Garfield, E. "Online retrieval in the transition from bibliographic to encyclopedic information," Increasing Productivity Through Library Automation (Helal, A.H., Weiss, J.W., eds.) Essen: Gesamthochschulbibliothek, 5 p.27-60, 1983. - Book Chapter

ONLINE RETRIEVAL IN THE TRANSITION FROM BIBLIOGRAPHIC TO ENCYCLOPEDIC INFORMATION

E U G E N E G A R F I E L D

INSTITUTE FOR SCIENTIFIC INFORMATION

PHILADELPHIA, PENNSYLVANIA

Online Retrieval in the Transition from Bibliographic to Encyclopedic Information

Eugene Garfield, Ph.D.
Institute for Scientific Information®
Philadelphia, Pennsylvania, U.S.A.

Although it's widely known that the printing press was invented in Germany, most people don't realize this country published some of the earliest scientific journals. In 1670 the Collegium Naturae Curiosorum, one of the oldest societies in Germany, began publishing Miscellanea Curiosa sive Ephemeridum Medico Physicorum Germanorum. Shortly after this medical journal came out, the Collegium Gellianum followed suit with one of the first review journals -- called Acta Eruditorum. These journals focused world attention on work being done in Germany. They also informed German scholars of research going on beyond their own borders.[1]

The editors of Miscellanea and Acta Eruditorum, like most editors of the period, were desperate for manuscripts. Papers often were rejected out of deference to censors, and editors were frequently obliged to print inferior material or to publish their own papers. For example, the first volume of Miscellanea included 14 articles by the editor.[2] Clearly, 300 years ago, anyone who abided by the censors' restrictions had a much easier time getting his work into print than do scientists today. Now, more than one million researchers around the world are producing a constant stream of literature in thousands of scientific and technical journals. Even though so many people manage to get into print, there is a lot of competition to do so. This means that editors must be highly selective in determining which papers to publish in the primary journals. Researchers, if

they're not to become overwhelmed by this information, must identify which of these thousands of papers are worth reading. This is a difficult task -- but one made much easier through secondary information retrieval organizations such as the Institute for Scientific Information® (ISI®).

ISI is a US-based firm that has been providing scholarly information retrieval services since the 1950s. Our earliest and best-known products are Current Contents® (CC®) and the Science Citation Index® (SCI®). These services were introduced more than 20 years ago to help scholars locate, out of the thousands available, those papers relevant to their research projects. In those days, we were working toward "bibliographic control". Today I think it is safe to say that, in most cases, we can obtain reasonably complete bibliographies on almost any topic. In fact, it is often easy to become overloaded with information. To cope with this problem, ISI is developing information services that go beyond the provision of mere bibliographies.

Suppose you need information about a specialty you're not very familiar with. Today, you may have to be satisfied just to get a list of current papers. But in the future, you will also have at your fingertips a summary of how that field developed -- an instant review which identifies the milestone papers in that field. You will also retrieve

information that can tell you how that field is related to your own, and to others. I call this type of information "encyclopedic information".

ISI is currently making a transition to encyclopedism through a series of online data bases. Since these information services are based on citation indexing, I will briefly reiterate the simple principle behind it. Almost all papers contain references, or citations. These cited publications support, illustrate, elaborate, or provide precedent for the author's arguments. Each one of these cited references symbolizes a particular subject the author is discussing in that sentence, paragraph, or in the main theme of the paper. I might point out that while there is a simple mechanical relationship between the cited and citing publications, an enormous literature has developed around the exact nature of this relationship.[3]

The SCI was ISI's first product in which citation indexing was used as a classification scheme. It is actually a system of three separate indexes. The Citation Index is used to find out who has cited a specific author or paper. For example, if you want to obtain the latest information on the topic of a paper Lottspeich published in 1977 in Hoppe-Seylers Zeitschrift fur Physiologische Chemie, you would turn to his name in the 1981 annual Citation Index, shown in Figure 1. There you'd find all authors who had cited Lottspeich during 1981.

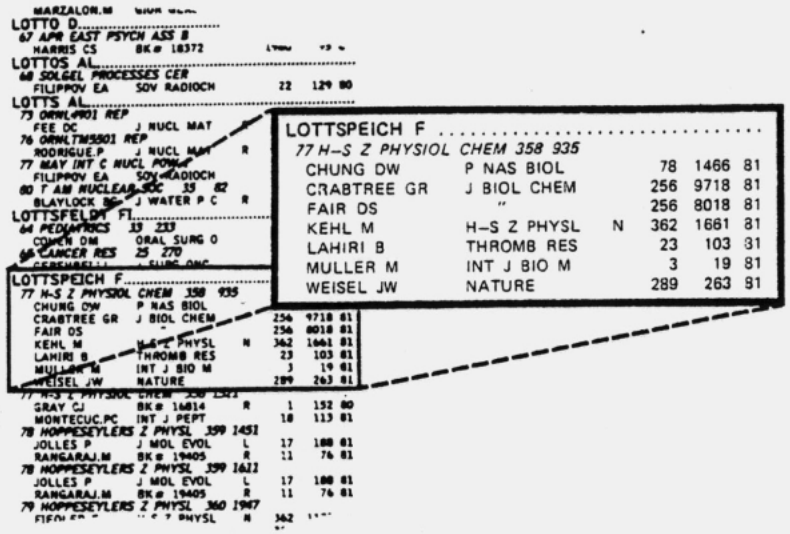


Figure 1: 1981 *Science Citation Index®* Citation Index entry.

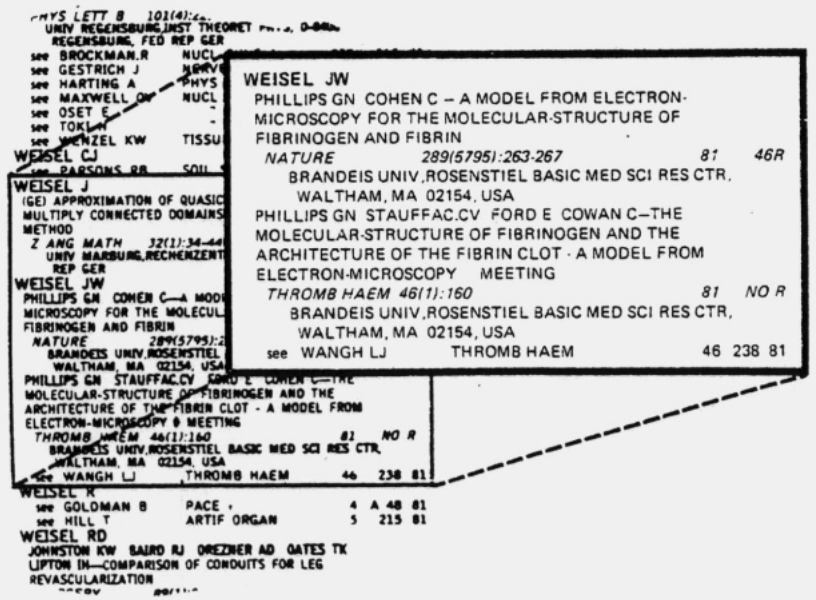


Figure 2: 1981 *Science Citation Index®* Source Index entry.

To obtain the title and other bibliographic information on each of the citing papers, you'd turn to the Source Index. This section of the SCI tells you what each author has published during an indexing period. Figure 2 is the 1981 Source Index entry for J.W. Weisel of Brandeis University, one of the authors who cited Lottspeich's 1977 paper. Under Weisel's name are listed all the papers he published in 1981 as first author. You're also referred to other articles he published as a coauthor during the same period.

When beginning a literature search, you often don't know or remember the author of a key paper. For this reason, the SCI also includes a component called the Permuterm® Subject Index (PSI). In this section, every significant word in every title covered by the SCI is permuted -- that is, paired -- with every other significant word in the title. To look for a paper on the molecular structure of fibrinogen, you could look up "fibrinogen" in the PSI and then glance through the co-terms until you found the word, "molecular" (Figure 3). There you would find the names of all the relevant authors, including Dr. Weisel. You would use the Source Index for the full title as with the earlier search in the Citation Index.

The SCI is issued bimonthly, and in annual and five-year cumulations. In the near future, we will also publish a ten-year cumulation, covering 1955 to 1964. When this is completed, we will have covered more than 25 years of the

FIBRINOGEN

MODIFICATL - NICKERSOJM
MOIETIES - - - NISHIBE H
MOLECULAR - CRABTREE GR
- FOWLER WE
GLADNER JA
WEISEL JW@ ←
MOLECULE - - ZIMA VL
MONITORING - BREHM R
MONOCHLORL - KNIGHT LC
MONOLAYER- - TOWNSEND RR

Figure 3: *Science Citation Index®, Permuterm®* Subject Index entry.

most important journals of science.

Since the print version of SCI comes out every two months, it is possible to update searches quite often. However, before the next annual index can be printed, it is more convenient to use the online version of the SCI for certain types of searches. That is why we were one of the first data bases to be made available through several data base vendors in the US. In Germany, this file is mounted at Deutsches Institut for Medizinische Dokumentation und Information (DIMDI) and is called MultiSci.

In Figure 4, we've done a search through MultiSci, called SCISEARCH® in the US, for papers published by authors at the Max Planck Institute of Biochemistry. We began by asking for the number of papers in the data base whose authors list Max Planck Institute of Biochemistry as their address. The computer tell us 406 papers in the data base include an

Request for number of documents that list Max Planck Institute for Biochemistry as address

```
? SS CS = MAX(W)CS = PLANCK(W)CS = INST(W)CS = BIOCHEM
          1   406 CS = MAX(W)CS = PLANCK(W)CS = INST(W)CS = BIOCHEM
```

Max Planck Institute for Biochemistry appears as an address in 406 papers

Command to print five of the papers that include an address from Max Planck Institute for Biochemistry

```
? T1/3/1-5

1/3/1
0989087  OATS ORDER#: PB460  35 REFS
  DISTURBANCES OF SELECTED PLASMA-PROTEINS IN HYPERDYNAMIC SEPTIC SHOCK
(ENGLISH)
  WITTE J; JOCHUM M; SCHERER R; SCHRAMM W; HOCHSTRASSER K; FRITZ H
  UNIV MUNICH, KLINIKUM GROSSHADERN, DEPT SURG, D-8000 MUNICH 70, FED REP GER/
  UNIV MUNICH, DEPT INTERNAL MED, D-8000 MUNICH 2/ FED REP GER/; UNIV
  MUNICH, INST CLIN CHEM & CLIN BIOCHEM, D-8000 MUNICH 2/ FED REP GER/; MAX PLANCK
  INST BIOCHEM, DEPT EXPTL MED D-8033 MARTINSRIED, FED REP GER/; UNIV MUNICH,
  KLINIKUM GROSSHADERN, DEPT EAR NOSE & THROAT DIS, D-8000 MUNICH 70/ FED REP GER/
  INTENSIVE CARE MEDICINE, V8, N5, P215-222, 1982

1/3/2
0985987  OATS ORDER#: PB594  22 REFS
  DERIVATIZATION OF POLYACRYLAMIDE GELS WITH PALLIDIN, A LECTIN OF THE CELLULAR
SLIME-MOLD POLYSPHONDYLIUM-PALLIDUM (ENGLISH)
  BOZZARD S
  MAX PLANCK INST BIOCHEM, D-8033 MARTINSRIED, FED REP GER/
  ITALIAN JOURNAL OF BIOCHEMISTRY, V31, N3, P173-182, 1982

1/3/3
0979147  OATS ORDER#: PA929  19 REFS
  SINGLE CHANNEL GATING EVENTS IN TRACER FLUX EXPERIMENTS .3. ACETYLCHOLINE
RECEPTOR-CONTROLLED LI+ EFFLUX FROM SEALED TORPEDO-MARMORATA MEMBRANE-
FRAGMENTS (ENGLISH)
  BERNHARDT J; NEUMANN E
  MAX PLANCK INST BIOCHEM, D-8033 MARTINSRIED, FED REP GER/
  BIOPHYSICAL CHEMISTRY, V15, N4, P327-341, 1982

1/3/4
0979146  OATS ORDER#: PA929  10 REFS
  SINGLE CHANNEL GATING EVENTS IN TRACER FLUX EXPERIMENTS .2. FLUX AMPLITUDE
ANALYSIS (ENGLISH)
  BERNHARDT J; NEUMANN E
  MAX PLANCK INST BIOCHEM, D-8033 MARTINSRIED, FED REP GER/
  BIOPHYSICAL CHEMISTRY, V15, N4, P317-325, 1982

1/3/5
0976948  OATS ORDER#: PA339  0 REFS
  RELATIONSHIP BETWEEN PRIMARY STRUCTURES AND ALLERGENICITY OF ASTHMA-
INDUCING INSECT PROTEINS (CHIRONOMID HEMOGLOBINS) (ENGLISH)
  BAUR X; ASCHAUER H; DEWAIR M; FRUHMANN G
  UNIV MUNICH, D-8000 MUNICH 2, FED REP GER ; MAX PLANCK INST BIOCHEM, D-8033
  MARTINSRIED, FED REP GER/
  CHEST, V82, N2, P254-254, 1982
```

Figure 4: Institutional search through MultiSci.

address at the institute. In this example, we've had the computer print out five of them. Since they're usually printed out in a "last in, first out" sequence, you see the most recent papers first.

SCI was developed as an information retrieval system, and its presence in more than 1,100 major libraries throughout the world attests to its value. However, even before the SCI

NAME	INSTITUTIONAL AFFILIATION	TOTAL CITATIONS 1965-1978
DRUCKREY H	MAX PLANCK INSTITUT FUR IMMUNBIOLOGIE	2597
LUDERITZ O	MAX PLANCK INSTITUT FUR IMMUNBIOLOGIE	2980
WESTPHAL OH	MAX PLANCK INSTITUT FUR IMMUNBIOLOGIE	3129
WEBER K	MAX PLANCK INSTITUT FUR BIOPHYSIKALISCHE CHEMIE	13,427
OSBORN M	MAX PLANCK INSTITUT FUR BIOPHYSIKALISCHE CHEMIE	10,376
THOENEN HF	MAX PLANCK INSTITUT FUR BIOCHEMIE	4506
KLEIN J	MAX PLANCK INSTITUT FUR BIOLOGIE	3677
CARDONA M	MAX PLANCK INSTITUT FUR FESTKORPERFORSCHUNG	3190
WITTMAN HG	MAX PLANCK INSTITUT FUR MOLEKULARE GENETIK	2776
FRANKE WW	GERMAN CANCER RESEARCH CENTER	3031
IVANKOVIC S	GERMAN CANCER RESEARCH CENTER	2485
PREUSSMANN R	GERMAN CANCER RESEARCH CENTER	2692
HUISGEN R	MUNICH UNIVERSITY	5087
KLINGENBERG ME	MUNICH UNIVERSITY	2548
FISCHER EO	TECHNICAL UNIVERSITY OF MUNICH	3472
SCHMIDBAUR H	TECHNICAL UNIVERSITY OF MUNICH	2576
SCHLEYER PV	ERLANGEN-NUREMBERG UNIVERSITY	5736
BOHLMANN F	TECHNICAL UNIVERSITY OF BERLIN	2640
REMMER H	TUBINGEN UNIVERSITY	3013
MULLER A	UNIVERSITY OF BIELEFELD	3508
GREINER W	UNIVERSITY OF FRANKFURT	3299

Table 1: Scientists from the Federal Republic of Germany who appear on ISI®'s 1,000 most-cited authors list.

came out, sociologists and historians of science realized that citation indexing had another role to play. Since highly cited papers are often the most important papers of science, they found citation analysis could be used to identify significant papers, journals, and even individuals within a field. ISI has done a number of studies using citation analysis. In one such study, of the 1000 authors most cited from 1965 to 1978, we identified 21 authors from the Federal Republic of Germany.[4] They are listed in Table 1.

Since a group of frequently cited papers generally includes the more significant papers of science, we can also identify the most active areas of science by identifying groups of highly cited papers. The fuller realization that citation indexing could reveal "fields of knowledge" or emerging new specialty areas brought ISI a step closer to the era of encyclopedism. We now knew it was possible to

define the literature, and individuals, who had played a significant role in the development of the specialty areas that comprise scholarly research. In fact, we had a system for automatic classification.

After some initial work at ISI, we realized that citation indexing could not only be used to identify highly specialized research areas, but also to create historical maps. Later, historian Derek J. de Solla Price of Yale University proposed that we use citation analysis to systematically diagram the structure of science. He believed that by studying citation relationships among documents, we could view the structure of science "in which the parts of science are conceived as mapped like a territory."[5]

In 1974, Henry Small of ISI, and Belver Griffith of Drexel University, found a way to create this map.[6] They discovered that papers could be "clustered" automatically through co-citation relationships. Each cluster they identified was a group of highly cited, closely related papers -- in effect, the core literature of a given specialty. Taken together, these clusters -- through their relationships with one another -- could be used to study the structure of science.

The map shown in Figure 5 represents the structure of biomedical research in 1973. Each of the boxes on this

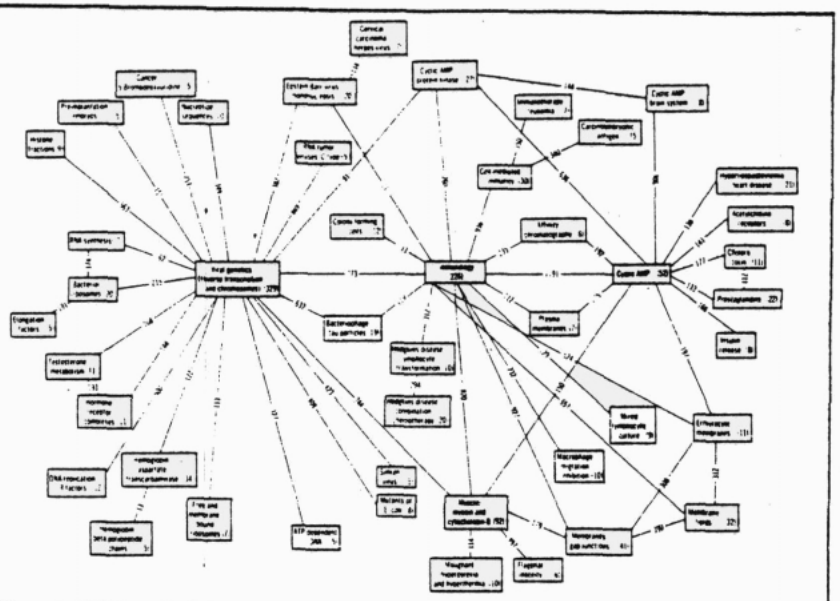


Figure 5: 1973 biomedical clusters.

map represents a cluster -- or highly specialized field of inquiry. By studying the highly cited, or milestone, documents in this cluster and the more recent papers that cite them, we can describe the specialties they represent. From this information, we can write comprehensive, encyclopedic reviews of each of the specialty areas identified through our clustering process.[7] Through clustering, we also can visually observe the relationships between the specialty areas.

We begin the clustering process by identifying the group of journals that are most important to scientists publishing in a broad subject field. This could be all of science or a specific field such as biomedicine or mathematics. Then, we

identify the papers in these journals that have been cited a given number of times. This limits us to the most significant papers of a field. For example, less than one percent of the papers cited in the 1981 SCI were cited 17 or more times.

The next step is determining which of these highly-cited papers have been cited together -- or co-cited. (In the 1981 SCI, this produced a list of about 24,000 co-cited pairs out of a theoretically possible 292 million pairs.) For example, the two papers shown in Figure 6 were co-cited by thirteen papers in 1977 and 1978. These two papers could be considered a cluster, albeit a small one. Through our computer programs, however, we can identify which other papers Cabantchik and Lepke are cited with, and add these to our cluster.

The cluster that includes the Cabantchik and Lepke papers is shown in Figure 7. The lines on this map, of 1978 research on the structure of red blood cell membranes, indicate which documents are co-cited. For example, the paper by Cabantchik has been co-cited with another paper he wrote, and with papers by Ho and Lepke. To find out what other papers are linked to this cluster, we looked for documents that are co-cited with Ho. Figure 8, the map representing this step, shows that Ho's paper has only been co-cited with one other paper, by Jenkins, and of course by the Cabantchik paper with which we started the cluster.

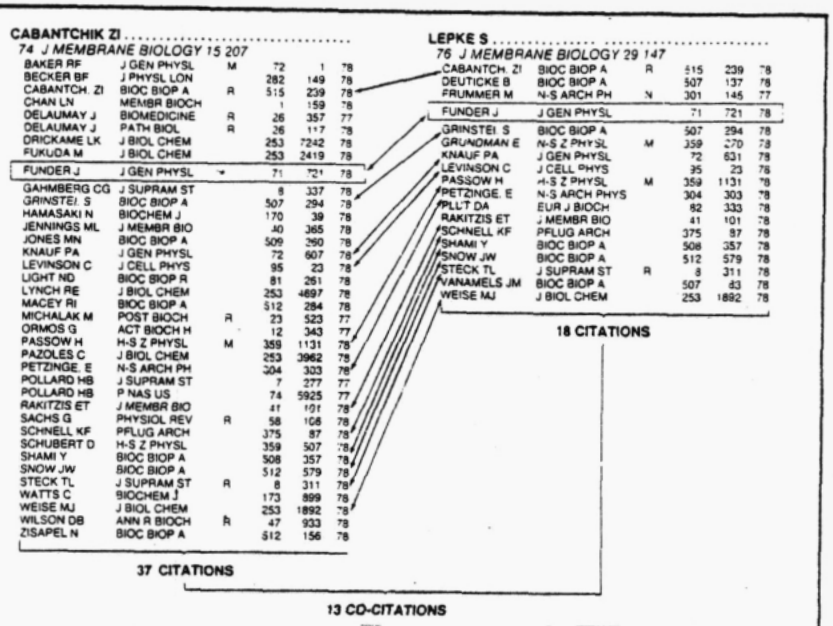


Figure 6: Co-citation links between papers. Red Blood Cell Membrane Structure.

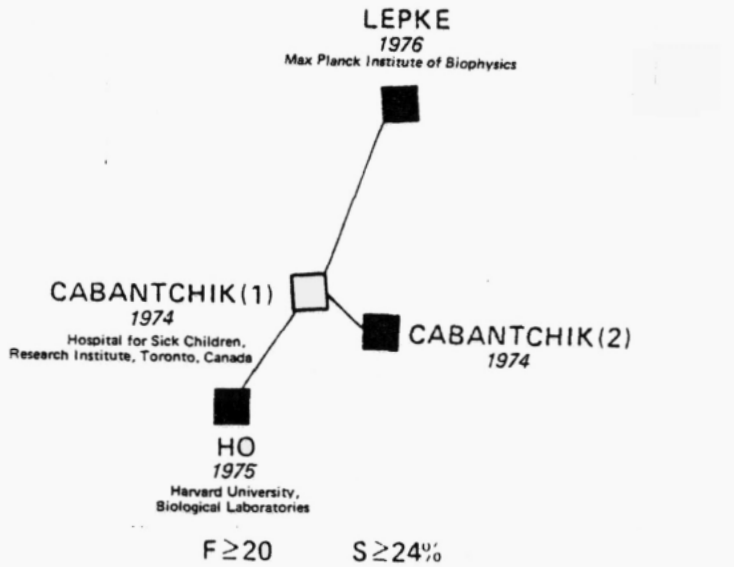


Figure 7: Steps in cluster development.

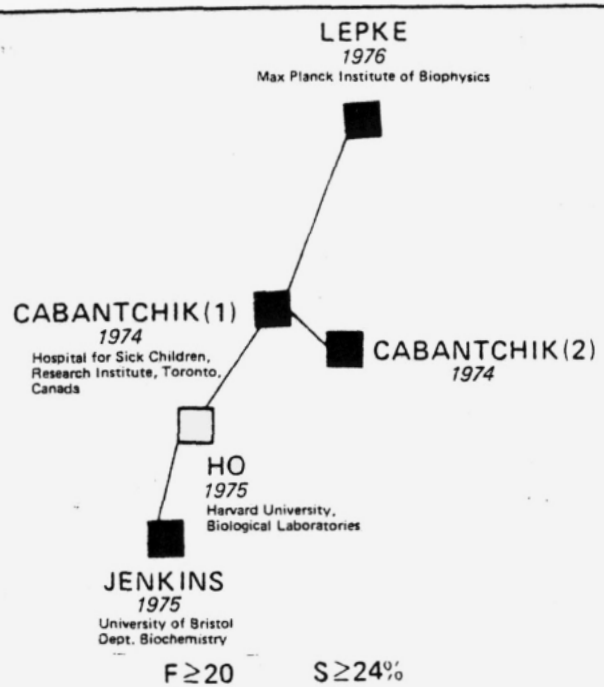
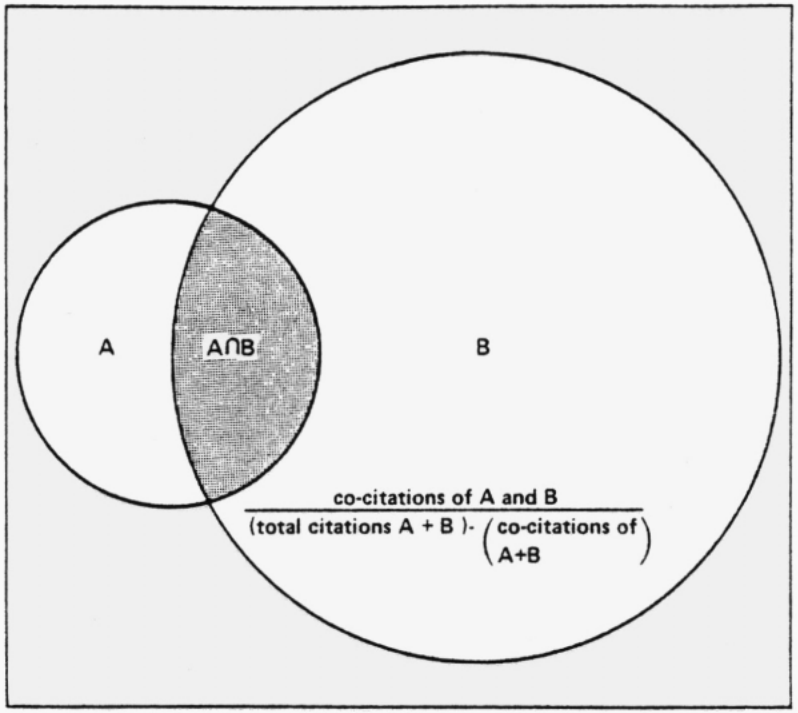


Figure 8: Steps in cluster development.

The computer continues this process until all papers that have been co-cited are identified. Since large groups of co-cited papers with rather tenuous relationships to one another could be formed through this process, we limit the papers in a cluster to those cited more than a given number of times. We also set a "strength threshold." This indicates how related two documents are in terms of the proportion of their total citations that are co-citations. The shaded area in Figure 9 represents co-citations of documents A and B. The formula on this figure is the calculation used to measure the strength of association between two co-cited documents. For example, suppose



Source: Garfield E. ABCs of cluster mapping. Part 1. Most active fields in the life sciences in 1978. *Essays of an information scientist.* Philadelphia: ISI Press, 1981. Vol.4. p. 634-41. (Reprinted from: *Current Contents.* (40) :5-12. 6 October 1980.)

Figure 9: Strength of association between co-cited documents. A and B, represented as two overlapping spheres. The shaded area of intersection represents co-citations of A and B.

document A is cited 20 times, document B is cited 50 times, and A and B are co-cited 10 times. To get the strength thresholds we would divide co-citations by the total number of citations to these papers minus their co-citations:

$$\frac{10}{(70) - (10)}$$

In this case we have a strength of association of 17 percent. The strength threshold can be increased to make a

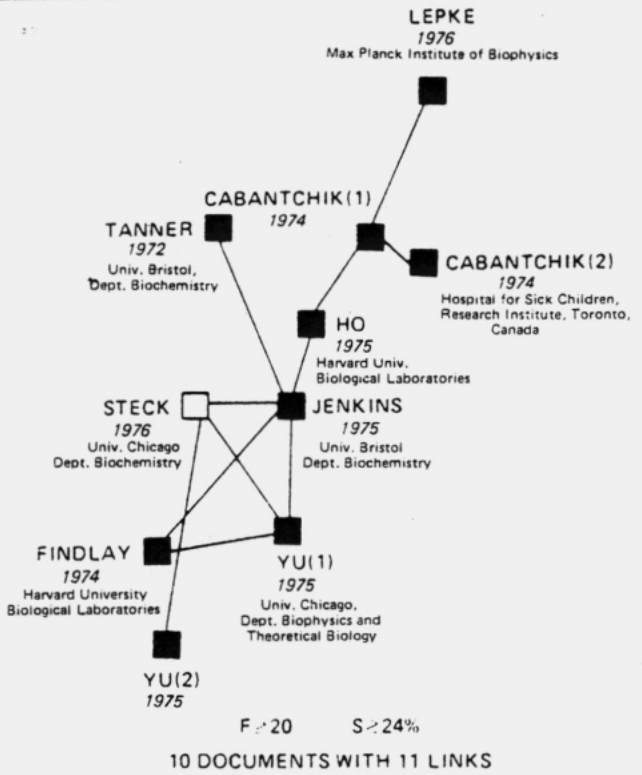


Figure 10: Steps in cluster development.

more sharply focused cluster. It can also be decreased to make the cluster larger and, therefore, less focused.

For example, the map in Figure 10 includes only papers cited more than 20 times. Also, 24 percent of all the citations to each of these papers are co-citations with another paper in the cluster. At these thresholds, the cluster includes ten documents with 11 links. If we lowered the threshold, the cluster would then include more papers.

Since research is an active, evolving process, the literature cited by scientists changes from year to year. These clusters also change from year to year to reflect this evolution. In the following paragraphs, these changes will be demonstrated.

In 1973, the six papers represented in Figure 11 announced the discovery of opiate receptors. All were fairly well cited. However, they did not appear in a cluster until 1974, after many scientists became aware of them and, in papers reporting their own work on opiate receptors, cited the documents. The 1971 paper by Goldstein provided the conceptual framework for physically demonstrat-

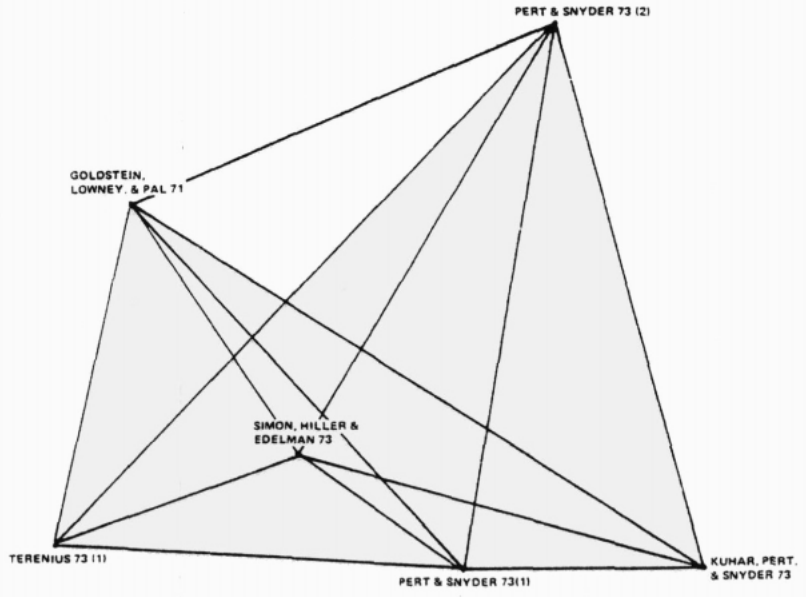


Figure 11: 1974 cluster map: Opiate Receptors.

ing the existence of opiate receptors. Papers announcing the discovery of such receptors were published more or less simultaneously, in 1973, by Pert and Snyder at Johns Hopkins, Simon and colleagues of New York University, and Lars Terenius at Uppsala University in Sweden.

The 1975 cluster map shown in Figure 12 reflects the increase of activity following the initial discovery of opiate receptors. Notice the Hughes paper was cited enough times in 1975, the year it was published, to make the cluster. It announces the isolation of the substance that Hughes and Kosterlitz would later name enkephalin. Papers by Pert and Terenius continue to appear in the cluster, as do papers by Goldstein. Later clusters continue to show these people publishing high impact work in the field.

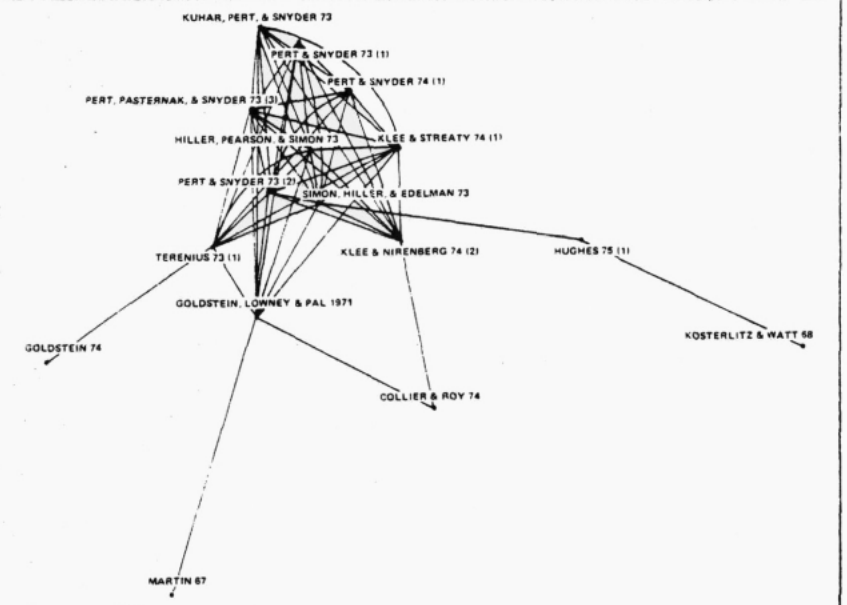


Figure 12: 1975 cluster map: Opiate Receptors.

The changes that occur within these clusters from year to year reflect the historical development of opiate receptor research. By looking at the papers that cite these clusters, you can also find out what researchers in that field are currently doing. So you can see that clusters provide a means of gaining a dynamic view as well as an encyclopedic perspective of any research area.

In the transition from bibliographic to encyclopedic information, thousands of such specialty areas will be defined through co-citation clustering. ISI is already offering disciplinary data bases that allow researchers to gain direct access to a bibliography of documents that cite the articles in these clusters. Since the clusters represent the most active research areas, the papers you obtain through these data bases are the most relevant to the specialty areas on which you decide to search.

At present, three clustered data bases are available through the ISI Search Network. The first of these, ISI/BIOMED™, serves the biomedical community. ISI/CompuMath® covers the literature of mathematics, computer science, statistics, operations research, and management science. ISI/GeoSciTech™ provides access to the literature of geology, meteorology, mineralogy, oceanography, and metallurgy, to name just a few of the disciplines covered. A completely separate ISI data base available through the ISI Search Network is an online version of our Index to Scientific &

Technical Proceedings®, to which has been added comprehensive coverage of multiauthored books. It is called ISI/ISTP&B™. Although items in this system are not yet clustered, we'll soon be offering other disciplinary data bases that are. These will cover various areas of chemistry, polymer and materials science, and arts and humanities. These will be followed as soon as possible by other data bases covering engineering, physics, neuroscience, plant science, agriculture, and many more. All of these data bases will offer virtually the same features. Therefore, the procedures I will describe in reviewing the following ISI/BIOMED search can be used with all the ISI Search Network systems.

ISI/BIOMED was the first of ISI's clustered online systems. Through this data base, the inexperienced searcher can obtain a bibliography of the most important, recent articles in almost 8,000 different research specialties. It lets you quickly review the literature on a subject and determine how deeply your research should go. ISI/BIOMED consists of articles from nearly 1,400 biomedical journals issued since 1979. The data base is updated monthly, with more than 270,000 new articles added each year. At present, it includes more than one million articles.

Each subscriber to ISI/BIOMED is given a guide to the almost 8,000 research front specialties in this data base. This guide is used to find the names of specialties that

VIRAL-INFECTIONS
ANTI-VIRAL CHEMOTHERAPY against HSV and other VIRAL-INFECTIONS 79-2275
ANTIBODY-DEPENDENT CELL-MEDIATED CYTOTOXICITY EVOKED by VIRAL-INFECTIONS 79-2151
DETERMINANTS of the SPECIFICITY of T-CELL IMMUNITY DIRECTED against INFLUENZA-VIRUS and other VIRAL-INFECTIONS 79-0531
VIRAL-INFECTIONS and BRONCHIAL-DISEASES in children 79-1745
Slow VIRAL-INFECTIONS and NEUROLOGICAL-DISEASE 79-2302
VIRAL-INFECTIONS in TERATOCARCINOMA-CELLS 81-2973

VIRAL-PROTEINS
CHEMICAL CROSS-LINKING of VIRAL-PROTEINS 80-1468
Structure of VIRAL-PROTEINS 81-0005
Modification and transport of VIRAL-PROTEINS across MEMBRANES 81-0786

VIRION
TRANSCRIPTION of NK VIRION DOUBLE-STRANDED RNA 80-0616

VIROIDS
VIRAL NUCLEIC-ACIDS and VIROIDS 80-1893
➡ Structure and function of VIROIDS . 79-1560
..... 81-1524

VIROLOGY
VIROLOGY of RIBAVIRIN 80-2706

VIRULENCE
Effects of IRON on BACTERIAL VIRULENCE 80-0913
..... 81-0989
MICROBIAL ADHESION, COLONIZATION and VIRULENCE 81-1598
MOLECULAR-GENETICS of REOVIRUS VIRULENCE 80-2432
VIRULENCE and PLASMIDS of YERSINIA 81-0815

Figure 13: Entry from *Index to Research Fronts in ISI/BIOMED™*.

reflect the searcher's interest. Each significant word or phrase appearing in a research front specialty is alphabetically arranged in the guide, followed by a list of the specialties in which it appears. A searcher looking for papers on the structure of viroids might turn to the portion of the guide shown in Figure 13. If the specialty entitled "structure and function of viroids" seemed relevant, he or she would simply key the numbers for this research front into the computer terminal.

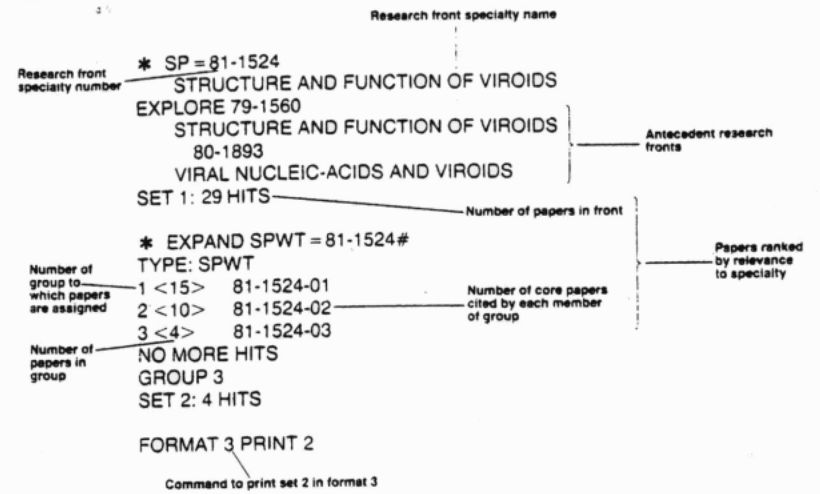


Figure 14: *ISI/BIOMED™* sample search. Structure and Function of Viroids.

As shown in Figure 14, the computer first prints out the name of the specialty to confirm that the correct number has been entered. You are also referred to any of the antecedent specialties that evolved into this one. Though closely related to the specialty you are already examining, searches in these other specialties often yield important papers. Next, the computer gives the number of hits, or papers, that appear in the 1981 specialty -- in this case 29. These papers are assigned to set one. When this is more than you care to see, you can ask the computer to rank these papers according to their relevance to the specialty -- that is, the number of core documents they cite. This is what is known as a specialty weight search.

In a specialty weight search, papers are grouped together according to the number of core documents they cite. The

KLEINSCHMIDT AK; KLOTZ G; SELIGER H
VIROID STRUCTURE
(ENGLISH/BIBLIOGRAPHY, REVIEW)
ANNUAL REVIEW OF BIOPHYSICS AND BIOENGINEERING 10:115-132 1981 70 REFERENCES
UNIV ULM, MIKROBIOL ABT, D-7900 ULM, FED REP GER
UNIV ULM, SEKT POLYMERE, D-7900 ULM, FED REP GER
ISI OATS ORDER #LS353
EXPLORE SPECIALTY 81-1524; 81-1553

DIENER TO
VIROIDS — ABNORMAL PRODUCTS OF PLANT-METABOLISM
(ENGLISH/BIBLIOGRAPHY, REVIEW)
ANNUAL REVIEW OF PLANT PHYSIOLOGY 32:313-325 1981 56 REFERENCES
USDA SEA, BELTSVILLE AGR RES CTR, INST PLANT PROTECT, PLANT VIROL LAB, BELTSVILLE, MD, 20705
ISI OATS ORDER #LT256
EXPLORE SPECIALTY 81-1524; 81-0818

HASELOFF J; SYMONS RH
CHRYSANTHEMUM STUNT VIROID — PRIMARY SEQUENCE AND SECONDARY STRUCTURE
(ENGLISH/ARTICLE)
NUCLEIC ACIDS RESEARCH 9(12):2741-2752 1981 30 REFERENCES
UNIV ADELAIDE, DEPT BIOCHEM, ADELAIDE, SA 5001, AUSTRALIA
ISI OATS ORDER #LV998
EXPLORE SPECIALTY 81-1524; 81-2483; 81-0800; 81-1553

ROHDE W; SCHNOLZER M; RACKWITZ HR; HAAS B; SELIGER H; SANGER HL
SPECIFICALLY PRIMED SYNTHESIS INVITRO OF FULL-LENGTH DNA COMPLEMENTARY TO POTATO-SPINDLE-TUBER VIROID
(ENGLISH/ARTICLE)
EUROPEAN JOURNAL OF BIOCHEMISTRY 118(1):151-157 1981 22 REFERENCES
UNIV GIESSEN, ARBEITSGRP PFLANZENVIROL, D-6300 GIESSEN, FED REP GER
UNIV ULM, BIOCHEM ABT, D-7900 ULM, FED REP GER
MAX PLANCK INST BIOCHEM, D-8033 MARTINSRIED, FED REP GER
ISI OATS ORDER #MB582
EXPLORE SPECIALTY 81-1524; 81-C170; 81-0950

Figure 15: *ISI/BIOMED*™ **sample search, Structure and Function of Viroids.**

first ranked group in Figure 14 includes 15 papers, each of which cite one core paper, while the ten papers in group two each cite two core papers. The four papers in group three are most relevant to the specialty, since they each cite three of the core documents. These are the papers you'd want to see first in a specialty weight search. So after the computer assigns these four papers to set two, we give the command to have them printed out in format three. You have a choice of formats in which to have papers displayed. They offer varying amounts of information so that users can obtain as much information as they need.

A great deal of information is given about each paper printed out for this search (Figure 15). For each paper,

you are given the authors' names, the title of the article, the language in which it's written, and the type of document it is. You're given the journal title and all other bibliographic information needed to retrieve the article, the authors' affiliations, and the number of references included in the article. You can also have bibliographic information about each paper's references printed out. Articles can be assigned to more than one specialty, so the numbers of all specialties to which they're assigned are included at the end of the entry. Each entry also includes an Original Article Text Service, or OATS®, number for ordering the article from ISI.

A number of options, in addition to specialty weight searching, are available for retrieving bibliographies through ISI Search Network data bases. You can select only the most current papers, or ask that only papers by a specific author be listed. You can also limit your bibliography to papers from a particular journal or institution, and can specify the type of document you want (proceedings, journal article, etc.) as well as the language in which it should be written. In any case, these systems will provide you with highly focused bibliographies.

All of these data bases offer several advantages over existing systems. The most obvious is that you needn't be a search specialist to use them. Through clustering and other innovations, we've been able to eliminate many of the

steps usually required in an online search. Until the advent of these systems, researchers had to rely on librarians to search data bases with their library's computer terminal. However, a large number of professionals already have terminals (microcomputers) on which they can search our systems. And many more will have them in the near future. Another advantage of these systems is that they permit you to take advantage of citation indexing without requiring that you begin with a particular author or paper. Instead, you begin with a research front specialty name. You can consider it a generic citation search.

Clustering has brought citation indexing a giant step forward in our march toward encyclopedism. Through our new clustered data bases, you retrieve groups of closely related papers that focus on narrowly defined specialties. This is clearly superior to retrieving individual articles, whose content is indicated by only one cited document. And it is certainly better than retrieving papers classified by what are usually outdated and inaccurate subject headings.

As I envision things, ISI's new data bases will be used by researchers when they require rapid access to a few highly relevant articles. However, our ultimate goal is to provide review-type information about fields of inquiry. We want to treat each scholarly discipline separately, providing incisive information about the discrete areas of inquiry that make up that discipline. The encyclopedic services we plan

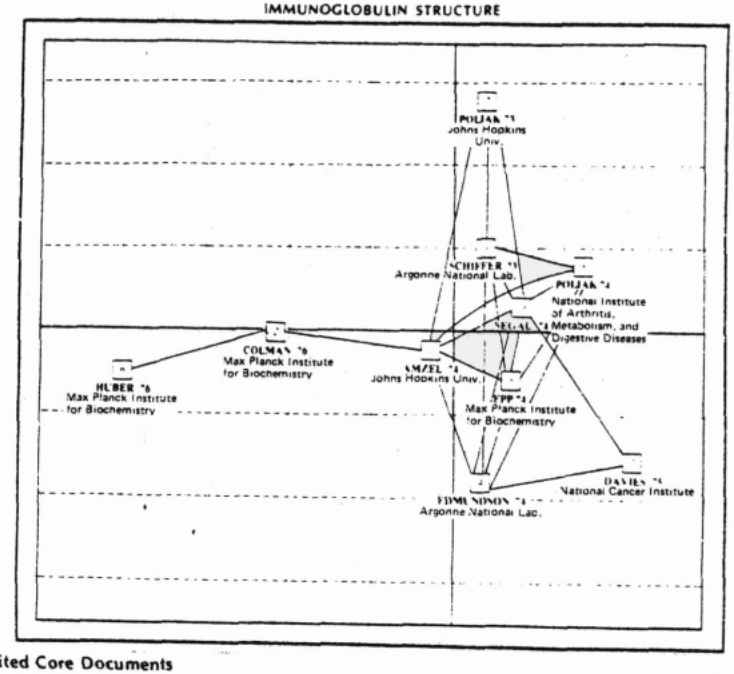
to offer will include bibliographies of the most important papers in each area, and a description of how these papers fit into the general scheme of research activity. Although ISI's total conversion to encyclopedism is still some years off, we have created a prototype of the product through which we intend to make this information available.

In this prototype, a print product called the ISI Atlas of Science™: Biochemistry and Molecular Biology, we've used clustering to identify the specialties that comprise the cutting edge of biochemical research. Using the documents identified in this manner, we've written minireviews that describe these specialties. These minireviews are presented in a format that also includes comprehensive bibliographies and visual portrayals of research relationships. Each of the 102 chapters in the Atlas is designed to acquaint researchers with the state of the art in a field by reviewing the important scientific events in the specialty and leading them to the newest research in that area.

A map from one of the Atlas chapters is shown in Figure 16. It is intended to provide some insight into the growth of the specialty and the names of some of the key scientists working in the field. This map consists of the cited core documents that comprise a cluster on the structure of immunoglobulin. The papers represented by the names on the map are those most frequently cited by the new scientific literature in this specialty. The distance between two

Specialty 23

IMMUNOGLOBULIN STRUCTURE



Cited Core Documents

		CF
1	AMZEL LM, POLIAK RJ, SALL F, VARGA JM, RICHARDS FF 3 dimensional structure of a combining region-ligand complex of immunoglobulin new at 3 5-A resolution *P NAS U S* 71(4):1427, 1974	30
2	COLMAN PM, DEISENHO J, HUBER R, PALM W Structure of human antibody molecule Kol (immunoglobulin-G1)—electron-density map at 5 degrees-A resolution *J Mol Biol* 100(3):257-278, 1976	19
3	DAVIES DR, PADLAN EA, SEGAL DM 3-dimensional structure of immunoglobulins *Ann R Biochem* 44:639-667, 1975 R	33
4	EDMUNDSON AB, ELY KR, GIRLING RL, ABOLA EE, SCHIFFER M, WESTHOLM FA, FALSCH MD, DEUTSCH HF Binding of 2, 4-dinitrophenyl compounds and other small molecules to a crystalline lamoda-type Bence-Jones dimer *Biochem* 13(18):3816, 1974	19
5	EPP O, COLMAN P, FEHLHAMM H, BODE W, SCHIFFER M, HUBER R Crystal and molecular-structure of a dimer composed of variable portions of Bence-Jones protein Rei *Eur J Biochem* 45(2):513-524, 1974	29
6	HUBER R, DEISENHO J, COLMAN PM, MATSUSHI M, PALM W Crystallographic structure studies of an IgG molecule and an Fc fragment *Nature* 264(5585):415-420, 1976	30
7	POLIAK RJ, AMZEL LM, AVEY HP, CHEN BL, PHIZACKE RP, SALL F 3-dimensional structure of Fab' fragment of a human immunoglobulin at 2 8-A resolution *P NAS U S* 72(12):3305-3310, 1973	29
8	POLIAK RJ, AMZEL LM, CHEN BL, PHIZACKE RP, SALL F 3-dimensional structure of Fab' fragment of a human myeloma immunoglobulin at 2 0-A resolution *P NAS U S* 72(10):3440-3444, 1974	28
9	SCHIFFER M, GIRLING RL, ELY KR, EDMUNDSO, AB Structure of a lambda-type Bence-Jones protein at 3.5-A resolution *Biochem* 12(23):4620-4631, 1973	30
10	SEGAL DM, PADLAN EA, COHEN GH, RUDIKOFF S, POTTER M, DAVIES DR 3-dimensional structure of a phosohorylcholine-binding mouse immunoglobulin Fab and nature of antigen binding-site *P NAS US* 71(11):4298-4302, 1974	43

Figure 16: Map of cited core documents from *Atlas of Science: Biochemistry and Molecular Biology 1978/80.*

documents on the map is a measure of their relatedness. For example, the cited document by Colman, number 2, is related to those by Huber and Amzel, as shown by the lines

describing their linkages. Information on each author's institutional affiliation can help you identify centers of activity on a specialty area. Bibliographic data on all the documents on the map, and the number of times each was cited, accompanies the map.

The essay portion of the chapter, shown in Figure 17, reviews research on the structure of immunoglobulins. It provides an historical summary of the specialty, describes papers that have played an important role in the specialty, and summarizes ongoing research. The numbers in the squares refer to the cited core documents found in the cluster map,

Specialty (23)

Immunoglobulin Structure

When a foreign and therefore potentially harmful substance capable of initiating an immune response (an antigen) enters the body of an animal (such as man) it is recognized by protein molecules (antibodies) which bind to it. This binding is an early stage in the process of eliminating the unwanted intruder. An antibody recognizes a specific antigen, although invasion by a single antigen usually induces the formation of several different antibodies which recognize it. Understanding the recognition process is of considerable interest in its own right as well as being of practical importance. It not suppressed by drugs, the immune response leads the body to reject transplanted organs, and in diseases like rheumatoid arthritis, antibodies react with the body's own tissues (autoimmunity).

X-ray crystallographers have determined the molecular structures of some antibody proteins (immunoglobulins) and are finding out how they recognize antigens. Small molecules that are part of an antigenic structure (haptens) bind to the combining site of specific antibodies. Hapten-antibody complexes have been studied by crystallographic techniques to investigate the basis of the recognition process. This is the same approach which has proved so successful in understanding how enzymes recognize their substrates.

If X-ray crystallography is to accurately determine the positions of atoms in molecules, all the molecules in the crystal must be the same. But immunoglobulins are usually heterogeneous. One way of overcoming this problem is to investigate "monoclonal" immunoglobulins of restricted heterogeneity, such as those characteristic of spontaneously occurring human myelomas (tumors) or of laboratory-induced mouse myelomas. Frequently, a component of the serum myeloma protein (light chains, see below) is secreted in the urine of the patients, from which it can be readily purified. This component is called a Bence-Jones protein [7, 9].

The commonest class of immunoglobulin molecule from myelomas (IgG) consists of four polypeptide chains linked by disulfide bridges—two identical light chains (molecular weight about 23,000) and two identical heavy chains (molecular weight about 50,000). In the amino-acid sequence of each chain there are both variable (V) and constant (C) regions. The light chain consists of one of each V and C, while the heavy chain contains one variable (V$_H$) and three constant (C$_H$1, C$_H$2, C$_H$3) regions. The ability of an immunoglobulin to recognize a specific hapten or antigen resides in the hypervariable segments (extremes variable amino acid sequences) of both variable regions V$_L$ and V$_H$. Bence-Jones proteins consist of light polypeptide chains only. Treatment of IgG with the enzyme papain yields an antibody-binding fragment (Fab) and a constant fragment (Fc) which is easily crystallized. Treatment with pepsin yields a dimeric fragment called F(ab')$_2$ which is converted by reduction to the monomer Fab'. Fab' is slightly larger (by about 10 amino acids) than Fab.

Poliak and his colleagues at Johns Hopkins University determined the molecular structure of the Fab fragment from a human myeloma protein (Fab New) first at a resolution of 5 Å, then at 2.8 Å [7] and subsequently at 2.0 Å [8]. The structure of the Fab fragment from a mouse myeloma immunoglobulin (Fab McPC 603) was determined at 3.1 Å resolution at the National Institutes of Health [5] and more recently the structures of both an intact IgG molecule (Kol) and its Fab fragment (Fab Kol) have been determined in West Germany [2, 6]. A detailed comparison by the German group shows that all the structures of the comparable regions in these fragments, and in a Bence-Jones protein whose structure had been previously determined, are very similar, but with sufficient variability for specific recognition of an antigen.

Figure 17: Essay from: *Atlas of Science: Biochemistry and Molecular Biology 1978/80.*

and in the bibliography following the map. Since this essay guides you directly to the documents most relevant to your specialty, you don't need to read dozens of papers. For example, if you wanted more information on Bence Jones protein, which is mentioned in the third paragraph of the essay, you might take a look at the studies reported by Edmundson and Schiffer, in boxes four and nine.

Key Citing Documents — RW

1. POLJAK RJ — RW 10
   Correlations between 3-dimensional structure and function of immunoglobulins
   Crc C R Bi 5 45, 1978 R
2. POLJAK RJ — RW 10
   Studies on 3-dimensional structure of immunoglobulins (Litman GW, Good RA, eds) *Immunoglobulins* New York: Plenum Medical Book Co. 1978 p 1 R
3. EDMUNDSO AB, ABOLA EE, ELY KR — RW 9
   Conformational flexibility in immunoglobulins (Reisfeld RA, Inman FP, eds) *Contemporary Topics in Molecular Immunology* Vol 7 New York: Plenum Press, 1978 p 95
4. FIRCA JR, DORRINGT KJ, EDMUNDSO AB, ELY KR, KREMSER P, WESTHOLM FA — RW 9
   Interconversion of conformational isomers of light-chains in Mcg immunoglobulins
   Biochem 17 148, 1978
5. ELY KR, ABOLA EE, EDMUNDSO AB, FENTON JM, FIRCA JR, PANAGIOT NC, SCHIFFER M, WILLIAMS KJ — RW 9
   Crystal properties as indicators of conformational-changes during ligand-binding or interconversion of MCG light chain isomers
   Biochem 17 158, 1978
6. PUTNAM FW — RW 8
   Immunoglobulins 1 Structure (Putnam FW, ed) *Plasma Proteins* Vol 3 Second Edition *Structure Function and Genetic Control* New York: Academic Press 1977 p 1 R
7. KABAT EA — RW 7
   Structural basis of antibody complementarity (Anfinsen CB, Edsall JT, Richards FM, eds) *Advances in Protein Chemistry* Vol 32 New York: Academic Press, 1978 p 1 R
8. CATHOU RE — RW 6
   Solution conformation and segmental flexibility of immunoglobulins (Litman GW, Good RA, eds) *Immunoglobulins* New York: Plenum Medical Book Co 1978 p 37 R
9. MATSUSHI M, BARTELS K, COLMAN PM, HUBER R, JONES TA, MARQUART M — RW 6
   Crystal-structure of human Fab fragment Kol and its comparison with intact Kol molecule
   Mol Biol 121 441 1978
10. MATTHEWS BW — RW 6
    X-ray structure of proteins (Neurath H, Hill RL, Boeder CL, eds) *Proteins* Vol 3, Third Edition New York: Academic Press, 1977 p 403 R
11. POTTER M — RW 6
    Antigen-binding myeloma proteins of mice (Kunkel HG, Dixon FJ, eds) *Advances in Immunology* Vol 25 New York: Academic Press 1977 p 141 R
12. RICHARDS FF, KONIGSBE WH, ROSENSTE RW, VARGA JM — RW 6
    Antibody combining regions (Litman GW, Good RA, eds) *Immunoglobulins* New York: Plenum Medical Book Co. 1978 p 117 R
13. SAUL FA, AMZEL LM, POLJAK RJ — RW 6
    Preliminary refinement and structural-analysis of Fab fragment from human immunoglobulin new at 2.0 A resolution
    J Biol Chem 253 585 1978
14. ZAVYALOV VP, ABRAMOV VM, TETIN SY, TROITSKY GV — RW 6
    Effect of salt concentration on immunoglobulin-G structure
    Bioc Biop A 533 496 1978

Supplementary Citing Documents — RW

1. ABOLA EE, EDMUNDSO AB, ELY KR — RW -
   Marked structural differences of the MCG Bence-Jones dimer in 2 crystal systems
   Biochem 19 432 1980
2. KABAT EA — RW 5
   Structural and genetic insights into antibody complementarity (Cohen EP, Kohler H, eds) *Membranes Receptors and the Immune Response* New York: Alan R Liss Inc 1980 p 1
3. MARQUART M, DEISENHO J, HUBER R, PALM W — RW 5
   Crystallographic refinement and atomic models of the intact immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 A and 1.9 A resolution
   Mol Biol 141 369 1980
4. DOWER SK, DWEK RA — RW 4
   Antibody-binding site — a combined magnetic-resonance and crystallographic approach (Shulman RG, ed) *Biological Applications of Magnetic Resonance* New York: Academic Press 1979 p 271
5. HUBER R — RW 4
   Spatial structure of immunoglobulin molecules
   Klin Woch 58 1217 1980
6. KABAT EA — RW 4
   Basic principles of antigen-antibody reactions (Vanunakis H, Langone JJ, eds) *Immunochemical Techniques* Pt A New York: Academic Press 1980 p 3
7. ATASSI MZ — RW 3
   Precise determination of protein antigenic structures has unravelled the molecular immune recognition of proteins and provided a prototype for synthetic mimicking of other protein-binding sites
   Mol C Bioch 32 21 1980
8. BERTRAM J, GUALTIER RJ, OSSERMAN EF — RW 3
   Amyloid-related Bence-Jones proteins bind dinitrophenyl L-lysine (DNP) (Glenner GG, Freitas AFD, Costa PP, *Amyloid and Amyloidosis* Amsterdam: Excerpta Medica 1980 p 351
9. BRAUN DC, Schalch W — RW 3
   Multifunctional properties of antibodies (Bisswanger H, Schminckeott E, eds) *Multifunctional Proteins* New York: John Wiley & Sons Inc 1980 p 293
10. KOHMA M, TAKENAKA T, ODANI S — RW 3
    Amino-acid sequence of the lambda-type light chain of a human IGGL myeloma protein MOT with unusual antigenicity — a possible new subgroup of lambda-chain having a unique N-terminal sequence
    Mol Immunol 17 1407 1980

Figure 18: Citing documents from *Atlas of Science: Biochemistry and Molecular Biology 1978/80.*

Specialty 23

A molecule of Bence-Jones protein contains two light chains and consists of two globular regions. One region contains the two variable stretches of amino-acid sequence, and the other the two constant stretches (7, 9). The two globular regions have very similar three-dimensional structures, both consist of two sheets of β conformation which are roughly parallel. But in the variable region there is a cavity about 15 Å in diameter and 10 Å deep which is bounded by the hypervariable stretches of both light chains. In the Fab McPC 603 and in the Fab. Fab New fragments there is a distinct groove in the corresponding region. In the former it is roughly 15 Å deep, and 20 Å long, while in the latter its dimensions are 15 Å × 6 Å × 6 Å. By diffusing haptens into the crystal and determining where they are bound, these cavities or grooves have been identified as the recognition sites. Portions of these grooves have been identified as participating as part of the recognition sites.

Now that the recognition site has been identified two problems remain: to map out its detailed stereochemistry in a number of hapten-IgG complexes, and to find out how the immunoglobulin molecule interacts with other serum and cellular components in the so called "effector functions" (such as complement fixation, binding to mast cells, etc.). Considerable progress is being made in determining the details of hapten binding, not only by X-ray crystallography but also by other physical techniques (1). The second problem is more difficult because it is not clear whether or not the effector function processes involve structure changes in the antibody molecule after antigen binding, i.e., an "allosteric" mechanism, or simply aggregation of antibody molecules at the antigenic surface. A major problem is to crystallize and study a monoclonal antibody whose site has been mapped by immunochemical methods so that a ligand filling the entire site becomes available.

1980 Supplement

Antigen binding triggers the effector functions of antibodies (S5). Occuring at the tips of the Y-shaped antibody, it should be noted that its effector functions are demonstrated at the stem. However, the question still remains as to whether conformational changes of the antibody play a significant role in trigger metabolism.

Studies performed by Marquart et al. (S3) showed the detail of fragments of atomic models crystallographically refined at 3.0 Å and 1.9 Å resolution. The analyses permitted detail not seen before.

Figure 19: Essay supplement from *Atlas of Science: Biochemistry and Molecular Biology 1978/80.*

The citing, or source, documents listed in the minireview (Figure 18) are those more recent articles which cited the core documents on the map. Each of these articles is followed by the number of core documents it cites. The first two articles on this list cite ten of those core documents, while the third and fourth cite nine. These would probably be among the first you would want to see. Review articles often appear at the top of the list because they would cite the most relevant core articles. The supplementary citing documents show the 1980 papers that have cited the cluster. These papers update the Atlas, which was originally compiled in 1978. They lend an added element of timeliness to the Atlas. A short supplement to the essay, shown in Figure 19, is also included. It mentions several of the important papers listed among the supplementary documents, and tells you about new methods and developments in this specialty since the prototype Atlas was compiled.

Encyclopedism can be applied to virtually every field of scholarly inquiry. Eventually, we hope to use this novel approach to make incisive encyclopedic information available in all the disciplines of science, social science, and the arts and humanities. Although we're starting out with print products, some day we'll be offering computerized access to minireviews of the thousands of specialty areas that comprise scholarly research. At a touch of a button, you'll be able to examine maps depicting the entire structure of any discipline, perhaps even the entire structure of scholarship. You'll be able to determine how specialties within a discipline relate to one another, or to research going on in seemingly unrelated disciplines. In minutes, you'll have a complete and unified view of any field of inquiry along with insight into how this field relates to the whole range of scholarly research. The era of encyclopedism will be an era of instant knowledge, an age in which every scholar can gain a fairly complete understanding of any field he or she desires. I hope you will join me in welcoming in this new and exciting era.

REFERENCES

1. Garfield E. Has scientific communication changed in 300 years? Essays of an information scientist. Philadelphia: ISI Press, 1981. Vol. 4. p. 394-400. (Reprinted from: Current Contents (8):5-11, 25 February 1980.)

2. Jarcho S. Seventeenth-century medical journalism, as exemplified by the Ephemerides Naturae Curiosorum. J. Amer. Med. Assn. 220:64-8, 1972.

3. Garfield E. Citation indexing - its theory and application in science, technology, and humanities. New York: Wiley, 1979. 274 p.

4. ----------. The 1,000 most-cited contemporary scientists. Part 3. Details on their current institutional affiliations. Current Contents (27): 5-20, 5 July 1982.

5. Price D J D. Why does science cumulate? Unpublished speech presented to the University of Pittsburgh Program in the Philosophy of Science. 12 February 1963. Pittsburgh, PA. 28 p.

6. Small H G & Griffith B C. The structure of scientific literatures. 1: Identifying and graphing specialties. Sci. Stud. 4:17-40, 1974.

7. Garfield E. ABCs of cluster mapping. Parts 1 & 2. Most active fields in the life and physical sciences in 1978. Essays of an information scientist. Philadelphia: ISI Press, 1981. Vol. 4. p. 634-49. (Reprinted from: Current Contents (40):5-12, 6 October 1980 and (41):5-12, 13 October 1980.)