CONSTRUCTION OF LITERATURE DATABASES FOR CLATHRATE AND INCLUSION COMPOUNDS

Akira Yamasaki* and Toschitake Iwamoto#
*Laboratory of Applied Chemistry, The University of Electro-Communications, Chofu, Tokyo 182, Japan
#Department of Chemistry, College of Arts and Sciences, The University of Tokyo, Komaba, Meguro, Tokyo 153, Japan

ABSTRACT. An attempt to construct the literature databases for inclusion complexes since the beginning of this century is illustrated. The original literatures were cited and cross-checked by several secondary sources including Chemical Abstracts. The existing database systems appear to be insufficient to obtain necessary information efficiently in such a newly and rapidly developing field as inclusion chemistry. The trends in research revealed in the course of the database construction are analyzed and discussed for some important categories of inclusion chemistry.

1. INTRODUCTION

Powell proposed the term "clathrate" to describe a particular form of molecular compound in which one component formed a cage structure imprisoning the other in 1948 [1]. Before that many inclusion compounds had been prepared, sometimes by chance, and described with various names such as adduct, addition compound, complex, molecular compound, hydrate, solvate, etc., since the 19th century. Even after the proposal by Powell, confusion in classification and nomenclature has still remained in the field of inclusion chemistry, the field which has been demonstrating rapid and remarkable progress, especially in the recent decade. As much novel information has been accumulating, proposals for the classification and nomenclature of inclusion compounds and/or inclusion complexes have attracted a keen interest of the chemists working in this developing field; an example is the recent proposal by Weber and Josel [2]. Initially we attempted to build up a personal literature database in this field based on the existing retrieval systems. Although we have the huge machine-readable literature databases Chemical Abstracts Condensate (CACon) and Chemical Abstracts Search (CAS) for the chemical literatures of late, our major premise that the key words appropriate to the field should be available in the systems is not always satisfied. For example, if one refers to CAS Vol. 92 (1980) using the most popular key words "HOST" and "GUEST", only 14 citations are retrievable among more than 300 ones related to the inclusion

Journal of Inclusion Phenomena **2**, 423–429. 0167-7861/84.15.

complexes in this volume. Furthermore, even among the 14 citations a few "noises" concerned with liquid crystal display systems are included, although it may be an issue whether the binary mixtures should be included in the field of inclusion chemistry or not. In any case, this fact shows that the key words "HOST" and "GUEST" are quite insufficient to collect the related literatures in the existing systems. It is necessary to build up our own literature databases for inclusion chemistry, and the results are expected to act as a feedback to the existing system.

Our view-point is that the inclusion complex is a chemical species in which one component, host, provides one-, two-, or three-dimensional space of molecular dimensions with another, guest, and that the complex species is formed first by the geometrical fitness between the host and the guest; the molecular inclusion phenomenon is the process related to the formation of inclusion complex. From this view-point, we report on the attempt of the database construction; the trends in research revealed in the course of working are also analyzed and discussed.

2. DATABASE CONSTRUCTION

The construction of literature databases has been carried out after the format of the personal database system TOOL-IR-PDB/Orion, which was initially designed as the sub-system of the database management system TOOL-IR in the Computation Center of the University of Tokyo. This personalliterature database system has been applied to construct several chemical databases including those for solvent extraction, nuclear magnetic resonance spectroscopy, abstracts and preprints of some important domestic symposia held in Japan, etc. [3].

The format of this database system is briefly shown as follows:

[1] Serial number
[2] Title
[3] Author(s)
[4] Journal name with journal code (CODEN)
[5] Volume, issue, page, publication date (year)
[6] Key-word phrases
[7] Chem. Abstr. citation number

The classification of literature was carried out with the controlled key-word phrases denoting broader groupings of the host-species concerned such as hexa-host, Hofmann-type, quinol, urea adduct, Werner-type, etc. Crystal structure analysis, spectroscopic technique, and other physico-chemical methods applied were also included in the key-word phrases. In the original literature various terms have been used to express the group of inclusion complexes, and the selection appears to depend on the author's preference. Such confusion or controversy causes severe difficulties to retrieve important information from vast amount of chemical literatures. Therefore, we applied rather loosely controlled groupings as cited above. The citation number in the Chemical Abstracts (Chem. Abstr.) was added to ease the reference of the original items written in various unfamiliar languages.

Review articles and books [4-14] were helpful to collect the relevant literature ; the recently published book series [15] should be the materials for reinforcing the databases. These and other secondary sources including Chem. Abstr. were used for cross-checking. Although more than 6000 references collected have not yet been compiled completely into the databases, the data input and output are exemplified as follows:

TOOL-IR-PDB/Orion INPUT

```
 1:
 2:MEM:
 3:     1 #820110.02:'Novel host molecules derived from hexakis(benzyl
 4:     2 aminomethyl benzene: x-ray analysis of a dimethylformamide adduct
 5:     3 of hexakis(N-benzyltrifluoroacetamidomethyl)benzene complex of
 6:     4 hexakis(N-benzyl-(2-methoxyethoxy)acetamidomethyl)benzene.',
 7:     5 Andrew A. Freer, James H. Gall and David D. MacNicol,
 8:     6 J. Chem. Soc. Chem. Commun.(JCCCAT),(12), 674-6(1982).
 9:     7 @hexahost clathrates, X-ray structure analysis@
10:     8 @
11:     9 @CASNO 97.21.181609b@
12:>>
```

TOOL-IR-PDB/Orion OUTPUT

```
---------------------------------------------------------------------------
(       1)
SER#:820110     TYPE:02
AUTH:Freer, Andrew A. / Gall, James H. / MacNicol, David D.
TITL:Novel host molecules derived from hexakis(benzyl aminomethyl benzene: x-
    ray analysis of a dimethylformamide adduct of hexakis(N-benzyltrifluoroac
    etamidomethyl)benzene complex of hexakis(N-benzyl-(2-methoxyethoxy)acetam
    idomethyl)benzene.
HTTL:J. Chem. Soc. Chem. Commun.     CODN:JCCCAT     ISSU:12     PAGE:674-6
DPBL:1982
CIDX:hexahost clathrates, X-ray structure analysis     UCOD:CASNO 97.21.18160
    9b
```

OUTPUT from CAS

```
-------------------------------------------
(       1)
CA: 097/21/181609B   SE: CA122013   TY: J
JR: J. Chem. Soc., Chem. Commun.   IS: 12   PP: 674-6   YR: 82   CD: JCCCAT
    LA: Eng
AU: Freer, Andrew A. / Gall, James H. / MacNicol, David D.
LW: Univ. Glasgow   DV: Dep. Chem.   CI: Glasgow   PC: G12 8QQ   NA: UK
TI: Novel host molecules derived from hexakis(benzylaminomethyl)benzene:
    x-ray analysis of a dimethylformamide adduct of
    hexakis(N-benzyltrifluoroacetamidomethyl)benzene and the 3 KSCN.H2O
    complex of hexakis(N-benzyl-(2-methoxyethoxy)acetamidomethyl)benzene
KW: benzylacylaminomethylbenzene hexahost inclusion compd; inclusion compd
    hexahost hexakisbenzylacylaminomethylbenzene; crystal structure
    hexaamide DMF adduct; thiocyanate adduct hexaamide crystal structure
```

For the sake of comparison the output of the same reference from CAS is listed. The TOOL-IR-PDB/Orion output can be retrieved from the database using any key words or key-word phrases in the record including author name, and journal name as CODEN.

3. KEY WORD PROBLEMS IN THE EXISTING SYSTEMS

As mentioned in 1., the key words "HOST" and "GUEST" are not effective to retrieve the appropriate references in Chem. Abstr., although "host-guest complex" may be an alternative to "inclusion complex." The key-word phrase "INCLUSION COMPOUND" had appeared effective, since the phrase has been adopted as a heading in the General Subject Index (GSI) of Chem. Abstr. However, only a part of the references could be collected by manual retrieval under the heading in the hard copies in comparison with the results of machine retrieval using a combined questionnaire "HOUSETSU" being comprised of an ".OR." (OR) set (sum) of "CLATHRATE", "INTERCALATE", and "INCLUSION"; each of these three was also an OR set to cover the relevant categories of inclusion complexes. In the machine retrieval, those cited by "HOUSETSU" were scarcely overlapped with those by "MACROCYCLES" comprised of an OR set of "CROWN ETHER", "CRYPTAND", etc. Figure 1 shows the annual distribution of abstracts on inclusion complexes based on the results of machine retrieval from CACon and CAS cross-checked by manual retrieval in the hard copies. In Chem. Abstr. Vol. 95 (1982), for example, the totally available 689 citations were the sum of the OR set among "HOUSETSU" (374), "CYCLODEXTRIN" (94), and "MACROCYCLES" (250). The arithmetical sum of each set scored 718 was larger only by 29 than 689, whereas in "HOUSETSU" the arithmetical sum of 447 was larger by 73 than 374. This fact shows that overlap of the citations in "HOUSETSU" is considerable but that overlap in total is very small. At the present stage, macrocyclic chemistry is classified outside of inclusion chemistry so far as the retrieval system in Chem. Abstr. is concerned. We have no objection against that the Chemical Abstracts are the most reliable and world-largest secondary information sources in the chemical literatures. We say that it is rather difficult to collect satisfactorily the references of our concern either by manual retrieval in GSI or by machine retrieval in CACon and CAS unless we are extremely careful in selection of headings or key words to be referred.

4. TRENDS IN RESEARCH

The rapid growth of the citation numbers in the field of inclusion chemistry is clearly seen in Fig. 1 in recent years. Since the total number of chemical references in each volume of Chem. Abstr. has not changed on the whole in these years, the trend of growth is not only absolute but also relative to other fields of chemistry. The trend is common to the categories we have tentatively assigned except that grouped under "CLATHRATE" with a wavy curve of small amplitude in Fig. 1. That grouped under "INCLUSION" shows six times increase from 1977 to

1983 in the annual rate, that under "INTERCALATE" is placed at the second with the growth factor 5.5, and that under "CYCLODEXTRIN" is the third with the factor 3.6. The increase of the literature related to applications, including patents, contributes to the growth of citation numbers in these areas.

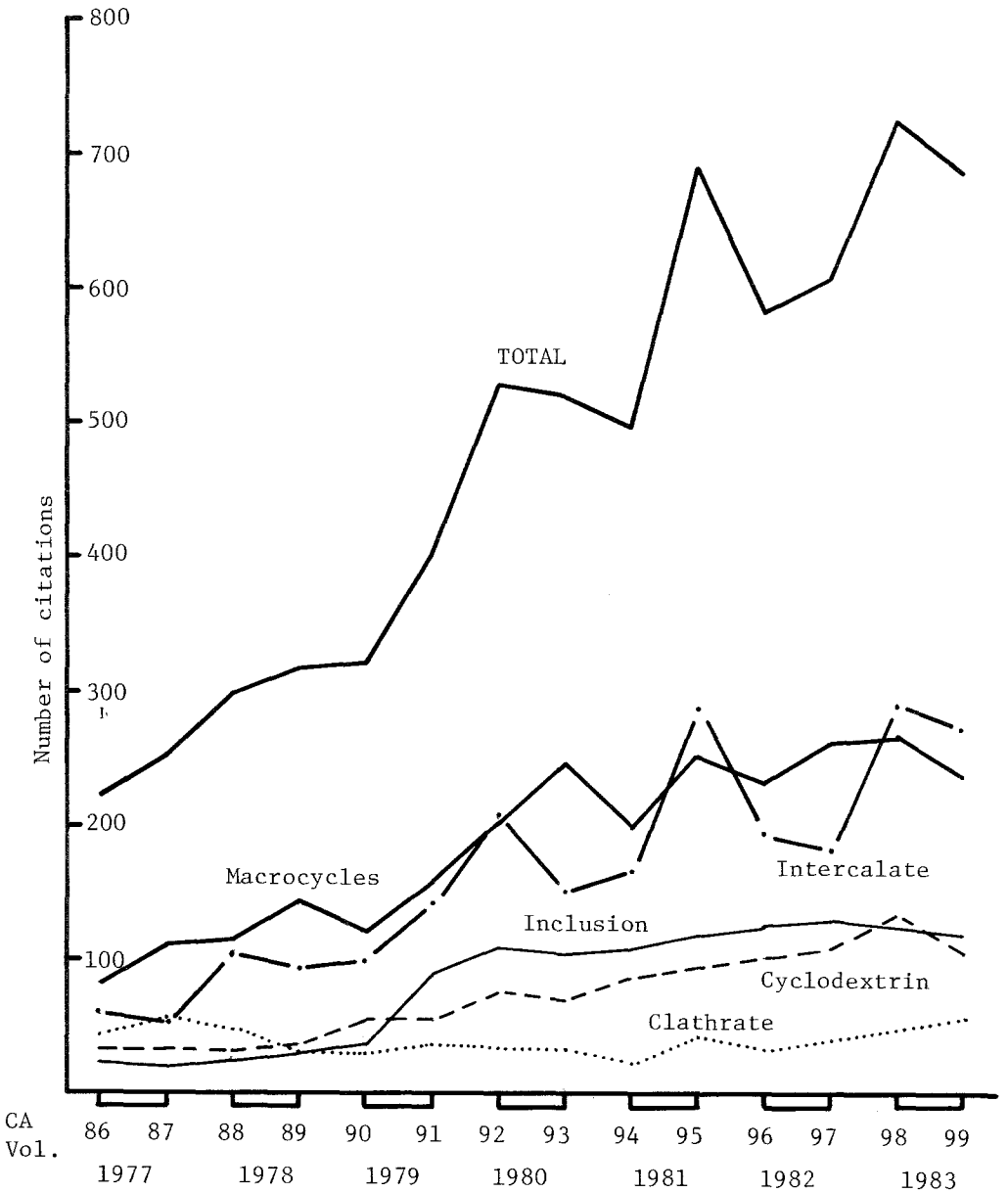

Fig. 1. The annual distribution of abstracts on inclusion complexes in total and classified items.

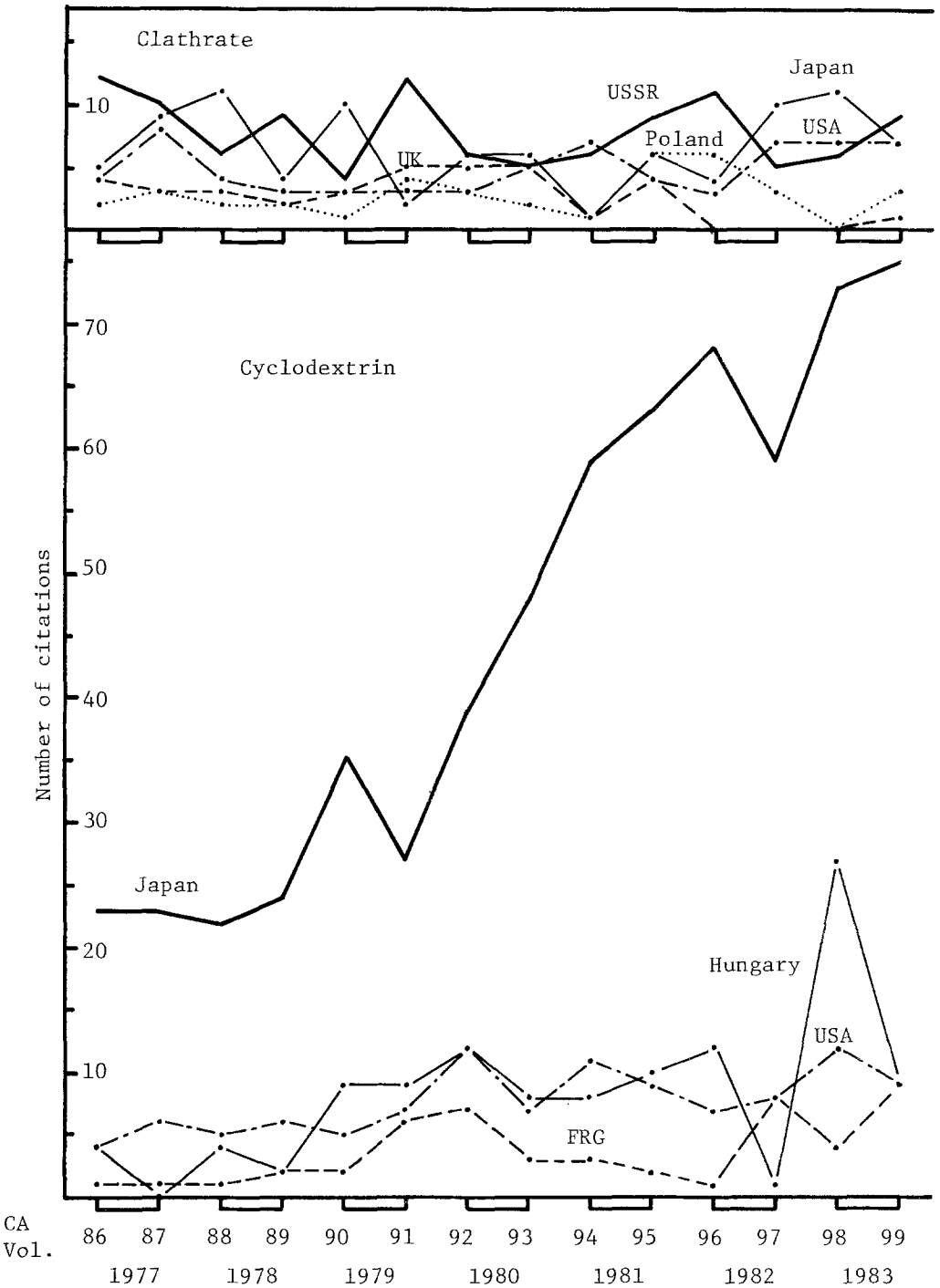

Fig. 2. The annual distribution of abstracts on "CLATHRATE" and "CYCLODEXTRIN" classified by the country of the first author's institution.

In Fig. 2, the numbers of citations are classified according to the country where the institution of the first author is affiliated for the categories under "CLATHRATE" and "CYCLODEXTRIN." As shown in Fig. 1, there is a distinguishable difference in trend between the two catogories: the trend in that under "CLATHRATE" shows a steady state of top-five countries in these years. It means that there are certain groups of researchers who like to put the key word "CLATHRATE" on their papers in these countries. On the other hand, for "CYCLODEXTRIN" the contribution from a country, Japan, shows an extremely rapid growth and big share. Although there are a number of active research groups in Japan, one reason is due to the increase of patents claiming industrial applications of cyclodextrin. Another is probably due to the preference of Japanese chemists that cyclodextrin reagents are available at relatively cheap prices in Japan.

ACKNOWLEDGMENT

One of the authors, T.I., acknowledges the support to this work from the Grant-In-Aid for the Special Project Research on the Properties of Molecular Assemblies (No. 58118007) from the Ministry of Education, Science and Culture.

REFERENCES

1. H.M.Powell: J.Chem.Soc., 61 (1948).
2. E.Weber and H.-P.Josel: J.Inclusion Phenom., 1, 79 (1983).
3. A.Yamasaki: Anal.Chim.Acta, 133, 741 (1982).
4. F.Cramer:'Einschlussverbindungen', Springer (1954).
5. L.Mandelcorn, ed.:'Non-stoichiometric Compounds', Academic Press (1964).
6. K.Takemoto:'Housetsu-Kagoubutsu no Kagaku' (Chemistry of Inclusion Compounds), Tokyo Kagaku Dojin (1969).
7. M.L.Bender and M.Komiyama:'Cyclodextrin Chemistry', Springer (1978).
8. M.Hiraoka:'Kuraun Kagoubutsu' (Crown Compounds), Kodansha (1978).
9. The Chemical Society of Japan, ed.:'Bunshi-Shugotai' (Molecular Assemblies), Gakkai Shuppan Senta (1980).
10. F.Vögtle, ed.: 'Host Guest Complex Chemistry I', Springer (1981).
11. T.Osa, ed.: 'Hosuto-Gesto no Kagaku' (Host-Guest Chemistry), Kyoritsu Shuppan (1979).
12. K.Takemoto:'Tsumiki no Kagaku' (Chemistry of Building Blocks), Kodansha (1981).
13. J.E.D.Davies, W.Kemula, H.M.Powell, and N.O.Smith: J.Inclusion Phenom., 1, 3 (1983).
14. M.Hiraoka, H.Yanagida, M.Obara, and K.Koga:'Hosuto-Gesuto Kemisutorii' (Host-Guest Chemistry), Kodansha (1984).
15. J.L.Atwood, J.E.D.Davies, and D.D.MacNicol, ed.:'Inclusion Compounds, Vols. 1-3', Academic Press (1984).