Diversity Analysis of 14 156 Molecules Tested by the National Cancer Institute for Anti-HIV Activity Using the Quantitative Structure−Activity Relational Expert System MCASE

Gilles Klopman* and Meihua Tu

Department of Chemistry, Case Western Reserve University, Cleveland, Ohio 44106

Received August 3, 1998

Using the MCASE program, a procedure to analyze the diversity of the large amount of available HIV-1 antiviral data was proposed. A subset of 1 819 chemicals was logically selected from the original 14 156 chemicals tested by NCI. This subset of chemicals was shown to contain most of the structural and the functional information of the original database. A full analysis of the 1 819 chemicals by the MCASE program produced a correlation between chemical structures and HIV antiviral activity. In our model, 74 fragments were identified as being responsible for all the chemical's HIV antiviral activity. These fragments may be related to different inhibiting mechanisms, some known and some probably still unknown. The expert system resulting from this analysis can be used to predict the activity of new chemicals and to design new agents that can target multiple enzymes. This was shown to be the case by using the model to predict the activity of 10 diverse chemicals whose activities were not known at the time of model development. Of these, 8 were predicted in agreement with experimental observations. As far as we can tell, this is probably the first project ever to attempt to create a quantitative model of activity for such a massive database of diverse chemicals.

Introduction

AIDS, the acquired immunodeficiency syndrome, is believed to be caused by a virus called "human immunodeficiency virus", or HIV. Billions of dollars have been spent worldwide to find an effective treatment for HIV infection. One of these projects is the AIDS Antiviral Screen, led by the Antiviral Evaluations Branch of the National Cancer Institute's (NCI's) Developmental Therapeutics Program. In this project thousands of chemicals were tested for their capability to inhibit the HIV virus.

QSAR refers to the statistical analysis of potential relationships between chemical structure and biological activity. QSAR can be viewed as a technique attempting to summarize chemical and biological information in a form that allows one to generate and test hypotheses to understand interactions between molecules. Indeed, when it is possible to elucidate quantitative relationships between chemical structure and biological activity, these relationships have been shown to be useful in describing possible mechanisms of interaction and predicting the activity of new structures with better properties than those used to formulate the original QSAR. Unfortunately, the results of such an analysis only serve to characterize trends and properties within the bounds of the learning set of data. Therefore, the larger the learning set, the better the QSAR's predicting ability. Recent advances in technology have made it possible to obtain and analyze large amounts of bioactivity data:

1. The development of automated synthesis capabilities along with the formulation of the combinatorial chemistry approach has enabled the rapid synthesis of large numbers of molecules.

2. The automation of in vitro bioassays affords high-throughput screening systems capable of generating massive amounts of data in a relatively short period of time.

3. The ever-expanding power of computers makes it possible to calculate hundreds or even thousands of descriptors to characterize chemical structures and to correlate these with specific biological or chemical properties.

Although such advances in technology provide more opportunities for QSAR practitioners, how to handle, summarize, and effectively use the huge amount of information generated by the high-throughput screening methods also brought a major challenge upon computational chemists. Molecular modeling and QSAR studies of anti-HIV-1 virus have been performed by a number of groups independently worldwide. However, all these approaches are based on a single inhibiting mechanism and a small number of congeneric molecules (less than 100) (Table 1).

In this paper, we use a somewhat different methodology capable of analyzing the activity of large sets of diverse molecules. The MCASE program is a quantitative structure−activity relational expert system (QSAR-ES) capable of learning automatically from data and organizing that knowledge into an expert system. In the course of this study, a new diversity analysis procedure was invented and used to reduce the data to a more manageable level. (The MCASE program is available from MULTICASE Inc., 25825 Science Park, Suite 100, Beachwood, OH 44122. URL address: http://www.multicase.com.)

* Corresponding author.

10.1021/jm980451i CCC: $18.00
Published on Web 03/09/1999

Table 1. Summary of Computational Approaches to HIV Antiviral Activity

inhibiting	methodology	refs
protease	traditional QSAR	1, 2
	molecular modeling pharmacophore searching	3−7
	3D QSAR	8, 9
reverse transcriptase	traditional QSAR	10, 11
	3D QSAR	12−15
	molecular modeling pharmacophore searching	16−17
integrase	traditional QSAR	18
	molecular modeling pharmacophore searching	5, 19−24
	3D QSAR	25
replication	3D QSAR	26, 27

The Database

The HIV antiviral data were obtained from the NCI archive (http://epnws1.ncifcrf.gov:2345/dis3d/aids_screen/aidspub.html). The AIDS Antiviral Screen methodology was developed by the NCI in the hope of discovering new compounds capable of inhibiting the HIV virus more effectively. The screen uses a soluble formazan assay to measure protection of human CEM cells from HIV-1 infection.[28] The maximum culture concentration for all the compounds is 0.25%, which had no apparent direct toxic effects on the cell lines used or the HIV-1 infection.

The results of the screening tests are evaluated and placed in one of three categories: confirmed active (CA), confirmed moderately active (CM), or confirmed inactive (CI). CA, CM, and CI are defined as follows: compounds that are able to provide at least 50% protection to the CEM cells are retested. Compounds that provide at least 50% protection on retest are listed as "moderately active". Compounds that reproducibly provide 100% protection are listed as "confirmed active". Compounds that do not meet these criteria are listed as "confirmed inactive" (note that the inability to provide protection could be due to toxicity).

The released AIDS screening results contain 24 110 compounds, of which 230 are confirmed active, 444 are confirmed moderately active, and all others are confirmed inactive. Because the current version of MCASE lacks the ability to handle chemicals containing metal atoms and some other uncommon inorganic atoms such as B, Si, Se, Te, etc., in our analysis, all the chemicals containing such atoms in the original database were removed. After this, there were 14 156 chemicals left, of which 175 are confirmed active, 318 are confirmed moderately active, and all others are confirmed inactive.

The MCASE Program

The detailed algorithm used in the MCASE program was described in a couple of papers previously published by Klopman.[29,30] Basically, after the user inputs a set of chemical structures and their respective experimentally determined biological activities, the program is capable of identifying substructural descriptors (fragments and/or distances) that may be associated with the biological activity of the chemicals. The molecular fragments are generated as a result of breaking down each individual chemical structure into its constituent parts. The fragments normally are linearly connected atoms including, if necessary, a side chain. They can be as small as two heavy atoms (non-hydrogen) and can be as large as required. Each fragment is "labeled" with an activity index that is associated with its parent compound. The distance descriptors are two-dimensional distances between atoms within a chemical structure.

MCASE will find descriptors that have the highest probability of being related to the observed biological activity (biophore) and/or inactivity (biophobe). The chemicals will then be divided into different groups according to the biophores they contain. After this, MCASE attempts to derive a "local" QSAR within each group of compounds in order to identify molecular features that control the degree of activity. These features, termed modulators, are selected from the pool of molecular fragments, distance descriptors, calculated electronic indices (molecular orbital energies, charge densities), and calculated transport parameters (octanol/water partition coefficient, water solubility). In this way, the program will generate the best correlation between the chemical structures and their observed biological activities.

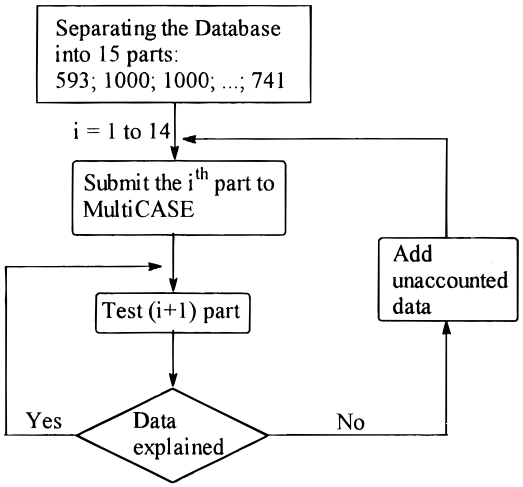


Figure 1. Database processing algorithm.

Database Processing

Because the MCASE program selects substructural descriptors by decomposing the learning set of chemical structures into fragments, when the total number of chemicals of the learning set is large, millions of fragments may be generated. The memory required and CPU time needed to manage such a vast array of data are beyond the current capability of most computers. Therefore, we propose a new method to process the information generated by such a large learning set.

The object of the analysis is to select the smallest and most representative subset of chemicals from the original learning set. This subset must possess all or most of the information contained in the original database. When applied to predict the activity of new chemicals, the subset must show similar predicting ability as would have been obtained using the full database. The paradigm to guide the selection of the subset is to minimize redundancies. In a large database, there are usually many combinatorial redundant chemicals. Such combinatorial redundant chemicals have different structures, but when decomposed into fragments, they possess similar substructures. Therefore, it is possible to select a subset of diverse chemicals from the full learning set that still contains all of the useful information within the boundary of the description of the substructures.

Here, we propose two new terms for diversity: structural diversity and functional diversity. The structural diversity refers to chemicals containing the maximum number of different substructures, and the functional diversity refers to chemicals containing the maximum amount of activity information. Hence, the optimal subset of the large database should have both maximum structural diversity and functional diversity.

Structural Diversity Selection. The following procedure (Figure 1) was used to select the subset of chemicals that contain all the substructural information of the full database.

The first step of the preprocessing consisted in dividing the original database into 15 parts. The first part contains all the active chemicals (493, both confirmed active and moderately active) and 100 inactive chemicals randomly selected from the original database. Therefore, the first part contains 593 chemicals. Each of the other parts contains 1 000 randomly selected inactive chemicals (except for the last part, made up of the remaining chemicals).

We use the first part of chemicals as a learning set to test the second part. The output of MCASE will indicate which chemicals of the test set (second part) contain fragments unknown to the learning set (first part). For the purpose of this study, MCASE will report an unknown structural group when it encounters a sequence of two or three "heavy" atoms that have not been seen in the learning set. A "heavy" atom consists of an atom other than a hydrogen atom, with hydrogen atoms and double-bonded oxygen or sulfur atoms attached. Examples of "heavy atom" sequences of sizes 2 and 3 are $-CH_2-OH$; $-NH-CO-OH$; $-O-CS-CH=$. The following is an example of output (biophores refer to fragments believed to be responsible for activity):

```
*********************************************************************
                    Molecule nr.  35;
*********************************************************************
*** WARNING *** The following functionalities are UNKNOWN to me:
      *** NH2-O  -C  =

MCASE-3 Prediction
------------------

         ** The molecule does not contain any known Biophore **
               It is therefore presumed to be INACTIVE
```

From this kind of output, we can identify the chemicals that contain fragments unknown to the learning set. Some of the tested chemicals contain several unknown fragments, while others only contain one. It is therefore possible to select the minimum number of chemicals that contain all the unknown fragments of the test set. The chemicals containing the best sample of unknown fragment information are selected and added to the learning set (first part). This new learning set now contains all the fragment information existing in both part one and part two chemicals. We then use the updated part one learning set to test the third part and continue until the structural diversity of all parts are accounted for.

The final learning set, obtained after executing the above steps, contained 1 048 chemicals. When this new learning set is used to test the original 14 156 chemicals, no more unknown fragments are found. Hence, we can declare that the new learning set of 1 048 chemicals contains all the substructural diversity information of the original database.

Functional Diversity Selection. The learning set obtained from the structural diversity selection only contains the substructural information of the original database. As mentioned before, the new learning set must also possess the same predicting ability as the full database. Therefore, we also need to do a functional diversity selection from the original database so as to include all the activity information.

The functional diversity selection can be performed in the same fashion as the structural diversity selection (Figure 1). The chemicals not included in the learning set obtained above (14 156 − 1 048 = 13 108) are randomly divided into 13 parts. All of these chemicals are inactive, because all the active chemicals had been made part of the learning set in the structural diversity selection step. Using this learning set, we test the first part of the remaining chemicals. MCASE will use its knowledge to predict the activity of the molecules of this part.

When this is done, we find that some of the inactive chemicals of the test set are predicted inactive because they do not contain any biophore, but some of them are erroneously predicted active because they contain a substructure deemed relevant by the learning set. It is therefore necessary to evaluate the inactive chemicals incorrectly predicted active and identify the biophores that were responsible for the erroneous conclusions.

For example, we find in one case that five experimentally inactive molecules contain the substructure that was selected to explain the activity of two molecules of the learning set. Thus even though both of the chemicals that contain this substructure are active in the learning set, there are five chemicals in the test set that contain it but are inactive. Therefore, it is clear that the statistical basis for declaring this substructure to be a biophore is flawed, and it will be necessary to add one or two of the inactive chemicals from the test set into the learning set in order to inform the program of this fact. When this is done, then this substructure will no longer be selected as a biophore and a better prediction will be obtained. We analyzed all results in this fashion and added chemicals that were predicted incorrectly from the test set to the learning set. The number of chemicals to be added for an ineligible biophore was somewhat arbitrarily set to be equal to the number of chemicals in the learning set containing such a biophore. In this way, we can continuously update the learning set so that better predictions will be produced. While the empiricism of this procedure may have an effect on the final model, we find that small modifications do not generally affect the results in any significant way.

We can now proceed to use the new learning set to test the next part of chemicals and, using the same process, continue to improve the quality of the learning set. These steps are repeated, until all parts of chemicals have been processed.

After the functional diversity selection, we obtained an expanded learning set that contained 1 819 chemicals. We postulate that this learning set contains most of the structural and activity information of the original database.

Validation Experiments. Two steps are required to validate the model. One is to make sure that it contains all the structural and activity information of the original database. The second is to make sure that it provides good predictivity.

Structural and Activity Validation. In this validation, we use the model based on the final learning set of 1 819 chemicals to test all the chemicals of the original database (14 156 chemicals). The MCASE results are summarized in Table 2.

From this table, we can see that with the reduced learning set, the program was able to identify 100% of the experimentally active compounds and 96.8% of the experimentally inactive compounds. Furthermore, there are no unknown fragments identified during the test.

Table 2. Test Results of the Full Database with MCASE Based on the Reduced Learning Set

	experimental active	experimental inactive
predicted active	493	438
predicted inactive	0	13225

Table 3. Predicted Results for "Unknown" Chemicals

	experimental active	experimental inactive
predicted active	32	17
predicted inactive	14	112

Therefore, we can declare that the reduced learning set contains all the structural information and almost all of the activity information of the original large database.

The erroneous results for the 438 inactive compounds predicted to be active is probably inherent to the methodology. For example, it may happen that during our functional diversity selection we find that 50 active chemicals of the learning set contain a given biophore but 10 inactive chemicals of the test sets also contain it. In such a case, none of the inactive compounds would be forced into the learning set because they will not affect significantly the statistics leading to the selection of this biophore.

Predictivity Validation. The object of this validation is to make sure that models created from a subset of molecules are capable of predicting the activity of the chemicals left out of the learning set. For this purpose, we randomly separated the 1 819 compounds into two subsets. One contained 10% of the chemicals (I/M/A = 129/0/46) and the other 90% of the chemicals (I/M/A = 1197/0/447). The large subset was then used as the learning set for the construction of a new model. It is to be noted that by doing this, we forego the benefits of the diversity analysis since 10% of desirable chemicals have been left out of the design of this "diminished" model. These 10%, i.e., 175 molecules, were then submitted as a test set to the "diminished" model. The results are shown in Table 3.

We can see that the model predicted 32 of the 46 active chemicals correctly (70%), while 112 of the 129 inactive chemicals were correctly (87%) identified. The total prediction rate is 82% (144/175). We have repeated this experiment two more times and found the total prediction rates to be 80% and 82%, respectively.

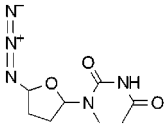
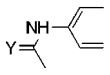
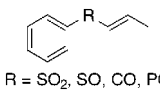
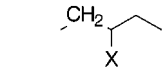
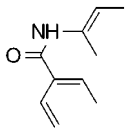
Considering the highly compact and diverse nature of our database, we believe that the results of the prediction are highly acceptable. Indeed, the exercise is particularly harsh because by removing 10% of the compounds, we basically voided our procedure leading to optimal structural and functional diversity.

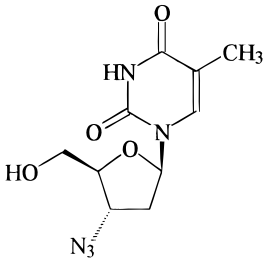
Results and Discussion

Biophore and Biophobe. Using the reduced but optimized learning set of 1 819 compounds, MCASE identified 78 different biophores for HIV antiviral activity. The most statistically significant biophores are listed in Table 4.

New molecules containing any of these biophores will have a high probability of being active. Furthermore, combining biophores may help design molecules that can inhibit HIV virus via several mechanisms. However, in general, the program was only able to generate a semiquantitative assessment of the importance of the biophores and modulators because the inputted data did not provide enough detailed information to create truly quantitative structure−activity relationships.

Table 4. Most Significant Biophores for HIV Antiviral Activity

chemical no.	biophores	total	I/M/A	average activity
1		52	0/0/52	64.0
2	 Y = S or O	50	0/0/50	60.0
3	 R = SO_2, SO, CO, PO	23	0/0/23	51.0
4	 X= F, Cl, I, epoxidic O	40	2/0/38	56.0
5		23	2/0/21	37.0



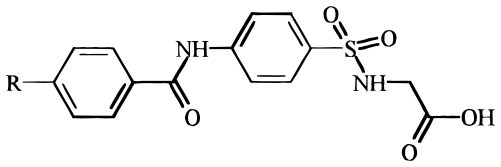
"AZT"

Figure 2. Structure of 3′-azido-3′-deoxythymidine.

From Table 4, we can see that the most significant biophore is a complex functionality encompassing much of the structure of 3′-azido-3′-deoxythymidine (AZT; Figure 2). Azido compounds are widely believed to be capable of inhibiting HIV replication. Indeed, AZT has been the anti-HIV-1 drug of choice for some time, and many azido compounds are active as well. However, because these compounds are very strong electrophilic agents, high toxic side effects have limited their usefulness.[31−36] There are no modulators associated with this biophore, as all the compounds that contain this fragment are very active. This probably also reflects the fact that relatively little diversity exists in all these molecules, and therefore little information is available about the modulating effect of the structures to which it is attached.

Biophore number **2** is a thioamide (or amide) of an aromatic aniline derivative. All 50 chemicals that contain such a biophore are very active, and therefore, there are no modulators associated with this biophore either.

It is interesting to note that this biophore contains an element of biophore **1**, namely a conjugated amide group. This may indicate that the molecules that contain



R = F or OCH_3

Figure 3. Inactive compounds that contain biophore **3**.

this biophore may act, at least partially, by a similar mechanism as that followed by the azido molecules.

Biophore number **3** is a hydrogen bond acceptor connected to two hydrophobic benzene rings. One of the most important modulators for this biophore is the coefficient of the atomic p_Π orbital of the oxygen atom in the highest occupied molecular orbital (HOMO) of the molecules (eq 1):

$$\text{activity} = 35.3 + 28.6S1 + 29.0S2 + 58.7S3 \quad (1)$$

where S1 is the presence of a tertiary methylamine (CH_3-N), S2 is the presence of a polycyclic aromatic group, and S3 is the coefficient of the atomic orbital of the oxygen in the HOMO. This coefficient determines the ability of the oxygen atom to form hydrogen bonds with a donor. Hydrogen bond capability will help the molecule bind to the macromolecule, which we suppose to be the HIV-1 integrase. Indeed all the active chemicals designed from the integrase pharmacophore contain this biophore.[22] There are 23 chemicals in our learning set that contain biophore **3**, all of them active.

Biophore number **4** is a halogen-substituted aliphatic fragment. There are 40 chemicals containing such a biophore. Two of them are inactive, and 38 are active. While we still do not know the exact mechanism associated with this and most subsequent biophores, these fragments significantly contribute to the observed activities and are very probably the pharmacophores for unknown mechanisms of activity.

Biophore number **5** again contains the amide of an aniline (see biophores **1** and **2**) with the additional proviso that the amide itself is now aromatic as well. The square of the logarithm of the partition coefficient, $(\log P)^2$, was identified as the most significant modulator. The overall QSAR found to describe the activity of the molecules containing this biophore is as follows (eq 2):

$$\text{activity} = 29.1 + 35.0S1 - 20.0S2 + 0.6S3 \quad (2)$$

where S1 is the presence of O−CH, S2 is the presence of $SO_2-NH-CH_2$, and S3 is the square of the log of the octanol/water partition coefficient.

The fact that, in some cases, the HIV antiviral activity depends on the agent's ability to partition into the lipid

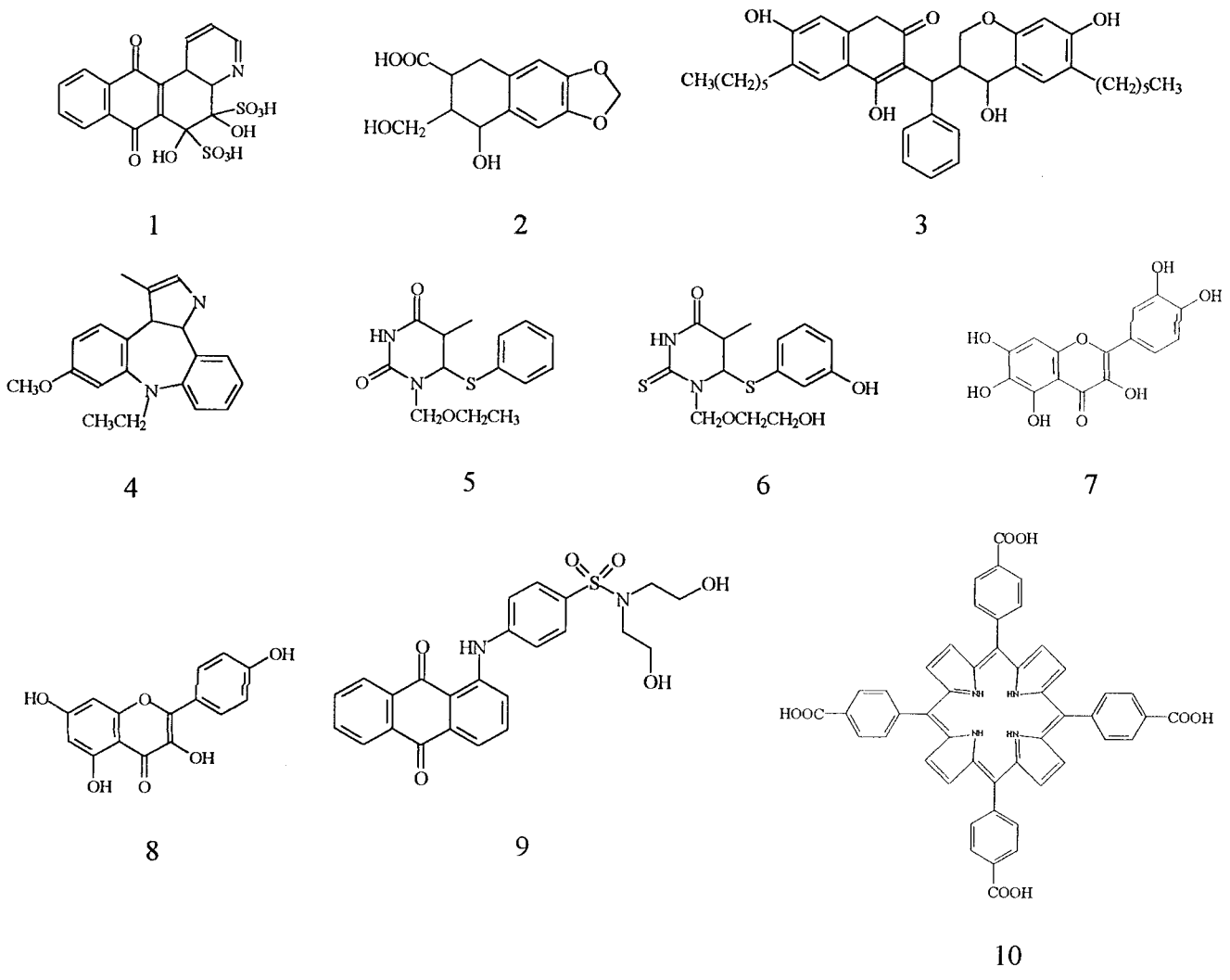


Figure 4. Ten chemicals used to test the MCASE model.

Table 5. Predicted and Tested Results of 10 Chemicals

chemical no.	inhibiting	ref	experimental result	prediction
1	protease	6	active	active
2	protease	6	inactive	inactive
3	protease	6	active	inactive
4	reverse transcriptase	14	active	active
5	reverse transcriptase	15	active	active
6	reverse transcriptase	15	inactive	inactive
7	integrase	25	active	inactive
8	integrase	25	inactive	inactive
9	integrase	22	active	active
10	replication	27	active	active

domain of HIV-1 protease was noted previously.[6] This dependence is reinforced by the observation that $SO_2-NH-CH_2^-$, which is strongly polar and hydrophilic, is identified as a biophobe (fragment that inhibits activation) by MCASE. In fact, both of the inactive chemicals that contain biophore **5** also possess the $SO_2-NH-CH_2^-$ group (Figure 3), while none of the active chemicals contain this group. Overall, $-SO_2-NH-CH_2-$ appears in nine molecules, all inactive although some of them also contain a biophore.

Predicting the Activity of Unknown Compounds. Ten compounds (Figure 4) from various sources (see references in Table 5) were selected for further testing of our model. None of the compounds existed in our learning set. Because the activity of these compounds was tested under different experimental conditions than those used by NCI, it is impossible for us to predict the exact IC_{50} value. Therefore, these chemicals were simply categorized as active or inactive according to the original experimental results. The test results and experimental observation of the activities of these 10 chemicals are listed in Table 5.

From Table 5 and in line with the results obtained above, MCASE predicts 5 of the 7 experimentally active compounds to be active. All 3 experimentally inactive compounds were correctly predicted to be inactive. Overall, we believe that this is a remarkable result (8 out of 10 molecules were predicted correctly), considering the high diversity of the structures of the tested compounds. Nevertheless, this is only a very small sample of chemicals, and it would be dangerous to extrapolate the results and claim 80% predictivity without further tests.

Conclusions

In this paper, using the MCASE program, we have successfully extracted a diverse subset of compounds from the large NCI HIV-1 antiviral database of 14 156 compounds. This subset of 1 819 compounds was shown to cover both the structural and activity information of the full database.

Our MCASE study shows that certain structure−activity relationships exist among the HIV-1 antiviral agents. Several substructural features believed to activate or deactivate HIV-1 antiviral activity have been identified. Furthermore, we found that log *P* and the HOMO coefficient of hydrogen bond acceptors are important factors for the activity of some biophores. While all the biophores are seen to lead to HIV-1 antiviral activity, we believe that they may be involved in different mechanisms, like replication inhibition, protease inhibition, integrase inhibition, and reverse transcriptase inhibition. With the help of the resulting model, we have tested 10 highly diverse chemicals that came from different sources, the overall accuracy of our prediction being 80%. This result provides a first glance at the possible predictivity of our methodology.

While there are a number of models proposed for HIV antiviral activity, all of them are based on a limited number of chemicals and work only for a single mechanism. Therefore, the results may lack generality. The diversity analysis technique we proposed here has made it possible to analyze very large amounts of high-throughput screening data, which is impossible to do by any traditional method. As far as we can tell, this is probably the first project ever to attempt to create a quantitative model of activity for such a massive database.

The biophores obtained from the MCASE analysis may not be as straightforward as those pharmacophores obtained from CoMFA and other molecular modeling packages, but they provide a better explanation of the large amount of screening data for which the exact mechanism of action is unknown. Furthermore, we believe that the biophores obtained from the MCASE analysis provide a good starting point for the process of pharmacophore search. Indeed, both CoMFA and all other molecular modeling packages need to align and superimpose molecules. This is impossible for diverse data unless a procedure such as MCASE is used to subdivide the database into logical subsets.

Acknowledgment. We thank the Center for AIDS Research at CWRU for the award of a Research Development Award, sponsored by NIH Grant AI 36219, for the partial support of this work.

References

(1) Wilkerson, W. W.; Dax, S.; Cheatham, W. W. Nonsymmetrically substituted cyclic urea HIV protease inhibitors. *J. Med. Chem.* **1997**, *40*, 4079−4088.
(2) Wei, D. T.; Meadows, J. C.; Kellogg, G. E. Effects of entropy of QSAR equations for HIV-1 protease: 1. Using hydropathic binding descriptors: 2. Unrestrained complex structure optimizations. *Med. Chem. Res.* **1997**, *7*, 259−270.
(3) Thaisrivongs, S.; Romero, D. L.; Tommasi, R. A.; Janakiraman, M. N.; Strohbach, J. W.; Turner, S. R.; Biles, C.; Morge, R. R.; Johnson, P. D.; Aristoff, P. A.; Tomich, P. K.; Lynn, J. C.; Horng, M. M.; Chong, K. T.; Hinshaw, R. R.; Howe, W. J.; Finzel, B. C.; Watenpaugh, K. D. Structure-based design of HIV protease inhibitors: 5,6-Dihydro-4-hydroxy-2-pyrones as effective, non-peptidic inhibitors. *J. Med. Chem.* **1996**, *39*, 4630−4642.
(4) Quere, L.; Wenger, T.; Schramm, H. J. Triterpenes as potential dimerization inhibitors of HIV-1 protease. *Biochem. Biophys. Res. Commun.* **1996**, *227*, 484−488.
(5) Mazumder, A.; Wang, S.; Neamati, N.; Nicklaus, M.; Sunder, S.; Chen, J.; Milne, G. W. A.; Rice, W. G.; Burke, T. R.; Pommier, Y. Antiretroviral agents as inhibitors of both human immunodeficiency virus type 1 integrase and protease. *J. Med. Chem.* **1996**, *39*, 2472−2481.
(6) Wang, S.; Milne, G. W. A.; Yan, X.; Posey, I. J.; Nicklaus, M. C.; Graham, L.; Rice, W. G. Discovery of novel, non-peptide HIV-1 protease inhibitirs by pharmacophore searching. *J. Med. Chem.* **1996**, *39*, 2047−2054.
(7) Doweyko, A. M. Three-dimensional pharmacophores from binding data. *J. Med. Chem.* **1994**, *37*, 1769−1778.
(8) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-dimensional QSAR of human immunodeficiency virus (I) protease inhibitors. 1. A CoMFA study employing experimentally determined alignment rules. *J. Med. Chem.* **1993**, *36*, 4152−4160.
(9) Oprea, T. I.; Waller, C. L.; Marshall, G. R. Three-dimensional quantitative structure−activity relationship of human immunodeficiency virus (I) protease inhibitors: 2. Predictive power using limited exploration of alternate binding modes. *J. Med. Chem.* **1994**, *37*, 2206−2215.

(10) Tronchet, J. M. J.; Grigorov, M.; Dolatshahi, N.; Moriaud, F.; Weber, J. A QSAR study confirming the heterogeneity of the HEPT derivative series regarding their interaction with HIV reverse transcriptase. *Eur. J. Med. Chem.* **1997**, *32*, 279−299.
(11) Mager, P. P.; De Clercq, E.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Baba, M.; Walther, H. Molecular simulation of 5,6-substituted 1-((2-hydroxyethoxy)methyl)uracils with anti-HIV-1 activity. *Eur. J. Med. Chem.* **1996**, *31*, 701−712.
(12) Hannongbua, S.; Lawtrakul, L.; Sotriffer, C. A.; Rode, B. M. Comparative molecular field analysis of HIV-1 reverse transcriptase inhibitors in the class of 1(2-Hydroxyethoxy)-methyl)-6-(phenylthio)thymine. *Quant. Struct.-Act. Relat.* **1996**, *15*, 389−394.
(13) Kireev, D. B.; Chretien, J. R.; Raevsky, O. A. Molecular modeling and quantitative structure−activity studies of anti-HIV-1 2-heteroarylquinoline-4-amines. *Eur. J. Pharmacol.* **1995**, *30*, 395−402.
(14) Gussio, R.; Pattabiraman, N.; Zaharevitz, D. W.; Kellogg, G. E.; Topol, I. A.; Rice, W. G.; Schaeffer, C. A.; Erickson, J. W.; Burt, S. K. All-atom models for the nonnucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: A 3D QSAR approach. *J. Med. Chem.* **1996**, *39*, 1645−1650.
(15) Kireev, D. B.; Chretien, J. R.; Grierson, D. S.; Monneret, C. A 3D QSAR study of a series of HEPT analogues: The influence of conformational mobility on HIV-1 reverse transcriptase inhibition. *J. Med. Chem.* **1997**, *40*, 4257−4264.
(16) Artico, M.; Silvestri, R.; Massa, S.; Loi, A. G.; Corrias, S.; Piras, G. 2-Sulfonyl-4-chloranilino moiety: A potent pharmacophore for the anti-human immunodeficiency virus type 1 activity of pyrrolyl aryl sulfones. *J. Med. Chem.* **1996**, *39*, 522−530.
(17) Fisher, M. A.; Yadav, P. N. S.; Yadav, J.; Kristol, D.; Arnold, E.; Modak, M. J. Identification of a pharmacophore for nucleoside analogue inhibitors directed at HIV-1 reverse transcriptase. *J. Mol. Recog.* **1994**, *7*, 211−214.
(18) Buolamwini, J. K.; Raghavan, K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. Application of the electrotopological state index to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors. *Pharm. Res. (New York)* **1996**, *13*, 1892−1895.
(19) Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S.; Milne, G. W. A.; Pommier, Y.; Burke, T. R. Hydrazide-containing inhibitors of HIV-1 integrase. *J. Med. Chem.* **1997**, *40*, 937−941.
(20) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke, T. R.; Pommier, Y.; Milne, G. W. A. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem.* **1997**, *40*, 930−936.
(21) Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M. C.; Milne, G. W. A.; Proksa, B.; Pommier, Y. Depsides and depsidones as inhibitors of HIV-1 integrase: Discovery of novel inhibitors through 3D database searching. *J. Med. Chem.* **1997**, *40*, 942−951.
(22) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W.; Pommier, Y. HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920−929.
(23) Zhao, H.; Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Milne, G. W. A.; Pommier, Y.; Burke, T. R. Coumarin-based inhibitors of HIV integrase. *J. Med. Chem.* **1997**, *40*, 242−249.
(24) Fesen, M. R.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W. Inhibition of **HIV**-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. *Biochem. Pharmacol.* **1994**, *48*, 595−608.
(25) Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. Three-dimensional quantitative structure−activity relationship (QSAR) of HIV integrase inhibitors: A comparative molecular field analysis (CoMFA) study. *J. Med. Chem.* **1995**, *38*, 890−897.
(26) Joao, H. C.; De Vreese, K.; Pauwels, R.; De Clercq, E.; Henson, G. W.; Bridger, G. J. Quantitative structural activity relationship study of bis-tetraazacyclic compounds. A novel series of HIV-1 and HIV-2 inhibitors. *J. Med. Chem.* **1995**, *38*, 3865−3873.
(27) Debnath, A. K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P.; Neurath, A. R. Three-dimensional structure−activity analysis of a series of porphyrin derivatives with anti-HIV-1 activity targeted to the V3 loop of gp120 envelope glycoprotein of the human immunodeficiency virus type 1. *J. Med. Chem.* **1994**, *37*, 1099−1108.
(28) Weislow, O. S.; Kiser, R.; Fine, D. L.; Bader, J. P.; Shoemaker, R. H.; Boyd, M. R. New soluble formazan assay for HIV-1 cytopathic effects: application to high flux screening of synthetic and natural products for AIDS antiviral activity. *J. Natl. Cancer Inst.* **1989**, *81*, 577−586.
(29) Klopman, G. MULTICASE 1. A hierarchical computer automated structure evaluation program. *Quant. Struct.-Act. Relat.* **1992**, *11*, 176−184.
(30) Klopman, G. Artificial intelligence approach to structure−activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(31) Larder, B. A.; Darby, G.; Richman, D. D. HIV with reduced sensitivity to ziaovudine (AZT) isolated during prolonged therapy. *Science* **1989**, *243*, 1731−1734.
(32) St. Clair, M. H.; Martin, J. L.; Tudor-Williams, G.; Bach, M. C.; Vavro, C. L.; King, D. M.; Kellam, P.; Kemp, S. D.; Larder, B. A. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science* **1991**, *253*, 1557−1559.
(33) Montaner, J. S. G.; Singer, J.; Schechter, M. T.; Raboud, J. M.; Tsoukas, C.; O'Shaughnessy, M.; Reudy, J.; Nagai, K.; Salomom, H.; Spira, B.; Wainberg, M. A. Clinical correlates of in vitro HIV-1 resistance to zidovudine. Results of the multicentre Canadian AZD trial. *AIDS* **1993**, *7*, 189−196.
(34) Richman, D. D.; Fischl, M. A.; Grieco, M. H.; Gottlieb, M. S.; Volberding, P. A.; Laskin, O. L.; Leedom, J. M.; Groopman, J. E.; Mildvan, D.; Hirsch, M. S.; Jackson, G. G.; Durack, D. T. The toxicity of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. *N. Engl. J. Med.* **1987**, *317*, 192−197.
(35) Dainiak, N.; Worthington, M.; Riordon, M. A.; Kreczko, S.; Goldman, L. AZT inhibits proliferation in vitro of human haematopoietic progenitor cells. *Br. J. Haematol.* **1988**, *69*, 299−304.
(36) Gallicchio, V. S.; Hughes, N. K.; Hulette, B. C. In vitro modulation of the toxicity associated with the use of zidovudine on normal murine, human, and murine retrovirus-infected hematopoietic progenitor stem cells with basic fibroblast growth and synergistic activity with interleukin-1. *J. Leukoc. Biol.* **1992**, *51*, 336−342.

JM980451I