Computer-Aided Selection of Compounds for Antitumor Screening: Validation of a Statistical-Heuristic Method

LOUIS HODES

National Cancer Institute, Silver Spring, Maryland 20910

Received February 3, 1981

Recent work helped confirm the validity of a previously published statistical-heuristic method for selecting compounds to be processed in the National Cancer Institute (NCI) mouse tumor prescreen. One study involved about 35 000 compounds which satisfied certain biological and chemical criteria. These compounds, taken from the NCI collection, were ranked according to predicted probability of activity. The results showed 34% of the active compounds in the top 10% of the ranking. In a second study the predictive ability of a chemist and the computer were compared by having each rate almost 1000 previously unselected compounds for activity. The results were about equal with respect to the yield of biologically active compounds, though their selection agreement was fairly low.

INTRODUCTION

This is the third in a series of papers dealing with the development and application of a statistical-heuristic method to select compounds for antitumor testing. The first paper[1] described the method and compared its performance to other prediction methods. The second paper[2] was a progress report on how the method applied to the large volume of National Cancer Institute (NCI) data to find new classes of drugs showing activity in the established in vivo prescreen, mouse lymphocytic leukemia or P388. This prescreen is the initial stage of a large-scale antitumor screening program.

Progressively, this work is becoming important for the acquisitions effort. Compounds are collected from many sources in the form of structures, which are fed into computer programs. These produce activity and novelty measures that aid the chemist in selecting compounds for screening. This paper covers two studies which helped verify the utility of the method.

The first study was a standard large-scale test of the method on the collection of about 35 000 compounds described earlier.[2] These were all the NCI compounds with definitive P388 results at that time except those belonging to well-known active chemical classes. A new model was constructed from a training set consisting of 80% of the compounds. The remaining 20% comprised a test set from which scores were compared to the actual activity.

The second study used a model constructed from the same 35 000 compound training set. A sample of almost 1000 previously unselected compounds was sent to P388 testing after being rated by both a chemist and the computer. A surprising outcome was that the chemist and the computer were roughly comparable in the yield of biologically active compounds though they differed somewhat in their prediction of compounds likely to be active. This study provided an estimate of the enrichment in the yield of active compounds one can expect by use of computer preselection.

STUDY I

Previous work had left a gap in testing the method on the NCI data. Two experiments had been used to evaluate different sets of features and other variations of the method. Neither experiment was completely satisfactory.

One experiment used a training set of about 15 000 compounds created by restricting NSC numbers[3] to below 260 000. The test sets were taken from later ranges of NSC numbers. This design was an attempt to simulate operating conditions. The main drawback was a relative lack of diversity in the compounds of the test set.

A second experiment used a greater variety of compounds by running the 15 000 compound training set itself through the model. In this way it was possible to test for relative performance, but the magnitude of the performance might be unrealistically high as will be presently observed.

Upon collection of the first complete training set to create and install the operational model, i.e., the 35 000 compounds previously mentioned, there was another reason for running the entire training set through the model. This was to establish a range of activity scores against which to evaluate potential acquisitions. The results were examined for performance on the active compounds and seemed too good to be true, with 45% of the active compounds falling in the top 10% of the ranking.

There was a question as to how much of the excellent performance was due to the sheer size of the feature set which consisted of more than 5000 features. Moreover, there were a large number of features which occurred only once. That occurrence was necessarily in an active compound since a criterion for acceptance as a feature was at least one occurrence in an active compound. The method gives these features quite high weight, which incidentally helps pick up new classes. However, when the training set was run as a test set these features practically guaranteed high activity scores for the compounds in which they occurred.[4]

Therefore it became especially important to perform a less biased test by separating a portion of the compounds into a test set. Thus, any feature that occurred just once would have no effect at all. If it occurred in the reduced training set, it would produce a weight which would not be used. If it occurred only in the test set it would be passed with an effective zero weight.

The first complete training set[5] consisted of about 120 highly active compounds (A's), 2000 moderately active compounds (C's), and 33000 inactive compounds (N's). A test set was formed by taking every fifth compound from each of the three sets beginning with the first compound in each set. The remaining 80% formed the reduced training set. Weights were computed from this training set, and the test set was run as if it were a set of compounds of unknown activity.

The test set was ranked by activity score and separated into tenths according to the ranking. Thus, the incidence of actives in each tenth can be compared to a random selection where each tenth would be expected to contain one tenth of the actives.

A second 80/20 cut was formed by beginning with the second compound in each component of the original training set. Five disjoint test sets were run in this manner, and the

Table I. Ranking of Actives in 20% Test Sets of Study I[a]

A compd						C compd				
cut 1	cut 2	cut 3	cut 4	cut 5	tenth	cut 1	cut 2	cut 3	cut 4	cut 5
18	15	19	22	20	10	130	113	125	121	135
2	1	0	2	2	9	64	52	53	52	44
2	2	1	1	0	8	41	46	45	48	38
0	3	0	0	0	7	30	27	30	45	31
0	0	1	0	0	6	29	32	24	26	30
2	1	2	0	0	5	28	34	36	25	25
0	0	0	0	1	4	27	29	32	30	31
0	1	0	0	1	3	19	25	22	16	30
0	0	0	0	0	2	15	19	15	14	16
0	0	0	0	0	1	12	15	17	16	14
24	23	23	25	24	total	395	393	399	393	394

[a] The A compounds are highly active and the C compounds moderately active. The cuts are disjoint so that each compound occurs once and only once.

Table II. Totals over Cuts in Table I

tenth	A compd	C compd	total actives
10	94	624	718
9	7	265	272
8	6	218	224
7	3	163	166
6	1	141	142
5	5	148	152
4	1	149	150
3	2	112	114
2	0	79	79
1	0	74	74
total	118	1973	2091

results on the active compounds are shown in Table I. The cuts are not more exactly equal in number of compounds because several compounds were disqualified for undefined structure[6] after the cuts were formed.

A look at the totals over the cuts in Table II shows that there were almost 10 times as many actives in the upper 10% of the ranking as there were in the lowest 10%. A test for statistical significance is hardly necessary. Nevertheless, a simple chi-squared test with 10 equal values expected gives a P value less than 0.00001.

As expected, the results were not as good as when the training and test sets were identical (see Table III). But there are still 34% of the actives, as opposed to 45%, in the highest 10% of the activity scores.

One may still consider the data biased since all these compounds have been selected, in one way or another, for antitumor testing. One effect of this kind of bias was observed when running the complete training set as a test set. Due to a quirk in the programming system used, namely, the limitation of the size of vectors to 32 768, the inactive compounds were run in two sections: NSC 260 000 was chosen as the dividing line.

Table III shows the results of running the training set as a test set. The ranking is broken into tenths with the active compounds, A and C, left intact and the inactive compounds split using N for those up to NSC 260 000 and N′ for those above NSC 260 000. The table illustrates that, independently of the computer model, selection was performed so that practically no compounds with NSC number showing acquisition later than NSC 260 000 fell into the lowest 10% overall and comparatively few in the lowest 20% of activity score. This shows some mirroring of the computer method by human selection.

Table III. Training Set Run as Test Set with Inactives Split at NSC 260000 into N and N′

tenth	A compd	C compd	N compd	N′ compd
10	101	827	1109	1458
9	6	141	818	2529
8	0	79	578	2837
7	2	80	503	2910
6	4	76	537	2878
5	1	91	755	2647
4	2	95	1057	2341
3	1	166	1737	1590
2	1	245	3037	211
1	0	142	3351	1
total	118	1942[a]	13482	19402

[a] The total number of C compound is lower than that in Table II because some compounds had their structures defined after running the complete training set but prior to generating the reduced models for the 20% test sets.

Ideally, one would wish to have data on the body of compounds likely to be obtained by NCI but not necessarily selected for testing. Unfortunately, it would be difficult to get biological data on a sufficient number of such compounds to form a useful training set. However, it is possible to test the performance of the current model on a collection of such unselected material for a more realistic study.

STUDY II

This study was designed to fulfill several objectives. The main objective was an appraisal of performance of the method in practice. That is, compounds were to be run through the model and then sent to biological testing. In this study all the compounds were to be tested so that biological activity could be correlated with prediction by the computer.

A second objective was to compare the performance of the computer with that of a medicinal chemist familiar with the file. For this purpose the chemist was asked to designate those compounds which were likely to be active. Not only can the relative performance of the chemist and computer be evaluated upon biological testing but also their agreement, first with respect to compounds predicted as likely to be active, and second, after testing, their agreement on the biologically active compounds.

Another objective was to familiarize the acquisitions contractor with the limitations as well as the power of the method. In practice the chemist would have the computer ratings available with the option of using them to make the selection. It was also of interest to see if the computer ratings would exert a major influence on what the chemist considered active.

Finally, one can estimate the degree of enrichment in the yield of active compounds that can be achieved by using the computer as an aid in selection.

An assortment of 988 compounds was provided by the acquisitions contractor for this study. These compounds varied over the required broad range and were unselected except for having been skimmed over for interesting compounds. The 988 compounds were reviewed by the chemist, who classified 298 as possibly active and another 14 as novel; the remaining 676 were thought to be inactive.

The 988 compounds were then run through the computer which produced an activity score for each compound. Also, a feature from the compound was selected, one which occurred least often in P388 testing. The incidence of this feature was given as an inverse measure of novelty, this value being zero if the feature had not occurred at all in P388 testing.

These runs took place at the time that the augmented atom (AA) version of the features was being replaced by the ganglia

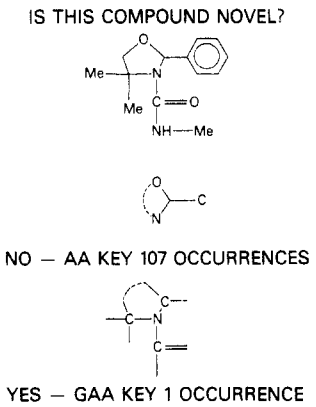

Figure 1. Repeated from ref 2. Typical structure features are the least occurring (rarest) AA and gAA keys for a representative compound. In this example the rarest AA key stems from a different central atom than does the rarest gAA key.

Table IV. Agreement on Selection between Chemist and Computer over 988 Compounds Subsequently Tested in P388

	computer		
chemist	active[b]	other	total
active[a]	120	178	298
other	178	512	690
total	298	690	988

[a] Compounds chosen by the chemist as more likely to be active. [b] Equal number (298) with highest activity score by computer.

augmented atom (gAA) version;[7] so both versions were applied to the 988 compounds. Figure 1 is taken from ref 2 to illustrate the AA and gAA keys or features. An AA key consists of a chemical atom, its bonds, and its neighboring atoms. A gAA key includes further bonds on the neighbors. Exhaustive generation of all keys for each compound accounts for the large number of features employed.

The computer results again showed the difference between the previously screened compounds of the training set and the 988 unselected compounds. In the AA model only 33 of the 988 had activity scores in the highest 10% of the range determined by the training set, while 467 of the 988 fell in the lowest 10%. The gAA model distributed the compounds similarly, with 23 of the 988 in the highest 10% and 308 in the lowest 10%.

It was now possible to compare the ratings performed by the chemist and the computer. The agreement was measured on the 298 compounds designated by the chemist as active. Since there was no corresponding sharp cutoff in the computer ratings, it seemed equitable to measure concurrence with the 298 compounds rated highest by the computer.

Of these compounds, 120 were among the 298 designated active by the chemist. This left 178 compounds chosen by the chemist that ranked from 299 to 988 by the computer and an additional 178 compounds with rank in the top 298 by the computer that were designated inactive by the chemist. The remaining 512 compounds were designated inactive by the chemist and ranked from 299 to 988 by the computer (see Table IV).

The amount of agreement can be expressed by the kappa measure of Fleiss[8] which factors out any agreement due to chance. In this example the observed agreement, p_0 is $(120 + 512)/988 = 0.640$, while the chance agreement, p_c is $(298^2 + 690^2)/988^2 = 0.579$. The measure of agreement, κ, is $(p_0 - p_c)/(1 - p_c)$ which equals 0.145 with a standard deviation[8] of 0.032.

Thus, the agreement is significantly away from 0, which would indicate complete independence, but is a great deal further from 1, which means complete agreement. The kappa measure will be used again later to quantify agreement on the compounds that proved to be biologically active.

Table V. Prediction Performance on P388 Active Test Results from 988 Compounds

	chemist[a]	comp 312[b]	agree[c]	comp 494[d]	all 988
confirmed active	11	13	8	20	26
presumptives	8	10	4	18	33
pass/fails	5	4	2	6	10
toxic	19	14	12	18	27
$A + 0.6P + 0.2(P/F) + 0.1T$	18.7	21.2	12.0	33.8	50.5

[a] Actual P388 results on 312 compounds selected by chemist. [b] Results on highest 312 compounds by computer activity score. [c] Agreement, or overlap, between chemist and computer. [d] Results on highest 494 compounds (50%) by activity score.

Next, the computer ranking was examined by a chemist to provide some experience of how the computer evaluates structures. Also, it was of interest to see if the original designations would be changed upon viewing activity scores from the computer.

The original chemist who had rated the compounds was not then available, but the computer results were reviewed by the chemist who had replaced him, temporarily, as it turned out. At that time the ratings had just been obtained by using the AA model.

Recall there were 33 compounds in the highest 10% of the scores according to the training set. Of these compounds, the chemist had designated 20 as active and 2 as novel, which is at least twice as many as would have appeared by chance. Moreover, all but two compounds were deemed possibly active or worthy of testing by the new review.

Thus, it was safe to assume the computer results would influence selection. It remained for the biological results to determine whether this influence is useful.

The 988 compounds were submitted to the P388 screen during the Spring of 1979. The results are reported as of April 1981. This should include all results where there was sufficient material for testing or additional material could be easily obtained. There remained 70 compounds requiring further testing as will be discussed shortly. The expected yield of confirmed actives from these compounds with incomplete testing can be estimated.

The biological criterion for activity in P388 as a prescreen is a T/C[9] of at least 120%, confirmed in a second test. Compounds which have passed one test at a T/C of 120% or greater are placed in a repeat status as presumed active and must fail two tests to be disqualified as inactive. Toxic tests are generally repeated at lower doses.

As of April 1981 the 988 compounds have yielded 26 confirmed actives. The 70 compounds requiring further testing consisted of 33 presumed actives, or presumptives, 10 presumed actives which have failed one test, or pass/fails, and 27 toxic compounds; see the last column of Table V. We can estimate the number of ultimately confirmed actives by multiplying the number of presumptives by 0.6, the number of pass/fails by 0.2, and the number of toxic compounds by 0.1, according to current estimates of the ultimate yield of confirmed actives from those categories of compounds. Thus, we estimate 50.5 expected confirmed actives from the 988 compounds for a yield of about 5.1%.

The 312 compounds designated by the chemist contained 11 of the 26 confirmed actives, 8 presumptives, 5 pass/fails, and 19 toxic compounds, which would add up to an estimated 18.7 expected confirmed actives; see the first column of Table V. This will be compared to the number of actives expected by choosing 312 compounds at random. That is, by chance

Table VI. Agreement between Chemist and Computer on the 53.5 Estimated Eventual P388 Actives Out of 988 Compounds

	computer		
chemist	active[a]	other	total
active[a]	12.0	9.2	21.2
other	6.7	22.6	29.3
total	18.7	31.8	50.5

[a] These results are the estimated eventual actives which were successfully predicted as shown in the bottom row of Table V.

there would be (312/988)50.5 or 15.9 expected ultimately confirmed actives.

The 312 compounds ranked highest according to the computer contained 13 confirmed actives, 10 presumptives, 4 pass/fails, and 14 toxic compounds. This would ultimately amount to 21.2 expected confirmed active compounds; see Table V, column 2.

First consider the 26 actual confirmed actives. Choosing a random 312 of 988 compounds gives an expected (312/988)26 or 8.21 confirmed actives with a standard deviation of 2.34. Therefore, normalizing the chemist's result of 11 gives (11 - 8.21)/2.34 or 1.19 which has a one-tailed P value of 0.12. The computer's achievement of two additional confirmed actives gives (13 - 8.21)/2.34 or 2.05 for a somewhat significant one-tailed P value of 0.021.

The significance of these results are reduced a bit by using the estimate for ultimate expected confirmed actives. The reason is both the chemist and the computer did better on the confirmed actives, than they did on the presumptives. Computing the variance of $A + 0.6\,P + 0.2(P/F) + 0.1T$ under the hypergeometric distribution with a selection of 312 out of 988 and $A = 26$, $P = 33$, $P/F = 10$, $T = 27$, we get a standard deviation of 2.79. Thus, the chemist result on the estimate is (18.7 - 15.9)/2.79 or 1.0 for a P value of 0.16 while the computer result of (21.2 - 15.9)/2.79 or 1.9 is statistically significant with a one-tailed P value of 0.03.

The computer ranking becomes somewhat more significant if one considers the actives included in the top 50%, i.e., the 494 highest ranking compounds; see the fourth column of Table V. Then the confirmed actives show (20 - 13)/2.52 or 2.78 for a P value of 0.003 and the estimated confirmed actives yield a normalized value of (33.8 - 25.25)/3.0 or 2.85 for a P value of 0.0022. The expected enrichment in the yield of actives by choosing merely the top 50% of the compounds in this study is 33.8/25.25 or 1.34 with a standard deviation of 0.12. This enrichment, also, is 2.85 standard deviations away from 1 (random selection or no enrichment) for a P value of 0.0022.

It is also possible to compare performance on the active compounds between the chemist and the computer. First, the agreement on the actives will be quantified using the same kappa measure that was used earlier to measure the agreement on choices. The concurrence in the top 312 was 8 confirmed actives, 4 presumptives, 2 of the pass/fail and 12 toxic compounds. Table VI separates the 50.5 estimated expected actives as was done earlier with the 988 compounds. The same kappa measure as before yields $\kappa = 0.34$ with a standard deviation of 0.14. So although the agreement on actives is more than twice the earlier value of 0.145 for agreement on choices, it is not quite statistically significantly better on the actives than on the choices. It is still significantly away from zero, or complete independence.

Second, it is possible to compare performance of the chemist and the computer on selecting actives. To use tests designed for paired data, it is necessary to impose a ranking on the chemist's choices. Lacking any further discrimination, we can consider the 312 compounds chosen by the chemist as tied in rank at 156.5 and the remaining 676 compounds at a tied rank of 650.5. This is not being completely fair to the chemist, who could have supplied finer gradations of activity predictions, but this will serve as an approximation.

Under this convention, performance on the 26 confirmed actives can be compared by the sign test and the more discriminating Wilcoxon signed rank test.[10] The computer ranked 17 of the 26 compounds higher than the chemist, and the chemist did better on the remaining nine. The sign test has a P value of 0.084 which is not significant. When the magnitude of the differences are taken into account by use of the Wilcoxon test, the P value becomes more significant at 0.032. However, in comparing the two methods, i.e., the chemist and the computer, it is more appropriate to use the two-tailed P value of 0.064. Thus, if this test were done at the 0.05 level, it would not show a significant difference between the computer and the chemist.

That is, although the computer predicted 13 confirmed actives to the 11 predicted by the chemist, this is definitely not significantly different. If we take into account the relative rankings in the manner just described on all 26 confirmed actives, then the computer almost has a significant edge over the chemist. For example, were we to set up a hypothesis that the computer performed worse than the chemist by one standard deviation, then our data is sufficient to reject such a hypothesis.

COMPUTER-ASSISTED SELECTION

To use the computer results for selection one would of course be interested in the high scoring compounds and those with unique features. However, the number of potential acquisitions and the relatively large number of compounds to be selected makes it more reasonable to eliminate low scoring compounds as well as compounds which seem to be already adequately tested as described soon.

Suppose we were to eliminate the 308 compounds in the lowest 10% according to the training set range. Among these compounds were 3 confirmed actives, 7 presumptives, 4 pass/fails, and 5 toxic compounds for a total of 8.5 expected confirmed actives. This would still keep 42.0 of the actives in the remaining 680 compounds, instead of the (680/988)50.5 or 34.8 one would get by random choice. When the difference of 7.2 is normalized by the standard deviation of 2.8, the significance is 2.57 standard deviations for a P value of 0.005. The analysis here does not take into account that the presumptives, etc., which score low are less likely to confirm, on the basis of the more striking separation of actual confirmed actives.

Now we examine the other important category of compounds which are candidates for elimination. These are compounds which are too similar to those which have already been tested, regardless of activity. A simple approximation is obtained by requiring the least occurring feature, and hence all features, to occur at least 30 times in P388 testing. The threshold of 30 was convenient at the time the 988 compounds were run.

It was another pleasant surprise that the compounds which would be eliminated because of adequate representation had a remarkably low number of actives. Among the 305 compounds which exceeded the threshold at 30 or more occurrences there were just 2 confirmed actives, 7 presumptives, 4 pass/fails, and 8 toxic compounds for a total of 7.8 expected confirmed actives; see Table VII, column 2. If these compounds were removed there would remain 42.7 actives in the 683 compounds, as opposed to 34.9 by random choice. The normalized difference is 7.8/2.8 or 2.79 standard deviations for a P value of 0.003.

It is also of note that the overlap between the low activity 308 compounds and the high occurrence 305 compounds was

Table VII. P388 Results on Certain of the 988 Compounds Which Were Candidates for Elimination by Computer Selection

	308 low active[a]	305 high occur[b]	agree[c]	combined 488[d]	all 988
confirmed active	3	2	1	4	26
presumptives	7	7	2	12	33
pass/fails	4	4	2	6	10
toxic	5	8	1	12	27
$A + 0.6P + 0.2(P/F) + 0.1T$	8.5	7.8	2.7	13.6	50.5

[a] These 30% compounds scored in the lowest 10% according to the range of the P388 training set. [b] These 305 compounds met the criterion for adequate representation in P388 testing. [c] Agreement or overlap between the compounds in the first two columns. [d] All the compounds that satisfied either or both of the first two columns.

Table VIII. Agreement between the 308 Low Activity Prediction and the 305 High Occurrence Compounds As Defined in Table VII

	occurrence		
activity	high	other	total
low	125	183	308
other	180	500	680
total	305	683	988

just 125 compounds; see Table VIII. This agreement quantifies to $\kappa = 0.142$, about the same as the 0.145 agreement between the computer and the chemist in their choices. Such an unforeseen, albeit low agreement, corresponds to the unexpectedly small number of actives among the 305 high occurrence compounds.

Thus, it is expected that compounds will be eliminated if they yield low activity score and/or high occurrence in the training sets. In the current study, this would exclude 488 compounds or almost half of the 988 compounds. The 488 compounds contain only 4 of the 26 confirmed actives, 12 presumptives, 6 pass/fails, and 12 toxic compounds for an estimated 13.6 expected confirmed actives. The remaining 500 compounds would have 36.9 expected confirmed actives for an enrichment of 36.9/25.6 or 1.44 with a standard deviation of 0.109. So there are 4.0 standard deviations for statistical significance with a P value below 0.0001. This compares with the enrichment of 1.34 and P value of 0.0022 achieved by selection of the top half by activity alone.

Again, the enrichment should be even better since the presumptives, etc., of the eliminated compounds have been given the same weight as their counterparts among the selected compounds. But the experience of actual confirmed actives indicates that the ultimate confirmed actives will continue to bias in favor of the selected compounds.

Early experience on the use of an operational model installed April 1979 to rate potential acquisitions continues to reinforce the trends found in the study just reported. When more biological data is accumulated this operational experience will be reported. Beginning March, 1980, the computer began an active participation in selection of compounds for screening in P388 from all potential acquisitions.

ACKNOWLEDGMENT

Study II owes its inception to Sidney Richman. The essential chemistry was performed by Kenneth Paull who was always encouraging. The successful conclusion was wrought by statistical help from Robert Tarone. Alfred Feldman and Ruth Geran read the manuscript and provided helpful criticism and advice.

REFERENCES AND NOTES

(1) Hodes, L.; Hazard, G. F.; Geran, R. I.; Richman, S. "A Statistical-Heuristic Method for Automated Selection of Drugs for Screening". *J. Med. Chem.* **1977**, *20*, 469-475.
(2) Hodes, L. "Computer-Aided Selection of Novel Antitumor Drugs for Animal Screening". *ACS Symp. Ser.* **1979**, *No. 112*, 583-602.
(3) NSC numbers are assigned in chronological order to compounds sent for biological testing. NSC 260 000 corresponds to the end of 1975.
(4) Similar difficulties with singly occurring features were reported in a previous work (Cramer, R. D., et al. *J. Med. Chem.* **1974**, *17*, 533-535) which was a precursor of the current work.
(5) Reference 2, p 598-600.
(6) Many compounds (2 or 3%) of the NCI file are sent to biological testing with partially or wholly undefined structure.
(7) The AA keys are described in ref 1 and the gAA keys are described in ref 2, p 597-598. All the computer results in Tables I-VIII are derived from the gAA version.
(8) Fleiss, J. L., "Statistical Methods for Rates and Proportions", Wiley: New York, 1973.
(9) The T/C ratio, in P388, is the median life span of the treated animals divided by that of the controls.
(10) Lehmann, E. L., "Nonparametrics: Statistical Methods Based on Ranks"; Holden Day: San Francisco, 1975.

Selection of Molecular Fragment Features for Structure-Activity Studies in Antitumor Screening

LOUIS HODES
National Cancer Institute, Silver Spring, Maryland 20910

Received May 4, 1981

The National Cancer Institute Developmental Therapeutics Program screens about 13 000 compounds per year for antitumor activity in a mouse prescreen (P388). A method for predicting activity in P388 uses molecular fragment features of potential acquisitions. This paper covers some details about how the set of features was chosen, filling a gap in earlier publications.

INTRODUCTION

Molecular structure fragments offer a rich body of features for the prediction of biological activity. One such predictive method[1-3] is being used to aid in the selection of compounds for antitumor screening in the National Cancer Institute Developmental Therapeutics Program (NCIDTP). An in vivo prescreen, mouse lymphocytic leukemia or P388, is used for antitumor testing of large numbers of compounds, about 13 000 per year.

Two or three times that number of compounds are collected in the form of structure diagrams as potential acquisitions. These are searched for duplicates in our chemical information system.[4] At the same time the structures are run through the following programs which help in the acquisitions process.