INSTITUTE FOR SCIENTIFIC INFORMATION
325 Chestnut Street
Philadelphia, Pa. 19106

SUB-STRUCTURE SEARCHING USING THE INDEX CHEMICUS REGISTRY SYSTEM

Eugene Garfield, Gabrielle S. Revesz, Hayes A. Dorr,
Maria M. Calderon, Andrea Warner

Presented at the 4th Middle Atlantic Regional Meeting,
American Chemical Society, Washington, D.C.
February 12-14, 1969

During the past 15 months, ISI has encoded approximately 200,000 new chemical compounds in Wiswesser Line Notation (WLN). This information is now being used in a service known as the Index Chemicus Registry System. The service is being regularly supplied, on a monthly basis, to a group of industrial, government and academic organizations in the form of printouts and magnetic tapes. The printout is a hierarchically arranged index in which classes of compounds are grouped to permit visual scanning for compounds containing specific sub-structure characteristics. The tapes are susceptible to both simple and complex searching, using a series of programs developed by ISI and others. Simple search programs such as one using the floating stem, permit location of code fragments embedded in WLN in various combinations. More complex programs involve conversion of the input notations to connectivity tables. The tapes are also used for SDI systems, employing "word" and other searching terms in addition to the WLN fragments. This paper will report on the use of the tapes by ICRS subscribers and on the status of more elaborate programs that will permit conversion of WLN notations to a subscriber's individualized fragment code.

SUB-STRUCTURE SEARCHING USING THE INDEX CHEMICUS REGISTRY SYSTEM

Eugene Garfield, Gabrielle S. Revesz, Hayes A. Dorr,
Maria M. Calderon, Andrea Warner

INSTITUTE FOR SCIENTIFIC INFORMATION
325 Chestnut Street
Philadelphia, Pa. 19106

The Index Chemicus Registry System is designed to solve the problem of providing chemists with current and retrospective chemical information reported in the Index Chemicus.

Since the Index Chemicus has been described elsewhere (1) we will not repeat here in detail the characteristics of that weekly publication. It is sufficient to state that IC provides detailed abstracts of journal articles which report new chemical compounds or chemical reactions.

On the other hand ICRS has not been described in the literature as yet and a brief description of its main characteristics is necessary to an understanding of our main topic -- searching for sub-structures, both currently and retrospectively, using the Index Chemicus Registry System.

ICRS consists essentially of four data files: 1) WLN Magnetic tapes, 2) IC Bibliographic Tape, 3) WLN Printouts, 4) Index Chemicus weekly issues.

The WLN Tapes (Slide #1) contain unique structural descriptions of all new compounds reported in the Index Chemicus. These descriptions are in the form of Wiswesser Line Notations. The WLN tapes also contain molecular formulas and IC registry numbers which identify an exact line in the numbered Index Chemicus abstract where a structural

diagram and other information is given.

The IC Bibliographic Tape (Slide #2) provides, in machine language, most of the data provided in the printed Index Chemicus. IC abstracts include molecular formulas, codes for new reactions and analytical instrumentation, subject-index terms assigned by chemists including terms related to the properties and biological activity of the compounds and their uses or use profiles. The bibliographic citation is also included in the IC tapes.

In Slide #3 you see a portion of the monthly WLN printout in which the information from the WLN tape has been alphabetized.

In Slide #4 you see a page from the corresponding IC abstract identified through the WLN printout.

The WLN has been extensively described elsewhere (2). It is sufficient to say that the WLN Tapes can be searched for both "parent" or generic structures or more specific sub-structures. While there are many instances, especially in SDI uses, in which the IC Bibliographic Tapes are used to augment sub-structure searches, the primary sub-structure searching capability is derived from the WLN encoding as reflected in the WLN Tapes. For example, a rather generic search could be conducted for all ortho-substituted aniline compounds. (Slide #5 shows several such compounds and WLN codes.)

It is important to note that many searches can be done by simply referring to the monthly or annual ICRS printouts. In Slide #6 a search for substituted adamantanes is shown.

Of equal importance is the use of the printouts to formulate machine-search questions. The printouts frequently enable the user to estimate the number of compounds which answer a given question. In this way he can determine how precise the question should be to receive a reasonable number of responses.

On the other hand the IC Bibliographic Tape can be used to further qualify the results of a WLN-Tape search. If a retrospective search of the WLN file produces a large number of candidate compounds, the IC Bibliographic file can be used to indicate those papers in which certain biological or other properties, activities, uses or analytical methods have been reported. The Index Chemicus itself can then be used for making the final selections from the candidates produced by the computer search. The most important selection criterion IC printed abstract provides is the structural diagram or the flow diagram of the reaction in which the compound occurs. As shown in Slide #7, IC's abstract usually also contains an author-prepared summary of the paper. Later on references will be made to programs for printing structural diagrams directly on the computer printer.

The WISWESSER LINE-FORMULA NOTATION (WLN)

To fully utilize the sub-structure search capabilities of ICRS, one must understand the Wiswesser Line Notation. In the time available today it is not possible to describe WLN in detail. However, as will be seen, it is not necessary for a chemist to master every detail of WLN in order to use the results of a search.

In Slide #8, you see the list of WLN symbols. As a further example of WLN, some substituted 3-pyrazolidones are shown in Slide #9. These are found in the printed version of the WLN tape under T5NNVTJ. Obviously any clerk can find this section of the printout once the term pyrazolidones is included in a search dictionary. Later on we intend to incorporate such common terms in the printout.

In a search of this kind, it is relatively easy to visually scan the remainder of each notation for substituents in the remaining available positions.

It is one thing to provide a machine readable file or tape of encoded compounds and subject descriptions--it is another thing to use that file to provide information to the ultimate user. To use such a file on a computer obviously requires programs or software--and this implies a system. The RADICAL System is the software component of the ICRS.

Retrieval and Automatic Dissemination of Information from Index Chemicus and Line Notations (RADICAL)

There have been many programs written for machine searching of a WLN file. For example, Hyde and Thompson (3) at ICI have described programs to convert WLN into connectivity tables. They have also written programs to convert line notations to structural diagrams. Fraction (4) at NBS and Kulpinsky of the University of Pennsylvania (5) have also written algorithms which convert WLN to connectivity tables. Barnard (6) of J. T. Baker & Co. and Finlay (7) of Pfizer Research Center developed programs for a "Permuted WLN" Index as did Gelberg, at al (8).

In Slide #8, you see the list of WLN symbols. As a further example of WLN, some substituted 3-pyrazolidones are shown in Slide #9. These are found in the printed version of the WLN tape under T5NNVTJ. Obviously any clerk can find this section of the printout once the term pyrazolidones is included in a search dictionary. Later on we intend to incorporate such common terms in the printout.

In a search of this kind, it is relatively easy to visually scan the remainder of each notation for substituents in the remaining available positions.

It is one thing to provide a machine readable file or tape of encoded compounds and subject descriptions--it is another thing to use that file to provide information to the ultimate user. To use such a file on a computer obviously requires programs or software--and this implies a system. The RADICAL System is the software component of the ICRS.

Retrieval and Automatic Dissemination of Information from Index Chemicus and Line Notations (RADICAL)

There have been many programs written for machine searching of a WLN file. For example, Hyde and Thompson (3) at ICI have described programs to convert WLN into connectivity tables. They have also written programs to convert line notations to structural diagrams. Fraction (4) at NBS and Kulpinsky of the University of Pennsylvania (5) have also written algorithms which convert WLN to connectivity tables. Barnard (6) of J. T. Baker & Co. and Finlay (7) of Pfizer Research Center developed programs for a "Permuted WLN" Index as did Gelberg, at al (8).

At ISI we have developed a software system which can be used in two modes--one is for ISI Source Tapes, a by-product of the Science Citation Index. We call that mode ISIS an acronym for ISI Search. The other mode is for ICRS Tapes. We call that mode RADIICAL--an acronym for Retrieval and Dissemination of Information by IC and Line-notations. ISIS and RADIICAL are completely compatible. The basic software is written for IBM 360/30 but is easily modified to a larger 360 system. RADIICAL uses Basic Assembly language and requires a 32K memory. I might add that at least one ICRS user plans to use the RADIICAL documentation and flow charts to write programs for another computer.

Formulation of the interest profiles for the RADIICAL system is quite flexible. Questions may comprise segments of names of authors, journal titles, (or portions thereof), subject indexing terms, words in article titles, portions of words or stems, word combinations, organizational or industrial names, etc. Search elements may be weighted. or combined in a multitude of logical (combinational) or syntactical relationships. Thus, one can specify the order in which certain search elements must occur, and the environment in which they must occur.

RADIICAL software consists of the following programs shown in the flow chart on Slide #10.

1. The question (profile) update program will delete, add, or change profile questions on the input tape. The resulting output tape is acceptable to the Search Program. It is in Account Number and Profile Number sequence.

2. The search program can search for:

(a) WORDS (one or more characters surrounded by a blank space).

(b) STRINGS (one or more characters in sequence) including spaces (such as two words). Strings include word phrases, floating stems, and segments of the notation.

(c) PREFIXES (initial stem) meaning a series of characters in sequence, but preceded by a space and not ending with a space.

(d) STEMS, a series of characters not preceded or followed by a space, nor containing a space.

The types of logic used with the RADICAL Search Program can include:

(a) AND logic

(b) OR logic (including exclusive or)

(c) NOT logic (indicating absolute NOT or weighted NOT)

(d) SAME WORD logic (looking for two Search Terms but only if in same word)

(e) FOLLOWED BY logic (two words, strings, stems or prefixes, but only if one follows the other, rather than precedes).

The search program results in a "hit" tape which must be sorted under control of the sort control card deck.

3. The Sort Control Program and Card Deck sorts the "hit" tape. The sort run arranges hits in order by account number. The sorted hit tape then is used with the printout program.

4. The Printout Program prints a separate report of one or more pages for each account number. Input to the Printout Program is the sorted "hit" tape, and an Account Number Address Tape. Account numbers with no hits during search runs are identified, and the statement "NO HITS FOR THIS ACCOUNT AND PROFILE" is printed. If for some reason there are hits without matching Account Numbers, the report prints these out without an address, under the message "NO MATCH IN ACCOUNT FILE FOR THE FOLLOWING HITS."

5. Card-to-Tape Program and Control Card Deck puts the Account Number Addresses on tape. This can be done with any card-to-tape program.

Users have a wide latitude for expressing their requirements. However, search results will depend directly upon the information specialist's knowledge of the file structure, his skill in designing profiles and familiarity with appropriate terminology. For sub-structure searches knowledge of WLN and chemistry is essential.

Neither software nor notation skill can replace imagination and ingenuity on the part of the user in designing intelligent questions.

To illustrate how the WLN can be used for sub-structure searching let us examine the following two slides.

Slide #11--A researcher interested in finding all phenothiazines with piperazine in the side chain will search for the string T C666 BN ISJ followed by the string T6N DM TJ or T6N DN TJ or T6N DK TJ.

Slide #12--A researcher interested in 17-ketosteroids will look for L E5 B666 followed by the word FQ.

These examples were aimed at illustrating the use of the system by suitably formulated search questions. Successful answers demand two prerequisites:

(a) the storage of as comprehensive a file of chemical compounds as possible,

(b) programs for interrogating and retrieving from the file.

The former is accomplished in ICRS by encoding new chemical compounds at an annual rate of over 150,000. The latter requires a continuing development of software. The foregoing descriptions indicate a beginning along these lines; it is hoped that more such programs will be developed in the future.

Returning to my opening remarks it may not be obvious that the ICRS RADICAL System permits one to provide sub-structure and other searches for continuous monthly SDI reports as well as retrospective one-time searches. Depending upon the amount of use made, each user may wish to make suitable modifications to take advantage of their own particular hardware configurations.

We mentioned before the problem of computer display of structural diagrams. In the time available it is not possible to detail what has been accomplished by others nor what ISI contemplates in its on-going development programs. Similarly, the use of connectivity tables for more sophisticated sub-structure searches and generation of fragmentation codes is under active study but we would prefer to report on these at the ACS Meeting in Minneapolis after we have completed some studies now in progress.

1) Revesz, G.S., and A. Warner, "An Important Tool for Chemical Information Processing and Retrieval," Presented at the 156th National ACS Meeting, Atlantic City, New Jersey, September, 1968.

2) a-William J. Wiswesser, "A Line-Formula Chemical Notation," Published by Thomas J. Crowell Company, New York, 1954.

b-Elbert G. Smith, "The Wiswesser Line-Formula Chemical Notation," published by McGraw Hill Book Company, New York, 1968.

3) Hyde, E., F. W. Matthews, L. H. Thomson and W. J. Wiswesser, "Conversion of Wiswesser Notations to a Connectivity Matrix for Organic Compounds," Journal of Chemical Documentation 7, 200 (1967).

4) Fraction, G.F., J.C. Walker, S.J. Tauber, "Connection Tables from Wiswesser Chemical Structure Notations--A Partial Algorithm," National Bureau of Standards Technical Note #432, issued September, 1968.

5) Kulpinski, S., N. London, D. Lefkovitz and A. Genarro, "A Study and Implementation of Mechanical Translation from Wiswesser Line Notation to Connection Table," Volume 1, October 25, 1967, 114pp. and Volume II, November 30, 1967, 44pp., Annual Reports to National Science Foundation on Contract #NSF C-467.

6) J. P. Barnard, Private Communication.

7) A. Finlay, Private Communication.

8) Granito, C.E., J.E. Schultz, G.W. Gibson, A. Gelberg, R.J. Williams and E.A. Metcalf, "Rapid Structure Searches Via Permuted Chemical Line-Notations (III): A Computer-Produced Index, Journal of Chemical Documentation 5, 229 (1965).

Slide #1

ICRS ®

ICRS TAPE Record

TAPE 1 PRINT FILE 1

```
          &
101318   1 03220 L66 B6 A B- C 1B ITJ B- ET5NVVOJ
                                        C 013H 015N 001D 003
          &
101318   2 04120 L66 B7 A B- C 1B J IXTJ H02 I-+ ET5MVVOXJ
                                        C 015H 021N 001D 004
          &
101318   3 02720 L66 B6 A B- C 1B ITJ BVMVVZ
                                        C 013H 018N 002D 003
          &
101318   4 02620 L66 B6 A B- C 1B ITJ BVNCO
                                        C 012H 015N 001D 002
          &
>

101318   5 02720 L66 B6 A B- C 1B ITJ BVMVMR
                                        C 018H 022N 002D 002
          &
101318   6 03025 L66 B6 A B- C 1B ITJ BYUM+OVVG
                                        C 013H 016CL001N 001D 003
          &
101318   7 03825 L66 B6 A B- C 1B ITJ B- ET5KVVOJ AR +G
                                        C 019H 020CL001N 001D 003
          &
101318   8 04420 L66 B7 A B- C 1B J IXTJ H02 I-+ ET5NVVOXJ AR
                                        C 021H 025N 001D 004
          &
101318   9 02725 L66 B6 A B- C 1B ITJ BVMVVG
                                        C 013H 016CL001N 001D 003
          &
>

101318 10 02925 L66 B6 A B- C 1B ITJ BVNR+VVG
                                        C 019H 020CL001N 001D 003
          &
101318 11 04325 L66 B7 A B- C 1B J IXTJ HG I-+ ET5NVVOXJ AR
                                        C 019H 020CL001N 001D 003
          &
101318 12 04025 L66 B7 A B- C 1B J IXTJ HG I-+ ET5MVVOXJ
                                        C 013H 016CL001N 001D 003
```

Slide #2

ICRS©

IC TAPE RECORD

TAPE 1 PRINT FILE 1

```
101318   ASASAKI T
101318   ATORU T
101318   AEGUCHI S
101318   GTETRAHEDRON LETTERS 1968(38),4135-8. RECD MAY 13, 1968.
101318   HINST APPL ORG CHEM, FAC ENGN, NAGOYA UNIV, CHIKUSA-KU, JAPAN.
101318   IGULX
101318   SSPIROHOMOADAMANTANES, REARR FROM ADAMANTYLOXAZOLINE
>

101318   SADAMANTANES, OXAZOLINYL-, REARR SPIROHOMOADAMANTANE
101318   SOXAZOLINES, ADAMANTYL-
101318   TA NOVEL REARRANGEMENT OF ADAMANTANE SKELETON TO SPIROHOMOADAMANTANE.
101318   TREACTION OF ADAMANTANE-1-CARBOXAMIDES WITH OXALYL CHLORIDE.
101318   TSYNTHESIS OF ADAMANTANE DERIVATIVES. V.
```

Slide # 3

ICRS®

INDEX CHEMICUS REGISTRY SYSTEM

WLN	ABSTR. NO.	CPD NO.	M.F.
L66ETJ CO1 GO H- AL5TJ	100154-	2	C16 H22 O2
L66ETJ CO1 GO H- AL5TJ		3	C16 H22 O2
L66ETJ CO1 G1 H- AL5TJ		18	C17 H24 O
L65ETJ CO1 G1 H- AL5TJ BO		21	C17 H24 O2
L66ETJ CO1 G1 H- AL5TJ CG1 CO1		29	C19 H26 O3
L66ETJ CO1 G1 H- AL5TJ CO		22	C17 H24 O2
L66ETJ CO1 G1 H- CL5UTJ		16	C17 H22 O
L66ETJ CO1 G1 H- CL5VTJ		27	C17 H22 O2
L66ETJ CO1 G2 H- AL5TJ		17	C18 H26 O
L66ETJ CO1 G2 H- AL5TJ BO		19	C18 H26 O2
L66ETJ CO1 G2 H- AL5TJ CO1 CO1		28	C20 H28 O3
L66ETJ CO1 G2 H- AL5TJ CO		20	C18 H26 O2
L66ETJ CO1 G2 H- AL5TJ CO		32	C18 H26 O2
L66ETJ CO1 G2 H- BL5VTJ		24	C18 H24 O2
L66ETJ CO1 G2 H- BL5YTJ AUNMVZ		26	C19 H27 N3 O2
L66ETJ CO1 G2 H- CL5UTJ		15	C18 H24 O
L66ETJ CO1 G2 H- CL5VTJ		23	C18 H24 O2
L66ETJ CO1 G2 H- CL5YTJ AUNMVZ		25	C19 H27 N3 O2
L66ETJ CO1 H- AL5TJ BOV1		4	C18 H24 O3
L66STJ CO1 H- AL5TJ BQ		6	C16 H22 O2
L66ETJ CO1 H- AL5TJ BQ		36	C16 H22 O2
L66ETJ CO1 H- AL5TJ COV1		5	C18 H24 O3
L66ETJ CO1 H- AL5TJ CO1 CO1		30	C18 H24 O3
L66ETJ CO1 H- AL5TJ CQ		7	C16 H22 O2
L66ETJ CO1 H- AL5TJ CQ		35	C16 H22 O2
L66ETJ CO1 H- BL5VTJ		9	C16 H20 O2
L66ETJ CO1 H- BL5XTJ A-E BT5OXOTJ .		31	C18 H24 O3
L66ETJ CO1 H- BL5YTJ AUNMVZ		10	C17 H23 N3 O2
L66ETJ CO1 H- CL5VTJ		8	C16 H20 O2
L66ETJ CQ H- CL5VTJ		49	C15 H18 O2
L66ETJ G G	100683-	6	C12 H16

Slide #4

ICRS®

100154

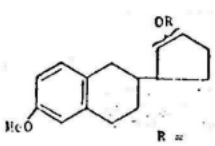
ANALOGS OF STEROID HORMONES. II.
6-(CYCLOPENTYL) DERIVATIVES OF 2-NAPHTHALENONE.
JUDAY R E, CUBBAGE L, MAZUR J, BUKWA B.
CHEM DEPT, UNIV MONTANA, MISSOULA, 59801.
J MED CHEM 11(4),872-5(1968). RECD DEC 23, 1967.

A number of 6-cyclopentyl derivatives of 2-naphthalenone have been prepared and tested for hormone and hormone antagonist activity. 3,4-Dihydro-2-(hydroxymethylene)-6-methoxy-1(2H)-naphthalenone was first alkylated with 3-bromocyclopentene. Further elaboration of the resulting cyclopentyl derivative was carried out by standard synthetic procedures to prepare the title compounds. None of the compounds showed significant activity when bioassayed for uterotropic, antiuterotropic, androgenic, antiandrogenic, and antigonadotropic activity.

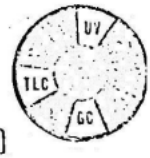
USE PROFILE

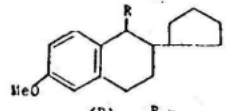
UTEROTROPIC ACTIVITY
ANTIUTEROTROPIC ACTIVITY
ANTIGONAPOTROPIC ACTIVITY
ANDROGENIC ACTIVITY
ANTIANDROGENIC ACTIVITY

1)	C16 H18 O2	(2)
2)	C16 H22 O2	(B3)
3)	C16 H22 O2	(B4)
4)	C18 H24 O3	(24A)
5)	C18 H24 O3	(24B)
6)	C16 H22 O2	(13A)
7)	C16 H22 O2	(13B)
8)	C16 H20 O2	(14)
9)	C16 H20 O2	(15)
10)	C17 H23 N3 O2	SEMICARBAZONE OF (15)
11)	C18 H22 O	(3)
12)	C18 H22 O	(3A)
13)	C17 H20 O	(4)
14)	C17 H20 O	(4A)
15)	C18 H24 O	(5)
16)	C17 H22 O	(6)
17)	C18 H26 O	(B1) (CIS & TRANS FORMS)
18)	C17 H24 O	(B2) "
19)	C18 H26 O2	(7A)
20)	C18 H26 O2	(7B)
21)	C17 H24 O2	(8A)
22)	C17 H24 O2	(8B)
23)	C18 H24 O2	(9)
24)	C18 H24 O2	(11)
25)	C19 H27 N3 O2	SEMICARBAZONE OF (9)

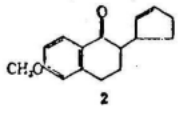


(24) A: 2-OAc
B: 3-OAc

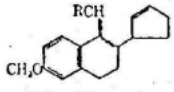




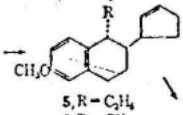
(B) R =
1: Et
2: Me
3: OH
4: OH



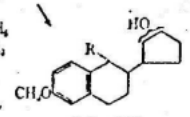
2



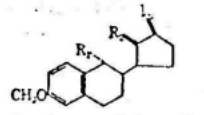
3, R = CH_3 (1-Et)
4, R = H (1-Me)
3A, R = CH_2 (1-Methylene)
4A, R = H (1-Ethylidene)



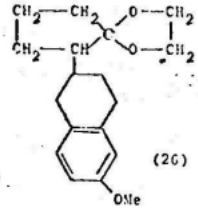
5, R = C_2H_5
6, R = CH_3



7, R = C_2H_5
8, R = CH_3
A: 2-Subst
B: 3-Subst

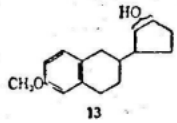


9, $R_1 = C_2H_5$; $R_2 = H_2$; $R_3 = O$
10, $R_1 = CH_2$; $R_2 = H_2$; $R_3 = O$
11, $R_1 = C_2H_5$; $R_3 = O$; $R_2 = H_2$
12, $R_1 = CH_2$; $R_2 = O$; $R_2 = H_2$

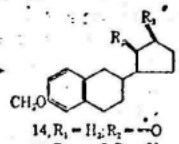


(26)

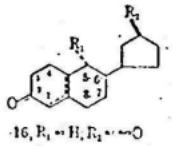
2 →



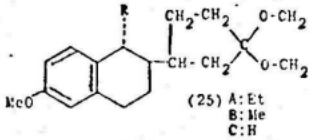
13
A: 2-Subst
B: 3-subst



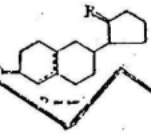
14, $R_1 = H_2$; $R_2 = O$
15, $R_1 = O$; $R_2 = H_2$



16, $R_1 = H$; $R_2 = O$
17, $R_1 = CH_2$; $R_2 = O$
18, $R_1 = C_2H_5$; $R_2 = O$
19, $R_1 = H$; $R_2 =$ OH, H
20, $R_1 = C_2H_5$; $R_2 =$ OH, OH



(25) A: Et
B: Me
C: H



Slide #4

ICRS®

100154

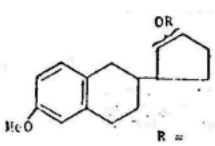
ANALOGS OF STEROID HORMONES. II.
6-(CYCLOPENTYL) DERIVATIVES OF 2-NAPHTHALENONE.
JUDAY R E, CUBBAGE L, MAZUR J, BUKWA B.
CHEM DEPT, UNIV MONTANA, MISSOULA, 59801.
J MED CHEM 11(4),872-5(1968). RECD DEC 23, 1967.

A number of 6-cyclopentyl derivatives of 2-naphthalenone have been prepared and tested for hormone and hormone antagonist activity. 3,4-Dihydro-2-(hydroxymethylene)-6-methoxy-1(2H)-naphthalenone was first alkylated with 3-bromocyclopentene. Further elaboration of the resulting cyclopentyl derivative was carried out by standard synthetic procedures to prepare the title compounds. None of the compounds showed significant activity when bioassayed for uterotropic, antiuterotropic, androgenic, antiandrogenic, and antigonadotropic activity.

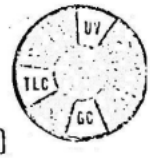
USE PROFILE

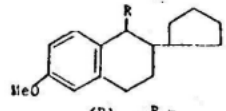
UTEROTROPIC ACTIVITY
ANTIUTEROTROPIC ACTIVITY
ANTIGONAPOTROPIC ACTIVITY
ANDROGENIC ACTIVITY
ANTIANDROGENIC ACTIVITY

1)	C16 H18 O2	(2)
2)	C16 H22 O2	(B3)
3)	C16 H22 O2	(B4)
4)	C18 H24 O3	(24A)
5)	C18 H24 O3	(24B)
6)	C16 H22 O2	(13A)
7)	C16 H22 O2	(13B)
8)	C16 H20 O2	(14)
9)	C16 H20 O2	(15)
10)	C17 H23 N3 O2	SEMICARBAZONE OF (15)
11)	C18 H22 O	(3)
12)	C18 H22 O	(3A)
13)	C17 H20 O	(4)
14)	C17 H20 O	(4A)
15)	C18 H24 O	(5)
16)	C17 H22 O	(6)
17)	C18 H26 O	(B1) (CIS & TRANS FORMS)
18)	C17 H24 O	(B2) "
19)	C18 H26 O2	(7A)
20)	C18 H26 O2	(7B)
21)	C17 H24 O2	(8A)
22)	C17 H24 O2	(8B)
23)	C18 H24 O2	(9)
24)	C18 H24 O2	(11)
25)	C19 H27 N3 O2	SEMICARBAZONE OF (9)

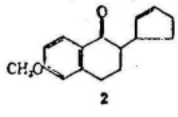


(24) A: 2-OAc
B: 3-OAc

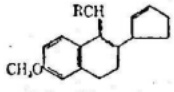




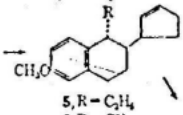
(B) R =
1: Et
2: Me
3: OH
4: OH



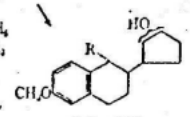
2



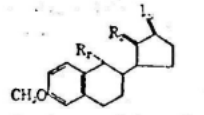
3, R = CH_3 (1-Et)
4, R = H (1-Me)
3A, R = CH_2 (1-Methylene)
4A, R = H (1-Ethylidene)



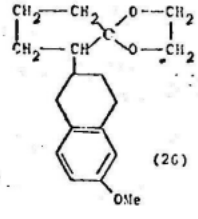
5, R = C_2H_5
6, R = CH_3



7, R = C_2H_5
8, R = CH_3
A: 2-Subst
B: 3-Subst

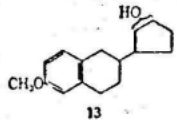


9, $R_1 = C_2H_5$; $R_2 = H_2$; $R_3 = O$
10, $R_1 = CH_2$; $R_2 = H_2$; $R_3 = O$
11, $R_1 = C_2H_5$; $R_3 = O$; $R_2 = H_2$
12, $R_1 = CH_2$; $R_2 = O$; $R_2 = H_2$

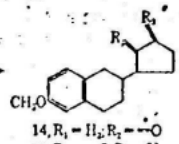


(26)

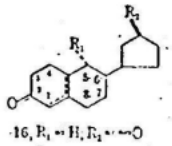
2 →



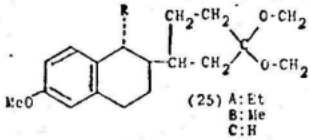
13
A: 2-Subst
B: 3-subst



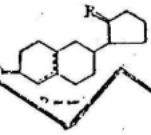
14, $R_1 = H_2$; $R_2 = O$
15, $R_1 = O$; $R_2 = H_2$



16, $R_1 = H$; $R_2 = O$
17, $R_1 = CH_2$; $R_2 = O$
18, $R_1 = C_2H_5$; $R_2 = O$
19, $R_1 = H$; $R_2 =$ OH, H
20, $R_1 = C_2H_5$; $R_2 =$ OH, OH

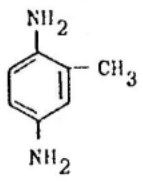


(25) A: Et
B: Me
C: H

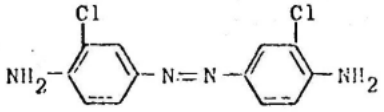


Vide 45

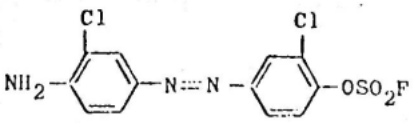
ICRS©



ZR DZ B



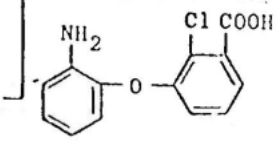
ZR BG DN 2U



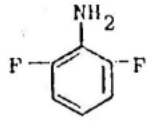
ZR BG DNUNR CG DOSWF

INDEX CHEMICUS REGISTRY SYSTEM

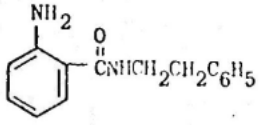
WLN	ABSTR. CPD NO.	M.F. NO.
ZR DZ B	100765- 7	C7 H10 N2
ZR BF FF	100143-30	C6 H5 F2 N
ZR BG DN 2U	100782- 6	C12 H10 CL2 N4
ZR BG DNUNR CG DOSWF	5	C12 H8 CL2 F N3 O3 S
ZR BOR BG CVQ	100140-27	C13 H10 CL N O3
ZR BVM2R	100131- 1	C15 H16 N2 O
ZR BVOVN1&1	101307- 1	C10 H12 N2 O3



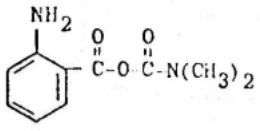
ZR BOR BG CVQ



ZR BF FF



ZR BVM2R



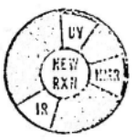
ZR BVOVN1&1

Slide #6

ICRS©

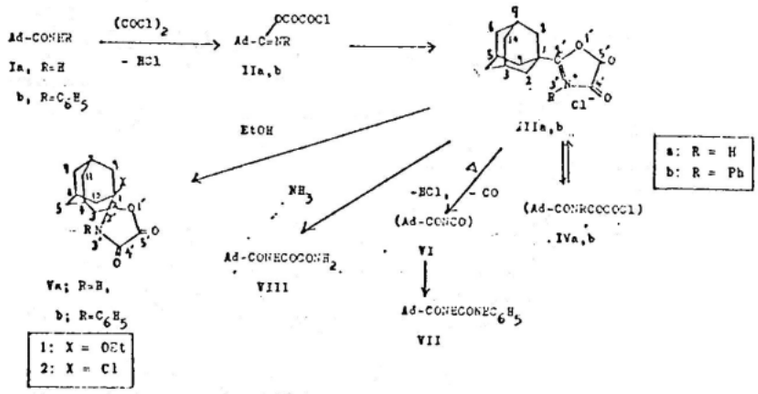
101318

SYNTHESIS OF ADAMANTANE DERIVATIVES. V.
REACTION OF ADAMANTANE-1-CARBOXAMIDES WITH OXALYL CHLORIDE;
A NOVEL REARRANGEMENT OF ADAMANTANE SKELETON TO SPIROHOMOADAMANTANE.
SASAKI T, EGUCHI S, TORU T.
INST APPL ORG CHEM, FAC ENGN, NAGOYA UNIV, CHIKUSA-KU, JAPAN.
TETRAHEDRON LETTERS 1968(38),4135-8. RECD MAY 13, 1968.



Ad = adamantyl

1)	C13 H15 N O3	(IIIA) (AS HCL)
2)	C15 H21 N O4	(VA-1)
3)	C13 H18 N2 O3	(VIII)
4)	C12 H15 N O2	(VI)
5)	C18 H22 N2 O2	(VII)
6)	C13 H16 CL N O3	(IIA)
7)	C19 H20 CL N O3	(IIIB)
8)	C21 H25 N O4	(VB-1)
9)	C13 H16 CL N O3	(IVA)
10)	C19 H20 CL N O3	(IVB)
11)	C19 H20 CL N O3	(VB-2)
12)	C13 H16 CL N O3	(VA-2)



INDEX CHEMICUS REGISTRY SYSTEM

WLN	ABSTR. NO.	CPD NO.	M.F.
L66 B6 A B- C 1B ITJ ANSUR DNVXFFF	100136-	5	C18 H21 F3 N2 O3 S
L66 B6 A B- C 1B ITJ ANVR DNW		1	C17 H20 N2 O3
L66 B6 A B- C 1B ITJ ANVR DZ		2	C17 H22 N2 O
L66 B6 A B- C 1B ITJ B- ETSKVVOJ AR &G	101318-	7	C19 H20 CL N O3
L66 B6 A B- C 1B ITJ B- ETSNVVOJ ..		1	C13 H15 N O3
L66 B6 A B- C 1B ITJ BNYZUNCN	100142-	11	C12 H18 N4
L66 B6 A B- C 1B ITJ BN2	100120-	74	C12 H21 N
L66 B6 A B- C 1B ITJ BN2&2Q		75	C14 H25 N O
L66 B6 A B- C 1B ITJ BVNVMR	101318-	5	C18 H22 N2 O2
L66 B6 A B- C 1B ITJ BVMVVG		9	C13 H16 CL N O3
L66 B6 A B- C 1B ITJ BVNVVZ		3	C13 H18 N2 O3
L66 B6 A B- C 1B ITJ BVNCO		4	C12 H15 N O2
L66 B6 A B- C 1B ITJ BVNR&VVG		10	C19 H20 CL N O3
L66 B6 A B- C 1B ITJ BYUN&OVVG		6	C13 H16 CL N O3
L66 B7 A B- C 1B J IXTJ NG 1-& ET5NVVOXJ		12	C13 H16 CL N O3
L66 B7 A B- C 1B J IXTJ NG I-& ET5NVVOXJ AR		11	C19 H20 CL N O3
L66 B7 A B- C 1B J IXTJ N02 I-& ... ET5NVVOXJ		2	C15 H21 N O4
L66 B7 A B- C 1B J IXTJ N02 I-& ET5NVVOXJ AR		8	C21 H25 N O4

Slide # 7

ICRS®

100051

CARBONYL PARTICIPATION IN THE SOLVOLYSIS OF KETONE DERIVATIVES. THE OBSERVATION AND ISOLATION OF INTERMEDIATES.

WARD H R, SHERMAN P D.
METCALF RES LAB, BROWN UNIV, PROVIDENCE, RI 02912.

J AM CHEM SOC 90(14),3812-7(1968). RECD JAN 17, 1968.

The solvolyses of γ- and δ-keto-*p*-bromobenzenesulfonates in formic or trifluoroacetic acid proceed with participation of the carbonyl through intermediates which have been directly observed by nmr. The rates of formation of these intermediate oxonium ions (1 and 13) have been measured for 4-(*p*-bromobenzenesulfonoxy)-butyrophenone (1.9×10^{-3} sec^{-1}, 30°) and 5-(*p*-bromobenzenesulfonoxy)valerophenone (4.9×10^{-4} sec^{-1}, 30°). The hexachloroantimonate salt of 1 has been isolated and found to be stable to 120°. The relative shifts in the nmr spectra of these intermediate ions compared to model compounds indicate that most of the positive charge resides on oxygen. Intermediates have not been observed for β- or ε-keto-*p*-bromobenzenesulfonates.

$HOCH_2(CH_2)_3$-Bz
(18)

$HOCH_2(CH_2)_4$-Bz

Slide #8

ICRS®

WLN SYMBOLS

All international atomic symbols except K, U, V, W, Y, Cl, and Br.

A generic alkyl	T heterocyclic ring
E bromine	U double bond
G chlorine	V carbonyl
J generic halogen	W nonlinear dioxo
K quaternary nitrogen	X quaternary carbon
L carbocyclic ring	Y tertiary carbon
M secondary nitrogen	Z amino or amido
N nitrogen atom, tertiary nitrogen	& punctuation mark
Q hydroxyl	- separator or connective
R benzene ring	/ multiplier stop

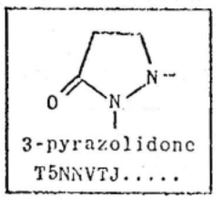
Numerals preceded by a space are multipliers of preceding notation symbols; or within ring signs L...J or T...J show the number of multicyclic points in the ring structure.

Numerals not preceded by a space show ring sizes if within the ring signs; elsewhere numerals show the length of internally saturated, unbranched alkyl chains and segments.

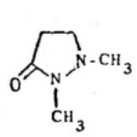
Letters following a space and hyphen are proposed as symbols with special meanings to denote stereoisomerism.

Slide # 9

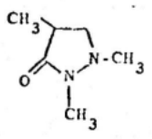
ICRS®



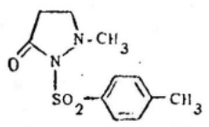
3-pyrazolidone
T5NNVTJ.....



....A B



....A B D



....A BSWR D

INDEX CHEMICUS REGISTRY SYSTEM

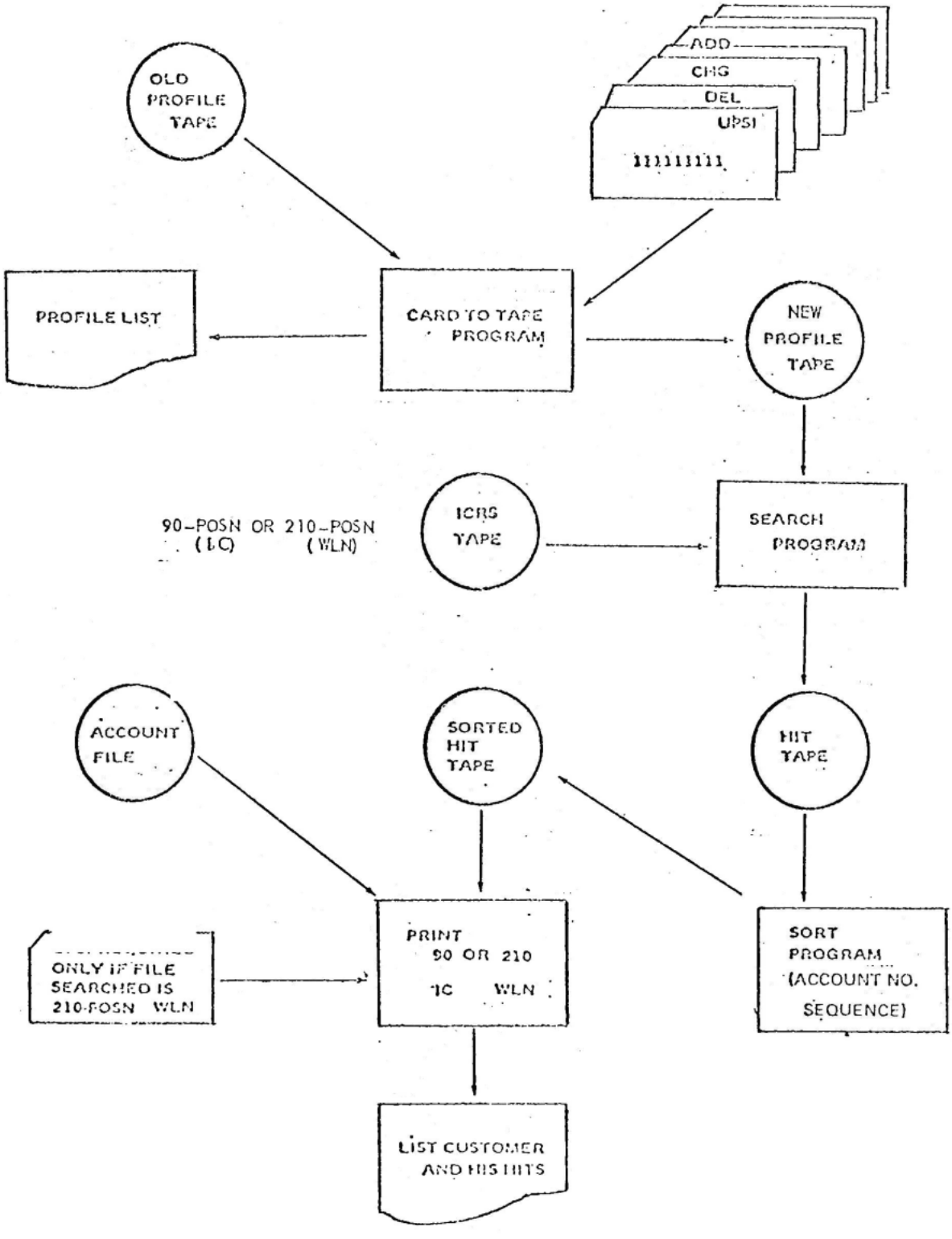
ULH	ABSTR. NO.	CPD NO.	N.F.
T5NNVTJ A B	99543-	9	C5 H10 N2 O
T5NNVTJ A B D		10	C6 H12 N2 O
T5NNVTJ A B E		17	C6 H12 N2 O
T5NNVTJ A BSWR D	99620-	3	C11 H14 N2 O3 S
T5NNVTJ ASWR D& B		2	C11 H14 N2 O3 S
T5NNVTJ AY BY	99543-	15	C9 H10 N2 O
T5NNVTJ AY BY D		16	C10 H20 N2 O
T5NNVTJ AY BY E		20	C10 H20 N2 O
T5NNVTJ A2 B2		11	C7 H14 N2 O
T5NNVTJ A2 B2 D		12	C8 H16 N2 O
T5NNVTJ A2 B2 E		18	C8 H16 N2 O
T5NNVTJ A3 B3		13	C9 H18 N2 O
T5NNVTJ A3 B3 D		14	C10 H20 N2 O
T5NNVTJ A3 B3 E		19	C10 H20 N2 O

Slide # 10

ICRS ©

ICRS RADICAL SOFTWARE

SYSTEM FLOW CHART



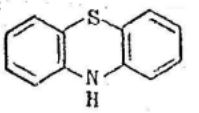
Slide #11

ICRS®

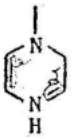
Query: All phenothiazines with piperazine in the side chain; that is, compounds encoded with the STRING T C666 BN ISJ, followed by the STRING T6M DNTJ, or T6N DNTJ, or T6K DNTJ.

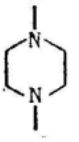
The search profile would be:

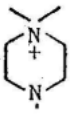
	Term	Type	Logic
1)	T C666 BN ISJ	STRING	AND
2)	T6M DNTJ	STRING	FOLLOWED BY
3)	T6N DNTJ	STRING	OR
4)	T6K DNTJ	STRING	OR



phenothiazine







Slide #12

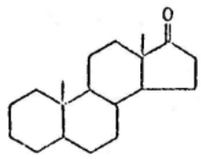
ICRS®

Query: All 17-keto steroids; that is, compounds whose WLN contains the STRING L E5 B666, followed by the WORD FV.

String

The profile would be:

Term	Type	Logic
1) L E5 B666	STRING	AND
2) FV	WORD STRING	FOLLOWED BY



WLN	ABSTR. NO.	CPD NO.
L E5 B666 FV JUTTTEJ B0 E L0 L0 0Q E23/H-3 E28/H-3 E31/H-3	100214-	4
L E5 B666 FV LUTJ A E GY01E01 0Q	100895-	2
L E5 B666 FV LUTJ A E GY01RE01RE 0Q		12
L E5 B666 FV LUTJ A E GY04E04 0Q ..		7
L E5 B666 FV LUTJ A E GY05E05 0Q		10
L E5 B666 FV LUTJ A E 0Q G- BT60 COTJ D D F		13
L E5 B666 FV OV NUTJ A E N01	100173-	2
L E5 B666 FVTJ A ECN	100807-	6
L E5 B666 FVTJ A EVZ		7
L E5 B666 FVTJ A I N N	100355-	8
L E5 B666 FVTTTEJ B0 E J0 KQ L0 L0 0DV1 E20/H-3 E31/H-3 E34/H-3	100214-	2
L E5 B666 FVTTTEJ COV1 E N	100166-	1
L E5 B666 FVTTTEJ E N1G 001 E231 ..	101412-	A
L E5 B666 FVTTTEJ E N1G 005 E231		B
L E5 B666 FVTTTEJ E N1N1E1 001		4
L E5 B666 FVTTTEJ E N1OV1 005		11
L E5 B666 FVTTTEJ E 001 N1- AT6NTJ		6
L E5 B666 FVTTTEJ E 001 P1- AT6NTJ		5
L E5 B666 FVTTTEJ E C01 P1G E232 ..		J
L E5 B666 FVTTTEJ E 001 P1N1E1		3
L E5 B666 FVTTTEJ E 001 01G E233		1
L E5 B666 FVTTTEJ E 005 N1- AT6N .. DOTJ		10
L E5 B666 FVTTTEJ E 005 N1- AT6NTJ		8
L E5 B666 FVTTTEJ E 005 P1- AT6N DOTJ		9
L E5 B666 FVTTTEJ E 005 P1- AT6NTJ		7
L E5 B666 FVTTTEJ E 005 P1G E232		K
L E5 B666 FVTTTEJ E 005 P1OV1		12
L E5 B666 FVTTTEJ E 005 01G E233 ..		2
L E5 B666 FVY LUTJ A E GU1OX 0Q	100895-	9
L E5 B666 FVY LUTJ A E GU1OX2 0Q		11

WLN	ABSTR. NO.	CPD NO.
L E5 B666 FVY LUTJ A E GU1OY 0Q ...		5
L E5 B666 FVY LUTJ A E GU1OY2 0Q		8
L E5 B666 FVY LUTJ A E GU1O1 0Q		1
L E5 B666 FVY LUTJ A E GU1O2 0Q ...		3
L E5 B666 FVY LUTJ A E GU1O3 0Q		4
L E5 B666 FVY LUTJ A E GU1O4 0Q		6
L E5 B666 FVY LUTJ A E GU1X 0DV1 ..	100510-	11
L E5 B666 FVY LUTJ A E GU1X 0Q		10
L E5 B666 FVY LUTJ A E GU1Y 0DV1		9
L E5 B666 FVY LUTJ A E GU1Y 0Q ...		8
L E5 B666 FVY LUTJ A E GU3 0DV1		6
L E5 B666 FVY LUTJ A E GU3 0Q		7
L E5 B666 FVY LUTJ A E 0DV1 GU1- ..		15
AL5TJ		
L E5 B666 FVY LUTJ A E 0DV1 GU1- AL6TJ		13
L E5 B666 FVY LUTJ A E 0DV1 GU1- BL5TJ A A D D		18
L E5 B666 FVY LUTJ A E 0DV1 GU1- .. DL55 ATJ C C		24
L E5 B666 FVY LUTJ A E 0DV1 GU1- FL E5 B666 LUTJ A E 0DV1		4
L E5 B666 FVY LUTJ A E CQ GU1- AL5TJ		14
L E5 B666 FVY LUTJ A E 0Q GU1- AL6TJ		12
L E5 B666 FVY LUTJ A E 00 GU1- BL5TJ A A D D		17
L E5 B666 FVY LUTJ A E 00 GU1- DL55 ATJ C C		19
L E5 B666 FVY LUTJ A E 00 GU1- FL . E5 B666 LUTJ A E 00		3
L E5 B666 FVY OV NUTJ A E GU1- FL E5 B665 OV NUTJ A E		5
L E5 B665 FVYTJ A E GU1Y 0DV1		21
L E5 B666 FVYTJ A E GU1Y 0Q		20