the database, many conformational isomers were built for structures having free rotors, but no attention was paid to the energies of the individual conformers. In many cases, the internal energy associated with a specific conformer is so high, relative to the global minimum, that the conformer, in effect, will not exist. It will still be added to the database and may be the reason why a specific compound is retrieved from a pharmacophore search. In a bioassay however, the conformation which contains the pharmacophore and which could therefore bind to the enzyme is essentially absent from the system, since its formation requires far too much energy. Thus the compound shows no activity. As a practical matter, a limit of 10 kcal/mol of excess conformational energy was adopted. If the conformation that fits the rigid pharmacophore is 10 kcal/mol or more above the global energy minimum for that structure, the compound was rejected from the search results.

The second reason is related to physical properties. Compounds like **5** or **6** certainly possess the pharmacophore



5

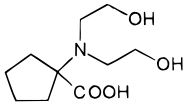


6

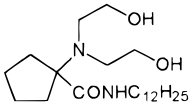
in a low energy conformation but are nevertheless inactive with respect to PKC. This enzyme is lipid-soluble, and it is known that the range of water solubility amongst PKC substrates is quite narrow. The relationship between K_i and the ligand water solubility (WS) is

$$\log (1/K_i) = -1.282 - 0.501 \log(\mathrm{WS}) - 0.0795[\log(\mathrm{WS})]^2$$

and the optimum value of log(WS) = −3.15.[19] The value of log(WS) for **5** or **6** is much higher and this explains their lack of activity. Support for this hypothesis was obtained from the carboxylic acid **7**, which essentially failed to bind to PKC. Its dodecylamide (**8**), however, was a moderately potent inhibitor of PKC, with a K_i value of 42 μM.[20] Solubility is a property which can be reliably estimated from the structure,[21] and so a solubility check can be incorporated into the drug discovery process.

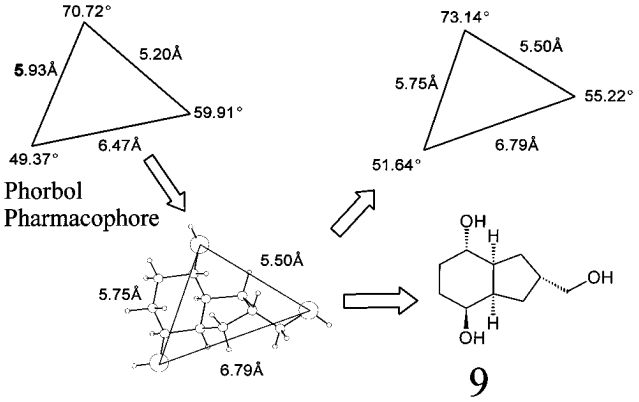


7



8

Scheme 2



9

ELECTRONIC SCREENING IN LEAD DRUG DISCOVERY

From experiments of this sort, the general procedure shown below has been developed to exploit the NCI Repository in the search for drug leads. As a first, administrative type of filter, the "open" database is divested of entries for which no physical sample remains. About 25% of the database is removed in this step. The second step, the pharmacophore search, typically eliminates >99% of the database leaving a fairly small number (50−500) of candidates for more detailed examination. In the third step, application of a single criterion, such as solubility or log *P*, generally eliminates more than half of the structures, leaving a fairly small number, manageable in a bioassay.

In practice, the pharmacophore search may have to be done repeatedly so as to establish the size of the "window" that is used. It also is often not clear *a priori* how narrowly the chemical specifications of the pharmacophore (atom type, functional group type, and so on) should be for optimal performance. This will depend upon the problem to a large extent. Finally, the precise nature of the last filter will also depend upon the problem. Solubility is a generally useful property in this connection, and it can be estimated by computer.[21] It is not difficult, however, to imagine conditions where economic or patent considerations might be invoked in this final filter.

DE NOVO SYNTHESIS

In the event that no satisfactory pharmacophore-containing compounds are found in the database, another option consists of designing compounds which contain the pharmacophore and which can be synthesized. This particular method has been explored by Johnson and his group at Leeds who have written a program called SPROUT[22] which can build structures around a predefined pharmacophore, seeking synthetically accessible possibilities. When the PKC pharmacophore was used to seed SPROUT and the appropriate structural constraints were defined, the program suggested only two structures, **9** and a stereoisomer (Scheme 2). As a consequence of these results, **9** and closely related compounds are currently being investigated as possible synthetic targets.

ACTIVE SITE

When a pharmacophore has been accurately characterized, it is possible, knowing the dimensions of hydrogen bonds,

Table 2

ligand	K_i (nM)	binding energy (kcal/mol) site A	site B	site C
PDBU	0.1	−9.68	−21.44	−18.72
4-deoxy-PDBU	10	−9.11	−18.30	−14.92
(S)-1,2-diacylglycerol	200		−12.03	−13.00
PDBU 4-methyl ether	200	−6.28	−17.49	−12.32
PDBU 20-methyl ether	>10 000	−9.41	−11.76	−10.57
4α-hydroxy PDBU	>10 000	−11.36	−18.90	−9.09
20-deoxy PDBU	>10 000	−8.50	−11.66	−7.89
(*R*)-1,2-diacylglycerol	>10 000		−9.70	−6.11

to define the geometry of the active site. If the overall structure of the enzyme is available from X-ray or NMR methods, one may search for an active site which corresponds to the dimensions that were derived from the pharmacophore. When this was done with PKC, using the NMR-derived structure,[23] 31 possible sites were found, but 28 of these could be eliminated from the extensive site-directed mutation data.[24] Calculation from the models of the binding constants of various ligands in these sites gave the data in Table 2 and showed one (site C) to be superior to the others.[25] When the total hydrogen bond energy between the ligand and the enzyme is calculated for each site, only site C gives values for the binding energy which are consistent with the measured binding affinities. This can be regarded as support for the correctness of this active site in protein kinase C.

SUMMARY

With an electronic screen like the pharmacophore search described here, it is possible to examine rapidly large databases and to extract the relatively few structures that meet some structural criterion. Bioassay of the few compounds that are retrieved can be readily conducted and this therefore offers a means of rescreening entire databases for activity with respect to a new target.

REFERENCES AND NOTES

(1) The starting for this effort is generally regarded as the authorization, in 1955 by the U.S. Congress 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) In the NCI database, about half of the compounds are treated as proprietary ("discreet") upon request of the supplier. The remainder, together with all the data associated with them, are open to the public.

(3) 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. P.; 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. P.; 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 Maintenance. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 193−197.

(4) 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.

(5) Jakes, S. E.; Watts, S. N.; Willett, P.; Bawden, D.; Fisher, J. D. Pharmacophoric Pattern Matching in Files of Chemical Structures: Evaluation of Search Performance. *J. Mol. Graph.* **1987**, *5*, 41−48.

(6) Wender, P. A.; Koehler, K. G.; 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.

(7) Nakamura, H.; Kishi, Y.; Pajares, M. A.; Rando, R. R. Structural Basis of Protein Kinase C Activation by Tumor Promoters. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9672−9676.

(8) See: Wang, S.; Milne, G. W. A.; Nicklaus, M. C.; Marquez, V. E.; Lee, J.; Blumberg, P. M. Protein Kinase C: Modeling of the Binding Site and Prediction of Binding Constants. *J. Med. Chem.* **1994**, *37*, 1326−1338 and references cited therein.

(9) Brandl, V. F.; Röhrl, M.; Zechmeister, K.; Hoppe, W. Röntgenstrukturanalysen von Neophorbol $C_{31}H_{35}O_9Br$, und Phorbol, $C_{20}H_{28}O_6$. *Acta Crystallogr.* **1971**, *B27*, 1718−1730.

(10) A program which attempts to characterize a pharmacophore from structure−activity data (Pharmacophore Identification in ChemDBS-3D) has been produced by Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, Oxon OX7 5SR U.K. Another such program, DISCO, was developed at Abbott Laboratories (Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Design* **1993**, *7*, 83−102).

(11) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of a Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, *246*, 1149−1152.

(12) Omichinski, J. G.; Trainor, C.; Evans, T.; Gronenborn, A. M.; Clore, G. M.; Felsenfeld, G. A Small Single-"Finger" Peptide from the Erythroid Transciption Factor GATA-1 Binds Specifically to DNA as a Zinc or Iron Complex. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1676−1680.

(13) Kazanietz, M. G.; Bustelo, X. R.; Barbacid, M.; Kolch, W.; Mischak, H.; Wong, G., Pettit, G. R.; Bruns, J. D.; Blumberg, P. M. Zing Finger Domains and Phorbol Ester Pharmacophore. *J. Biol. Chem.* **1994**, *269*, 11590−11594.

(14) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219−1224.

(15) Hendrickson, M. A.; Nicklaus, M. C.; Milne, G. W. A. CONCORD and CAMBRIDGE: Comparison of Computer-Generated Chemical Structures with X-ray Crystallographic Data. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 155−163.

(16) 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.

(17) Chem-X is produced by Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, Oxon OX7 5SR U.K.

(18) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.

(19) Wang, S.; Milne, G. W. A.; Nicklaus, M. C.; Marquez, V. E.; Lee, J.; Blumberg, P. M. Protein Kinase C: Modeling of the Binding Site and Prediction of Binding Constants. *J. Med. Chem.* **1994**, *37*, 1326−1338.

(20) Wang, S.; Zaharevitz, D. W.; Sharma, R.; Marquez, V. E.; Lewin, N. E.; Du. L.; Blumberg, P. M.; Milne, G. W. A. The Discovery of Novel, Structurally Diverse Protein Kinase C Agonists Through Computer-3D Database Pharmacophore Search. Molecular Modeling Studies. *J. Med. Chem.* **1994**, *37*, 4479−4489.

(21) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of Aqueous Solubility of Organic Molecules by the Group Contribution Approach. Application to the Study of Biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474−482.

(22) Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: Recent Developments in the de novo Design of Molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207−217.

(23) Hommel U.; Zurini, M. Solution Structure of a Cystein-Rich Domain of Rat Protein Kinase C. *Nature Struct. Biol.* **1994**, *1*, 383−387.

(24) Kazanietz, M. G.; Wang, S.; Milne, G. W. A.; Lewin, N. E.; Liu, H. L.; Blumberg, P. M. Residues in the Seconf Cystein-Rich Region of PKCδ Relevant to Phorbol Ester Binding. *J. Biol. Chem.* **1995**, *270*, 14679−14684.

(25) Wang, S.; Kazanietz, M. G.; Blumberg, P. M.; Marquez, V. E.; Milne, G. W. A. Molecular Modeling and Site-Directed Mutagenesis Studies of a Phorbol Ester Binding Site in Protein Kinase C. *J. Med. Chem.*, in press.

CI9500849