The NCI Drug Information System. 5. DIS Biology Module

M. T. ZEHNACKER, R. H. BRENNAN, and G. W. A. MILNE*

Information Technology Branch, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20205

J. A. MILLER

Fein-Marquart Associates, Baltimore, Maryland 21212

Received April 21, 1986

The NCI drug screening program tests over 10000 chemicals per year for activity against cancer. The associated Drug Information System (DIS) captures all the raw testing data and provides for its validation. The large quantity of numeric data gathered during testing is maintained within the DIS in a database that is interactively searchable and automatically updated at regular intervals.

INTRODUCTION

In the earlier papers in this series, we have described the components of the National Cancer Institute's Drug Information System (DIS) that deal with selection, acquisition, inventorying, and shipping of chemicals that are examined by the NCI screening program. The testing of these compounds against cancer, the raison d'être of the program, is the subject of this paper.

Following acquisition, a sample of each compound is shipped at NCI request to one of the five laboratories currently under contract[1] to NCI to carry out the biological testing. Each shipment, as has been seen, is reflected by changes in the shipping and inventory records for the compound(s) shipped. All compounds tested in the NCI program are identified by an NSC Number[2] and every sample of a compound by an NSC/Sample Number. Each screening laboratory is sent samples of compounds, along with a shipping list that contains the NSC/Sample Number of the chemical, the amount shipped, information concerning storage and preparation, and specifications as to the testing to be performed along with any special instructions. The screening laboratories measure the quantity of material, and receipt of the shipment is acknowledged through the DIS. This updates the Shipping History database to reflect the date of receipt of the shipment and the amount received.

With the material in hand, the screening laboratory completes the testing requested and enters the results into a terminal that writes the data to local storage, such as a cassette tape. At intervals of 1 or 2 days the data are uploaded to the computers at NIH, and then they are passed through validation and error-checking routines to a staging file. The staging file is merged into the working DIS file upon a fortnightly basis. Once in the DIS, all the screening data can be searched and displayed in the standard DIS manner.

TESTING OF COMPOUNDS

In the early 1960s, when the screening program was in its formative years, it became clear that rigorous standardization of animals, tumors, and cell lines was necessary if results were to be reproducible from one laboratory to another. Consequently, in spite of the formidable logistics, test animals are now supplied to all screeners by NCI's own Mammalian Genetics and Animal Production Section. This NCI unit, in turn, uses accredited sources, whose management methods meet or exceed the standards of the Institute of Laboratory Animal Resources[3] for breeding, care, and management of laboratory animals. Likewise, tumor strains used in testing are maintained and supplied upon request by NCI, which no longer supports or accepts testing done with nonstandard tumor strains or animals.

The approach generally used for screening is to test a compound against increasingly refractory tumors until it fails to demonstrate a predefined level of activity.[4] Different strains of mice are shipped on demand[5] to the screening contractors[1] and are quarantined for 1 week. They are then moved into holding rooms where they stay for as long as 3 weeks, during which time the healthy animals, those that gain weight slowly, are identified. These individuals are moved to the screening area where they are tested according to priorities established by NCI. In all tests which require parenteral administration, the compound is prepared as a solution or a suspension in one of a few selected vehicles.[6] Saline (0.85% NaCl in water, without preservatives), is used for solutions, for example, and Tween-80 may be used to stabilize a suspension. Solutions or suspensions of synthetic compounds are injected intraperitoneally once daily unless a different route or frequency of administration is requested by NCI.

A. The Preliminary Screen. Initial in vivo testing of all compounds is carried out in mice inoculated with P388 lymphocytic leukemia and proceeds according to a standard NCI protocol.[7] The experiment employs six animals per test group; an inoculum of 1 million leukemia cells is implanted in each mouse on day zero. A test group is a group of six mice, each of which receives the same specific dosage of test material, and typically four dose levels and thus four test groups are used for each compound. A control group of inoculated but untreated mice is maintained along with the test groups. In the P388 leukemia screen, the drug is administered by means of daily intraperitoneal injections for 5 days. Animal weights are measured on days 1 and 5. For all test groups in which more than 65% of the mice survive through day 5, the ratio (T/C) of the mean survival time of the test animals (T) to that of the corresponding controls (C) is determined on the evaluation day for that experiment. In the preliminary screen, this is usually day 30. In this way a T/C value is developed for each acceptable test group, i.e., for each dose level. A T/C value of less than 0.85 (% T/C of 85%) suggests the presence of significant toxicity, and the test results from such groups are not used in the evaluation of the compound except to help establish lower dose levels for subsequent tests. An initial T/C of 127% or greater is considered indicative of activity, and compounds that demonstrate such behavior pass to the next stage, confirmatory testing.

B. Confirmatory Testing. Screening data are reviewed on a daily basis by NCI biologists, and compounds that demon-

0095-2338/86/1626-0186$01.50/0

strate activity in the preliminary P388 testing are scheduled for confirmatory testing. Prior to 1980, confirmatory testing was done by a different screening contractor, but this measure has been dropped in an attempt to economize and the additional testing is now carried out in the same laboratory. The confirmatory testing duplicates the preliminary testing and is carried out in exactly the same way. The results are passed back to the DIS in the same way.

C. Tumor Panel. After initial and confirmatory testings against P388 leukemia are completed, the results are reviewed by a NCI committee of chemists and biologists. A compound that has shown confirmed activity in the first round of screening may be scheduled by this committee for testing against a group of more refractory tumors. The compound will be tested against every member of this group, which is known as the "Tumor Panel". At the same time, a material classification code[8] is assigned to the compound on the basis of its reported activity, the perceived uniqueness of its structure, its availability, and its status in the NCI screening program. Some 15 000 compounds have material classification codes, and this parameter is an important management device. Accordingly, the DIS captures the material classification codes as they are assigned or adjusted at committee meetings and carries the data in a field in the Chemistry database.

The current NCI Tumor Panel is comprised of four primary and five secondary tumor types, which are considered to be representative of major human cancers. Ovarian and breast cancers are included, as are melanoma, leukemia, and a strain of adriamycin-resistant leukemia. A great deal more testing is done in each of these tumors than was done in the preliminary screen. The criteria for activity vary from one tumor to another, but the testing proceeds much as before, and the results are passed back to the DIS in just the same way. While 10 000 compounds per year are examined in the preliminary screen, only some 200–300 are tested in the Tumor Panel. The data from such testing is reviewed with great care by NCI, and about 20 compounds per year are recommended for further development toward use in human cancer.

Since the various screening contractors have different capabilities and are geographically dispersed, it is often necessary to transfer the supply of a compound from one screening laboratory to another. Appropriately authorized NCI staff can invoke the TRANSSHIP command in the DIS to accomplish this task. This command is essentially identical with the normal material shipping command except that in this case the sender is the first screener while the recipient is a second screener. Essentially, the TRANSSHIP command does not accomplish any shipment but simply instructs the first screener to send the sample to the second screener and also records the transshipment activity in the Shipping History database. A copy of the shipping list that results from this command is forwarded electronically to each of the screeners by means of messages which appear as they logon. A receipt acknowledgment must be provided by the second screener when the material arrives at that laboratory.

D. Cell Lines. The NCI is currently involved in an investigation of the feasibility of a disease-oriented approach to drug discovery. In this approach, each compound is tested in vitro vs. a large number of cancer cell lines. The role of the DIS in this new system is much as it was with the traditional in vivo testing. The acquisition and supply of compounds for testing is unchanged, and the biological data, while very different from the mouse-derived data, reach the DIS in just the same way.

DATA ACQUISITION

The flow of biology data into the Drug Information System is depicted schematically in Figure 1. The responsibility of

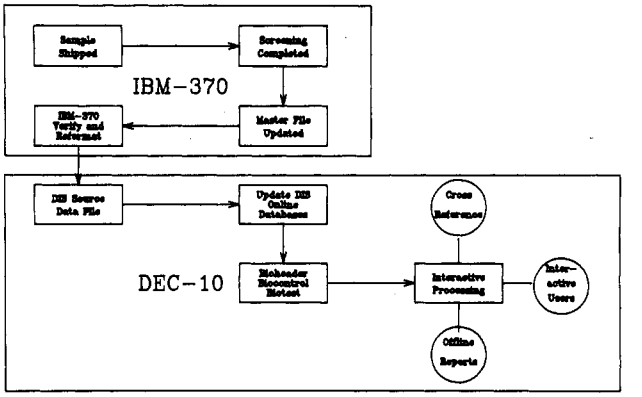


Figure 1. Flow of data in the DIS Biology module.

the screening laboratories is to receive the sample, carry out the screening, and enter the raw data into the data validation area on the NIH IBM mainframe. Programs on this computer take the new data and assemble a "DIS Source Data File". This is passed from the IBM computer to the DEC-10 computer where the DIS resides.[9] At the next update of the DIS Biology module, all new data in the Source Data File are merged into the searchable DIS. Each of these steps is discussed in more detail here.

A. Data Entry via a Passive Terminal. The screening laboratories collect and transmit data to a mainframe IBM computer at the NIH Computer Center in batch mode. The programs that support this were written by NCI and run on an on-site terminal consisting of a Hewlett-Packard 2645A terminal equipped with a keyboard, a dual magnetic tape cassette unit for storing data, and a thermal printer. Utilizing full screen edit programs, the screening laboratories can record the testing results during the course of the experiment, prepare the data in machine-readable form, and store the formatted data on magnetic tape cassettes for transmission to the NIH IBM 370 computer at nonpeak hours.[10]

B. Data Entry via a Microprocessor. The tape cassette based system described above is obsolete in terms of hardware and is currently being superseded by a data collection system which runs on an IBM PC(AT) microprocessor. The computation, edit, and entry facilities that are possible on such microcomputers enable a screener to transmit fully processed and edited experimental results without delay, capabilities which are not available with the 2645A Hewlett-Packard terminal. Incorporating these enhancements, as well as the networking of the NIH computers,[10] into the screening laboratories' procedures improves efficiency and also allows DTP staff to monitor an experiment at a screening laboratory on a continuous basis, obtaining at any time a status report on the progress of that experiment. This increases the turnaround time in scheduling the next experiment and leads to financial savings as well as higher quality and reliability of data transmission.

C. Transmission to and Editing at the NIH Mainframe. Data transmitted from a screening laboratory to the NIH enter an IBM 370 resident data system where they are checked and edited. This system reviews data and, if they are acceptable, transmits them into the Biological data processing system directly. Such acceptable data are also returned to the screening laboratories for their own records and to selected NCI staff in the form of preliminary reports for the screener's review. Unacceptable data are returned to the screening laboratory in the form of an error report, which is presented at the next logon. Corrections are handled by the screener through an online computer procedure, and the corrected data is reprocessed through the Transaction Input processing system and recycled in this manner until all errors have been validated.

D. Master File Update. On a fortnightly basis, all ac-

Developmental Therapeutics Program
Division of Cancer Treatment
National Cancer Institute, Bethesda, MD, 20205
Leukemia Screen (3PS31) Test Results

For Supplier 629P

NSC #	Dose Tested	T/C%	Scr	Exp. #	Date on Test	Status
379571	240.00 mg/kg	87	03	74873	12-Nov-84	Inactive
	120.00	92				at dose levels
	60.00	93				tested
379534	240.00 mg/kg	89	23	74873	12-Nov-84	Inactive
	120.00	102				at dose levels
	60.00	108				tested

Figure 2. DIS test status summary report.

ceptable data from screeners are merged into a staging file known as the Biological Update and Reporting System. The format for entries in this staging file consists of eight fixed-length records identified as 100, 200, 401, 600, 610, 630, 670, and 680 series. An experiment with a single test group of six mice leads to (a) a 100 or header record containing general descriptive and status information pertaining to the compound; (b) a 200 record containing evaluative information on the tumor systems against which a compound had been tested; (c) a 401 record containing the cross-reference record for compounds that are merged together, permitting the original information to be saved; (d) a 600 record containing the in vivo test record; (e) a 610 record containing death-pattern and weight-change information for the test group; (f) a 630 record containing comments for the test group; (g) a 670 record containing death-pattern and weight-change information for the control group; and (h) a 680 record containing comments concerning the control group. A set of records, 600 and, optionally, 610, 630, 670, and 680, completely describes one test. Multiple numbers of these record sets exist for each compound. All the data records from a single test group are placed under a "super-identifier" called the "Test Set Number", and this is an important index term in the Biology module of the DIS because it allows retrieval of data from a single test group rather than all data for an NSC/Sample Number.

E. Link between Master File and DIS. After a fortnightly update of the master file has been completed, a final job is to create a DIS update file. This contains every record that, during the prior 2 weeks, has been added to the master file, deleted from it, or modified. The test set number for every such record is carried forward into this update file along with a flag designating the appropriate type of record—addition, deletion, or modification. In the cases of addition or modification, the new data are also included in the update file. The DIS programs use the data in this update file to generate increments to the Bioheader, Biocontrol, and Biotest files in the searchable DIS.

If new or changed test data for any compound appear in this fortnightly master file, then a summarized status report is prepared for the compound and distributed to the supplier of the compound. More detailed reports such as that shown under Display of Data, Specialized Output, are provided to NCI staff and to suppliers who request them. The summarized status report includes all test data measured in the P388 Leukemia Tumor Test System (3PS31) during the 2-week period. A typical report of this type is shown in Figure 2.

The tape that was created from the master file update is processed by the DIS biology module which reformats the data for input to the standard TDRS database programs of the DIS described in earlier papers of this series. Further validity checking on the data is carried out at this point to ensure the integrity of the online database as the DIS update proceeds. The file resulting from this final conversion is the DIS source file, which reflects all additions, modifications, or deletions of data in search and display requests.

Once per year, the entire Biology database is reviewed in a batch process. The entire biology record for any compound that is (biologically) inactive and whose data have not been subject to any change for at least 18 months is archived. Such data become inaccessible to the DIS, and if they are requested during a DIS session, the system will inform the requestor that the data in question are archived. The DIS will also offer to initiate a batch job for the retrieval of the data if the user wishes.

DATA RETRIEVAL

A. Search and Display Capabilities. For management purposes, Biology data in the DIS are subdivided into three categories, in the same way the master file is broken down. The "Bioheader" database carries the 100 and 200 records for every compound. The "Biotest" database contains the 600, 610, and 630 records, and the "Biocontrol" database has the 670 and 680 records. Each of these databases is updated independently during the DIS file updates, but the appropriate entries in each are cross-referenced, and the fact that there are three databases is essentially transparent to the searcher.

The three Biology databases contain a total of 104 fields, 28 of which are searchable. The proportion of searchable fields in the Biology module is the lowest in the DIS and reflects the end-point nature of the Biology data. It is clearly important, however, to be able to display all the biology fields, and this is supported by the DIS. Many types of searches have little meaning, however. For example, there is little purpose in retrieving all compounds that were tested at a dose of 12.5 mg/kg because this would amount to virtually the entire database.[11] Searching on tumor system, measured activity, experiment number, or screener identity is necessary for management, and searches of these types are supported. As an example, it is quite simple to retrieve all compounds with a specific T/C value

OPTION? PTC/166
2610 hits were found and stored in file 1 associated with the BIOTEST database

or even a range of T/C values
OPTION? PTC/120 TO 130
81 145 hits were found and stored in file 2 associated with the BIOTEST database

but because no tumor system was specified, the retrievals will include all compounds giving the appropriate T/C in any system.

A more useful search would be for compounds giving a specific T/C in one tumor system, such as the Lewis Lung tumor (3LL39):

OPTION? TSY/3LL39 AND PTC/>150
865 hits were found and stored in file 3 associated with the BIOTEST database

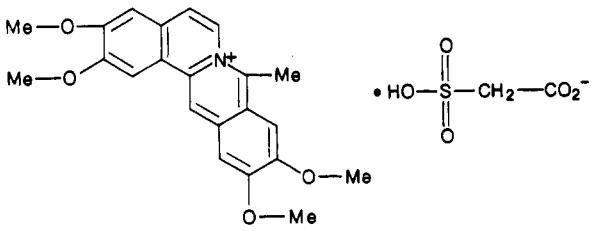
Search results are returned in the form of temporary results files that contain accession numbers of the records which satisfied the search criteria. Each search command produces a new temporary results files.

OPTION? TSY/3LL39 AND NSC/154890
17 hits were found and stored in file 4 associated with the BIOTEST database

Displays of results files can be produced upon request by using a format established by the user, as has been described in earlier papers in this series:

OPTION? FORMAT STRC CNAM TSY DAM PTC
OPTION? T 1
File: 1 Entries: 1 to 17

NSC Number 154890-R



(CNAM) Chemical Name: Berbinium,
5,6,7,8,13,13a-hexadehydro-2,3,10,11-tetramethoxy-8-methyl-,
salt with sulfoacetic acid;
Coralyne sulfoacetate;
Coralyne, sulfoacetate;
CORALYNE SULFOACETATE;
Dibenzo[*a*,*g*]quinolizinium, 2,3,10,11-tetramethoxy-8-methyl-,
salt with sulfoacetic acid (1:1) (9CI);
monoanhydride with sulfuric acid

Biology Test Record – NSC/Sample Number 154890-R/4
(TSY) Test System: 3LL39
(DAM) Dose per Injection: 00100.00
(PTC) Percent *T*/*C*: 122

Biology Test Record – NSC/Sample Number 154890-R/4
(TSY) Test System: 3LL39
(DAM) Dose per Injection: 00050.00
(PTC) Percent *T*/*C*: 113

Biology Test Record – NSC/Sample Number 154890-R/6
(TSY) Test System: 3LL39
(DAM) Dose per Injection: 00012.50

Biology Test Record – NSC/Sample Number 154890-R/6
(TSY) Test System: 3LL39
(DAM) Dose per Injection: 00006.25

B. Numeric Fields in the DIS Biology Module. Many of the fields throughout the DIS are numeric. Some of these merely carry alphanumeric or pure numeric identifiers, but many of them carry pure numeric data. This is particularly true in the Biology module, where some very important fields such as dose amount, treatment schedule, and measured activity, to name three, are numeric. In all-numeric fields, ranged search expressions are honored by the DIS. Thus one may retrieve all records in which compounds tested against the melanoma 3B132 showed an activity (% *T*/*C*) of 180

OPTION? TSY/3B132 AND PTC/180
12 hits were found and stored in file 5 associated with the BIOTEST database

or *T*/*C* values lying between 180 and 190

OPTION? TSY/3B132 AND PTC/180 TO 190
86 hits were found and stored in file 6 associated with the BIOTEST database

or simply *T*/*C* values greater than 180

OPTION? TSY/3B132 AND PTC/>180
266 hits were found and stored in file 7 associated with the BIOTEST database

All the test systems have activity criteria, or % *T*/*C* values, that, if equaled or surpassed, trigger various program responses, such as termination or expansion of testing of the compound. This means of searching allows a rapid count of the numbers of compounds that have met activity criteria against a specific tumor system and thus provides an immediate measure of the financial commitment that faces the program. As an example, the search

OPTION? TSY/3C872 AND PTC/<10
679 hits were found and stored in file 8 associated with the BIOTEST database

readily retrieves the 679 test records in which significant activity has been observed against the colon carcinoma 3C872. This is an inhibition test, in which activity is signaled by low % *T*/*C* values.

C. Sorting of Search Results Files. In any search in the Biology module, the results file will be organized by test set number. Such an ordering, when used for every results file, is useful because it permits rapid intersections and merging of different results files. From the point of view of the user, however, organization by test set number has no special value. It is possible, upon command, to re-sort files by some other value. The newly sorted file can no longer be used in interfile manipulations, but it is often very useful for the generation of intelligible reports.

In the following example, the 1662 test records of compounds that at some dose level gave a % *T*/*C* in excess of 150 (the criterion for further testing) against the M5 sarcoma, 3M531, can be retrieved.

OPTION? TSY/3M531 AND PTC/>150
1662 hits were found and stored in file 9 associated with the BIOTEST database

These records can now be arranged in increasing order of % *T*/*C*, so that the less active appear first. This is done with the SORT command:

OPTION? SORT 9/PTC
1662 hits were found and stored in file 10 associated with the BIOTEST database

*** The sorted result is not in ascending database ***
*** identifier order. The file cannot be used in ***
*** subsequent search operations.***

The sorted file can now be displayed, and from the partial listing below, it can be seen that this ordering is helpful to a reading of the data.

OPTION? FORMAT DAM PTC
OPTION? T 10

Biology Test Record – NSC/Sample Number 363813-I/1
(DAM) Dose per Injection: 00001.00
(PTC) Percent *T*/*C*: 150

Biology Test Record – NSC/Sample Number 740-K/65
(DAM) Dose per Injection: 00005.00
(PTC) Percent *T*/*C*: 151

Biology Test Record – NSC/Sample Number 287513-W/1
(DAM) Dose per Injection: 00020.00
(PTC) Percent *T*/*C*: 155

Biology Test Record – NSC/Sample Number 95441-W/20
(DAM) Dose per Injection: 00003.00
(PTC) Percent *T*/*C*: 181

Biology Test Record – NSC/Sample Number 119875-I/32
(DAM) Dose per Injection: 00002.00
(PTC) Percent *T*/*C*: 219

Biology Test Record – NSC/Sample Number 95441-W/20
(DAM) Dose per Injection: 00006.00
(PTC) Percent *T*/*C*: 310

Biology Test Record – NSC/Sample Number 95441-W/9
(DAM) Dose per Injection: 00024.00

(PTC) Percent T/C: 315

D. Combination of Biology Data with Non-Biology Data. Irrespective of its source, any DIS results file can be combined in a Boolean sense with any other DIS results file. This is a very powerful capability because it allows the user to answer quite complex questions that require reference to diverse types of data.

As an example, the passage of a chemical through the entire DTP program, from submission, through pre-selection and acquisition, to testing, is tracked by the DIS. A display of the various dates in the record for the compound provides a simple check as to the functioning of the entire system. The entire NCI experience with a compound such as NSC 600278 can be displayed with a pair of commands.

OPTION? FORMAT DTEN DACK DNSC DSUP DACQ DRCD DLWG AMT DLSH DSHP DRCV SHIP DTN
OPTION? TYPE (NSC/600278)
NSC Number 600278-K
(DTEN) Initial Entry Date (IDE): 11-FEB-85
(DACK) Acquisition Verification Date: 02-APR-85
(DNSC) Date of NSC Assignment: 02-APR-85

Inventory Record - NSC/Sample Number 600278-K/1
(DSUP) Date of the Supplier's Letter: 25-JAN-85
(DACQ) Date Acquired by Acq Contractor: 08-FEB-85
(DRCD) Date Received by Storage Contractor: 03-APR-85
(DLWG) Date of the Last Weighing: 09-APR-85
(AMT) Amount in Inventory: 0.7 GM
(DLSH) Date of Last Shipment: 15-APR-85

Shipping Record - NSC/Sample Number 600278-K/1
(DSHP) Date the Shipment Originated: 25-JAN-85
(DRCV) Date Shipment Received: 08-FEB-85
(SHIP) Shipment Type: Receipt by the Acquisitions contractor

Shipping Record - NSC/Sample Number 600278-K/1
(DSHP) Date the Shipment Originated: 03-APR-85
(DRCV) Date Shipment Received: 03-APR-85
(SHIP) Shipment Type: Receipt by the Storage contractor

Shipping Record - NSC/Sample Number 600278-K/1
(DSHP) Date the Shipment Originated: 15-APR-85
(DRCV) Date Shipment Received: 18-APR-85
(SHIP) Shipment Type: Storage contractor to a screener

Biology Test Record - NSC/Sample Number 600278-K/0
(TSY) Test System: 3PS31
(DAM) Dose per Injection: 00240.00
(DTN) Date on Test: 08-MAY-85

Biology Test Record - NSC/Sample Number 600278-K/0
(TSY) Test System: 3PS31
(DAM) Dose per Injection: 00120.00
(DTN) Date on Test: 08-MAY-85

Biology Test Record - NSC/Sample Number 600278-K/0
(TSY) Test System: 3PS31
(DAM) Dose per Injection: 00060.00
(DTN) Date on Test: 08-MAY-85

From this printout it can be seen that the supplier sent the compound, along with a covering letter, on Jan 25, 1985 (DSHP, DSUP, respectively). On Feb 8 (DACQ) the sample reached the Acquisitions contractor who created a Pre-Registry record on (DTEN) Feb 11. The structure was processed through the pre-selection process during March, and it was decided formally to acquire and test the compound on (DACK) April 2. An NSC Number was assigned to the compound on the same day (DNSC). On the following day

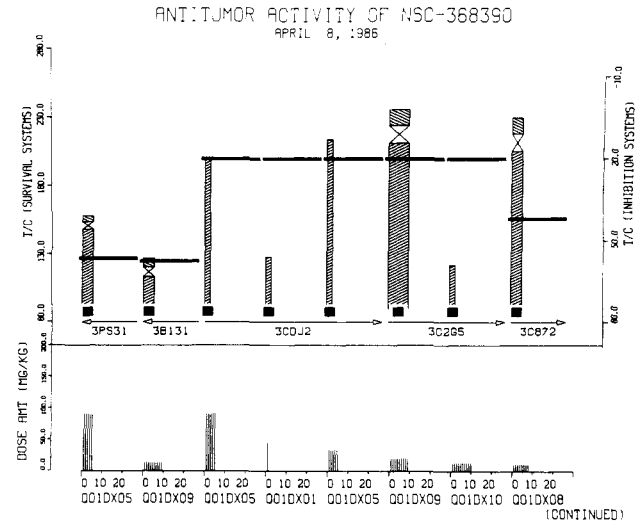


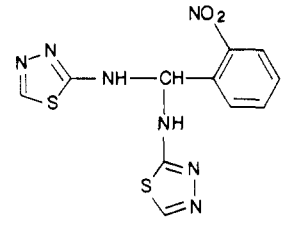
Figure 3. Bar-plot representation of Biology data.

(DRCD), the sample arrived at the repository, and on April 15 (DSHP), it was shipped to a screening laboratory, where it was received on (DRCV) April 18. Testing of the compound against 3PS31 leukemia at three dose levels began on (DTN) May 8, 1985.

E. Use of the DIS in Support of NCI Decision Making. Detailed searching in the Biology module is an important capability for the NCI biologists who manage the actual testing programs. For NCI managers working at other levels, however, the combination of screening information with data from the Chemistry or Inventory files is very important because the information that is derived in this way forms the basis of program decisions.

With a different example, it can be seen how the DIS permits NCI staff to retrieve and review background data very quickly in the course of making a planning decision. The bis(thiazole) derivative NSC 383500 (I) was confirmed as active against P388 leukemia in November 1985. It was therefore placed on the agenda of a review committee during that month.

NSC Number 383500-H



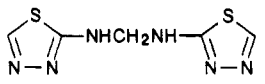
I

Biology Test Record - NSC/Sample Number 383500-H/O
(TSY) Test System: 3PS31
(DAM) Dose per Injection: 00200.00
(PTC) Percent T/C: 154

Biology Test Record - NSC/Sample Number 383500-H/O
(TSY) Test System: 3PS31
(DAM) Dose per Injection: 00100.00
(PTC) Percent T/C: 121
(MCC) Material Classification Code: 1D

A review of the DIS revealed that 33 related compounds, all containing the methylenebis(2-amino-1,3,4-thiadiazole) moiety of I, had been tested in the program. Of these, at least seven were equally active. Two of them (II and III) had completed antitumor testing and were being readied for possible clinical use. From such a perspective it was clear that this type of structure is quite commonly associated with activity

against P388 leukemia; there was no unusual discovery here, and since two related compounds (NSC 143019 and NSC 379892) were well advanced in the program, it was a simple decision to "table" NSC 383500. Accordingly, it was assigned a Material Classification code of "1D"—known to have activity, but not currently being studied—and with that the matter was settled.



II

NSC Number 143019-0

(MCC) Material Classification Code: 2B

(DMCC) Date of MC Code Assignment: 01-NOV-71

(SAC) Selected Agent Compound: YES

(NNZS) Number of Non-zero Sample Amounts: 3

(CAMT) Total Amount in Inventory: 4.05 GM

Biology Test Record - NSC/Sample Number 143019-0/2

(TSY) Test System: 3PS31

(DAM) Dose per Injection: 00100.00

(PTC) Percent *T/C*: 240

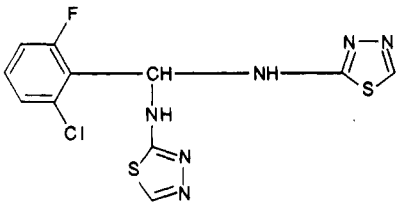
Biology Test Record - NSC/Sample Number 143019-0/2

(TSY) Test System: 3PS31

(DAM) Dose per Injection: 00100.00

(PTC) Percent *T/C*: 235

(MCC) Material Classification Code: 2B



III

NSC Number 379892-K

(CNAM) Chemical Name: Methanediamine, 1-(2-chloro-6-fluorophenyl)-*N*,*N*'-bis(1,3,4-thiadiazol-2-yl)-

(MCC) Material Classification Code: 2B

(DMCC) Date of MC Code Assignment: 07-AUG-85

(SAC) Selected Agent Compound: YES

(NNZS) Number of Non-Zero Sample Amounts: 1

(CAMT) Total Amount in Inventory: 70 mG

Biology Test Record - NSC/Sample Number 379892-K/1

(TSY) Test System: 3PS31

(DAM) Dose per Injection: 00240.00

(PTC) Percent *T/C*: 187

Biology Test Record - NSC/Sample Number 379892-K/1

(TSY) Test System: 3PS31

(DAM) Dose per Injection: 00120.00

(PTC) Percent *T/C*: 171

(MCC) Material Classification Code: 2B

DISPLAY OF DATA

A. DIS General Output Capability. The TYPE and PRINT commands provide for the display of information from the DIS databases. All or part of any DIS temporary results file can be TYPEd or PRINTed. The type of information that is so displayed may be controlled by the user by means of a FORMAT command. The file that is to be displayed may be created by any search command, such as

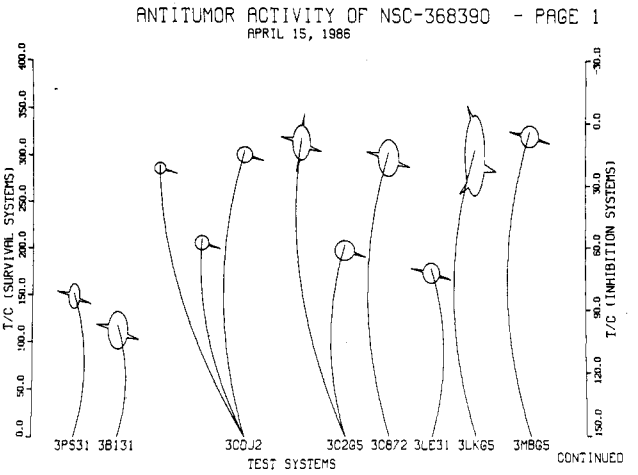


Figure 4. Flower-garden representation of Biology data.

OPTION? TSY/3LE31 AND (NSC/350000 TO 351000)

581 hits were found and stored in file 11 associated with the BIOTEST database

Then, once an appropriate FORMAT is defined

OPTION? FORMAT TSY DAM PTC CSC DTN

those fields from the 53rd and 54th entries in this file can be displayed at the user's terminal with the command

OPTION? TYPE 11//53-54

The TYPE command produces the display directly on the terminal, while the syntactically identical PRINT command directs the output to the system printer with the option of writing a disk file in the user's working area. The PRINT command is intended to be used in situations where the user does not wish to tie up the terminal with a lengthy output or where a higher quality print is needed. All requests for printed listings, whether or not they contain graphics, are made by means of the PRINT command and are automatically entered into the computer's print queue. No further action is needed to have the information printed.

B. Specialized Output. There is a great deal of data associated with every test that is completed in the NCI screening program. Many of the data are of interest only to the biologists who manage the screening, but they need to be able to see all the data for any experiment. This would require the DIS to list 45 fields for every dose level, and since such a voluminous output would be essentially unreadable, various specialized output formats have been devised.

The Screening Data Summary, which was developed a number of years ago by NCI staff, is a formatted report containing all the data associated with the testing completed for a specific compound. The report can be generated on command, online, and is prepared within the Biology module. It can also be generated in a batch procedure, and this is generally done routinely for compounds that are actually in test.

Practical considerations mandate massive abbreviation in reports of this sort, and to assist the user, every field has a descriptive "HELP" message online. These messages can be retrieved with the field name in full or abbreviated form. A Screening Data Summary for a specific compound can be

obtained by using the NSC Number of the compound.

```
OPTION?  TYPE (NSC/154890/SDS)



 NSC        MC  /DATE                                   QNS   /DATE       PAGE  1
------      ------------                                -------------     2-Jan-86
 154890-R    3T 30-JUL-76                DAY
                                       R 1ST TOT
 TSTSYS SCR  EXPT#   DATEON   TIS LVL T INJ INJ SAMP VEH FED TED SEX TSC SSC CSC
 ------ --- ------- --------- --- --- - --- --- ---- --- --- --- --- --- --- ---
 3LL39   08  00223  11-DEC-79  2   6  1   1   9   4   0T  60   5  F  22F      E
                                                      TUMOR            CONTROL
  TRTMT                                  W/  W/     -------------   -------------
  SCHED                                  D1  D2     LINE CODE GEN   EVAL HOST  BWC
 --------                                --  --     ---- ---- ---   ---- ---- ----
  Q01DX9                                  1   5      060   R  012   18.8  1A   0.8
        S   SURVS
  DOSE  0 --------
  mg/kg L  C NT TS   KE   TXSUR      DEATH PATTERN (DAY/DTHS)       EVAL T/C  BWD
 ------ - -- -- -- ------ ----- --------------------------------- ---- --- ----
  100.00 4  0  0  0       10/10 18/2 19/2 21/1 23/1 25/2 28/1 30/1 23.0 122 -0.5
   50.00 4  0  0  0       10/10 18/3 19/1 21/2 22/1 23/1 24/1 28/1 21.3 113 -0.4
   25.00 4  0  0  0       10/10 17/3 20/2 21/1 22/1 24/2 25/1      21.0 111  0.1
   12.50 4  0  0  0       10/10 16/1 17/1 18/1 19/2 20/1 22/3 24/1 20.0 106 -0.4
    6.25 4  0  0  0       10/10 11/1 17/1 18/2 20/3 21/1 24/1 26/1 20.0 106  0.1
  CNTRL                      30 13/2 16/2 17/4 18/5 19/8 20/4 21/3
                                22/2
```

The online SDS contains the user's search criteria, the chemical structures of the compound identified, the detail lines, and a summary of the detail information.

Searches which identify archived NSC's will cause a message to that effect to be displayed to the user, export a list of archived NSC's to the IBM computer, and submit a batch job for data retrieval and report generation. The user will be notified electronically when the report has been produced.

C. Graphics Output. The Screening Data Summary is a very rich report, and its major liability is that rather few people can read them well. Few of the suppliers of compounds to the screening program are familiar enough with the format to be able to use the summary report effectively, and so an effort has been made to take advantage of graphics in presenting the biological data.

A format involving bar plots has been developed, and an example is shown in Figure 3. This graph can be requested by an online user, and the hard copy is produced automatically, but offline, since a high-resolution plotter is used. The bar plot carries a bar for every test system for which data are available. Thus in the example in Figure 3, it can be seen at a glance that NSC 368390 has been tested in 3PS31, 3B131, 3CDJ2, 3C2G5, and so on. The height of the bar in each case designates the highest % T/C measured for the compound in that tumor system, and the center of the "X" in the bar denotes the median % T/C found, i.e., about 150 in 3PS31. Which of the two axes is appropriate is indicated by the horizontal arrow below the shaded bar. The vertical height of the X is proportional to the standard deviation associated with the mean % T/C (i.e., the "spread" of the data), and the width of the bar is proportional to the total number of tests recorded for that tumor system and that compound. For every bar there is, below the abscissa, a device denoting the dose level and dose regimen involved. Finally, the solid horizontal bars indicate the NCI activity criteria for the various tumor types. There is therefore a considerable amount of data packed into this graph, and it can be regarded as an alternative means for the representation of screening data.

A more adventurous approach to the same problem resulted in the "flower garden" display, an example of which is shown in Figure 4. Here the height of the flower, measured on the axis toward which the flower is leaning, provides a measure of the best % T/C. The height of the flower's head is proportional to the spread of the data, the width to the reciprocal of the dose level. The number of petals is proportional to the number of completed tests, and the route of administration is provided as a mark on the flower's stem. In the example shown in Figure 4, NSC-368390 has been tested in the leukemia 3PS31 and the melanoma 3B131 and gave % T/C values generally below 160. Its activity was somewhat higher against the mammary 3CDJ2 and the colon 3C2G5 and against the colon tumor 3CD72, the lung tumor 3LKG5, and the mammary 3MBG5, but the number of completed tests in each of these was small. The compound's best performance was against the mammary xenografted tumor 3MBG5. Here two satisfactory experiments were recorded for an average % T/C of about zero, provided the route of administration, as indicated by the absence of a device on the flower's stem, was intraperitoneal, which is the normal route.

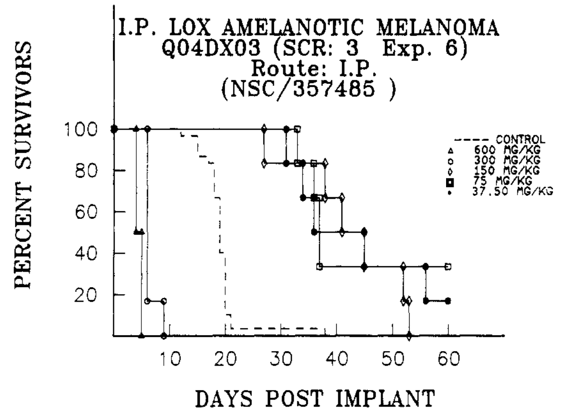


Figure 5. Survival-plot representation of Biology data.

Finally, for detailed display of the behavior of a drug in a single tumor system, survival plots such as that shown in Figure 5 may be used. In this case, with NCS-357485 administered 4 times per day for 3 days, it can be readily seen that at the higher doses (600 and 300 mg/kg), the animals' survival rate is far poorer than that of the controls. At lower doses, survival is improved relative to the control, and in fact at 75 or 37.5 mg/kg, some "cures" (60-day survivals) can be seen in Figure 5.

The survival plot offers considerable detail, but allows only one tumor system per graph. The bar graph or the flower garden allow much more data in a single plot, but are necessarily more superficial. An important point, however, is that use of graphics allows the display of very large amounts of data for a given area of paper. In conventional printed form, using the SDS format, for example, the data in Figure 3 or Figure 4 would require hundreds of pages.

SUMMARY

The scale of the NCI Drug Screening Program ensures that large volumes of data will flow into the Biology databases of the DIS. The challenges in the data management area therefore are, first, capture of this flow of information and then retrieval, analysis, and display of the data. Acquisition of the raw data is handled independently of the DIS. Retrieval of information from within the DIS Biology databases is carried out by using the normal inverted file searches, and for display, various more or less concise output formats are used as are several graphics approaches.

REFERENCES AND NOTES

(1) The in vivo screening laboratories that are currently under contract to NCI are The Southern Research Institute (Birmingham, AL), Battelle Columbus Labs (Columbus, OH), Illinois Institute of Technology Research Institute (Chicago, IL), Mason Research Institute (Worcester, MA), Institut Jules Bordet (Brussels, Belgium), Arizona State University (Tucson, AZ), and the University of California (Los Angeles, CA).

(2) The acronym "NSC" stands for National Service Center, a short form of Cancer Chemotherapy National Service Center, the early name for the program. The "NSC Number" is used by NCI as a Registry Number. Other Registry Numbers, such as the CAS Registry Number, are not useful for NCI because CAS Registry Numbers cannot be assigned to the confidential structures which constitute about half of the NCI database.
(3) These standards are outlined in "Guide for the Care and Use of Laboratory Animals", Institute of Laboratory Animal Resources, National Research Council. *NIH Publ.* **1985**, *No.* 85-23.
(4) The appropriate level of activity varies from one tumor system to another and depends upon whether animal survival or tumor inhibition is being measured. Activity criteria for all standard tumor systems are published by NCI, see: Geran, R. I.; Greenberg, N. H.; Macdonald, N. M.; Schumacher, A. M.; Abbott, B. J. "Protocols For Screening Chemical Agents And Natural Products Against Animal Tumors And Other Biological Systems", 1st ed.; *Cancer Chemother. Rep.* **1972**, *3*. The third and current edition of this report is published as NIH Publication 84-2635, available from the U.S. Government Printing Office, Washington, DC, 20402.
(5) Mice are not long-lived animals, and consequently all the screening laboratories must be supplied with mice at a rate and in quantities which match the flow of chemicals. Assuring an appropriate match is a major task, and it is not undertaken by the DIS. Rather, an independent computer system, loosely articulated with the DIS, manages the flow of animals.
(6) A total of 52 standard vehicles is permitted in animal screening, and this group is, in fact, one of the minor DIS databases.
(7) Full details of all screening procedures are published by NCI in "Instruction 14", 1985. Copies of this booklet may be obtained from the Information Technology Branch, DTP, DCT, NCI, Bethesda, MD 20205.
(8) Some 70 Material Classification codes exist. These are assigned to any active compound and range from 0D (active in P388, but insufficiently so to merit further testing) to 7A (commercial drug developed by NCI). The full list of Material Classification codes is published in "Instruction 14" (see ref 6).
(9) The first paper in this series contains some detail concerning the organization of the NIH Computer Center. Of note here is that the IBM 370 System and the DEC System 10 are "hard-wired" together. Passing large files from either of these computers to the other is very simple and can be very fast.
(10) Non-peak-hour transmission of data was used traditionally to minimize telephone charges. With the networking of the DCRT computers, which began in January 1986, the practice has become unnecessary, and it may transpire that true online operation will be feasible for the screening laboratories.
(11) It is relatively easy to convert a display-only field into a searchable field. The penalty that must be paid, however, is in the form of storage costs for the associated index files and overhead when files are updated. The decision as to whether a field should be searchable devolves therefore upon a trade-off of benefits vs. costs.

The NCI Drug Information System. 6. System Maintenance

M. T. ZEHNACKER, R. H. BRENNAN, and G. W. A. MILNE*

Information Technology Branch, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20205

J. A. MILLER and M. J. HAMMEL

Fein-Marquart Associates, Baltimore, Maryland 21212

Received April 21, 1986

The NCI Drug Information System (DIS) is a collection of 24 interactively searchable databases which contain all the data associated with NCI's drug screening program. Data flow into all of these databases upon a daily basis, and maintenance procedures have been developed which provide a high degree of currency to the files. An extensive security system controls both write access and read access to the DIS and matches both to the authorization possessed by each specific user. Detailed usage statistics are collected automatically. The cost of the overall system in terms of both manpower and machine time is discussed briefly.

INTRODUCTION

The NCI drug screening program began in 1955 and is continuing at a pace that, although a reflection of fiscal limits, is still considerable. Currently, over 30000 new structures are considered for screening each year, and between 5000 and 10000 of these are actually acquired and tested. For any chemical that is acquired, a chemistry record is created at acquisition time, and then inevitably, records concerning its inventorying, shipments, and, finally, its biological test data begin to appear. In order that the NCI Drug Information System (DIS) databases be kept current, the system must be able to assimilate such data readily, and searches in the DIS must reflect the presence of the new data promptly. Given the flow of data and the number of databases, it follows that the updating programs must be essentially automatic, and this has been a design goal of the DIS.

Data integrity in the DIS has been supported, at least in part, by carefully restricting the number of individuals allowed to write to files in the system. To ensure this, a security system is used to control such write access to DIS databases. The same security system controls all read access to the DIS as well and, as will be described, monitors all use of the DIS.

An important guide to the continuing development effect in the DIS is found in the patterns of usage of the system. For this reason, detailed usage statistics are gathered by one of the management programs in the system, and these are processed into a report on a monthly basis. This report shows the frequency with which any command is used and also records the average cpu time required to honor the command. The statistics-gathering routine also records and reports any attempt to broach the DIS security on the part of any user of the computer itself, and such reports, when encountered, can be used to ascertain that data security was not compromised.

DATABASE UPDATING

A. General Considerations. When a new datum is acquired and must be added to one of the DIS databases, a number of steps are necessary. Modification of an existing entry is regarded as deletion of the old entry and addition of the new one, and so the same steps are involved. The first of these steps requires data entry, for which the DIS has an "update" program associated with every database. Thus CHEMUPD allows additions to the Chemistry database, NAMUPD to the Namecodes database, and so on.

With any of these programs, the user must first identify an old record that is to be changed or a new record which is to be entered. Once this is done, the new or changed data are

0095-2338/86/1626-0193$01.50/0