Computer-Assisted Data Management in Medicinal Chemistry—The Use of a General-Purpose Text Editor

S. JOEL KAUFMANN, ARTHUR E. JACOBSON, AND WILLIAM F. RAUB
National Institutes of Health, Bethesda, Md. 20014

Received March 4, 1970

WYLBUR, an interactive computer-based text-editor for the IBM 360/65, has been utilized as a chemical/biological information-handling system. Its applicability is illustrated with chemical and biological data on the 3-benzazocines (6,7-benzomorphans), a class of strong analgetics in which a separation of analgetic activity and physical dependence liability has been achieved in animal species. By providing an open-ended, completely user-defined format, a simple means to match character strings and retrieve data records, and a convenient entree to the batch computation stream, WYLBUR appears to be a relatively inexpensive and useful data management tool to study chemical compounds and their biological effects.

The recent proliferation of multiple-access computer systems is bound to spawn new applications for information-handling technology in every field of science. This seems especially true in the case of medicinal chemistry. Given the capability for "hands-on" interactive use of computers via remote terminals, scientists studying the relationships between molecular structure and biological activity can be expected not only to alter the style in which they use these machines but also to expand the range of appropriate tasks.

One obvious avenue for new applications involves personalized data files and the facilities for computer-assisted management of them. Medicinal chemistry is still largely a descriptive science; it has little in the way of a well-established theoretical underpinning. Individual scientists in this area spend much, if not most, of their investigative effort compiling empirical observations on molecules and the biological effects apparently associated with them and searching for new relationships among these data. It is not at all uncommon to find data on scores or even hundreds of different chemical compounds generated in a single project. On *a priori* grounds alone, it seems safe to predict that individual medicinal chemists will eventually look to the digital computer for aid in perusing, analyzing, and interpreting the data bases they amass.

Though future relief is in sight,[1] powerful data management systems are not yet generally available. This is especially true concerning capabilities for interactive access and user definition of file specifications. In the vast majority of computer installations, each user still must have a computer program specifically designed to handle his personal data files. While high-level languages such as FORTRAN, COBOL, and PL/1 provide considerable flexibility in dealing with tasks of this sort, the programming chores are often time-consuming at best, and substantial duplication of effort among various users is inevitable. Moreover, it is generally difficult if not impossible to alter the structure, content, or style of querying the data bases. And, of course, a file cannot be created and manipulated by one who does not know a programming language.

The study described in this paper clearly does not offer complete release from these limitations; only truly general-purpose data management systems will do that. However, this report does demonstrate that presently available computer technology for interactive text-editing can be a tolerable first approximation to the desired tools, at least for the relatively small data files normally used by individual medicinal chemists. Without embellishing the text-editor at all, we were able to achieve a fairly sophisticated and useful level of personalized file-handling. "While we're waiting," so to speak, for the emergence of more powerful data management systems, sophisticated text-editors may be able to fill some of the void.

BASIC EXPERIMENTAL PROCEDURE

WYLBUR is an interactive, computer-based text-editing system developed at Stanford University[2] for the IBM 360/67 and subsequently implemented on the IBM 360/65 at the National Institutes of Health.[3] As originally conceived, the project reported here was one means to assess the medicinal chemistry community's requirements for computer-based information-handling technology. The research plan was to establish a close working relationship with an active laboratory group and to exploit maximally computer capability already operational at the NIH. Our intent was simply to learn about the interests, ideas, and behavior of scientists interacting with currently available technology. Any insights gained were to have been used by NIH staff, advisors, and contractors to design automated information-handling systems for the study of molecular structure/biological activity relationships.[4]

Of the various computer capabilities available at the NIH, the WYLBUR system proved best suited for the above purpose. While WYLBUR was never intended to be a data management tool (and hence is far from ideal in this context), it seemed sufficiently powerful for a clever and tolerant scientist to use productively in this mode. However, experience soon indicated that WYLBUR, though limited, is nowhere near as awkward as originally

feared. (As a consequence, this project not only offers some insights as to the type of information-handling technology that the medicinal chemistry community requires, but also demonstrates that a general-purpose text-editing system can be a valuable "poor man's" information-handling tool.)

The medicinal chemist in our group (AEJ) was purposely kept at a distance from the machine by a "human interface" (SJK). The latter was by no means a passive conveyor of information. On the contrary, he was both reflective and introspective, and often responded to the chemist's inquiries with, "What is it you really mean?" or "Why do you want to know just that?" In this way, we not only made an initially strange and confusing technology palatable to the chemist; we also provided an opportunity to study both the science being practiced and the scientist.

This style of systems design was helpful in building the WYLBUR file structure, in discovering and inventing modes of interaction with the system, and in piecing together some requirements for a more sophisticated information-handling system for this community. The medicinal chemist rapidly understood the WYLBUR system itself; in fact he often had to be asked to clarify the reasons behind his queries, for he soon began to phrase these in terms of COPY and LIST commands rather than more general chemical or biological requirements.

THE DATA BASE

The data chosen for our sample file concerns morphine-like analgetic compounds. For some years, E. L. May[5] and his colleagues have been searching for drugs which possess strong (morphine-like) analgetic activity without morphine's dependence-producing characteristics or other side effects. One such general class of compounds is the 6,7-benzomorphans[6,7] which we chose for our file. The basic structure of these compounds is shown in Figure 1.

This series of analgetic compounds was chosen because it represents a relatively small file (approximately 200 compounds) for which structure-activity relationships had previously been derived by primarily intuitive reasoning.[6] We were interested to see whether the WYLBUR computer system, with its fairly extensive data-handling capabilities and its ready entree to mathematical calculations, could be used to find generalizations which had previously been overlooked in these data.

THE WYLBUR SYSTEM

As with most text-editing systems, each individual numbered line in WYLBUR constitutes one "piece of information" for the computer. Once entered via the IBM 2741 typewriter or standard Teletype terminal (conveniently in response to easy-to-understand prompts), individual lines or groups of lines can be listed, copied, moved, changed, or deleted. These lines can be specified by line number or by content—that is, a string of characters contained within a line can serve to specify that line.

The COPY command can be used to form subfiles of pertinent data. By establishing successive subfiles, Boolean file partitioning—e.g., (A *or* B) *and* C—and substructure searching via explicitly-encoded fragments can be performed. A COPY command (which returns the total number of lines copied) can also be used to count the lines containing given information, before LISTing lines of interest. Note that WYLBUR does not distinguish between letters and numbers: "A" and "99" and "?*&%" are all treated in analogous fashion.

WYLBUR will save working data sets for future use and distinguishes between public and private files, the latter accessible only through a keyword. File format, both on input and output, is free-form and easily modifiable, giving the user essentially complete freedom of file definition. In addition, the system has extensive built-in error-handling capability. The actual operation of WYLBUR should become clear by studying the examples presented in the body of this paper and in the appendix.

In addition to its file-handling capabilities, WYLBUR has direct access to the Remote Job Entry facility of the main IBM 360/65 computer system. Thus, data from a WYLBUR file can be used as input to a standard job—e.g., a program calculating a correlation coefficient—and the output can be readily inspected from the remote terminal. This Remote Job Entry feature puts the full computational power of the IBM 360/65 at the disposal of the WYLBUR user and thereby provides an added dimension in information-handling capability.

THE FILE

The benzomorphan file is designed to be a self-contained unit, such that someone with a basic knowledge of WYLBUR and little or no knowledge of the specific file can manipulate it successfully. Figure 2 shows the first few pages of the actual WYLBUR data set, obtained simply by issuing a LIST OFFLINE command. This introductory section describes the general content and organization of the ensuing file. Much of it is self-explanatory. We point out some of the more important characteristics for emphasis.

The name of the file and the date and time of the last revision (the last SAVE of the data set) are included at the beginning. The initial section (lines 0.5/1.0) contains a brief description of the over-all file and, importantly, the names and phone numbers of both a "computernik" and a medicinal chemist. We suggest that new users first

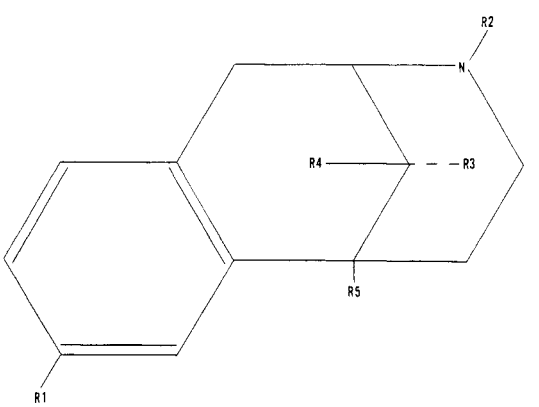

Figure 1. 6,7 Benzomorphan (the R-groups specify substituents)

```
0.001   PUBLIC FILE MORPHINE*LIKE*ANALGETICS
0.002
0.003
0.1
0.2     October 31, 1969  --  3:00 p.m..
0.3
0.4
0.5          The following data deals with the chemistry and pharmacology of morphine-like drugs.  The information
0.6     was compiled and arranged at the National Institutes of Health by Dr. Art Jacobson, NIAMD (301-496-5403)
0.7     and Joel Kaufmann, DCRT (301-496-3403).  In lines 100/3000, the NIH compound number is implicit in the last five
0.8     digits of each line number;  the first digit(s) acts as a key to the type of information, as specified below.
0.9     The NIH number is repeated in columns 1/8 of each of these lines.  We suggest that new users peruse the entire
0.91    file on a printout before initiating any queries.
1.
1.1
1.2                                                INDEX OF LINES
1.3
1.4      1/100                   Index and Notes
1.5      100.001/ 120.000        COMPOUND NAME:  begins in col 15; levo-rotatory specified by (-), dextro- by (+)
1.6      200.001/ 220.000        DATE RECEIVED, SOURCE, and MISC. I.D. NUMBERS
1.7      300.001/ 320.000        ANALGESIC ACTIVITY:  ED-50, mg/kg s.c. in mice (number in col 22/26, activity index in col 40/45);
1.8                                                   LD-50, mg/kg s.c. in mice (number in col 59/62, activity index in col 74/78);
1.9                                                   OD, mg/kg oral ED-50 in mice(number in col 89/92, activity index in col 104/108
2.                                   (numbers in parentheses represent 95% confidence intervals)
2.1      400.001/ 420.000        Analgesic Activity:  ONSET, PEAK, and DURATION in minutes...
2.2                                                   for s.c. administration:  O(col 26/29); P(col 36/39); D(col 46/50)
2.3                                                   for oral administration:  O(col 68/71); P(col 78/81); D(col 88/92)
2.4      500.001/ 520.000        PHYSICAL DEPENDENCE CAPACITY:  most recent activity index in col 21/26;
2.5                                 source:  Committee on Problems of Drug Dependence--Nat'l Res. Council, Nat'l Academy of Sciences
2.6      600.001/ 620.000        SUBSTRUCTURES, based on 6,7-Benzomorphan:  levo(-) or dextro(+) specified in col 10/12;
2.7                                 Substructures are:  R1(col 17/34), R2(col 39/60), R3(col 65/74), R4(col 79/88), R5(col 93/104);
2.8                                   these are substituents at:  R1=C2', R2=N, R3=C9, R4=C9, R5=C5, on the benzomorphan nucleus,
2.9                                   where R3 and R4 specify the stereochemical configuration of the molecule.
3.9                                 Also entered are the activity indices of PDC(col 105/110), ED-50(col 111/116), LD-50(col 117/121)
4.       700.001/ 720.000        SUBSTRUCTURES, based on other compounds
4.1      800.001/ 820.000             "
4.2      900.001/ 920.000             "
4.3     1000.001/1020.000             "
4.4     1500.001/1520.000        PARTITION COEFFICIENTS
4.5     1900.001/1920.000        THREE-DIMENSIONAL COORDINATES:  x,y,z in Angstrom units based on standard Dreiding model & x-ray data
4.6     2000.001/2020.000        Results of MO CALCULATIONS using INDO routine
4.601
4.602
4.603
4.604
4.605
4.606
4.607
4.608
4.609
4.61
4.69                                               COMPOUND NAMES
4.691
4.692        Names are stated with radicals in alphabetical order, with the first radical capitalized and the base
4.693   (6,7-benzomorphan) appearing last.
4.694        Stereochemical configurations are indicated by 'alpha'(R3 substituent) or 'beta'(R4 substituent) in parentheses
4.695   following the first occurrence of the 9-position in the name, e.g., 2,9(alpha)-Dimethyl-9-hydroxy-2'-methoxy-
4.696   5-propyl-6,7-benzomorphan.
4.697        n-Chains are written without the n-, e.g., n-amyl = amyl.
...
4.9
5.                                              ACTIVITY RANGES AND INDICES
5.1
5.2            ED-50                         LD-50                         OD                              PDC
5.3
5.4        < 0.10       ++++         0-100.00        +++        < 1.00         +++         high                 ++++
5.5        0.11-1.00    +++          100.01-200.00   ++         1.01-5.50      ++          intermediate         +++
5.6        1.01-3.50    ++           200.01-300.00   +          5.51-15.50     +           low                  ++
5.7        3.51-6.50    +            300.01-400.00   -          15.51-30.50    -           very low             +
5.8        6.51-15.00   -            400.01-500.00   --         30.51-50.00    --          none                 -
5.9        > 15.00      --           > 500.00        ---        > 50.00        ---         antagonist:
6.                                                                                           <<Nalorphine      --  (i.e. <=half)
6.1                                                                                          Nalorphine-like ---
6.2                                                                                          >>Nalorphine      ----(i.e. >=twice)
6.3
6.4                                                        note:  The OD is not predictable from mouse to man.
6.5
6.6
6.61
6.62
6.7
6.8                                         TYPICAL ACTIVITY FIGURES
6.9
7.              DRUG                      ED-50            LD-50               OD                PDC
7.1
7.2             morphine                   1.0              575                 2                high
7.3             meperidine                 4.5              270                21                high
7.4             codeine                    7.5              180 (oral)         25            intermediate
7.5
```

Figure 2. The introductory section

examine the entire file on a printout. This offline listing furnishes a convenient index and can be used to peruse the data for both format and content, to make notes, and to mark compounds or areas of interest.

The next section of the introduction (lines 1.1/4.6 in Figure 2) is an index to the entire file, by line number. Since WYLBUR does not recognize numerical values (an obvious shortcoming), specific ranges of activity were encoded into activity indices, as described in lines 5.0/6.6. In this way, it is possible to ask meaningfully for all compounds within a given range—e.g., to find all compounds with an ED-50 between 0.11 and 1.00 mg/kg

```
7.6
7.7
7.8
7.9
8.1                          BIOLOGICAL TESTING OF POTENTIAL ANALGETIC COMPOUNDS
8.2
8.3
8.4     Stage I
8.5
8.6     A.  Exploratory Dose--Analgetic activity is determined in mice using the procedure of N. B. Eddy and
8.7     D. Leimbach (J. Pharmacol. Exptl. Therap., Vol. 107, 385 (1953)).  An exploratory dose of 20 mg/kg
8.8     is administered subcutaneously to each of about ten fresh or previously used mice (see Ref. 9 in
8.9     A. E. Jacobson and E. L. May, J. Med Chem. Vol. 8, 563 (1965) for a description of these Caesarian-
9.      Derived General-Purpose (CDGP) mice).  If little or no effect is observed, higher doses are used, and
9.1     conversely.
9.2
9.3     B.  Assay--An assay range is chosen based on data from the exploratory dose (e.g., 10, 5, 2.5, 1.25,
9.4     0.625 mg/kg).  At each of these dose levels, a set of six to ten fresh (unused) CDGP mice and six to
9.5     ten previously used CDGP mice are tested for analgetic activity using the Eddy-Leimbach technique.  An
9.6     individual mouse is first tested without drug to obtain its normal response to the heat stimulus, that
9.7     is the time it takes, in seconds, for the mouse to raise its hind feet.  After subcutaneous injection of
9.8     the drug at a particular dose, the drug's effect is noted after 5, 10, 20, etc., minutes, until the mouse's
9.9     response time returns to within two seconds of its normal value.  The onset, peak and duration of action of
10.     the drug are obtained by a trained observer's visual inspection of the data obtained in the assay.  The number
10.1    of mice effected by a drug at each dose level is tallied and sent to the NIH Computer Center for probit
10.2    analysis.  Thus, the assay, followed by probit analysis, results in an ED-50 with 95% confidence limits.
10.3    An oral ED-50 is determined similarly.
10.31        The Eddy-Leimbach method is limited to morphine-like analgetics; narcotic antagonists give poor
10.32   (non-reproducible) results although the antagonist may itself be an effective analgetic in man.
10.4
10.5
10.6    Stage II
10.7
10.8    An LD-50 (acute) is obtained at NIH, in mice, for compounds of interest.  Chronic toxicities are determined
10.9    elsewhere in three species, in accord with FDA regulations.
11.
11.1
11.2    Stage III
11.3
11.4    Analgetically active and relatively non-toxic compounds may then be sent to Drs. J. E. Villarreal and
11.5    M. H. Seevers, Department of Pharmacology, University of Michigan, Ann Arbor, for evaluation for morphine-
11.6    like physical dependence in the Rhesus monkey.  The compounds are first studied in single-dose suppression
11.7    tests (SDS) in morphine-dependent monkeys where their capacity to suppress the manifestations of abstinence
11.8    is evaluated.  Some of the drugs are also administered chronically to non-morphine-dependent monkeys for a
11.9    month to determine their ability to induce primary physical dependence.
12.
12.1    Compounds that produced exacerbation of morphine abstinence are suspect of having morphine-antagonist
12.2    (nalorphine-like) properties, and are subsequently studied in abstinence-precipitation tests.  (For
12.3    details of the techniques used see the Minutes of the 25th Meeting in the 1963 Annual Report to the
12.4    Committee on Problems of Drug Dependence, National Academy of Sciences--National Research Council,
12.5    Division of Medical Sciences, Washington, D. C.).
12.6
12.7
12.8    Stage IV
12.9
13.     Compounds which are still of interest after undergoing Stages I-III may be sent to Dr. W. R. Martin at
13.1    the Addiction Research Center, Lexington, Kentucky for testing in man.
13.2
13.3    Data derived from the various stages are evaluated by the Committee on Problems of Drug Dependence and a
13.4    determination is made as to whether the drug in question should be put under control as a narcotic.
13.5
13.6
13.7
13.8
13.9
14.                                  SUBSTRUCTURE NOTATION
14.1
14.2
14.3    Ring Structures
14.4
14.5         Aromatic rings are indicated by 'ar' followed by the formula in <brackets>.  Other rings are indicated by
14.6    'cyclo' followed by the formula in <brackets>.
14.7         Substituents on benzene rings are preceded by '-o-' (ortho), '-m-' (meta), or '-p-' (para).  In heterocyclic
14.8    rings the hetero atom is numbered 1, and the smallest numbers are then used to specify substituents.  An initial
14.9    number indicates the place of attachment of the benzomorphan nucleus to this moiety.  For example, a
15.     2-pyridylmethyl substituent is indicated by CH2ar<2-C5H4N>.
15.1         The placement of a double bond in unsaturated non-aromatic rings is indicated by '-delta' appended to the
15.2    ring formula, followed by a number specifying the bond starting location.  The point of attachment to the benzomorphan
15.3    nucleus is numbered 1, and again the smallest numbers are used.  For example:  CH2cyclo<C6H9-delta1>.
15.4
15.5    Other Structures
15.6
15.7         Multiple bonds not in rings are omitted, e.g., -CH=CH2 is indicated CHCH2.
15.8         All chains are completely extended, e.g., CH2CH2CH2CH3.
15.9         A dimethylallyl substituent is written CH2CHC(CH3)2. (In ACS nomenclature this is actually 3-methyl-2-butenyl.)
16.          An acetoxy substituent is written OCOCH3.
16.1
```

of the WYLBUR data set

s.c. (+++). The final sections indicated in the index (lines 4.0/4.6) have not yet been entered; it may be more convenient to form separate files for some of these data.

Organization within lines is indicated in the index by noting the column in which a given piece of information is contained. Since a WYLBUR search can be confined to specific columns, this convention allows a greater diversity in the queries. Again, this should become clear by referring to later examples in this paper.

Following the index in Figure 2 are sections providing information on each type of data in detail: the format for compound names; the various activity ranges; some

```
? list 106.048,107.569,107.949,108.092,108.324

 106.048  NIH 6048          2,5-Dimethyl-6,7-benzomorphan
 107.569  NIH 7569      (-)-2'-Hydroxy-2,5,9(alpha)-trimethyl-6,7-benzomorphan
 107.949  NIH 7949      (+)-2'-Hydroxy-2,5,9(beta)-trimethyl-6,7-benzomorphan
 108.092  NIH 8092          2'-Amino-5,9(beta)-diethyl-2-methyl-6,7-benzomorphan
 108.324  NIH 8324      (-)-5,9(alpha)-Diethyl-2'-hydroxy-2-methyl-6,7-benzomorphan -d-mandelate salt


? list 206.048,207.410,207.558,208.104,208.445

 206.048  NIH 6048          12/16/53                    NIH - E. L. May
 207.41   NIH 7410          7/17/56 & 5/28/64           NIH - E. L. May & T. Ishimaru
 207.558  NIH 7558          10/23/57                    NIH - E. M. Fry (0-59-9) & E. L. May
 208.104  NIH 8104          5/16/63                     NIH - A. E. Jacobson
 208.445  NIH 8445          2/8/68                      NIH - J. H. Ager      (#A-960)


? list 306.048,307.775,307.781,307.892,308.345

 306.048  NIH 6048      ED-50: 11.03(9.64-12.60) (-)        LD-50: 148.(143.-153.)(++)       OD: 21.1(19.6-22.6)(-)
 307.775  NIH 7775      ED-50:  0.08(0.07-0.09)  (++++)     LD-50: not measured              OD:  3.9(3.3-4.6)  (++)
 307.781  NIH 7781      ED-50:  0.87(0.75-1.00)  (+++)      LD-50:  18.(16.7-18.5)(+++)      OD: not measured
 307.892  NIH 7892      ED-50: inactive          (--)       LD-50: not measured              OD: not measured
 308.345  NIH 8345      ED-50: 57.5 (51.1-64.7)  (--)(?)    LD-50: not measured              OD: not measured


? list 406.048,407.327,407.783,407.964,408.392

 406.048  NIH 6048      s.c. -- O:  2.8   P: 12.5   D:  79.4      oral -- not measured
 407.327  NIH 7327      s.c. -- O:  4.4   P: 28.9   D: 165.3      oral -- O:  4.1   P: 26.0   D: 156.5
 407.783  NIH 7783      s.c. -- not measured                      oral -- not measured
 407.964  NIH 7964      s.c. -- O:  6.7   P: 24.3   D:  94.5      oral -- O:  3.5   P: 25.1   D: 124.5
 408.392  NIH 8392      s.c. -- O:  5.5   P: 30.2   D: 131.4      oral -- O:  3.4   P: 22.4   D: 153.3


? list 507.327,507.410,507.569,508.392,508.408

 507.327  NIH 7327      PDC:  (-)   =none, 1962
 507.41   NIH 7410      PDC:  (+)   =very low, 1960          (-)   =none, 1960
 507.569  NIH 7569      PDC:  (--)  =antagonist, 1/50-1/30 x Nalorphine, 1969     (+)   =very low, 1960   (-)   =none, 1960
 508.392  NIH 8392      PDC:  (--)  =mild antagonist, 1969
 508.408  NIH 8408      PDC:  (+)   =slight (questionable) suppression of abstinence signs, 1969


? list 607.327,607.613,607.785,607.964,608.194

 607.327  NIH 7327     R1: OH    R2: CH3                  R3: H       R4: H       R5: CH3          (-)   (++)  (++)
 607.613  NIH 7613 (-)R1: OH    R2: CH2CH2ar<C6H5>       R3: CH3     R4: H       R5: CH3          (++++)(++++)(++)
 607.785  NIH 7785     R1: OH    R2: CH2CH2CH2CH2CH3      R3: CH3     R4: H       R5: CH3          (+)   (++)
 607.964  NIH 7964     R1: H     R2: CH3                  R3: H       R4: H       R5: CON(CH3)2          (-)
 608.194  NIH 8194 (+)R1: OH    R2: CH3                  R3: CH2CH3  R4: H       R5: CH2CH3       (+++) (-)   (+)
```

Figure 3. Some typical file entries

```
? list "5,9(alpha)-Diethyl-2'-hydroxy-2-(2-phenethyl)-6,7-benzomorphan" in 100/120
 107.864  NIH 7864          5,9(alpha)-Diethyl-2'-hydroxy-2-(2-phenethyl)-6,7-benzomorphan

? list "5,9(alpha)-Dimethyl-2'-hydroxy-2-(2-phenethyl)-" in 100/120
 107.613  NIH 7613      (-)-5,9(alpha)-Dimethyl-2'-hydroxy-2-(2-phenethyl)-6,7-benzomorphan
 107.614  NIH 7614      (+)-5,9(alpha)-Dimethyl-2'-hydroxy-2-(2-phenethyl)-6,7-benzomorphan
 107.775  NIH 7775          5,9(alpha)-Dimethyl-2'-hydroxy-2-(2-phenethyl)-6,7-benzomorphan
```

Figure 4. Retrieving the NIH number of a particular compound

typical activity figures; a scientifically accurate and complete description of the methods used to gather the data, including references and examples; and the format conventions for structure information. The layout of this information on the page is fairly free-form, including notes and comments, as contrasted with the more highly structured organization of the remainder of the file.

Figure 3 shows some of the entries (lines of text in WYLBUR) which comprise the body of the file. We found it convenient to use one series of lines (600/620) as "overview lines," containing both structure and rough activity data. In this way, we could easily locate compounds with a given substituent and a given activity range.

Lines 600/620 contain complete structure information as well as activity indices for the ED-50, the LD-50, and the OD. The method used for structure notation is adequate for the present purposes but obviously is not optimal, especially when compared with currently available graphics display capabilities.[8] However, since WYLBUR can easily accommodate other linear notation methods—e.g., the Wiswesser line notation[9]—instead of or in addition to the *ad hoc* one used here, a substantial chemical information-handling capability is readily achieved.

USE OF THE SYSTEM: EXAMPLES

Example 1. Biological studies and the filing of benzomorphan samples use NIH numbers rather than specific compound names. The retrieval of the NIH numbers of specific compounds is illustrated in Figure 4. As in the other figures, lines preceded by "?" (the prompt) were typed by the user; the other lines are WYLBUR's responses.

```
? list 607.550,307.550,607.410,307.410

 607.55   NIH 7550    R1: OCH3              R2: CH3                    R3: CH3       R4: H         R5: CH3          (+)   (+)
 307.55   NIH 7550      ED-50:  4.88(4.13-5.77)  (+)        LD-50: not measured                 OD: 10.9(9.8-12.0) (+)

 607.41   NIH 7410    R1: OH                R2: CH3                    R3: CH3       R4: H         R5: CH3          (+)   (++)  (++)
 307.41   NIH 7410      ED-50:  1.34(1.18-1.52)  (++)       LD-50: 175.(161.-190.)(++)          OD: 12.0(10.9-13.2)(+)

? list 607.625,307.625,607.519,307.519

 607.625  NIH 7625    R1: OCH3              R2: CH2CH2ar<C6H5>         R3: CH3       R4: H         R5: CH3          (+)   (++)
 307.625  NIH 7625      ED-50:  3.26(2.84-3.74)  (++)       LD-50: not measured                 OD:  5.6(4.5-6.2)  (+)

 607.519  NIH 7519    R1: OH                R2: CH2CH2ar<C6H5>         R3: CH3       R4: H         R5: CH3                    (++++)(-)
 307.519  NIH 7519      ED-50:  0.10(0.08-0.11)  (++++)     LD-50: 332.(278.-398.)(-)           OD:  3.2(2.8-3.6)  (++)

? list 608.445,308.445,607.327,307.327

 608.445  NIH 8445    R1: OCH3              R2: CH3                    R3: H         R4: H         R5: CH3                    (-)
 308.445  NIH 8445      ED-50:  7.22(5.73-9.10)  (-)        LD-50: not measured                 OD: 19.5(15.2-25.1)(-)

 607.327  NIH 7327    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH3          (-)   (++)  (++)
 307.327  NIH 7327      ED-50:  3.28(2.55-4.24)  (++)       LD-50: 175.(164.-186.)(++)          OD: 11.0(8.9-13.7) (+)

? list 608.486,308.486,608.484,308.484

 608.486  NIH 8486    R1: OCH3              R2: CH3                    R3: H         R4: H         R5: H                      (--)  (+++)
 308.486  NIH 8486      ED-50: inactive           (--)       LD-50: toxic at 20    (+++)         OD: not measured

 608.484  NIH 8484    R1: OH                R2: CH3                    R3: H         R4: H         R5: H                      (+)   (++)
 308.484  NIH 8484      ED-50:  4.46(3.33-5.97)  (+)        LD-50: ca. 180        (++)          OD: not measured
```

Figure 5. Comparison of the activities of 2′-hydroxy and 2′-methoxy compounds

```
? list 608.484,308.484,607.327,307.327,607.910,307.910,607.974,307.974,608.142,308.142

 608.484  NIH 8484    R1: OH                R2: CH3                    R3: H         R4: H         R5: H                      (+)   (++)
 308.484  NIH 8484      ED-50:  4.46(3.33-5.97)  (+)        LD-50: ca. 180        (++)          OD: not measured

 607.327  NIH 7327    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH3          (-)   (++)  (++)
 307.327  NIH 7327      ED-50:  3.28(2.55-4.24)  (++)       LD-50: 175.(164.-186.)(++)          OD: 11.0(8.9-13.7) (+)

 607.91   NIH 7910    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH2CH3       (++)  (++)  (++)
 307.91   NIH 7910      ED-50:  1.17(1.01-1.35)  (++)       LD-50: 171.(162.-180.)(++)          OD: not measured

 607.974  NIH 7974    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH2CH2CH3    (++)  (++)  (++)
 307.974  NIH 7974      ED-50:  1.05(0.90-1.23)  (++)       LD-50: 130.(118.-142.)(++)          OD:  7.4(6.5-8.3)  (+)

 608.142  NIH 8142    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH2CH2CH2CH3       (++)  (++)
 308.142  NIH 8142      ED-50:  2.01(1.73-2.32)  (++)       LD-50: 130.(118.-143.)(++)          OD: 54.3(51.1-57.8)(---)

? list 608.157,308.157,608.150,308.150

 608.157  NIH 8157    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH2CH2CH2CH2CH3(-)(++)  (+++)
 308.157  NIH 8157      ED-50:  3.42(2.94-3.98)  (++)       LD-50:  93.(83.-105.) (+++)         OD: not measured

 608.15   NIH 8150    R1: OH                R2: CH3                    R3: H         R4: H         R5: CH2CH2CH2CH2CH2CH3(-)
 308.15   NIH 8150      ED-50: 10.78(9.60-12.11) (-)       LD-50: not measured                 OD: not measured


? list 607.410,307.410,607.951,307.951,607.938,307.938,608.039,308.039,607.990,307.990

 607.41   NIH 7410    R1: OH                R2: CH3                    R3: CH3       R4: H         R5: CH3          (+)   (++)  (++)
 307.41   NIH 7410      ED-50:  1.34(1.18-1.52)  (++)       LD-50: 175.(161.-190.)(++)          OD: 12.0(10.9-13.2)(+)

 607.951  NIH 7951    R1: OH                R2: CH3                    R3: CH2CH3    R4: H         R5: CH3          (-)   (+++) (++)
 307.951  NIH 7951      ED-50:  0.75(0.67-0.84)  (+++)      LD-50: 134.(125.-144.)(++)          OD: not measured

 607.938  NIH 7938    R1: OH                R2: CH3                    R3: CH3       R4: H         R5: CH2CH3       (++)  (++)  (-)
 307.938  NIH 7938      ED-50:  2.47(1.95-3.12)  (++)       LD-50: 309.(273.-349.)(-)           OD: 15.9(14.2-17.7)(-)

 608.039  NIH 8039    R1: OH                R2: CH3                    R3: CH3       R4: H         R5: CH2CH2CH3    (-)   (++)  (-)
 308.039  NIH 8039      ED-50:  1.46(1.26-1.69)  (++)       LD-50: 400.(356.-449.)(-)           OD: 36.1(32.2-40.4)(--)

 607.99   NIH 7990    R1: OH                R2: CH3                    R3: H         R4: CH2CH2CH3 R5: CH2CH2CH3            (+++) (+++)
 307.99   NIH 7990      ED-50:  0.43(0.37-0.50)  (+++)      LD-50:  62.(59.-65.)  (+++)         OD: not measured
```

Figure 6. Activity *vs.* substructure

Example 2. Before creation of a computer file of the benzomorphan data, it was presumed that a change in the aromatic ring's substituent (R1) from hydroxy to methoxy will usually decrease the analgetic activity and the physical dependence capacity of the benzomorphan. Figure 5 presents this comparison between the two R1 substituents (OH *vs* OCH3). In each set of two compounds all of the other substituents (R2, R3, R4, and R5) are identical.

Figure 5 is the final result of a search of the entire file by structure substituent. As initial command—LIST 'R1: OCH3' IN 600/620—isolated all methoxy-substituted compounds. Then, successive COPY and LIST commands located corresponding hydroxy-substituted compounds:

COPY 'R1: OH' IN 600/620 TO 3700 (system responds: 3813 - LAST LINE); COPY 'R2: CH3' IN 3700/3813 TO 3900 (system responds: 3960 - LAST LINE); COPY 'R4: H' IN 3900/3960 TO 4000 (system responds: 4043 - LAST LINE); LIST 'R3: CH3' IN 4000/4043 (system lists all appropriate compounds). After other similar searches, corresponding results were grouped together to produce the output of Figure 5, which tends to corroborate the aforementioned presumption.

Example 3. Using data bases of from two to two thousand compounds, medicinal chemists inevitably try to relate biological activity with some structural fragment of the parent molecule. One such approach is shown in Figure 6; R1 = OH, R2 = CH3 and R4 = H are maintained as constants and comparisons are made using R3 and R5. Data derived from this have indicated that an attempt should be made to obtain at least one other compound in this series. If successful, this work will be reported elsewhere. However, even a simple comparison of this sort, when clearly presented by WYLBUR, can lead to further useful work.

As in the previous example, Figure 6 is the final result, or "pretty print," of successive searches and correlations by structure substituent. WYLBUR, of course, lacks a more comprehensive report generation capability.

Other Examples. Other queries have been made of WYLBUR, but space precludes their complete presentation here. They were resolved similarly to those presented above. For example, a very common question determined whether a particular compound had been prepared and, if it had, what was its biological activity. A more complex question located those benzomorphans for which electronic correlates of biological activities were to be sought using molecular orbital (MO) calculations. A series of compounds were needed for this purpose, since the results of MO calculations are quite relative in nature. Thus, WYLBUR was asked to choose benzomorphans in various analgetic activity ranges (high, medium, low) and to cross-correlate them with a similar range in physical dependence capacity. The compiled list will be used for the MO calculations.

Examples of miscellaneous useful commands as applied to the benzomorphan file are presented in the Appendix. These represent a "users guide" to WYLBUR as a data-management tool.

DISCUSSION

WYLBUR's usefulness in data management derives from the simultaneous presence of four distinct features: its facilities for easy entree to high-volume, direct-access storage from inexpensive remote terminals; its facilities for locating any given string of acceptable symbols within a prescribed set of columns in a prescribed set of lines; its facilities for altering, copying, and moving entire data sets or prescribed portions thereof; and its facilities for direct communication between data sets and the batch computation stream. Absence of any one of these features would seriously impair WYLBUR's performance as a text-editor and almost preclude realistic application for data management purposes. By exploiting all four in combination, however, one acquires a limited but useful unit record search and retrieval system having an especially facile interplay between data files and the full range of programming languages and procedures supported under Operating System 360.

One blanket qualification on WYLBUR's capabilities needs to be made. As a text-editor, WYLBUR is highly user-oriented; indeed, that is one of its hallmarks. As a data management tool, however, it is often somewhat clumsy to operate. Inspection of the foregoing examples and the appendix should make it clear that imagination and patience are often necessary for successful use. Nevertheless, the repertoire of special maneuvers and command sequences is actually quite short and can be mastered in much less time than it takes to become reasonably proficient with most high-level programming languages. Until sophisticated, general-purpose data management systems become commonplace, interactive text-editors like WYLBUR may be able to relieve medicinal chemists and others of many of their file-handling burdens.

IMPLICATIONS

The WYLBUR experiments clearly indicate the value of, if not the absolute requirement for, convenient, dynamic interaction between the user and the computer system. Questions concerning molecular structure/biological activity relationships are often complex, tentative, and interrelated, demanding that the scientist interact with and direct the data-handling process. An interactive system like WYLBUR provides an extension of the individual's cognitive powers, readily helping him order and interpret information otherwise understandable only with extreme difficulty, if at all. This cognitive assistance is especially crucial in empirical sciences like medicinal chemistry, where there is little theoretical base and where ordering and interpreting large volumes of experimental observations is the rule rather than the exception. In addition, the scientist is able to pursue questions, thoughts, guesses, and feelings that arise peripherally in his day-to-day work or in the perusal or query of his file. A cumbersome file system would preclude this undirected but vital browsing which often results in unexpected important discoveries.

The WYLBUR experiments also indicate the utility of "personalizable" data files. A data-handling tool designed to extend the cognitive powers of individuals must be readily modifiable in the hands of the individuals themselves. Pre-defined data types and data organization frameworks at best will be inconvenient for many users; at worst (and we believe this is the most common case) they will preclude effective use of some types of information vital to a given individual, either by providing no means to handle this information or by making it extremely difficult to do so. Ideally, the system should initially appear to the user as an "empty notebook." That is, the individual investigators should easily be able to tailor their personal data bases to their own needs and tastes by defining the file structure as well as the file content. Within the limits of its unit record orientation—i.e., its emphasis on handling *lines* of information and its relatively limited provision for dynamic storage allocation—WYLBUR offers the user considerable freedom to determine the character of his personal data file in this way.

APPENDIX

Some Examples of WYLBUR Commands Useful for Data-Management Purposes

The complete set of WYLBUR commands and their formats are explained fully in the WYLBUR reference manual.[2]

We list here some examples of miscellaneous useful commands as applied to the benzomorphan file. They are merely illustrative of the power of the system; WYLBUR can locate, compare, and list any of the information in the file.

To locate and list all compounds with a benzene ring as the R2 substituent:

```
LIST 'R2: ar <C6H5> 'IN 600/620
```

To locate and list all compounds with an aromatic ring present somewhere in the R2 substituent (columns 39/60):

```
LIST 'ar <C6H5> '39/60 IN 600/620
```

To locate all compounds with a high PDC, and to list the PDC values:

```
LIST '(++++)'21/26 IN 500/520
```

To locate all compounds with a high PDC, and to list their structures:

```
LIST '(++++)'105/110 IN 600/620
```

To locate all compounds with a PDC of low, intermediate, or high and to list the PDC values:

```
LIST '++'21/26 IN 500/520
```

To locate and list all compounds not made at NIH:

```
MOVE 'NIH'10/130 IN 200/220 TO 1000
LIST 200/220
```

To locate and list all compounds with a low PDC value, a high ED-50, and at least a three-carbon chain on the R5 substituent:

```
COPY '(++)'105/110 IN 600/620 TO 1000
(System responds: 1050 - LAST LINE)
COPY '+++'111/116 IN 1000/1050 TO 1100
(System responds: 1128 - LAST LINE)
LIST 'CH2CH2CH'93/104 IN 1100/1128
```

To list all information on a given compound:

```
LIST 'NIH 7410' IN ALL
```

To change data in the file—e.g., to modify substructure nomenclature of a phenyl ring from *ortho*-phenyl to 2-phenyl:

```
CHANGE '-o-' to '-2-' IN 600/620
```

To obtain a count of the compounds before listing them, one would replace the LIST command with a COPY. . .TO <some high number>. Then all or some of the resulting subfile could be LISTed.

ACKNOWLEDGMENT

We thank Everett L. May and Arnold W. Pratt for their kind help in this project.

LITERATURE CITED AND FOOTNOTES

(1) Minker, Jack, and Jerome Sable, "File Organization and Data Management," in *Annual Review of Information Science and Technology*, Carlos A. Cuadra, Ed., American Documentation Institute, Annual Review Series, Vol. **2**, Chap. 7, Wiley, Interscience, New York, N. Y., 1967.

(2) WYLBUR was written at the Stanford University Computation Center in 1967, principally by Roger Fajman and John Borgelt. Fajman is presently at the National Institutes of Health Computer Center in Bethesda, Md.

(3) Barden, Leslie, Ed., "WYLBUR Reference Manual," Computer Center, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Md., July, 1969.

(4) This study is a component activity of the Chemical/Biological Information-Handling Program, Division of Research Resources, Bureau of Health Professions Education and Manpower Training, National Institutes of Health, Bethesda, Md. The program is under the direction of William F. Raub. It employs both staff investigations and outside contracts to develop and evaluate new information-handling systems for pharmacology research.

(5) Chief, Section on Medicinal Chemistry, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Md.

(6) The chemical and biological data were obtained from the files maintained at the Section on Medicinal Chemistry, NIH. Many of these data have been published in various journals and some have been compiled in texts. For compilations and leading references to the original literature, see: Jacobson, Arthur E., E. L. May, and L. J. Sargent, "Medicinal Chemistry," Vol. **II**, Chap. 49, A. Burger, Ed., 3rd. ed., Wiley, New York, N. Y., in press (about July 1970); Eddy, N. B., and E. L. May "Synthetic Analgesics," Part II, J. Rolfe, Ed., Pergamon Press, London, 1966; May, E. L., and L. J. Sargent, in "Analgetics," Chap. IV, p. 123, G. de Stevens, Ed., Academic Press, New York, N. Y., 1965.

(7) Only those compounds which have been prepared at NIH or which have been explicitly released for publication by the various pharmaceutical companies are included in this "public file."

(8) Corey, E. J., and W. T. Wipke, "Computer-Assisted Design of Complex Organic Syntheses," *Science* **166**, 178–92, 1969.

(9) Smith, E. G., "Wiswesser Line-Formula Chemical Notation," McGraw-Hill, New York, N. Y., 1968.