TREND ANALYSIS IN SCIENTOMETRICS

ELIZABETH ROWENA JANE BELL

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

October 1988

The University of Aston in Birmingham

Trend Analysis in Scientometrics

Elizabeth Rowena Jane Bell
Doctor of Philosophy
1988

THESIS SUMMARY

The thesis investigates the properties of two trends or time series which formed a part of the Co-Citation bibliometric model "X-Ray Crystallography and Protein Determination in 1978, 1980 and 1982". This model was one of several created for the 1983 ABRC Science Policy Study which aimed to test the utility of bibliometric models in a national science policy context. The outcome of the validation part of that study proved to be especially favourable concerning the utility of trend data, which purport to model the development of specialty areas in science over time. This assessment could have important implications for the use of such data in policy formulation. However one possible problem with the Science Policy Study's conclusions was that insufficient time was available in the study for an in-depth analysis of the data.

The thesis aims to continue the validation begun in the ABRC study by providing a detailed examination of the characteristics of the data contained in the Trends numbered 11 and 44 in the model. A novel methodology for the analysis of the properties of the trends with respect to their literature content is presented. This is followed by an assessment, based on questionnaire and interview data, of the ability of Trend 44 to realistically model the historical development of the field of mobile genetic elements research over time, with respect to its scientific content and the activities of its community of researchers.

The results of these various analyses are then used to evaluate the strengths and weaknesses of a trend or time series approach to the modelling of the activities of scientific fields. A critical evaluation of the origins of the discovered strengths and weaknesses in the assumptions underlying the techniques used to generate trends from co-citation data is provided. Possible improvements to the modelling techniques are discussed.

Keywords

Co-Citation Analysis
Bibliometric Modelling
Scientometrics
Trends
Time Series

DEDICATION

In fond memory of Beryl Marsden, secretary of the TPU, and Peter Lucas, my prospective brother in law, who both died from cancer before their time.

ACKNOWLEDGEMENTS

I would like to thank the following people:
My supervisors Dr. David Collingridge and Dr. Harry Rothman for their advice and encouragement;
Johann Riedel for his monumental patience and perseverance in producing the figures and tables and for teaching me how to use the Harris word processor;
Beryl Marsden, late secretary of the TPU, who assisted in the mailing of the questionnaires;
The staff of Aston computer centre for their help over the years;
The staff of the computer centre at Hull University for allowing me to access the Harris computer from their terminals, and for their freely given assistance;
The statisticians at Humberside College of Higher Education for their advice concerning the applicability of statistics in Chapter 6;
The SERC for their studentship and for providing the funds necessary to interview scientists abroad;
Prof. Michael Brown, Pro-Vice Chancellor of Aston University for his assistance in obtaining a 5 month Aston studentship;
Prof. Douglas Berg for coming to Aston to be interviewed, Dr. Christine Haenni and her family who kindly took me into their home in Basel, Switzerland, and all those scientists who gave up their valuable time to fill in the questionnaire or be interviewed;
Last but by no means least, my family and friends, and especially my fiance Alick, for their support;

CONTENTS

Contents (cont)

LIST OF TABLES

List of Tables (cont)

List of Tables (cont)

List of Tables (cont)

List of Tables (cont)

List of Tables (cont)

LIST OF FIGURES

Declaration

Fig. 4.1 is derived from the trend diagram supplied for Trend 44 in the X-Ray Crystallography and Protein Determination Models, 1978, 1980 and 1982 produced for the 1983 AERC Science Policy Study. Fig. 5.1 and Appendix D.1 are copied directly from the above model.

This page has been left intentionally blank

Chapter 1: Introduction

In the last decade there has been a problem of economic recession in the UK and other OECD countries. These conditions have led to static budgets and even cuts in the funding of science, see Irvine and Martin, 1984, (40). This time of financial stringency has coincided with increasing complexity and capital intensity in science typified by the emergence of many new, often interdisciplinary fields and highly expensive projects (for example research at CERN). The combination of these factors has made resource allocations increasingly more complex, see J. King, 1987, (42).

The traditional way of allocating resources, the peer review system, has come increasingly under criticism with respect to its ability to adapt to these new problems. Amongst these criticisms are accusations of lack of impartiality and the presence of vested interests, see J. King, 1987, (42). These faults are exacerbated in times of financial stringency with resultant intense competition for funds and may result in the protection of established but declining fields at the expense of new, emerging areas. Thus the peer review system may often be ineffective as a mechanism for restructuring scientific activity in a period of retrenchment. This situation has led to demands for more effective science policies and more information for the allocation of funds.

This provided the opportunity for the work of researchers in a sub-discipline of information science called scientometrics to enter the policy arena. Scientometrics is a term that covers a variety of techniques dealing with the quantitative analysis of the properties of the scientific literature, such as its pattern of citations, in order to evaluate the scientific fields that produced it. The different techniques available and their various shortcomings and advantages are dealt with in H. Rothman, 1987, (71). Work from the 1970's onward showed that these techniques, and in particular one developed by Henry Small called Co-Citation Analysis, appeared to be able to show the structure of science.

This drew them to the attention of sociologists and philosophers of science interested in determining how structure related to the development of scientific fields. The potential of these techniques to map science and follow its development also brought it to the attention of policy makers who thought that their application might remedy the perceived weaknesses in the peer review system. However the subsequent entry of such techniques into the policy arena resulted in their having to fulfil a new set of requirements. Amongst these requirements was that data from these sources should be capable of disaggregation (so that small units such as individual research groups can be examined), internationally comparable, publicly accessible and able to examine time trends in scientific research.

One study that tested the ability of scientometric databases to meet these requirements was that conducted by the Advisory Board of the Research Councils of Great Britain (ABRC) in conjunction with the research councils in 1983. The study evaluated and compared several different scientometric modelling techniques such as co-word modelling and co-citation analysis (CCA). Some of the conclusions of the study, particularly those relating to Co-Citation Analysis, were reported by H. Roberts Coward, 1984, (14), and examined in P. Healey et al, 1986, (38). One of the conclusions was that out of all the techniques presented, the data produced using CCA, the product of research at the Institute for Scientific Information (ISI), seemed to be the most promising. However although the results of the ABRC study showed CCA in a favourable light the technique has come under criticism from sociologists of science with respect to its ability to accurately represent scientific fields. This has implications at the policy level where the accuracy of the data upon which decisions are based is important.

The research council experts who validated the data in the ABRC study were not able to examine it in any great detail due to the huge volume of data presented. In my opinion this makes the study's conclusion concerning the potential of CCA as a promising policy formulation tool a problematical one. In view of the potentially far reaching consequences of using such a tool at the

level of national science policy decisions I considered that a deeper analysis of the validity of some of the data was necessary.

This provided the inspiration for this present research topic which concentrates on the validation of the time series or trend data presented in the CCA models for the ABRC study. It was decided to concentrate on a closer examination of the properties of trends because these were considered to be a particularly promising source of information in the ABRC study. This research aims to examine one or two trends in detail in the context of the scientific fields that they supposedly represent, firstly in order to answer the question as to whether they accurately represent the development of the fields well enough for policy decisions to be made from them, and secondly to study the techniques used to generate trends in order to criticise them and suggest modifications. It is hoped that the combination of these two lines of enquiry might also eventually assist in the further development of theories concerning the mechanisms of specialty development in scientometrics.

Chapters 2 and 3 of this thesis review the literature underpinning this research. Chapter 2 is concerned with an analysis of research into co-citation analysis in general and also reviews relevant literature from the area of social studies of science. Chapter 3 specifically addresses the use of CCA in the construction of bibliometric models, focusing on the creation of the models used in the ABRC study. Chapter 3 also explores the development of time series or trend analysis in CCA with particular reference to those presented in the ABRC study. Chapters 4 to 7 deal with the analysis of the properties of the two trends selected for study. The research aimed to develop an alternative methodology for trend validation and analysis to that used in the ABRC study. This approach was taken in order to test the conclusions of that study from a different vantage point, and to provide in-depth information concerning the characteristics and limits of reliability of time series data. The research firstly developed a methodology for examining the ability of a trend to represent the content of the scientific literature that it models. This is presented in Chapter 4 which provides details of the

methodology used and results obtained from a close analysis of the characteristics of some of the scientific literature contained in Trend 44 from the ABRC study. Chapter 5 tests the soundness of the methodology used in the analysis of Trend 44 by using it to examine the second trend selected, Trend 11. Chapters 6 and 7 then extend the validation of Trend 44 by surveying the opinions of the scientists who had authored the literature that it attempts to model. This was done firstly in Chapter 6 by compiling a questionnaire to send to the scientists concerned which tested the validity of data derived from the trend. The questionnaires were then followed up by in-depth interviews with scientists which enabled the accuracy of the trend's portrayal of the historical development of the field to be examined. The implications of the results of these various analyses for the use of time series data in science policy is discussed in Chapter 8.

Chapter 2: Survey of Literature Relating to the Development of Co-Citation Analysis

Section 1: Introduction - The Development of Citation Indexing

This chapter will deal mainly with some of the relevant studies pertaining to the development of work on co-citation analysis (CCA) and the structure of science, which form the foundation for the large scale bibliometric models and trend modelling techniques that are the subject of this thesis. Since a discussion on these areas alone is necessarily lengthy the roots of these techniques in the various forms of citation analysis will be dealt with only briefly.

George Lester's thesis, 1977, (46), which examined the history and development of citation analysis up to 1975, described the contribution to citation studies of work on the retrieval of scientific information for the use of scientists. These concerns led to the development of the information sciences which attempted to retrieve documents by classifying them using measures such as subject indexing, keywords or abstracts. These initial efforts were followed by attempts to classify and retrieve documents utilising the standard scientific practise of referencing earlier work, i.e. citation, which led to the development of citation indexes.

Martyn, 1965, (51) described several early examples of manually constructed citation indexes for the scientific literature in the late 1950's and early 1960's. However this type of indexing became more visible to information scientists in particular and the scientific community in general with the production of the first volume of the computer generated Science Citation Index, (SCI) which was produced by the Institute for Scientific Information (ISI), in 1961. E.Garfield, 1970, (30) claimed that each annual file of this index cites between 35 and 50% of the 5 to 10 million books and papers published since its inception, and shows the citation links between them. According to Garfield its original purpose was to deal with the problem of

subject / index barriers in information science. However it proved also to be of intellectual interest to sociologists and historians of science, (and also of great concern to the scientific community), when its potential as a controversial aid to research policy formulation and performance evaluation was realised.

The literature resulting from this debate relating to the properties of the SCI was examined in detail by Lester, 1977, (46). Lester pointed out that the use of citations as an indexing tool proved to be controversial even within the information science community, some of whom regarded citations as being unreliable as objective information measures. Thus the development of the SCI proved to be not only controversial in its role as an indexing tool, but also because it was capable of providing a detailed and comprehensive database for citation analysis studies, another highly controversial area.

Citation analysis in its various forms, (for some examples see H. Rothman, 1984, (73)), had always been controversial because of its potential as a tool for evaluating the "quality" of institutions, papers and even individual scientists in relation to others in their fields. Criticism of these techniques come from two main sources; the scientific establishment, and sociologists and historians of science, all of whom are concerned with its validity as a source of useful information relating to scientific fields. It was the successful development in 1961 and the commercial production of the SCI in 1964, which by providing comprehensive citation data suddenly made citation analysis a serious issue for these groups.

Since then there have been many studies looking at the value of using citations as indicators. Various studies looked for example at the correlation of numbers of citations and honorific awards, a correlation that was generally found to be positive. Francis Narin's doctoral thesis, 1983, (58) reviews previous studies of this type and contains some examples of his own.

The success of these studies, and of the SCI as an information retrieval tool, has fuelled the debate over the role that citations play in the scientific literature and hence the validity of using citation measures as indicators. One author who

has closely examined this debate is Cronin, who in 1984, (21), pointed to its increasing polarisation in the last two decades. He described one pole as being represented by E. Garfield and the information scientists at ISI who champion the use of citation indexing and are concerned that it should not be unfairly dismissed due to some limitations and occasional irresponsible use, and at the other pole Chubin (e.g. see Chubin and Studer, 1979, (9)) who regards it as being a meaningless numerology of doubtful relevance to the way that the institution of science operates.

Lester, 1977, (46) described the criticisms from this camp of mainly being of two types. Firstly, with citation not being regarded as an accurate representation of information, and secondly that it is subject to social pressures and interpretations. However Lester argued against the assumption that the social and institutional functions in science hindered the effective retrieval of information and showed from his work on the retrieval of information in the specialism of inert gases that they may instead enhance its utility.

Garfields contribution to the debate is typically shown in his paper, 1979, (30) where he dealt with some of the more common criticisms, such as the phenomenon of self citation, and defended the basic assumptions of citation analysis using evidence drawn from sociology of science. Garfield in his defence of citation analysis claimed that citation counts were best used as a measure of the utility of scientific papers to later researchers, and thus should not be taken as an indicator of a papers quality. However he still had to concede to the critics that comparisons of different scientific fields using citation techniques would not be easy due to inter-field variation in citation habits. Cawkell, 1974, (6) presented a similar point of view when he talked of the surprisingly small effect of expected distortions such as self citation in the aggregate but admitted in Cawkell, 1976, (7) that they would cause problems in assessing the work of individuals.

J.Ronayne, 1983, (70), like Cronin, also saw the debate as a clash between rival schools of thought. He used the sociologist David Edge to represent the anti-bibliometric

position. Edge believes that informal communication is more significant than published findings, which places him in sympathy with Chubin's position that measurements of citations to such publications are meaningless. Edge, 1979, (24) claimed that quantitative measures of science only reveal the rational reconstruction of findings after the event and not the social construction of scientific results, which he regards as being the more important phenomenon. As champion of the pro-bibliometric position Ronayne put forward Henry Small who countered the suggestion of the insignificance of published findings by claiming that citations act as concept symbols for scientists (see Section 3 of this chapter). Other papers examining some of the theoretical problems associated with citation analysis are A.L.Porter, 1977, (61) who for example looked at problems involved with citations to collaborative research, and J.Martyn, 1975, (52) who claimed that a citation can be regarded as indicating the relevance of a previous work to the citing article but that the nature of that relationship cannot be assumed. Such assumptions are risky as it might be assumed that citation would tend to occur to good papers, whereas the citer might have referred to the paper in order to criticise it. Generally the debate can be said to be between information scientists who do not dwell on questions of the social reality of citation, and sociologists questioning the fundamental validity of citations as an indicator. Cronin, after examining the various issues in the controversy, called for the development of a comprehensive theory of citing in his book (and also in an earlier paper, 1981, (22),) in order to resolve these theoretical questions and effectively evaluate citation analysis. He regarded this as being of paramount importance due to the potential of citation analysis becoming a powerful analytic tool.

This controversy is still very much alive today, with much criticism originating from this quarter being brought to bear on studies attempting to use citations to elucidate the structure and dynamics of science, which are the main subject of this review.

Section 2: Studies on the Structure of Science

Data provided by the SCI made it possible to use citation behaviour to learn something about the structure of the scientific literature, and by implication, something about the way that science is organised, and how this affects its development.

Lester, 1977, (46) described studies in the sociology of science that he regarded as having considerably advanced understanding of how science develops. Lester regarded the most fundamental idea from this quarter as being the concept that scientific disciplines are sub-divided into specialisms or specialties, with progress in science being viewed as the sum of progress in many of these small units. Lester defined these units as scientific communities with a common intellectual focus of a shared problem, approach or technique. He regarded their structures as typically consisting of a prestigious elite, prominent in all the social and informational processes of the group. Lester emphasised the important contribution made by techniques such as citation indexing to the mapping of the specialty structure of science.

One of the most influential researchers in this area was Derek de Solla Price, who in 1963 in his influential book "Little Science, Big Science" (62), provided impetus to the study of scientific research and drew attention to the potential of a quantitative approach to the study of the structure of science. This work was followed by that of Garfield, Sher and Torpie, 1964, (32) who showed from an examination of the history of DNA research that it was possible to use citation analysis to demonstrate the intellectual relationships between past research events.

In 1965 Price, (63), working from these foundations used citation patterns to explore the structure of one of the specialties of physics. He proposed that networks of scientific papers are linked by citations that reveal the research front mostly drawing upon very recent work but also showing some links to more classic papers. He described this phenomenon as the immediacy factor, characterised by recent papers tending to be the most frequently cited with the result that the literature tends to

be composed of two parts, the ephemeral and the classic (the one representing the dynamics of research front activity and the other the archive of past achievements). He believed that research fronts and networks were very tightly organised with individual research fronts corresponding to only a few hundred researchers at any time.

Price was also concerned with the need to effectively plan science, an issue that was first put on the agenda by J.D. Bernal. Bernal, 1939, (5), examined the apparent roles of science in society and stressed the need for the effective planning of science if it was to reach its full potential.

In order to make the planning of science feasible, Price, 1965, (65) put the case for the development of a science of science. By this he meant a need for a scientific basis for knowledge about science and technology which would supply information such as the geographical locations and economics of research. He claimed that a science of science was needed to generate a special body of scientific knowledge that could be used as a basis for whatever policies are required, so circumventing the danger of ad hoc decisions. He related this to the work on the structure of science by asserting that scientific change is centred around the activities of informal groups of scientists, which he called invisible colleges. He derived the name for these groups from an informal group of London scientists in the seventeenth century called the invisible college (the visible one being Gresham college) who met before the formal organisation of the Royal Society. He described the new invisible colleges or specialties as existing at the research fronts of modern science and composed of groups of competing, important people who send each other pre-prints of papers, meet at small select conferences and collaborate as co-authors. His linking of these ideas pioneered the use of citation bibliographic data in science policy studies.

Price, 1966, (68) then hypothesized that the structure of science could be mapped if it was regarded as a 3-dimensional spatial system. He produced one of the earliest descriptions of how a spatial model of science might look when constructed from

bibliographic citation data. He achieved this by combining the 2D model (which is based on the commonsense theory that each new paper is laid down on a foundation of earlier work), with the mapping hypothesis that units of information can be arranged on a plane so that the closer the units are together the more strongly they are related, and that these units correspond to subspecialties or invisible colleges.

The hypothesis that science is composed of small interlocking units gives rise to many questions, such as how do they relate? what is their nature? and can analysis of the structure of science help make science policy decisions?, all questions that had been put on the agenda by Price. Price's work thus paved the way for sociologists of science to move towards the study of the smaller aggregations of science variously known as research fronts, invisible colleges or specialties. The study of these being considered important because sociologists postulate that it is at this level, not at the level of disciplines, that scientists communicate.

Work in this area in the early 1970's includes that of S.Crawford, 1971, (18) who described how the specialty of sleep research is built around communications between a small group of key individuals and research centres. Griffith and Mullins, 1972, (35) who showed that small,socially coherent groups are capable of producing major changes in disciplines, and D.Crane, 1972, (16) who examined specialties in rural sociology and mathematics and concluded that specialties could be mapped.

Chubin, 1985, (8) described Crane's book as being theoretically bold. Crane argued that growth in science was a social diffusion process with invisible colleges acting as the main driving force. Chubin states that although Crane received accolade for many of her ideas, she was criticised for paying scant attention to the personalities or the intellectual content of the mathematics specialties she studied. He claimed that her work generated so much interest that this research area became a field in its own right, known as social studies of science, although this might be considered by some people to be a rather narrow definition of the field. Work in this area included that of

Mulkay et al, 1975, (54) who described the growth, emergence and decline of specialties. They proposed that specialties undergo three phases. The first is an exploratory phase characterized by a lack of effective communication and imprecisely defined problems. The second is a rapid growth phase characterized by increased social and intellectual integration, consensus being gradually achieved. In the final phase the area becomes less fruitful, declines and the network disbands as participants move to new areas. Mulkay et al, 1975, (54) maintained that the creation of scientific knowledge involves a complex web of social relationships cutting across formal scientific boundaries, making the boundaries of specialties unclear.

Sullivan et al, 1977, (92) also described what they believed were the phases of development of a specialty, which are similar in concept to Mulkay's. In its primary phase of development they believed it to be mainly intellectual, the informal social organization growing up around it afterwards as scientists realise that they are members of a group with a common interest. The specialty then undergoes a process of institutionalisation, ending with it being taught at undergraduate level. Sullivan et al concluded this from looking at the specialty of weak interactions in physics, examining its patterns of article production, demography and referencing for the period 1950 to 1972.

Chubin and Studer, 1979, (9) found evidence for Price's conjectures on the existence of invisible colleges by studying networks of authors in biomedicine. They showed that most of the highly productive and highly cited authors in the network worked together, and that the movement of researchers around elite groups accounted for the high degree of co-authorship and connectivity. They found that members of the central groups cited each other at a high level, whilst authors in the satellite groups cited other authors as well as the ones in the central group. Other examples of this type of research can be found in Chubin, 1985, (8).

The work in social studies of science has implications for those attempting to place these units, or invisible colleges, into a mappable structure. Early attempts at this include that of

Cawkell, 1974, (6) who constructed citation networks for the specialty of amorphous semiconductors and concluded from the results that nodal articles and authors could be identified and socio-scientific interactions understood by inspection of the network.

E.Garfield, 1979, (31) concluded from this and from research in the social studies of science that evidence had been found showing that specialties are the basic intellectual and social unit of science. Lester, 1977, (46) on examining the work of Mullins, Crane, Crawford and other sociologists also concluded that their work supported the existence of specialties and showed their role in science.

Then with the development of co-citation analysis (CCA) in 1974, (see Section 3 of this chapter) a new technique became available to study the structure of science. Price expressed great interest in it from the first and finally endorsed it in Price, 1979, (64) saying that it had shown his 2D model of science to be very workable. He believed that in the future algorithms based on CCA would show the map of scientific knowledge in motion and expressed hopes that it would develop a workable "science of science". CCA owes much of the justification for its view of the structure of the scientific community to him. A document produced by the Centre for Research Planning (CRP) in 1985, (20), that was probably authored by Len Simon and Roberts Coward acknowledged Price's contribution to the field of scientometrics. They credited Price's work with having played a critical theoretical role in the development of bibliometric modelling techniques. They claimed that his ideas about the structure of the scientific literature explained the meaning of the clusters of highly co-cited papers. However the ability of CCA to accurately represent and show the interaction of specialties remains controversial.

Section 3: The development of Co-Citation Analysis (CCA)

(a) Co-Citation Analysis in General

The technique of CCA was developed independently in the U.S. by Henry Small of I.S.I. in Philadelphia, 1973, (84) and in

the U.S.S.R. by I.V. Marshakova, 1973, (48) at the Central Soviet Information Service. They proposed that the greater the number of times a pair of documents are cited together the more likely they are to be related in content. The idea occurred to Small when he observed the high incidence of co-citation in the SCI, on comparing the lists of citing documents in it and counting the number of identical entries. Small, using an example drawn from particle physics, argued that this measure of relationship between a pair of documents could be used as a tool for elucidating the specialty structure of science. Small observed that the patterns obtained by co-citation differed considerably from the patterns obtained by another established measure of document linkage, bibliographic coupling, which was developed by Kessler as an information retrieval tool in 1963, (41).

Bibliographic coupling differs from co-citation in that it links source documents whereas co-citation links cited documents. Small, 1973, (84) theorised that co-citation strength (the number of articles citing two publications together, used as a measure of the strength of their association) measures the degree of relationship between papers as perceived by the citing authors. So the patterns change as the fields evolve, whereas bibliographic coupling depends on references in coupled documents and is therefore fixed. Small hypothesized that frequently cited papers represented key concepts, methods or data and that co-citation could show the relationships between these ideas, could thus be used to understand the mechanisms of specialty development, and hence to study the dynamics of science. This new technique made it possible to identify a specialty literature for the first time, and raised hopes that it could answer questions posed by sociologists relating to the structure of science.

Since co-citation proved to be so successful, H.G. Small in 1974, (86) introduced the concept of tri-citation, (which represents co-citation between 3 documents as opposed to 2), as its logical extension. He devised geometrical models (the circle and the hill) to account for these and all forms of multiple citation. However in spite of the development of these other types of multiple citation analysis, CCA was the one that was

subsequently most used, presumably because the other forms proved to be too unwieldy.

Then followed the mapping studies of Small and Griffith, in which they further developed Small's co-citation technique and which Garfield, 1979, (31), said were the most sophisticated attempts of the time to define the structure of science at the level needed for science policy purposes. Small and Griffith, 1974, (89) devised a method for identifying clusters of papers linked by specific levels of co-citation strength. They extracted from the SCI lists of highly cited documents and the papers that cited them, then listed the co-cited pairs of papers and ran this list through a single link clustering programme that aggregated clusters of documents by sequentially linking together all pairs having at least one document in common. Using this methodology they looked at the first quarter of the 1972 SCI and found that it succeeded in identifying papers linked by a common interest in the natural sciences, and that the clusters seemed to represent the specialties believed to form the structure of science (see Section 2 of this chapter). On examining the titles of the citing papers they found that from the pattern of word usage the co-citation clusters seemed to be linguistically consistent. However they observed that the correspondence between specialties and clusters was not perfect because of the highly cited methodology papers that appeared in many specialties. E. Garfield in 1979, (31) said that this study provided the first large scale evidence that science is a network of specialties that can be seen through the literature, and was also important because of the development of a method using co-citation as part of an automatic clustering procedure.

Then Griffith, Small et al., 1974, (36) extended their clustering methodology by examining the intra- and inter-cluster links of their co-cited paper clusters. They defined the intra-cluster links as representing the contents of individual clusters and inter-cluster links as giving a measure of the relationship between clusters. Using these definitions they introduced a measure called cluster CCA which assumed that when clusters are formed at a specified level of co-citation strength a linkage

measure can be calculated for each pair of clusters and weak links omitted by the application of a threshold value. They also introduced multi-dimensional scaling (MDS) into their methodology which pictorially showed the relationships between specialties by placing those with many links close together on the map and less related ones further away. Using these measures they obtained a network diagram showing how the clusters were linked and so obtained what they deemed to be a reasonable structural map of the literature on reverse transcriptase.

E. Garfield, 1979, (31) described the work done by Small in 1975, (77) in an unpublished report for the National Science Foundation (NSF). In this report Small introduced a technique called normalization of co-citation strength into his clustering methodology, where the measure of Co-citation strength was normalized to take account of the total citation rate of the pair. This technique was designed to overcome the problem that had been encountered previously in their study, 1974, (89), where groups of specialties had tended to aggregate into macro-clusters because of the effect of methodology papers that tended to be co-cited with many documents. The new method obtained a greater number of clusters and made fields of relatively low activity more visible. Using this improved methodology, Small was the first to produce a large scale map of the social sciences for the years 1972 to 1974 from the data in the SSCI. He used the data to compare the basic structural characteristics of the social sciences with those that had previously been obtained for the natural sciences.

Support for some of Small's assumptions came in 1976, when Price and Gursey published two papers, (66,67). In them they looked for the laws of emergence, survival and disappearance of the names of publishing scientists, and identified the phenomena of transience and continuance. Transients tending to be the least productive scientists who form the large proportion of those working at the research front, and continuants who tend to be the highest producers and also the most continuously cited. They thus concluded that Small's assumption that there exists a highly cited and highly prolific core group in science was justified.

Small's idea that connections drawn between papers by CCA indicate the presence of a specialty or invisible college is controversial in social studies of science. Chubin, 1985, (8) criticised the idea by saying that such a specialty may be an artificial construct and thus that membership in it may not mean that much. Chubin said that the idea of a research circle was probably more realistic in that it was less deterministic than Crane's invisible colleges. Such research circles would have no clear boundaries, with researchers running in several circles simultaneously, and be very fluid, ephemeral structures.

This pioneering work in the development and refinement of CCA was then followed by a number of validation studies, the purpose of which was to determine how closely co-citation clusters corresponded to specialties. H.G. Small in 1976, (87) found that the extension of his procedures to annual SCI files was straightforward. He produced a map of 5 natural science disciplines which showed chemistry as the integration point for the natural sciences and looked at biomedical clusters to see if their structure and behaviour from year to year matched general perceptions of what had occurred in the biomedical specialties for 1972 and 1973. Further details of this study can be found in Chapter 3 in the section on trend studies.

Then Small, 1977, (81) produced a more complete picture of cluster behaviour over time and how this matched real world events, by examining the literature on collagen research between 1970 and 1974. This paper also introduced a technique called variable level clustering, where threshold levels can be varied over a wide range to see the effect on the size of each cluster. In a cluster, representing a relatively discrete area, as the co-citation strength threshold is lowered, papers are only added gradually until at some point a sudden increase in cluster size occurs which is the level at which the specialty joins with others to form a macro-cluster. Small considered that the ideal threshold clustering level for a specialty was just above this point. Further details of this study can be found in the section on trend analysis studies in Chapter 3, but overall he produced a map of clusters of collagen over time which he claimed compared well with

historical data supplied by collagen workers in questionnaires. So the questionnaire data from this study supported Small's hypothesis that co-citation clusters correspond to specialties, and showed that the highly cited documents were the significant papers from the point of view of the specialists in the current or citing literature, and that their authors were the leading researchers. Small thus assumed that the highly cited papers represented the loci of current consensus.

In 1977, N.C. Mullins et al, (55), drew co-citation data from Griffith et al's study, 1974, (36), on the structure of the literature on reverse transcriptase and Australia antigen. Mullins et al sent questionnaires to a sample of 105 scientists from the file and by using a technique called blockmodelling (which although it is beyond the scope of this thesis to describe it, the relevant details can be obtained from their paper), showed that the social structure of the area appeared to correspond to the cognitive map of the area supplied by Griffith et al. They found dense connections between the most productive researchers (who were often co-workers and colleagues), attached to which were groups of students and younger co-workers. They described this as a centre-periphery structure. They showed that the authors of the two clusters of papers showed denser patterns of contact than do the members of other scientific specialties studied. They concluded that this supported the claim that highly co-cited papers are produced by members of scientific specialties. T. Lenoir, 1979, (45), went on from this to suggest that CCA and blockmodelling should be linked.

In 1977, Sullivan et al, (91) studied the field of weak interactions in particle physics for the years 1964, 1965 and 1966 to test some of the claims made for CCA made by Small and his colleagues using more traditional approaches to the study of the history of science. In support of these claims they found that the cluster corresponded well with the intellectual history of the field and that CCA succeeded in identifying its core literature or intellectual focus. However they also found that the authors citing the clusters were unrepresentative of the population of authors active in the field, that the papers in the citing or

current literature did not represent the full bibliography of the subject and that the cluster sizes were not a constant fraction of the specialty literature in each year. The latter criticism undermines the assumption that the growth of a specialty can be gauged in terms of the growth of its cluster (see the papers by Small, (81, 87)). They concluded from this that CCA needed to be used in conjunction with other indicators such as questionnaires, in order to overcome these weaknesses.

Then in 1979, Sullivan et al, (90), looking again at the field of weak interactions, performed a month by month CCA in order to test the techniques ability to follow changes in fast moving fields. CCA requires a years accumulation of citations to get enough data for precise analysis, so Sullivan et al used a long series of moving twelve month periods, in which the next period began with articles published one month later than the beginning month of the previous. They found that this gave an accurate portrayal of the rapid changes as they occurred, which they concluded gave credence to the assumption that shifting citation patterns are capable of reflecting shifts in the state of a specialty. However it was also concluded that due to important changes often occurring on a monthly time scale, annual CCA plots could obscure much of the dynamics of theory change.

Apart from this important work many other studies testing the technique of CCA and using it to examine the structure of the sciences and / or the social sciences were carried out. One of these was B.C. Griffith, 1979, (34) who looked to see whether CCA could be applied as well to the social and behavioural sciences literature as it had been to the natural sciences. He did a co-citation clustering on a three year cumulation, 1972 to 1974, of the SSCI to try to outline the structure of the social sciences. He found that CCA apparently shed light on the operation of the area by showing that it differed from science in that it was most influenced by small groups of researchers with old documents not being readily displaced by new research.

In 1979 H. Small and D. Crane, (88) in a study similarly concerned with comparing the sciences and social sciences, used CCA to examine a three year file of the SSCI (1972 to 1974) and

the 1973 SCI. They compared the clusters obtained for a natural science (particle physics) and for three social sciences (economics, sociology and psychology) to compare the growth and structure of these disciplines. Their results differed from the Griffith study, 1979, (34) in their finding that certain common structural characteristics between the science and social science clusters, suggested that knowledge was developing in parts of the social science disciplines in a manner similar to the natural sciences.

A slightly later study using CCA to examine the structure of particular scientific disciplines was J. and S. Crawford, 1980, (17) who used CCA to examine psychiatry research. Using data from the SSCI for the years 1971 to 1973, they found that CCA identified 155 clusters representing psychiatry related areas. They used this data to report on the distribution of publications, the rate of growth in research activity, and observed inter-relationships among the psychiatry disciplines.

Nadel in 1981, (56) compared an intellectual account of the history of the field of the physics of superconductivity with citation and co-citation data, using for data lists of articles on this topic in Science Abstracts for the periods 1957 to 1959 and 1960 to 1964 when the central explanatory theory for the area was introduced. He found that CCA supported his hypothesis of impact phases in the effect of the theory on the cognitive organisation of the specialty. He also observed that citation and co-citation data are separate types of information which sometimes give different results. He concluded that CCA obtained extra information to that which can be gained from citation frequencies or intellectual history, but, in sympathy with Sullivan et al., 1977, (91) thought that it would still need to be used in collaboration with other measures to get a complete picture.

Irina Marshakova, 1981, (49) used CCA to build citation networks in information science, using data from the years 1961 to 1973 of the leading Soviet journal in the field. She concluded that the network gave a detailed picture of the structure of the field in that period.

H.G. Small, 1981, (78) did a CCA of a three year cumulation of the SSCI, 1975 to 1977, in order to examine the discipline of information science. He identified and clustered information science documents and elucidated the internal structure of the field. He then explored its linkages to other fields in the social sciences, and drew conclusions about their relationships.

E. Nadel, 1983, (57) used CCA to test the opposing positions of a philosopher of science, Popper, 1972, (60) and a historian of science interested in its philosophy, Kuhn, 1962, (43) who were concerned with how the members of the scientific establishment accept or reject theories. Nadel looked at historic examples of theory competition taken from the fields of superconductivity and weak electromagnetic unification in physics. As Kuhn expected theories in cognitive competition to be incommensurable, whereas Popper expected all scientists to be involved in objectively testing them, Nadel predicted that a high degree of incommensurability would result in a low degree of co-citation. He found that in both cases when the competing theories were introduced they were cited together, then one of the theories lost favour and was no longer co-cited with the other, until it finally disappeared, which he concluded seemed to support Popper's position. This study was interesting as it demonstrates how CCA may be used to shed light on a highly controversial area in the philosophy of science.

(b): Co-Citation Context Analysis

The studies in the previous section on CCA had shown that the structure of the clusters used to indicate events such as conceptual shifts, could also show details of the social structure of specialties.

Then H.G. Small in 1978, (85) proposed a way to extend the capabilities of CCA by making it capable of showing the cognitive structure of clusters. To this end he used an idea obtained from work done by sociologists of science who examined the context of citation passages in order to analyse citations by their role or function. However Small modified the technique by

analysing the content of a citation rather than its function. He did this by looking at a set of highly cited chemistry documents and examined the text surrounding references to these highly cited papers. He found that the cited papers were generally cited in connection with one central idea introduced, and so claimed that cited documents represent concept symbols to the citing literature. Small called this technique co-citation context analysis, and his work in this area acted as a further validation of citation usage in the form of CCA.

It then became apparent that co-citation context analysis could be used to produce machine written reviews of the scientific literature. The first step on this road occurred in 1979, when Small, (83) used context analysis to examine the cognitive connections between cited documents. He achieved this by substituting sentences for each co-citation link on the map rather than for the cited documents, in order to derive a cognitive structure in the form of a network of interlocking statements. The statements were obtained by analysing the textual passages in the citing papers in which the cited documents in a cluster are co-cited and then, by substituting the statements for inter-document links on the map, translating the co-citation map into linguistic form. The statements thus obtained were taken to represent the intellectual links between cited papers as perceived by the citing authors. Using this technique he examined the specialties of recombinant DNA and opiate receptors and found a high degree of consensus. As one of the potential uses of the technique he proposed that it might be possible to trace the evolution of a specialty by examining changes in the system of interlocking statements.

The upshot of this work was that co-citation clusters could now be viewed as highly specific cognitive structures as well as social structures. In 1980, Small and Greenlee, (80), applied these techniques to a cluster of frequently cited papers in recombinant DNA, to translate the co-citation map into a linguistic structure. Citation context analysis was used to label each of the documents in the cluster with its specific cognitive meaning for the citing authors, and to show the relationships

between the concepts symbolized by the highly cited documents to obtain the cognitive equivalent of the co-citation links on the map. The results confirmed Small's earlier conclusion that this technique could open up new ways of exploring conceptual change at specialty level. Comparing the data obtained with more traditional approaches to the study of science, e.g. participant accounts and questionnaire data, they concluded that the map was like a snapshot of the field circa 1976 and thus that citation context analysis could be a way of probing the structure of paradigms.

The success of these studies enabled Small in 1979, (82), to assume that co-citation frequencies act as concept markers, with the frequency as a measure of the degree of consensus regarding them among the citing papers. In looking at the 50 most cited documents in the field of chemistry, displaying the data in the form of a graph with the concepts as nodes and their relationships as lines connecting the nodes, they again found a high degree of consensus in the citation contexts. These co-citation context analysis studies provided additional evidence for the idea of highly cited documents being concept symbols for the citing literature, an idea that was originally put forward by Small in his paper proposing the concept of CCA in 1973 (84).

In 1981, Small, (76) proposed that co-citation context analysis could be used to generate knowledge bases for use in information science. He proposed that if these were linked with on-line bibliographic databases a user would be able to move back and forth between document and knowledge retrieval.

In 1985, H. Small, (79) attempted to use co-citation context analysis to generate reviews of scientific fields. Using the technique he obtained the most representative passage of a citing document for a core document (which he now termed the consensus passage) which formed the nodes in the co-citation network, and the links between the nodes showing why the documents were associated (the transitional sentences).The final narrative was thus composed of consensus passages ordered by the network and connected by transitional sentences. Small demonstrated the technique using a co-citation cluster in the field of cancer

virology, and speculated on the possibilities of generating synthetic reviews of science as a whole from higher level maps.

Lenoir, 1979, (45) said that CCA might show shifts in the foci of interest, but not necessarily the cognitive structure of the area. This problem was dealt with in the work done by Small and others on citation context analysis (see Section 3(b)). However Lenoir made a valid criticism in stating that the core literature may not represent the origin of important contributions because of the technique favouring the clustering of documents after a line of research has gained wide acceptance. A. Rip and J.P. Courtial, 1984, (69) said that co-citation ties were indicators of a set of accepted authorities at best, and that it was an additional assumption to take co-citation clusters as reflecting cognitive structures, and were dismissive of Small and Greenlee's attempts, 1980, (80), to verify this using co-citation context analysis. However they did point out that as it examines scientific fields through the social institution of citing practices, CCA could only be applied where this occurs and has stability.

Related to Small's work on the word structure of specialties is research on co-word analysis. This rival technique to CCA is concerned with elucidating the structure of science by grouping papers together on the basis of the co-occurrence of keywords in their titles. Work of this type was presented for the ABRC Science Policy Study (see Chapter 3, Section 3). This overlap between co-word and co-citation context analysis opens up interesting possibilities for the creation of hybrid maps of science.

(c): Author Co-Citation Analysis

Another development in CCA studies occurred in 1981 when White and Griffith, (96), proposed a new approach to co-citation mapping by showing that authors, instead of documents, could be used as the units of analysis in mapping science, with co-citations of pairs of authors as the variable indicating their "distances" from each other. The basic assumption was that the

more two pairs of authors are cited together, the closer the relationship is between them. White and Griffith regarded author in this context as representing an oeuvre or body of writings by a person. They looked at sets of documents associated with 39 information science researchers, and found that a pair of co-cited authors names could represent a subject description, whereas names taken separately were found to have a much broader meaning.

Then White and Griffith, 1982, (95), used the technique of author CCA to test the coherence of the field of studies on science, technology and society from 1972 onwards. They looked for the number of times authors oeuvres were co-cited by later writers, the idea being that densely interconnected oeuvres represent a field of research, that can be used to map subject specialties. They concluded from their results that studies on science, technology and society was a coherent field.

Katherine McCain, 1983, (63), used author CCA incorporating clustering and multi-dimensional scaling to examine the work of 42 prominent macro-economists from the years 1972 to 1981. She concluded that the technique could identify specific links, and also found that co-citation of authors acted as concept symbols.

H.D. White, 1983, (94) used author CCA to map the literature of the social indicators movement, and found that the map conformed well with independent reviews of the field. He suggested that author co-citation mapping could be used as graphics to illustrate published reviews.

P. Lenk, 1983, (44), proposed a technique called co-nomination analysis which assumes that the more frequently two researchers are nominated together by other scientists the more closely related they are, so sampling the structure of the invisible college. He compared co-nominational data of the field of information science with the author CCA map derived by White and Griffith, 1981, (96). He looked at researchers appearing in both studies and found a high degree of concordance between the two groups, and so concluded that this indicates a relationship between the invisible colleges and the researchers citatory

behaviour. He concluded that the two interact and reinforce each other.

Section 4: Concluding Remarks on CCA in General

Small's work can be regarded as representative of the "strong" school of thought which advocates the use of CCA to study and plan the structure of science. This has brought CCA to the attention of those interested in the debate concerned with the validity of citations in general, see Section 1 of this chapter.

An example of criticisms from this quarter aimed at CCA is a paper by D. Edge, 1977, (25) who deemed citation generally to be a trivial behaviour, and was concerned that CCA would destroy the evidence of individual variations. Edge criticises the ability of CCA to elucidate the structure of science. E. Garfield, 1979, (31) countered this by claiming that Edge had ignored the advantages of analysing the general and average, and that there was room for both sorts of studies.

Another criticism that can be directed at CCA is that of E. Noma, 1984, (59) who discussed the problems of the assumption in co-citation analysis that all citing articles view the literature from a common point of view. This casts doubt on the validity of the assumption in CCA that the citing literature of a specialty represents the research front of a particular research area. Noma suggested that in order to overcome this, before a citation matrix is constructed, citing articles should be limited to those written by the members of an invisible college.

Irrespective of criticisms of this type, Garfield, 1979, (31) claimed that the evidence produced by studies on CCA has led to the development of a methodology that can open up the underlying processes of science. He predicted that the work would lead to the development of theories of specialty behaviour, and that science could be characterized in a way useful to science policy makers. The role of CCA in producing maps for policy making has also drawn criticism, for example from Irvine, Martin and Hicks. These issues are examined in Chapter 3, where the use of CCA in generating large scale maps of science is examined.

Chapter 3: Bibliometric Modelling and Trend Analysis Studies

Introduction

The previous chapter was concerned with the development of CCA in its various forms and its concern with the analysis of the structure of science. Also examined was the concern of academics such as Bernal and Price with the necessity of planning science in order for it to attain its full potential. This chapter examines how these various concerns have come together with the development of the large, comprehensive and detailed co-citation maps of science known as bibliometric models.

The development of bibliometric modelling is outlined against the background of increasing government interest in new ways of allocating resources within restricted budgets. Finally the ability of such models to follow the development of scientific fields over time is examined with particular reference to data of this type that was produced in the ABRC science policy study. The validation of the time series data in this study being the primary concern of this thesis.

Section 1(a): Co-citation Bibliometric Models

According to Gibbons and Georghiou, 1986, (33), the evaluation of research within OECD member countries became a ubiquitous activity after the oil crises of 1973 and 1974. In the recession that followed, economic activity slowed down while inflation and unemployment increased. Thus government research agencies entered a period of austerity and so began to critically review expenditure. The argument was put forward that new activities can only be initiated at the expense of existing ones. The necessity of deciding which areas are worth keeping coupled with the explosive growth of science led to resource allocation problems.

Thus the stage was set economically for the proponents of CCA to attempt to map science on a scale useful to policy makers.

This required whole disciplines to be mapped, not just specialties, and for the data to be presented in ways useful to an analyst, who is perhaps not familiar with CCA techniques. Such models (involving the large scale processing of citation data) were expensive to produce, but the demand by governments for such data meant that a market now existed for them. In response to this the Centre for Research Planning (CRP), a private company in Philadelphia with a license to use the Institute for Scientific Information (ISI) SCI and SSCI databases, developed bibliometric modelling.

In general bibliometric models are constructed on the following lines. CCA is used to define the subject area in terms of the specialties making up the research front, using data drawn from a source such as the SCI or SSCI. From previous validation studies (see Chapter 2) the model builders assume that the clusters of highly co-cited papers represent the intellectual core or base of concepts, methodology and data driving work in a given specialty of the research front, and that the papers citing these represent the current work of a specialty. Then the model builders place these two sets of papers into a computer to get a detailed descriptive model of the research front, the names of the specialties (or their intellectual focus) being derived from the titles of their current literature papers. The intellectual interaction between specialties is ascertained by calculating the numbers of times authors of current papers in the model co-cite papers in the base literatures of other specialties. The result is a hierarchical mapping that shows specialties within sub-disciplines within disciplines which places each research area within the context of surrounding research.

The models would appear to be particularly useful to policy makers as they also contain information in addition to the citation data. This is derived from the scientific papers in the models which, in addition to theory and data, also carry institutional and national affiliation, autobiographical and keyword data. This information is recorded in the SCI in association with the current literature papers and provides a detailed record of activity on intellectual, social,

institutional, national and international levels which is displayed along with the specialty data in the models.

L. Simon, an ex vice president of ISI now president of the Centre of Research Planning, first suggested the idea of using CCA as a descriptive model for generating inputs for research management and planning in 1981 (see CRP, 1984, (23)). The time was ripe for this type of work as several technical developments had come together which made mapping on the scale needed feasible. These included the various technical refinements to CCA (see Chapter 2) and the development of computing capable of handling data on the scale required. In order to facilitate work in this area, ISI gave Simon a licence to work on commercial planning studies, since when he has directed a series of field tests to develop the concept. Some of these studies are described below.

H. Roberts Coward, 1981, (12), also of ISI (and later of CRP), was given the task of developing the public science planning market for CRP. He clustered the entire SCI database for 1979 to see if CCA modelling could be applied to the planning and operation of international scientific exchanges of personnel. The aim was to match U.S. weaknesses with strong scientific fields abroad, which he claimed to be successful.

In 1983, Simon and Coward, see CRP (19), constructed a model of ceramics research activity from 1974 to 1981. The subject was depicted as a network of specialties, the interaction between them being used to aggregate specialties into higher regional levels. They described this as a pilot study to test these capabilities against a set of real world planning requirements. They found that although the interpretation of the model was not complicated it was labour intensive and concluded that more machine analysis was needed in order to overcome this.

Various governmental agencies showed interest in the work at CRP. One study commissioned by such an agency was that carried out by H. Roberts Coward et al, 1984, (15) who described a model of 1982 research in solid state processes and devices. This model was built by CRP in response to a request for data from the Division of International Programs of US (NSF) for a study in assessing the likely impact of expanding technology export

controls on research done in areas of epitaxy and microlithography. Franklin and Coward, in a paper submitted in 1987, (28) examined the science / technology interface by matching patent data to this CCA model and concluded that it could be useful for identifying strategic areas of research.

CRP produced a literature based model of Australian science in 1984, (23), using the 1982 edition of the SCI for data, which was produced in response to enquiries by Professor Jarlath Ronayne as part of the Australian Science and Technology Indicators feasibility study. This model was the first to use normalized citation counts, that had previously been developed in smaller scale CCA studies.

One of the most ambitious attempts to construct bibliometric models and validate them was that commissioned by the Advisory Board of the Research Councils of Great Britain (ABRC) for their science policy study in 1983, (14). This study is examined in detail in Section 3 of this chapter, since this thesis examines data from it. Five models From that study were later used by L. Simon et al, 1985, (75) as part of a study For the U.S.-Italy Programme of Cooperative Research. A sixth model was obtained from a study on 1982 research on solid state processes and devices, (15). The aim was to identify and evaluate areas of Italian scientific activity by identifying specialties with Italian activity. They concluded that the results demonstrated the use of bibliometric models for programme evaluation and planning.

Section 1(b): Other Bibliometric Modelling Techniques.

There are other methods of building bibliometric models apart from CCA. One of the alternative methods, which was examined in the ABRC Science Policy Study, is that based on co-word analysis. This originated and was developed largely in France by Turner et al (see Coward et al., 1984, (14)) and seeks to derive cognitive networks from the words used by scientists in identifying their work. The technique uses words that professional indexers ascribe to articles when compiling bibliographic

databases. The papers are then linked by the degree of co-occurrence of their key words.

Another approach is that of converging partial indicators which takes the view that bibliometric and other indicators are partial and imperfect, each reflecting a different facet of research performance. Martin and Irvine, 1984, (50) of the Science Policy Research Unit (SPRU) developed the approach in which they searched for convergence between several indicators. They regard the information derived from these indicators as only being reliable when all point in the same direction. Examples of these two alternative approaches to modelling are examined in more detail in Section 3 of this chapter.

Section 2: The Development of Trend Analysis Studies

The move from static (i.e. single time period slices arranged through time) models of the structure of science to longitudinal, dynamic ones (i.e. that link different time periods together), occurred early on in Small's development of his CCA concept. In H.G. Small, 1976, (87) he used data from the 1973 and 1974 SCI, clustered them at identical thresholds and linked the different time periods by looking for shared base literature (co-cited) papers between the clusters. In this he assumed that specialties in different time periods sharing citations to co-cited papers were intellectually linked. He noted a great degree of change of documents common to the cluster sets of each year and noted that the mean size of continuing clusters, i.e. those which contained one or more documents in the next time slice, were bigger than new or dropping clusters. He found that the most common pattern was the passing of documents from one cluster to another single cluster, with merging and splitting less common. Small proposed that clusters with all new documents may signify new areas of research and that continuing clusters with a higher percentage of new documents may signify an important change in the specialty. He observed that new specialties appeared to emerge and disappear rapidly within a fairly stable framework. This work was important as it provided an opportunity to move on from static

analyses which could only show a snapshot of scientific activity at any one time, to a dynamic one where the development of scientific fields could be followed. This opened up possibilities for examining the mechanisms of specialty growth and development with respect to the structure of science.

H.G. Small, 1977, (81) further developed his technique for the dynamic analysis of CCA models by looking at the development of collagen research. He carried out an identical clustering procedure on successive accumulations of the SCI, using the same technique as before in using the same thresholds to assess the growth of a specialty in terms of cluster growth, and tracing the continuance of clusters by the re-occurrence of highly cited documents. He further refined his technique by introducing a measure called the stability index (SI) or degree of continuity, calculated by dividing the number of documents that survive n years by the number of unique documents that appear in the cluster over the same period of time. He found that as the span of time increases, the SI decreases, until after three years the SI = 0, i.e. not one document survives over all the time periods, although some continued from 1971 to 1972, or from 1972 to 1973. He tested the resulting time series maps by sending a questionnaire on the recent history of collagen research to 24 collagen researchers, the results of which showed a history matching that produced by the citation data. He found also that the research area was composed of a small group of people with very close informal communication and frequent exchanges of personnel between a small number of research centres. Small concluded from this that co-citation seemed to have identified a research specialty with the attributes of an invisible college. He noted the tendency of methodological papers to persist as markers for a specialty through a shift, with theoretical contributions being more quickly forgotten, and speculated that specialty mergers may mark research areas in periods of rapid growth. For further details of these two studies see Chapter 2, Section 3.

Then in 1979, H. Roberts Coward, (13), proposed that time series co-citation data could act as the dynamic representation of research fronts, reflecting the changing problems and priorities

of the research community. He clustered the years 1972 to 1977 of the SCI, and obtained a string of related co-citation clusters by searching for shared cited documents. He said that it was possible to get a variety of data to describe these cluster strings, such as the average age, size, connectivity, geographic focus, relative stability etc.

Then almost inevitably, time series analyses of co-citation databases was combined with bibliometric modelling to lend a dynamic aspect to otherwise static models. In 1983, the CRP, (19), as part of their bibliometric model, looked at a time series view of ceramics research activity. This study was conducted for the Allied Corporation in order to test the descriptive powers of a time series of models. As part of this validation they posed some general questions about the properties of cluster strings or trends (as they were now known), such as are the short or long lived trends found in particular regions of the models? They concluded that the existing criteria for defining trends by searching for shared co-cited papers was too narrow since it implied that the disappearance of all parts of an intellectual base always signaled the departure of a specialty from the research front. They suggested that it could instead sometimes signify a change in the intellectual focus of a specialty remaining at the research front but disguised by a completely new set of base literature papers. They hypothesised that this could be due to transient changes of focus that might be overcome by searching for the re-appearance of some part of the intellectual base at a later date. So they recommended that the criteria of continuity should be broadened to include any time slices, not just adjacent ones, which they believed would lengthen some trends and integrate and identify others. However they also concluded that the events and patterns shown in the trend were characterised by the standard set of activity parameters, for example the size of the intellectual base, a conclusion that will be checked in the current thesis.

A different approach to lending a dynamic aspect to co-citation models was that taken by Peter Weingart et al, 1988, (93), who used a combination of interdisciplinary co-citation

model data from ISI to assess the strengths and weaknesses of West German Science. They combined co-citation data of the disciplines in each year with publication and citation counts to get a time series view of fluctuations over time. Their time series data only involved publication and citation counts, and they found difficulty in matching the co-citation data and the time series data together. This would seem to support Small's method of time series analysis as being the most viable at present. However the apparent success of Small's technique has led to some bold claims for its utility in bibliometric models. For example L. Simon, year unknown, (74), claimed that a time series picture of the evolutionary development of research fields gives a systematic survey of significant changes in emphasis, focus, momentum, relationships and direction. Jarlath Ronayne, 1983, (70) claims that it is these very properties that as they are responsible for the development of a research field, make trend analysis important for planning purposes. L. Simon, year unknown, (74) stated that the CRP's long range objective in trend studies was to develop a general theory of specialty behaviour to lend a predictive capability to the models.

One study involving bibliometric co-citation models that included time series data, was the ABRC science policy study described below. This study will be examined in some detail as trend or time series data derived from it is closely examined in this thesis in order to evaluate the validity of these claims.

Section 3: The ABRC Science Policy Study

The Advisory Board of the Research Councils of Great Britain is a government body which plays a co-ordinating role among the five organisationally separate UK research councils. According to Healey et al, 1986, (38) the ABRC is the body mainly responsible for providing advice on overall research allocations to the Secretary of State for Education and Science. However it has only limited personnel and financial resources of its own and so has relied heavily on information from the individual councils.

In order to overcome this problem of the lack of an independent data source that could be used to evaluate alternative investment possibilities for science policy makers, the AERC became interested in bibliometric modelling. In 1983 / 84 the AERC persuaded the Department of Education and Science (DES) to provide money to supplement that already given by the Economic and Social Research Council (ESRC) to fund a study that aimed to evaluate the contribution of several approaches to analysing the output of scientific research. The AERC commissioned studies from groups in France and the USA to study the structure of five scientific disciplines, one for each research council. These were; Ocean Currents, X - Ray Crystallography and Determination of Protein Structures, Mathematics, Cognitive Psychology and Protein Digestion in Ruminants. Bill Turner and his French colleagues at SERPIA (the Centre for the Study and Realization of Advanced Information Products) in Paris used co-word analysis, and CRP in Philadelphia, USA, used CCA to build models of all five disciplines. It was agreed that the data resulting from these efforts would be owned by the ESRC and freely available for secondary analysis. The work started in June 1983.

The French groups co-word method of analysis was based on identifying the co-occurrence of keywords in documents. They used keywords assigned to documents by skilled indexers, and modelled the five assigned topics by identifying appropriate databases, extracting the keywords and listing them in order of their frequency. Inter-relationships between words were then examined. The research fields were then characterised by using dominant word association patterns to show the hierarchical structure, and association of low frequency words to define problem areas.

John Irvine and Ben Martin of SPRU (the Science Policy Research Unit) at Sussex University in the U.K., were brought into the study in December 1983, to examine using their converging partial indicators approach, the areas of Protein Crystallography and Ocean Currents. Irvine and Martin used a statistical software package, SPSS, to examine publication and citation data from the Computer Horizons Inc. database used for the National Science Fdn. (NSF) publication Science Indicators. They intended to demonstrate

that an ad-hoc process of bibliographic file building, analysis and interviewing of scientists concerned, could provide a cost-effective alternative to the rather costly co-word or co-citation modelling techniques.

Peter Healey and Harry Rothman, 1984, (37), discussed the outcome of the work at the annual meeting of the DES in London in July 1984, and also later in Healey et al., 1986, (38). Validation of the models produced by each group took place in December 1983 and March and May 1984. The validation was unashamedly pragmatic in approach, in the form of workshops in which policy makers examined their usefulness and scientific experts in the fields represented by the models examined the extent to which they captured known cognitive and social relationships. On the whole they concluded that CCA accurately reflected the state of a discipline and described activity in the current literature, although they agreed that the CCA models needed more predictive capabilities. Co-word analysis, although considered potentially a powerful tool because of the wide range of document databases it could be applied to, was found to have generated an incomplete picture on four of the topics. Irvine and Martin's approach, although not as comprehensive as the other two techniques, was found to be more sharply focussed with respect to science policy, e.g. by identifying the centres producing the highly cited papers. However full analysis of Irvine and Martin's work was not possible in the study since it was not completed by the end of the workshop sessions.

Healey et al, 1986, (38) said that more rigorous validation studies would become possible when the models were made available for secondary analysis. One of these subsequent validation studies is the subject of this thesis which will restrict itself to examination of the CCA models, the construction of which will now be examined in more detail.

All of the CCA models were built in three steps, using data from the ISI citation index databases, SCI and SSCI. Firstly the subject of each model was defined by the relevant journal literature identified by the scientific experts, then specialty areas of research generated for each model from the papers in the

journal literature using CCA. The papers were subjected to citation and co- citation frequency thresholds to weed out low impact papers and weak co-citation links. The co-citation strength measurements were normalized and single link clustering used to cluster the reference pairs of papers into specialties. Then finally the specialties were aggregated into regional areas of research. A fourth step was then introduced for two of these models, Ocean Currents and Protein Determination research. The ESRC decided to test the ability of bibliometric models to show the dynamics of research activity over time, so additional models were built for these two subjects to get a time series view of their activity. This thesis will restrict analysis of this study to one of the CCA models produced by the CRP, Protein Determination in 1978, 1980 and 1982, and specifically to the trend data supplied in it. The time series or trend data was considered to be particularly worthy of secondary analysis in order to test the conclusion in the AERC study, see Healey et al, 1986, (38), that time series analysis was a particularly promising tool. This conclusion resulted from the observation in the study that the trends seemed to match their perspectives more closely than did the specialties, and from the suggestion that identifying trend areas of interest might be a good starting point for analysing the vast amount of data contained in the models.

Further details of the co-citation modelling procedures can be found in Coward's report, 1984, (14). The only aspect of these models that will be looked at in any detail is the methodology used to obtain time series.

3(a): Time Series Analysis of the Protein Determination Models

In the AERC study a time series analysis was performed using Small's technique to link specialties in different time periods that shared a significant percentage of base literature papers. After Small this was assumed to indicate specialties that are linked by some degree of continuity of intellectual focus and effort, described in this study as a trend area of research.

A total of 333 trend areas of research were identified in the Protein Determination models. These contained 1,060 specialties in all. The 1987 model had 39.3% (285) of its specialties in trends, the 1980 model had 60.9% (424) and the 1982 model had 34.4% (351). The high value for the 1980 model was an artifact of its position in the series of three models as its central location would allow it to pick up all the trends. Most of the trends were found to be very simple with only two specialties in adjacent time periods, a few were more complex showing the fusion and fission patterns of several specialties. Some of the trends were found to finish in 1980, some began in 1980 and some continued through all three time periods.

Once the component specialties had been linked into trends, the CRP described the properties of the specialties using the following indicators;

(i) The names of the trends were derived from the names of their component specialties, which in turn were derived from the titles of their current literature papers. The name was then taken to indicate the intellectual focus of the trend or specialty.

(ii) The size of the intellectual base or base literature of each of the component specialties was taken to equal the number of the co-cited or reference papers driving current research activity. This is often referred to by Small as co-citation clusters.

(iii) The average age of the intellectual base equalled the average age of the reference papers and was assumed to indicate the development potential of the research area. An alternative way of measuring the age of the intellectual base was taken to be the age of its youngest paper, which the modellers regarded as being a more sensitive indicator of new concepts, methodology or findings being injected into the research area. In both cases the lower the average age, and the lower the age of the youngest paper, the more rapidly advancing the specialty was assumed to be.

(iv) The density of the intellectual base of a specialty was measured by the number of co-citation links between its reference papers and was normalized by the number of papers available to be

linked. This is assumed to represent the cognitive cohesiveness of the field.

(v) The level of activity at the research front of a specialty (its current activity) was assumed to equal the number of current papers. This is referred to as the current literature in the models.

(vi) The level of interaction between one specialty and others in its discipline was measured by the number and strength of its citation links to other specialties. This measure utilises citation of one specialty's base literature by the current literature of another. If a particular specialty had many such links it was assumed to indicate that it was mainstream to the field of research, and if few that it was periphery.

(vii) The demographics measured national participation in a specialty by counting the number of current papers produced by each country. This was expressed as a percentage of all current papers in the area and also by national rank. For the national rankings, data was presented for the top two nations and, since the models had been commissioned by the British Research Councils, if the UK was not among them it was also shown. In order to get a measure of UK activity in each specialty subject area relative to the overall performance of the UK in each model, an expectation ratio (ER) was calculated. This was obtained by normalizing the percentage of British papers in each specialty by the percentage of British papers in the whole model. If the ratio was found to be greater than one in any particular specialty it was assumed that British activity in the area was higher than the model average, and if less than one the converse.

(viii) The degree of change in the intellectual focus between linked specialties within a trend area of research was measured by the number of reference or base literature papers that a pair of specialties have in common. This was normalized by the total number of papers in their base literatures.

Using these indicators the average trend base literature paper age was found to have decreased or stayed the same for 17.1% (57) of trends. These were assumed to represent the "hottest" research areas. British and American participation in terms of

numbers of publications was found to have steadily increased over the years, but both lost ground relative to the rest of the world in terms of percentage shares of total publications.

Attempts were made in the validation workshops to evaluate trends as a measure of on-going research activity. One of the techniques used to analyse trends was to use each specialty in a trend as a starting point for showing a selective region of research activity. This was done by using interaction linkages to draw other specialties into the diagram and aimed to show the place of individual trends in a broader context of research. The workshops found that this made it possible to see patterns of parallel and divergent development. On examining sequential regions in the different time periods, the scientific experts concluded that it showed the capability of time series models to capture the dynamic development of research subfields and give insights into the origins of current research. The patterns were found by the experts to be intuitively appropriate to the evolution of the research fields involved.

British performance was then analysed using the performance indicators associated with the trend data, the results being displayed in the form of quadrants. The upper left part of the quadrant contained trends for which the current activity greatly increased or the average base paper age greatly decreased. This was assumed to indicate areas of growing scientific importance where British performance appeared to be high. The lower right quadrant contained trends where current activity greatly decreased and the average base paper age greatly increased which were assumed to be areas of both diminishing scientific activity and British participation. The lower left quadrant contained trends for which the current activity greatly decreased and the average base age greatly increased and where British activity greatly increased. The upper right quadrant contained trends where current activity greatly increased and average base paper age greatly decreased and where British activity was low. The potential application of such indicators in science policy was clearly demonstrated by the subsequent use of the data which was

used to assess Britain's performance in growing or declining fields to help make judgements in resource allocation decisions.

Overall it was concluded by the study that the results obtained in the trend validation were encouraging, but were limited in scope as only a few of the large number of trends in the models could be analysed in any detail. Coward et al, 1984, (14), claimed that they were found to accurately reflect the intellectual evolution of scientific activity and that network diagrams could show parallel trends passing through a given region or sector of the sequential models, linked by continuing interaction among the trend specialties. They concluded that further analysis would be useful, to perhaps get a better idea of the ways that research trends fuse or fission. They claimed that research along these lines could lead to adding a predictive capability to bibliometric models. Coward et al also thought that another area to be explored was the proportion of continuing research trends, and asked what was the explanation for the specialties that completely pulled up their intellectual roots (i.e. none of their base literature papers continued into the next time period), and so disappeared from the trend networks.

Section 4: Subsequent Studies on Trends

The positive results of these various studies encouraged CRP, 1984, (23) to extend their claims for the validity and utility of trends by saying that the consensus among these studies was that this type of model data was complete, accurate and rich, particularly with respect to the time series. The only problem they acknowledged was the sheer size of the models. CRP described attempts to overcome this by making the data more manageable with machine readable databases and software. In this study they seem to have gone beyond the testing of trend data to seeming to take it for granted. Therefore there seems to be a need for the present thesis to re-assess these claims particularly as the system that they propose would routinely provide time series view of the demographics of research areas to identify active organisations and nations and their activity patterns over time.

A very recent study looking at trends is that by Diana Hicks, 1987, (39), a post-graduate student of Martin's at SPRU, who compared co-citation data with a manually generated bibliography of a physics specialty, spin glass, over time. The co-citation data used was supplied by Henry Small and spanned the period from 1976 to 1983. The clusters or specialties were linked from year to year via the sharing of base literature papers. The linked clusters were referred to as cluster strings rather than trends. Hicks claimed that CCA was inconsistent when used to track national participation and relevant material in a specialty over time and that it under-represented experimental work. Her observations caused her to express concern, in line with the general reservations expressed by Irvine and Martin, over such a tool entering the science policy arena.

J. Jeffrey Franklin, 1988, (28), responded to her paper. He criticised Hicks' reliance on the judgements of one scientist in the evaluation of the co-citation data. He also regarded the under-representation of citing authors as not being significant. He claimed that the purpose of the data was to show relationships between research areas, the technique necessarily being highly selective about the publications that it includes due to it having to meet co-citation thresholds. His critique of Hicks' work led to a debate between them, which he describes in his unpublished paper of February 1988 (27). In this paper he describes Hicks as having responded to his criticisms by alleging that he had misread her paper. In answer to this he confirmed that his previously held opinions of her work still stood and that he had additional criticisms to make. Amongst these was his opinion of her attempts to classify papers as being either experimental or theoretical from the basis of a personal judgement. He regarded Hicks as being unqualified to make this type of judgement due to her lack of expertise both in the field of spin glass research and in the use of bibliometric model data. He claimed that Hicks had also made the mistake of concentrating on a single specialty cluster and then attempting to extrapolate her conclusions to criticise all model specialties. Franklin also criticised her assumption that publication counts and bibliometric models were directly

comparable, whereas he regarded them as being different tools suitable for different purposes. He believed that the techniques she used to select items for her bibliography, for example the use of a single database and two key terms, were no more objective than those compiled by bibliometricians. However in spite of all these objections he supported Hicks' questioning of the objectivity claims made on behalf of bibliometric models and her reservations concerning the methodologies used to construct them. He summed up by saying that more research was needed in this area.

L. Leydesdorff, 1987, (47) criticised policy orientated studies such as that conducted by the ABRC by saying that too much emphasis was placed on the validation of information derived from CCA models and too little attention paid to the methodological assumptions underlying the model building. Leydesdorff argued that some of these assumptions, particularly those concerning cluster analysis, have been basically wrong with little re-examination since their inception, and that co-citation maps in their present form do not accurately represent the structure or dynamics of science. Leydesdorff's criticism of cluster analysis resulted from his own work on journal-journal citations where he observed that different clustering techniques led to markedly different results. Leydesdorff proposed that as proponents of CCA believe that invisible colleges can be described by it, the detailed maps obtained by CCA should be validated at the level of the scientific enterprise. Leydesdorff claimed that the real question concerns what CCA clusters actually represent. Leydesdorff also considered time series analysis to be unreliable from examining co-citation data from the Dutch Advisory Council for Science Policy Study in 1983, which attempted to examine the use of science and technology indicators in national research and development projects.

Section 5: Aims of the Present Research:

The aim of the present study was to extend the trend validation aspect of the ABRC Science Policy study by examining two of the trends contained in the Protein Determination models in detail. Time was not available for an in-depth analysis of the

properties of individual trends in the ABRC study, an omission which this thesis intends to rectify.

By examining these trends in detail, it was hoped to evaluate claims made for the ability of current trend modelling techniques to portray the dynamics of continuing research. In order to do this a methodology of trend analysis was developed to closely examine the properties of the literature composing the trends, to study the validity of the relationships over time that had been imposed on it by the technique. This was intended to expose any problems with the methodological assumptions underlying time series analysis. The next stage of the analysis involved the use of questionnaire and interview data to see how well the dynamics of research in one of the trends corresponded to the experiences of scientists who had worked in the field. The use of questionnaire and interview data enabled CCA to be evaluated at the level of the scientific enterprise, in accordance with Leydesdorff's suggestions (see Section 4).

That this work is important is highlighted by the claims subsequently made by CRP in 1984, (14) that the validation of the CCA models produced for the ABRC Science Policy Study by the scientific experts showed that they were capable of providing inputs to science policy questions. Since trends would presumably form an important part of this process it would seem essential to look at them in more detail than was the case in the ABRC study. In the same document CRP claimed that they hoped to build single multi-disciplinary models of the whole of science in any given year. Judging from the display of models created by Henry Small from ISI at the British Association for the Advancement of Science Conference at Bristol University, UK, in 1986, they seem to have gone some way towards achieving this.

Chapter 4: Analysis of the Base Literature of Trend 44 "Transposons and Insertion Sequences of Bacterial Plasmids and Chromosomes"

Section 1: Introduction to the Analysis of Selected Trends from the Protein Determination Models.

In the ABRC Science Policy Study (see Chapter 3), the trends, time series or cluster strings from the co-citation Protein Determination models were analysed by examining their links with regional areas of scientific activity or by using the indicator data supplied in the Specialty Summary Reports to characterise the trends and analyse British performance. This validation appeared to show that trends accurately portray the dynamics of research activity, a conclusion of some importance considering that the same study showed them to be of interest to policy makers. Concern over the use of this type of citation data in policy making has come from many quarters, notably from Edge, Leydesdorff and Irvine and Martin (see Chapter 3). Bearing in mind the arguments from these workers, it appeared to me that the limited time available for the examination of trends in the ABRC study meant that their validation was essentially superficial and that an in-depth analysis of their properties was needed in order to evaluate some of the claims made for trends as a result of that study. One of the claims made by CRP, that it was hoped such an analysis would test, was that trends (and especially the indicator data supplied with them) could be used to develop theories concerning the mechanisms of specialty development. In order to test these claims it was decided to select two trends from the CCA X-Ray Crystallography and Protein Determination models (generally referred to in this thesis as Protein Determination for the sake of brevity) produced for the ABRC study and develop a novel methodology for their analysis. The analysis was restricted to two trends because time was not available to study in depth the properties of all the 333 trends from the models. The aim was to design a methodology that was capable of an in-depth analysis of

these trends in order to elucidate their properties, and which was sufficiently different from the analyses used in the AERC study to be able to test that studies conclusions from a different perspective. The details of the methodology that was designed in line with these aims is described below.

Section 2: Rationale Behind the Trend Analysis Methodology

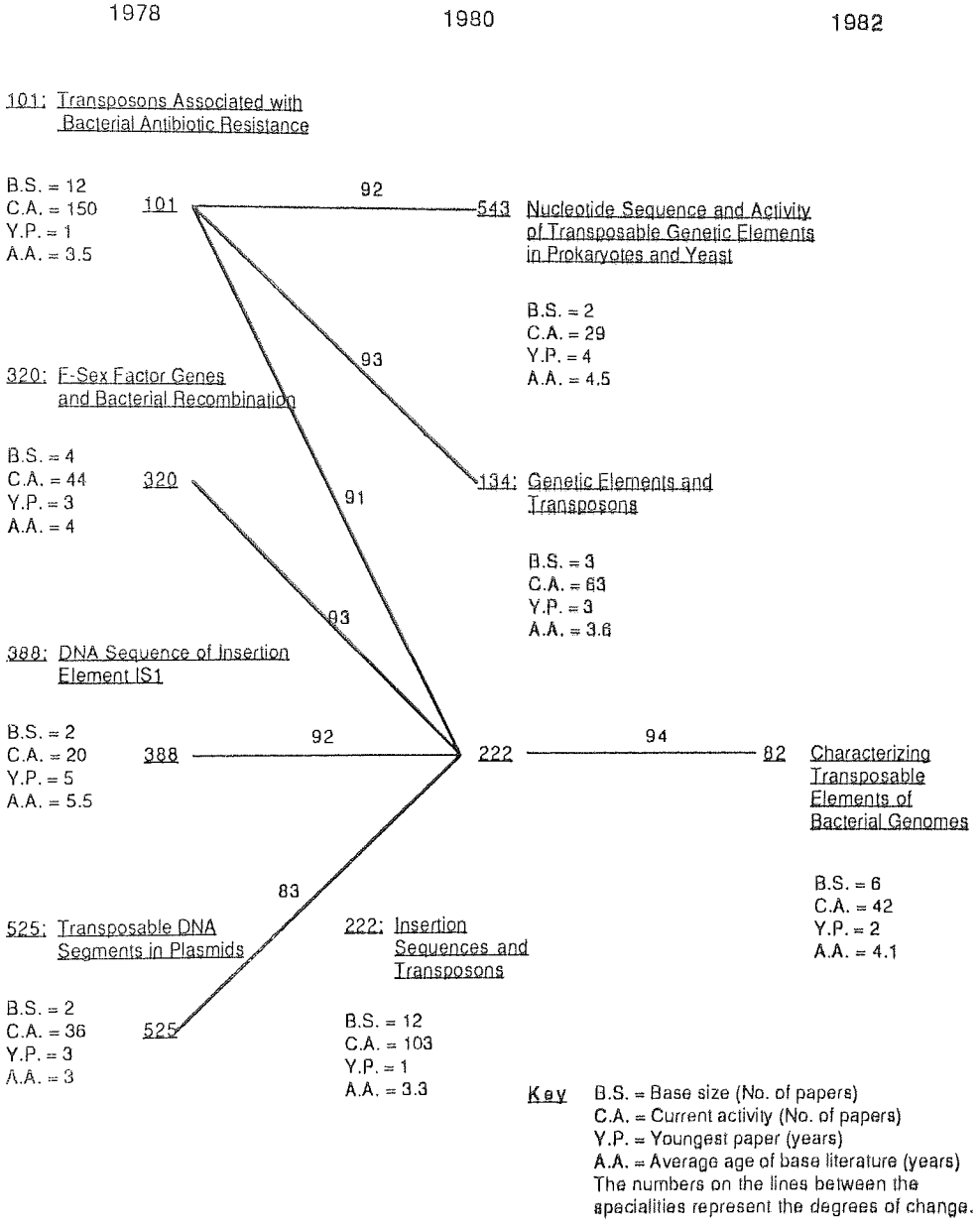
The main criteria for the selection of the trends was the complexity of their structure. Trends from the Protein Determination Models can be considered to roughly fall into two categories, simple and complex. The "simple" trends only show a single link from a specialty in one time period to one in the next. This is in contrast with the "complex" trends which show the patterns of merging and dispersion, or fission and fusion, of several specialties over time. Most of the trends examined in the AERC Science Policy Study (see Chapter 3, Section 3 (a)) were found to be very simple in that they only consisted of two specialties in adjacent time periods. Only a few of the trends were found to have a more complicated structure than this. Since complex trends can be considered to potentially provide more information concerning mechanisms that may underlie specialty development and change than the more commonly occurring simple trends, it was considered that examining this type of trend in order to test some of the claims made by CRP might prove fruitful. The key characteristics of trends and their component specialties are described in Chapter 3 Section 3(a). The two trends selected because of their interesting, highly complex structures, which made them among the most complex of all the 333 trends in the Protein Determination models, were those denoted by the numbers 11 and 44 in the models. Trend 44 was selected as being feasible for in-depth analysis within the time constraints of this study because of its manageable size. The large size of Trend 11 meant that an analogous in-depth analysis of it was not possible, and so analysis of the literature that it contained was restricted to verifying certain aspects of the methodology developed for the

examination of Trend 44. The structure of Trend 44 is shown in Fig.4.1 and the structure of Trend 11 in Fig.5.1.

H. Rothman, 1984, (72), posed some questions about specialty behaviour that he believed could be answered by dynamic or trend analysis. These questions included queries concerning the size of the specialties, which organisations are active in them, which specialties have changed their intellectual focus, which have had a turnover of dominant organisations, and which are growing, decaying or falling behind the research front. Modellers at CRP assume that these questions can be answered through examination of trend data sources such as the Specialty Summary Report, or the organisational Indexes for the current literature. Therefore it is important to ask whether this data can provide useful information capable of answering these types of questions, which are likely to be typical of those asked by policy makers. In order to evaluate this, knowledge of whether the trend accurately portrays the scientific research that it is supposed to represent is essential.

The development of a methodology for the analysis of the properties of trends 11 and 44 in this thesis, in order to answer some of these questions, aimed to produce a methodology different from that used by previous researchers (see Chapter 3). This was intended to overcome the apparent superficiality of some of these earlier studies. The methodology was designed to examine the literature of the trends in detail, in order to test the validity of the relationships shown between specialties over time by the trends with respect to information derived from the papers they contain. In the case of Trend 44 the validation was later extended by testing information derived from the literature analysis in a questionnaire sent to, and interviews with, the scientists identified from the papers, to see how accurately the trend portrayed the development of the field of transposable elements. This analysis attempts to provide answers to some of Rothman's questions, 1984, (72), something which has to be done before time series analysis can be considered a reliable tool for science policy.

Figure 4.1 - Trend Survey Data Combined with the Trend Diagram for Trend 44.



In the analysis of the literature of Trend 44, detailed analysis of the actual papers was restricted to the base literature due to the size of the undertaking. The base literature contains 43 papers (if each specialty is examined separately), and 34 when the trend is examined as a whole, (and the duplicated or linking papers are taken into account). It was thus a more manageable size for analysis than the current literature which contains 487. Definitions of the terms base and current literature can be found in Chapter 3 Section 3(a). Analysis of the properties of the base literature was considered to be especially important because it is at this level that links between the specialties in different time periods are drawn by the model builders, see Chapter 3 Section 3(a). Therefore it was considered likely that any problems with the assumptions underlying the techniques of trend modelling would be located at this level.

Thus the overall aim of the analysis was to test the underlying assumptions and hypotheses associated with trend modelling by close examination of literature associated with an area of on-going research that it claims to represent. The aim of the literature analysis was three-fold; to validate a trend as it stands, locate any weaknesses in the modelling techniques and to develop possible remedies.

The analysis began with papers from the base literatures of each specialty in Trend 44 being obtained from library sources and examined in detail. This was necessary because the CCA model database only gives the first named author and reference for these papers, in contrast to the wealth of information supplied for the current literature, such as paper titles, co-authors, organisational address and national ranking. The disparity between the base and current literatures occurred because of limitations in the original SCI database where details of the base literature or co-cited papers are only obtained from the reference sections of the papers in the current literature. The analysis aimed to get the same information for the base literature in order to develop an understanding of the characteristics of its papers and their relationship to the trend as a whole.

One of Rothman's questions asked which specialties have fallen behind the research front. This study assumed the research front of a field to be represented by the specialties present in the most recent time period of the trend, and specialties that have fallen behind to be those that fail to continue into this time period. In the case of Trend 44, specialties 543 and 134 do not continue into 1982, so these were closely examined to see whether they have fallen behind. This question was considered to be particularly important since it concerns the survival and mortality of specialties, which allows us to examine a trends ability to elucidate any underlying mechanism of specialty development.

The analysis used the following information which was obtained from each base literature paper: The title, the names of the co-authors, the institutional affiliations of the authors, a summary of the papers cognitive content, details of any acknowledgements, and sources of financial support. The following information was obtained from the papers reference lists: Citations to other papers in the base literatures of their own specialties; citations to the work of other first named authors in the base literature of the trend, and the total number of references in each list. A card was made out for each first named base literature author on which was recorded which other base literature authors, and in which specialties, cited them. The citation analysis was restricted to the first named authors because it was performed as the papers were looked up in the library. Thus as the co-authors names could not be known until all the papers were examined, analysis of citations to them would have necessitated a secondary examination of the papers. Time was not available for this.

Using this basic data the analysis proceeded on the following general lines. Firstly the specialties were examined individually to look for evidence of the existence of any invisible college, to examine the groupings of papers into co-cited pairs and to gain information with respect to their characteristics, e.g. the institutions involved. Then information obtained from the specialty analysis was analysed in the aggregate

to test relationships shown between specialties by Trend 44. This stage of the validation concentrated on testing the criteria of continuity, i.e. the continuation of base literature papers, that is used by the modellers to link the trend specialties together. The detailed methodology and results of these analyses for Trend 44 are examined below, firstly at the level of the individual specialties, and secondly at the level of the trend. The analysis of the base literature of Trend 11 is contained in Chapter 5. The scientific content of Trend 44 is examined in Chapter 7.

Section 3: Methodology and Results for the Individual Specialty Analysis.

The specialties were examined individually to see what the base literature papers showed about the relationships between the ideas and people who provided the intellectual foundations for each specialty in Trend 44. For each of the component specialties the following items were analysed: the base literature and the correspondence between its cognitive content and the cognitive content of the research front of the specialty as indicated by the specialty title; institutional origins of the base literature papers; agencies that funded the base literature research work, links between papers through common authors and the journal structure of the base literature.

The cognitive analysis of the papers examined the correspondence of the titles of the base literature papers with the titles of their specialties. The specialty title is generated from the current literature paper titles using a computer algorithm which selects the most commonly occurring keywords in the paper titles and uses these to synthesize a title that is assumed to represent the common areas of interest of the papers concerned. So this was done to test whether the specialty title was also an accurate indicator of the content of the base literature. It was hypothesised that if this was not found to be the case, could this property act as a possible indicator that the research front is progressing rapidly and moving away from its research base? The cognitive analysis was extended to test the

validity of the grouping of papers in co-cited pairs by the citing actions of the current literature. Two methods were used to approach this. Firstly the titles of individual papers grouped together were checked against each other for similarity of meaning. It was hypothesised that if by looking for shared keywords no correspondence between the paper titles could be found could this indicate that two unrelated areas of research have been brought together in a novel way?, or alternatively that the co-cited pairs do not make cognitive sense? If the former is the case could this measure be used as an indicator to point out areas of interest in a computer analysis? Secondly citation of base literature papers by other papers in the same specialty base literatures, were analysed to see whether citation between such papers corresponded to the co-cited pairings. This was designed to test the assumption in CCA that the current literature co-cites papers that are intellectually related in some way, by seeing whether these links are also perceived by the base literature authors.

Personal links between base literature papers that might be attributable to the presence of an invisible college were firstly examined by searching for authors common to them. Institutional affiliations were also examined because it was considered likely that personal links were more probable in papers originating from the same institutions. Secondly it was considered interesting to see whether the base literature research originated in the public or the industrial sector, and to prepare national rankings for the base literature that can be compared to those already given for the current literature by the model builders. Similarly sources of finance were examined to see from which sector the funding for each research project had originated.

The journal structure was examined for each specialty to see if the problem area concerned had a dominant journal(s). It was considered that this might indicate a representative or "paradigm" journal for the topic.

Thus for each specialty the results of these analyses are summarised as follows. Firstly the co-cited pairs of papers are examined for similarity of subject matter as determined by

their titles and for citation links between them and the other papers in their specialty base literatures. Links between these papers through shared co-authors are also shown. Then the institutional origins of the papers together with their sources of funding and the journals in which they were published is examined. The institutional and funding data is then used to prepare a national ranking for the base literature showing the position of the UK, which is compared to the rankings already available for the current literature. The first specialties analysed in this manner were in the 1978 Protein Determination model.

(1) Specialty 320: F-Sex Factor Genes and Bacterial Recombination.

This specialty had four papers in its base literature. On analysis of the properties of the co-cited pairs of papers it was observed that the papers were apparently very closely matched with respect to the cognitive content of their titles. All of the papers were observed to be on virtually the same subject, i.e. electron microscope heteroduplex studies of sequence relations among the plasmids of E. Coli, showing good cognitive links. However the titles of the papers did not correspond well with the title of the specialty derived from the current literature, but from the authors own knowledge of biochemistry it was considered that the base literature papers would make a plausible intellectual base for studies on genes and recombination since the techniques of sequence analysis would be important to it. The papers were observed to cite their co-cited partners, and other papers within their own specialty base literature. Where citation of a co-cited partner did not occur this was observed to be due to the date of publication, i.e. the "citing" paper preceded the other. As far as institutional affiliations are concerned it was observed that all the papers originated in the Department of Chemistry, California Institute of Technology, and were funded by U.S. public funds. The only journal to have more than one paper was the Journal of Molecular Biology. See Table 4.1.

In the co-author analysis it was observed that every paper had been co-authored with other authors in the specialty,

see Table 4.2. It was concluded from the co-authorship, citation and institutional data that the base literature authors of this specialty form a very close knit group.

The national rankings obtained for the base literature papers were compared with the national rankings in the current literature, see Table 4.3. It can be seen from this that all the base literature work was carried out in the public sector in the USA, this country also being the leader in the current literature. The UK is not represented in the base literature and is a poor third in the current.

(2): Specialty 101: Transposons Associated with Bacterial Antibiotic Resistance.

This specialty contained 12 papers in its specialty base literature. Analysis of the co-cited pairs of papers showed that cognitive similarity between co-cited pairs of papers as deduced from their titles was found to be good. The titles of the base literature papers matched very closely the specialty title given by the current literature. It was also noted that they do generally cite their co-cited partners. When this did not occur it was again observed that the "citing" paper was either received for publication or actually published before the other. Some of the papers did not supply information on the sources of their financial support, but those that did indicated that it came only from public sources in the countries concerned. Examination of the organisations that were involved showed the dominant ones to be the Royal Postgraduate Medical School in the UK, which produced three papers, and the Department of Microbiology, University of Washington, with three papers. It is also interesting to note that these two organisations were linked in one of the papers. All of the institutions were observed to be in the public sector. The journal distribution of the specialty showed that the Journal of Bacteriology was the dominant one with five papers, closely followed by the Proceedings of the National Academy of Sciences with four papers. The other journals appearing in the base

literature had only one paper each. The data for this is presented in Table 4.4.

When co-author links between the papers were examined it was observed that only four papers were co-authored with other people in the specialty base literature, see Table 4.5. When the national rankings were examined, see Table 4.6, it was seen that there was a total of 5 countries with the UK and the USA ranked equal in the base literature, but with the UK falling behind in the current.

(3): Specialty 388: DNA Sequence of Insertion Element IS1

There were only two papers in this specialty, so there was only one co-cited pair. As both are concerned with IS DNA sequencing the cognitive link between them is very good. There was no sharing of co-workers and no organisation appears more than once. Financial support came from the public sector in both cases. The Saedler paper cited the Fiandt one, but not vice versa, however as the Fiandt paper preceded its co-cited partner this is not surprising. Only one journal was involved, Molecular and General Genetics. The titles of the base literature papers matched the specialty title given to them by the current literature. All of the base literature work was done in the public sector. This data is presented in Table 4.7.

The national rankings, see Table 4.8, show that the USA and W. Germany are equally strong in the base literature, although the USA dominates in the current. The UK was not represented in this specialty.

(4): Specialty 525: Transposable DNA Segments in Plasmids

Again there were only two papers in the specialty base literature and therefore only one co-cited pair, however they were cognitively very related due to both being concerned with insertion sequences in plasmids, see Table 4.9. There were no shared co-authors. No organisation predominates, each has only one paper; three organisations were involved, two being linked in one

paper showing US and W. German collaboration. There was evidence of strong citation links between the two papers. The Hu et al paper cited the Ptashne et al paper in its abstract, indicating that it must have been regarded as being very important. The paper by Ptashne et al cites the other in its reference section. As both papers were received for publication on the 16th of December 1974 there is evidence of pre-publishing collaboration and therefore strong conceptual and personal links between them. Since they turned up as a co-cited pair the current papers must also recognize this link. The titles of the base literature papers are concerned specifically with insertion sequences. The specialty title is much broader in scope, but they do not seem unrelated. Only one paper was represented in the base literature, the Journal of Bacteriology. Financial support for both papers came from the public sector in the countries involved.

In the national rankings, see Table 4.10, the US leads in the base literature with W. Germany in second place, a state of affairs that also occurs in the current literature. The problem of the two countries being linked in one of the papers was overcome by giving a score of one paper to each country for ranking purposes. The UK was non-existent in the base literature and not very well represented in the current.

The following specialties were in the 1980 time frame of the models:

(5): Specialty 222: Insertion Sequences and Transposons

There were twelve papers in the base literature of this specialty. Comparison of the paper titles showed that there were strong cognitive similarities between the papers and their co-cited partners, see Table 4.11. The titles of the papers also matched the specialty title derived from the current literature. Financial support for all the papers originated from the public sectors of the countries involved. The dominant organisation was seen to be the Institut fur Genetik in Cologne with three papers, (one of which showed collaboration through co-authorship with a US

institution), followed by Harvard University with two papers. The papers were seen to have a tendency to cite their co-cited partners and other papers in the specialty that they were not directly paired with. Where citation of co-cited partners did not occur this was invariably found to be due to the date of publication. Thus it was concluded that the co-cited pairs show strong citation links. The most frequently occurring journal was Cell with four papers.

In the co-author analysis, see Table 4.12, four papers were co-authored with other authors found elsewhere in the same specialty. Analysis of the national rankings, see Table 4.13, showed that the USA was the leader in both the base and current literatures. Germany, although in second position in the base literature, does not get ranked in the current. However this may not be significant as the figures are generated to show the position of the UK if possible, but no more then three countries are ever shown by the model builders. Interestingly enough, Switzerland jumps from a lowly position in the base literature to second position in the current.

(6): Specialty 543: Nucleotide Sequence and Activity of Transposable Genetic Elements in Prokaryotes and Yeast

This specialty only contained two papers, both by the same people, see Table 4.14. Since this has occurred it is highly likely that the two are closely cognitively related. This is borne out by examination of the paper titles that showed that they are both about recombination in bacterial plasmids, which was observed to be cognitively in line with the current literature specialty title. The only organisation involved was the Department of Medicine at Stanford University, so the work was done in the public sector. Although one paper showed no grant data, the other showed financial support from the public sector. Analysis of shared co-authors showed that Kopecko worked with Cohen on both occasions. It was also observed that one paper cited the other, however even though this reflects a co-cited pair it has to be remembered that as this is an instance of self citation it is

hardly surprising. The papers were published in different journals.

It can be seen from the national rankings shown in Table 4.15 that the USA is the only country represented in the base literature and is also the leading one in the current. The UK is only poorly represented in the current literature.

(7): Specialty 134: Genetic Elements and Transposons

This specialty contained three papers in its base literature, see Table 4.16. Comparison of the paper titles (which all dealt with resistance determinants), showed that the linking of the papers into co-cited pairs was reasonable since they appeared to be cognitively similar. The base literature titles also appeared to be similar to the specialty title given by the current literature. An analysis of the co-authors showed that Botstein and Kleckner co-authored with each other on their two papers. The third paper by D. Berg had no co-authors. It could be seen from the institutional affiliation analysis that Botstein and Kleckner were at the same institution, the Massachusetts Institute of Technology, so their links were very close. Their two papers made this institution the dominant one in the base literature. Douglas Berg's paper came from the university of Wisconsin. All of the papers were produced in the public sector. Financial support for all the papers came from the public sector. Looking at the journals involved, the dominant one (which was actually a book) was DNA Insertion Elements with two papers. Looking at the way that the base literature papers cite each other it was observed that the Botstein paper cited the Kleckner one, although this is really an instance of self citation. The Kleckner paper did not cite either of the others but this can be explained by the fact that it pre-dated them. The Berg paper did not cite the Kleckner paper although it was published later.

As far as the national rankings are concerned the USA was the only country represented in the base literature and was top in the current. Japan and the UK lagged far behind in the current literature, see Table 4.17.

The last specialty occurred in the 1982 time period.

(8): Specialty 82: Characterising Transposable Elements of Bacterial Genomes

This specialty had six papers in its base literature. Comparing the titles of the papers in the co-cited pairs showed that they were cognitively related since they were generally concerned with the characterisation of transposable elements (and thus also generally in agreement with the specialty title). All the papers drew their financial support from the public sector. The dominant institution was the University of Basel which produced three papers, the other institutions only produced one paper each. All the institutions were in the public sector. No obviously dominant journal was observed as most had only one paper each with the highest one having only two. Each paper generally cited its co-cited partner(s), where this did not occur it was usually because the "citing" paper was older than the other. The only exception to the rule was the paper by S. Iida, which although it was more recent than either of the papers that it was paired with, cited only one of them. See Table 4.18.

Looking at shared co-authors among the base literature papers, see Table 4.19, it was seen that only four were co-authored with other authors in the same specialty. The national rankings shown in Table 4.20 showed that Britain did not appear in the base literature and was only weakly represented in the current. Interestingly enough Switzerland was as strong in the base literature as the USA, which it took second place to in the current.

Overall Results and Conclusions for the Individual Specialty Analysis.

The results showed that in all the specialties except 320, the base literature paper titles corresponded well with the specialty titles derived from the current literatures. The criteria used for determining correspondence were shared keywords and my own knowledge of biochemistry. In specialty 320 they did not match, but it was considered that the papers content as deduced from their titles and abstracts would make a plausible intellectual base for the current work of the specialty as deduced from its title. The authors of the base literature papers were later questioned about this in the questionnaire, see Chapter 6. It was concluded however from this preliminary inspection that the specialty title could still be regarded as a reasonable indicator of the content of the base literature.

In checking the validity of the co-cited pairs it was found that the cognitive content of papers grouped together by CCA matched well from a key word analysis of their titles. So it seems possible that perhaps the co-cited pairs could represent cognitive links between the base literature authors themselves and not just represent links created between them by the current literature. However the cognitive links between the papers were often observed to go beyond the co-cited pairs since they cite papers in their own specialty that they are not directly paired with. Also it was observed that in all the specialties each paper generally cited the papers it was paired with, allowing for the dates of publication of the papers involved. There were only two instances in the whole of the base literature of the trend where this did not occur. The papers also cited papers in their own specialty base literatures that they were not directly paired with. It was concluded that the papers show cognitive debts to each other that pervade the whole base literature of a specialty, but is particularly seen in the co-cited pairs. Thus it appears that when the current literature co-cites papers together it is picking up on relationships already demonstrated by the base literature papers. It is possible that this could occur because of current

literature authors selecting papers from a base literature papers reference list and then linking them all together in its own. It was concluded that co-citation analysis could be exposing relationships between papers and also about the way that scientists use information that is more fundamental than was previously imagined. This supports the idea presented in previous CCA validation studies that the co-cited pairs represent cognitive links drawn between papers (see Chapter 2).

This idea was also supported by the results for the co-author analysis which demonstrated that generally the specialties showed some personal links between their base literature papers by the sharing of authors. The results for this varied from Specialty 320 where every paper was co-authored by authors found in other papers in the base literature, to specialty 388 where there were no shared authors. From the the high level of citation and personal links it was concluded that the authors of the base literature may form an invisible college. This possibility was further examined in the analysis of the properties of the entire trend (see Section 4 of this chapter), and in the questionnaire (see Chapter 6).

Analysis of the national rankings for the base and current literatures showed that the USA dominated the base literature (the UK being generally poorly represented), in the same way that it dominates the current. Of great interest in the analysis of the institutional affiliations of the base literature was the finding that all the papers in Specialty 320 originated from the same department, indicating a very close knit specialty. All the base literature research was carried out in, and financed by, the public sector. So apparently there was no industrial interest at this level.

Section 4: Analysis of the Properties of the Overall Trend

After the relationships between, and the properties of, papers in the individual specialties had been examined, the data was studied in the aggregate to elucidate what it could show of the relationships between specialties relative to the structure of

the trend. The methodologies, results and conclusions of these various analyses are outlined in the appropriate sections below. The analyses aimed to expose any faults or inconsistencies in the trend that may have originated in the criteria used to generate them.

(1): Analysis of the Specialty Summary Report

The indicator data given to describe the properties of the specialties in Trend 44 in the Protein Determination Models in the form of the Specialty Summary Report, was analysed to see what it showed about their properties, development and behaviour. This was examined in order to test certain hypotheses about the type of information that can be derived from this type of data that have been advocated by the model builders. In this study only those that appear to directly be on on the properties of trends will be considered. The model builders assume that the size of the current literature, taken as the number of its papers, is supposed to represent the size of the research front of a particular specialty, and that the number of base literature papers represents the size of the paradigmatic base of the specialty. They also assume that the average age of the base literature and also its youngest paper can be used as an indicator of the development potential of a specialty. The hypothesis is that rapidly advancing fields with a high rate of turnover of papers would have a low value for both and that conversely, stagnating fields should have a high value for both. The model builders propose using this type of data to make judgements about the development of research fields over time in science policy studies, so it is worthy of further consideration. Also as some of the more vital decisions as to whether a specialty is stagnating or not is made at the level of the base literature, the need to focus on the properties of the literature at this level in this study is emphasised. Bearing in mind that CRP's ultimate aim is to develop theories relating to the development of specialties over time with a view to lending a predictive capability to future models, it was decided in this study to examine the Trend Summary

data against the background of the relationships shown between specialties by the trend diagram itself. This information is shown combined in Figure 4.1.

As can be seen in Figure 4.1, the trend apparently shows the development of the field characterized by the trend title, Transposons and Insertion Sequences of Bacterial Plasmids and chromosomes, from 1978 to 1982. It apparently shows that there were four specialties in 1978, all of which fed into one specialty in 1980 that continued into the 1982 time frame, and one of which fed into two specialties in 1980 that failed to continue into 1982. So according to this trend this field has suffered a steady contraction since 1978 and presumably would die out altogether if another model was built after 1982.

One of the most interesting questions is why Specialties 543 and 134 failed to continue into 1982. The Specialty Summary Report shows that their current activity levels are low, containing only 29 and 63 papers respectively and that their youngest papers are 4 and 3 years old. Since they both stem from Specialty 101 in 1978, a specialty with a high current activity level of 150 papers and a very low value for its youngest base literature paper of 1 year, i.e. apparently active and fast growing, this would supposedly mean that Specialties 543 and 134 are declining areas of research, and that the more dynamic aspects of Specialty 101 joined with elements of the other three specialties of 1978 to enter Specialty 222 in 1980, (which had a large current activity of 103 and a youngest paper of 1 year). So apparently compared to Specialty 222, specialties 543 and 134 are not very good.

The data also shows that Specialties 320, 388 and 525 in 1978 all have low current activity levels and high values for the age of their youngest papers. The model builders would assume that they were revitalised by linking up with Specialty 101 to form the promising specialty 222 in 1980. However this surge of activity did not appear to last since in Specialty 82 in 1982 the current activity level drops to 42 papers and the youngest paper age increases to 2 years. So the indicators would suggest that it is possible that the least promising aspects of Specialty 101 fed

into Specialties 543 and 134 and that the best of it found its way into Specialty 222.

The validity of the indicator data and the hypotheses underlying it were firstly tested by seeing which base literature papers carried over from Specialty 101 into the specialties in 1980, the idea being that if the same papers from 101 occur in the declining specialties 543 and 134 as in 222, then this is unlikely to be the case. From the data in Table 4.21 and 4.22 it can be seen that Specialty 101 passed a paper by Kopecko to Specialty 543, a paper by Kleckner to Specialty 134 and the papers by Gottesman and MacHattie to Specialty 222. Furthermore it was discovered that no base literature paper appeared in more than one specialty in the same time slice. If they had it could have been concluded that some component of them was incorporated in a continuing specialty with the resulting implication that they might not be as extinct as they seem from the trend diagram. Thus there was no overlap of cognitive bases (base literatures) between the specialties in 1980, so the Specialties 543 and 134 may well be isolated from the mainstream of research represented by Specialty 222.

An examination of the sizes of the base literatures of these specialties shows that Specialty 222 has a base literature size of 12 papers compared with Specialty 543 with 2 papers and Specialty 134 with 3 papers. Thus it can be seen that Specialties 543 and 134 have small base literature sizes compared to Specialty 222. So it was speculated that in cases where the base literature size of a specialty is small and the specialty does not appear to continue it may be that the research area is concentrating on one small problem, which when solved the authors move rapidly on to other problems. A specialty with a high base literature size could cover a broader range of topics and hence have more people working on it and last longer. So are Specialties 543 and 134 defunct because their particular small problem was solved? This becomes more probable in Specialty 543 which has only two papers in its base literature, both by Kopecko. The apparent exclusion of 543 and 134 from 1982 is further explored in the following sections. The reasons for the apparent mortality of these two specialties

were explored more fully in Chapters 6 and 7 using questionnaire and interview techniques, where the modellers view of the development of the field and the viability of individual specialties was also tested.

(2): Analysis of Base Literature Papers Linking Specialties into Trend 44

This was examined because trends are constructed by looking for shared base literature papers in adjacent time periods. The degrees of change information supplied on the diagram of the trend, see Fig.4.1, shows that there is a high level of change in the composition of the base literatures from year to year, i.e. not many papers continue into subsequent time periods.

From Table 4.21 it can be seen that the only paper that appeared in all three time periods was by S. Hu, J. Bacteriology V.122 p.764. Out of a total of 34 papers in the base literature of the trend, it was observed that 12 papers occurred only in 1978, 9 only in 1980 and 5 papers only in 1982. Of those papers that survived into a subsequent time period, 7 papers occurred only in the years 1978 and 1980, and no papers occurred exclusively in 1980 and 1982. Thus the paper by S. Hu was the only paper that linked the trend through all three time periods, (and in fact Specialty 82 was only linked into the trend by this one paper). Thus this way of linking specialties appears to be too tenuous as the continuance of the trend into 1982 could have been too easily lost.

Looking at the papers linking the specialties into the trend it can be seen that Specialties 543 and 134 are included in the trend on the strength of one paper each, therefore the chances would seem high that they would be easily lost from the next time period, especially since they only contained a total of 5 base literature papers between them to choose from. So in an attempt to find ways of supplementing the usual method of searching for shared base literature papers to generate a trend, the origins of papers in the base literature were searched for.

Since no data was available for specialties before 1978, only specialties from 1980 onwards were analysed. For these, it was noted which papers came from previous specialties, which new papers appeared in their base literatures, and whether any of these newcomers originated from the current literatures of previous specialties. Details of current literature papers were derived from the Specialty Detail: Current Literature reports supplied in the models. It was discovered, see Table 4.22 that some of the current literature papers in a particular time period appeared in the base literatures of subsequent specialties. It was concluded that these papers had assumed importance fairly quickly. It was observed that Specialties 543 and 134 do not pass any current literature papers into the base literature of Specialty 82, so they do not continue in the trend in this fashion.

It was also noted that of the current papers of Specialties 101, 525 and 388 that carried over into Specialty 222, two occurred both in the current literatures of Specialties 388 and 101 and one occurred in the current literatures of all three. This phenomenon of current papers turning up in more than one specialty is well known to the model builders who use this property to generate regions of related specialties in a particular time period. The three current literature papers examined here show that Specialties 101, 525 and 388 had links in the current literature, although like all the specialties in Trend 44, shared no papers in the base literature. So it was not considered to be surprising that they merged to form Specialty 222. Specialty 320 was only linked to Specialty 222 via its base literature, and had no links via its current literature to Specialties 101, 525 and 388. What is interesting is that the supposedly extinct specialties of 543 and 134 obtained no current literature from any specialty in 1978 and donated none to 1982, so again they remain unconnected to the rest of the trend. However out of all 250 papers in the current literature of the specialties in 1978, only 3 carried over into 1980, and out of the 103 current papers in 1980 only one carried over into 1982. Since only one base literature and one current literature paper went through to the 1982 time period it would appear that although the links in

the trend would be strengthened and perhaps more specialties picked up than at present by tracing the current literatures movement into subsequent base literatures, (the commonsense theory being that some of the work at the research front will form the paradigm for the next period), it still does not appear to have much effect. So another possible way of linking trends in the form of searching for the re-occurrence of authors, was examined.

(3): Analysis of Authors Occurring in the Base Literatures of Trend 44

The distribution of authors was shown with respect to the specialties that they appeared in, see Table 4.23. N.B. some of the authors re-occur by virtue of being associated with the papers linking the trend. Once these had been removed from the analysis, the extra links supplied by re-occurring authors were examined.

Since it was previously observed in Section 4 (1) of this chapter that no base literature paper appeared in more than one specialty in the same time slice, it was possible to look for shared authors in the same time periods directly. All of the base literature authors names were used for this analysis, not just the first named authors supplied in the model. It was found that the authors occurring in more than one specialty in the same time slice were; S.N. Cohen who appeared in Specialties 101 and 525 in 1978 and Specialties 543 and 222 in 1980, N. Davidson who appeared in Specialties 320 and 525 in 1978, S. Hu in Specialties 320 and 525 in 1978, N. Kleckner in Specialties 134 and 222 in 1980 and E. Ohtsubo in Specialties 320 and 525 in 1978.

Then the authors appearing in subsequent specialties were examined. It was found that D. Berg appeared in Specialty 101 in 1978 and Specialty 134 in 1980, N. Kleckner appeared in Specialty 101 in 1978 and Specialty 222 in 1980 and E. Ohtsubo appeared in Specialty 320 in 1978 and Specialty 82 in 1982. The question was asked whether these extra links corresponded to links shown in the trend diagram (Fig.4.1.). From this it can be seen that Specialty 101 in 1978 is linked with Specialties 134 and 222

in 1980, and that Specialty 320 in 1978 is indirectly linked to Specialty 82 in 1982 via Specialty 222 in 1980. So it can be seen that the links between specialties in the trend are verified by these author links. However only a few such author links were found, so not much improvement could be made to the modelling procedure for generating trends by adding them. As all these novel linking measures tested failed to provide much improvement, perhaps generating the trends in the usual way and then adding other specialties by looking for shared authors and the movement of current literature papers is the best answer to the problem of mapping trend areas of research.

It was then asked whether the extinct side-branch specialties of 543 and 134 were linked by authors to earlier specialties or donate any to Specialty 82 in 1982, and whether they shared any authors with other specialties in 1980? Only Specialty 134 was found to have an author link, via D. Berg, to Specialty 101 in the previous time period, and neither it or Specialty 543 donated any authors to 1982. However looking at the same time slice Specialty 543 was linked to Specialty 222 via S. Cohen, and Specialty 134 was linked to Specialty 222 via N. Kleckner. Thus it is possible that elements of these extinct specialties share common ground with the research in Specialty 222.

(4): Analysis of the Occurrence of Base Literature Authors in the Current Literature of Trend 44.

The occurrence of base literature authors in the current literature was examined to see how involved they were at the research fronts of their specialties. Does the "invisible college" that was suspected to exist at the base literature level also occur in the current, or is the current literature dominated by different people? All of the authors names identified in the base literature, not just the first named, were searched for. For each specialty not only its own authors were traced into its current literature but also any authors from elsewhere in the base literature of the trend. The number of papers authored by at least

one of the base literature authors was divided by the total number of papers in the current literature of each specialty and expressed as a percentage to get the level of base literature author participation in the production of the current literature. The number of individual appearances in the base and current literatures is recorded, which can sometimes exceed the total number of papers in the specialty. This occurs as a result of co-authorship.

It was found that in Specialty 101, (Table 4.24), that 16 of its base literature authors entered the current literature, together with 15 base literature authors from the rest of the trend. The percentage of papers in the current literature that were authored by base literature authors was 29.3%.

In Specialty 525, Table 4.25, only one author from its specialty base literature entered the current literature, together with 14 from the rest of the trend base literature. The percentage of current papers authored by base literature people was 50%.

In Specialty 320, Table 4.26, two authors from its own base literature and 6 from the rest of the trend entered the current literature. The percentage of current papers authored by base literature people was 18.18%.

In Specialty 388, Table 4.27, one author from its own base literature and 9 from the rest of the trend appeared. The percentage of current papers authored by base literature people was 40%.

In Specialty 134, Table 4.28, two of its own base literature authors entered the current literature with 7 from the rest of the trend. The percentage of current papers authored by base literature people was 15.87%.

In Specialty 222, Table 4.29, 10 of its own base literature authors and 9 from the rest of the trend appeared in its current literature. The percentage of current papers authored by base literature people was 33%.

In Specialty 543, Table 4.30, there was one author from its own base literature and 4 from the rest of the trend. The percentage of current papers authored by base literature people was 24.13%.

In Specialty 82, Table 4.31, there were 4 authors from its own base literature and 5 from the rest of the trend. The percentage of current papers authored by base literature people was 33.3%.

Thus all of the specialties had some of their own base literature authors, and some from elsewhere in the trend, working in their current literatures. This shows continuity of effort between the cognitive bases and the research fronts of each specialty. The fact that so many of the current literatures contain authors originating from the base literatures of other specialties in the trend suggests that these specialties are not discrete entities but that there is much interaction between them. The proportion of papers authored by base literature people in the current literature was found to vary between 15.87 and 50%. Thus in the current literature there are large numbers of papers at the research fronts authored by people not represented at base literature level. However, subjectively it was noticed that there was much interaction between the "old" and the "new" hands, since the new names often appeared as co-authors with the base literature people in the current literature as well as producing papers independently from them. However there was not enough time to quantify this.

The large numbers of base literature authors that turn up in the current literature is interesting. Table 4.32 shows that out of a total of 59 base literature authors in the trend, 37 turned up as authors in the current literature. Out of these only 6 appeared in all three time periods, W. Arber, D. Berg, S. Iida, J. Meyer, H. Ohtsubo and H. Saedler. Thirteen appeared in two time periods. When this occurred they were mostly in consecutive time periods, but in one case N. Datta appeared in 1978 and then in 1982. Eighteen appeared in one time period only. Thus comparatively few continue in the field over its whole span, although a high proportion do enter the current literature even if only once. The number of specialties that each author appears in can be quite high, e.g. E. Ohtsubo, H. Ohtsubo, H. Saedler and P. Starlinger appeared in all the specialties in 1978, and P. Bennett, M. Calos, S. Cohen, J. Miller and M. Richmond appeared in

all the specialties in 1980. So these specialties again do not appear to be discrete entities, perhaps the overall structure of the trend is the important thing and not the individual specialties.

In order to examine the apparent extinction of Specialties 543 and 134, the occurrence of the authors of the base literature papers in these specialties in the current literatures of other specialties was examined. In order to do this the authors names were found from Tables 4.14 and 4.16 and then searched for in Tables 4.24 to 4.31. The reasoning behind this was that if connections to other specialties could be found in this manner then it is possible that ideas from these supposedly extinct specialties would be found in other parts of the trend.

Thus it can be seen that in Specialty 543, Jean Brevet was active in the current literatures of Specialties 101, 525 and 320 in 1978. S.N. Cohen was active in Specialties 101, 525 and 320 in 1978 and in Specialties 134, 222 and 543 in 1980. This last bit is especially interesting since it links the two supposedly redundant specialties of 1980, 543 and 134 not only with each other, but also with the continuing specialty in that time period, 222. D.J. Kopecko appeared in the current literatures of Specialties 101, 525 and 320 in 1978. These links with earlier specialties are interesting as the trend diagram, Fig. 4.1, shows that specialties 543 and 134 are apparently only linked to specialty 101, whereas this analysis shows links to other specialties in 1978.

Similarly for Specialty 134, D.E. Berg appeared in the current literatures of specialties 101 and 525 in 1978, specialties 134 and 222 in 1980 and specialty 82 in 1982. Thus demonstrating extra links to earlier specialties and to the continuing specialty in 1980, 222, but also showing a link into the 1982 time period. D. Botstein and N. Kleckner appeared in specialty 101 in 1978, Kleckner also appearing in the current literature of 134. It was observed that R.K. Chan and B.K. Tye did not appear in the current literatures of any specialty.

It was also observed which authors from other specialties in the trend appeared in these "extinct" specialties.

These are shown in Tables 4.28 and 4.30. It was observed that P.M. Bennett appears in the current literatures of specialties 101 in 1978, and 543, 134 and 222 in 1980, M.P. Calos appears in 101 and 388 in 1978 and 134, 222 and 543 in 1980, S. Falkow appears in 101 in 1978 and 134 and 222 in 1980, J.H. Miller in 101 and 388 in 1978 and 134, 222 and 543 in 1980, M.H. Richmond in 101 in 1978 and 134, 222 and 543 in 1980 and P. Starlinger in 101, 525, 320 and 388 in 1978 and 134 and 222 in 1980. Thus further links to earlier specialties and to the continuing specialty of 222 were observed, although no more links into 1982 were found. Only those authors names that were familiar from the base literature papers were analysed because of the large numbers of authors in the current literature. It was considered likely that if the names of the other authors in the current literature were similarly analysed, further links would probably be found. Nevertheless, there seems to be a wealth of documentary evidence that specialties 543 and 134 are not really as isolated as they seem. This was further explored using the questionnaires and interviews in Chapters 6 and 7.

Then specialties in which authors appeared more than once in the base literature were examined to see in which specialties people were most influential, see Table 4.23. This was carried out in an attempt to identify the leaders in each research area. Thus for specialties in 1978 the dominant authors in specialty 320 (4 papers) were N. Davidson and E. Ohtsubo with 4 papers each and S. Hu and R.C. Deonier with 2 papers each. For specialty 101 (12 papers) they were S. Falkow, R.W. Hedges and F. Heffron with 3 papers each and A. Jacob and C. Rubens with 2 papers each. For specialties 388 and 525 no author appeared more than once. For specialties in 1980, in specialty 222 (12 papers) E. Ohtsubo authored 3 papers and M.P. Calos, N. Davidson, S. Hu, L. Johnsrud, J.H. Miller and H. Saedler each had two papers. In specialty 543 (2 papers) S.N. Cohen and D.J. Kopecko authored both papers and in specialty 134 (3 papers), D. Botstein and N. Kleckner each had two papers. For the one specialty in 1982, 82 (6 papers), W. Arber produced 3 papers and S. Iida and E. Ohtsubo 2 papers each.

These names were then compared with their performance in the current literatures of their specialties in order to identify those that continue to be leaders over a period of time. It was observed that for specialty 320 only E. Ohtsubo remained in the current literature, see Table 4.26. For specialty 101 all of them except Rubens appeared in the current literature, see Table 4.24. For specialty 222 Calos, Miller, E. Ohtsubo and Saedler survived, see Table 4.29. For specialty 543, there was Cohen, see Table 4.30, and for specialty 134 there was Kleckner, see Table 4.28. In specialty 82, Arber and Iida survived, see Table 4.31.

This analysis of the occurrence of base literature authors in Trend 44 shows that there is much sharing of authors in the specialties linked by the trend, in both the base and current literatures. Assuming that it is the people interested in a certain subject that create a research field, their activities in these specialties show that the distinctions between them are somewhat blurred and that it is possible that the trend is more important then its component specialties and could represent the overall interests of a group of people in an invisible college. So it might not be very significant that some small specialties are observed to disappear. Perhaps they only represent small problems of transient interest, or that the modelling technique fails to pick up how they feed back into the trend. The base literature authors habit of turning up in other specialty current literatures appears to suggest that all the specialties represent problems in their overall research, which could be represented by the trend title. The activities of these authors over the years could account for the linking up of the specialties in a trend.

(5): Patterns of Citation Between the Base Literature Authors

Details of how the base literature authors cite the papers in their own specialties is covered in the specialty analysis section. In this section the way that they cited all the other base literature authors contained in the trend was examined. This was examined by looking at the reference sections of their papers and recording the act of citing another base literature

authors name. The details of whether the papers these names were associated with appeared in other parts of the trend was ignored because of the time factor involved. Similarly since this analysis was carried out as the papers in the base literature of the model were being looked up in the library, only the first named authors were searched for since the names of all the co-authors were not yet known. If this had been done a second search of the papers would have been needed once their names had been obtained from a first search of the entire trend. The number of papers that the trend base literature authors appeared in was recorded together with the total number of papers in the reference lists, to get some idea of how prominent the work of other first named base literature authors were in their work. This was expressed as a percentage in Table 4.34. From this it can be seen that the percentage of papers authored by base literature people in base literature papers reference lists varied between 3.57 and 53.33. The mean value was 27.1%. Thus it can be seen that these papers do apparently draw on work produced by other members of the invisible college outlined by the base literature, the figure would probably be much higher if all the base literature authors names had been searched for.

There were 28 first named base literature authors in Trend 44. It was noted how many of these were cited in each base literature paper in order to see what proportion of base literature authors in the trend were cited. The number of first named trend base literature authors cited was divided by the total number of first named base literature authors in the trend and expressed as a percentage. The results are shown in Table 4.33. It was found that on average 22.8% of other trend authors are cited in each base literature paper. It is probable that this figure would have been much higher if co-authors had been included in the analysis. It was also observed from this table that out of a total of 34 papers there were 20 instances of self citation, i.e. a minimum of 58.8% of papers cite their own authors, a figure that again would probably be much higher if citations to co-authors were considered. In addition to the other calculations the number of times, expressed as a percentage, that individual authors were

cited by the base literature papers was examined. From Table 4.33 it can be seen that only Rubens was never cited. The others ranged from between 2.94 and 47.06%. The average number of times that each author was cited was 22.79%.

Thus it was concluded that the base literature authors do seem to show significant citation links to each other. Thus patterns of citation between base literature authors seem to lend credence to the idea that the trend could represent an invisible college.

(6): Occurrence of Organisations Within Trend 44

The occurrence of organisations in the specialty base literatures are shown in Table 4.35. From this table it can be seen that there were a total of 18 institutions involved in the production of the base literature papers. Out of these 9 appeared in only one specialty, and 8 had more than one paper in some of the specialties they appeared in, (which was examined in detail in the earlier section on specialties). Out of the 9 that appeared in more than one time period, 6 only re-appeared as a result of being associated with recurring papers. Only the University of Wisconsin, the State University N.Y., and the University of Geneva made new appearances. These provided a new link between specialties 388 and 134 and also between 222 and 82 and 101 and 222. Thus strengthening some links shown by the trend and showing another link, not shown in Fig.4.1, between a specialty in 1978, 388 and one of the apparently extinct specialties in 1980, 134. No organisation made new appearances in all three time periods, therefore little organisational continuity can be identified in the trend.

(7): Distribution of Journals Throughout the Base Literatures of Specialties in Trend 44

The details of predominant journals in individual specialties was analysed in the earlier section on specialties. In this section the intention was to provide an overview of their

occurrence throughout the trend. Data for this is shown in Table 4.36. The occurrence of these journals in the core journal set used in the 1978, 1980 and 1982 Protein Determination Models was also examined. These core journals are the base on which the models were built, the citing papers that formed the current literature being drawn from them. As the core journal sets were chosen because they were considered to be relevant to the field of the analysis of protein structures (see Chapter 3), it was thought that perhaps these journals did not adequately represent the field of mobile genetic elements covered by this trend, so leading to possible errors of omission. This was tested by seeing whether the base literature papers of this trend came from this set of journals. According to CCA definition these are important papers, therefore if they come from journals other than these then great errors of omission would seem likely to have occurred.

It was found that a total of 7 journals were associated with the base literature of Trend 44. Out of these 5 journals also belonged to the core set of the models. Two books were found to be in the base literature as well (these were comprised of collections of individual papers), which with the other two base literature journals were not found to be in the core journal set. This occurs because the citing papers drawn from this set are free to reference whatever journals they wish. Thus on the face of it there is a ratio of 5 core set journals to 4 non core set sources, which would seem to imply that nearly as much material outside the core set was being used as within. However if the number of papers involved throughout the base literature of the trend are considered, then the core journal set had a total of 28 papers, and the non core set had a total of 15. No allowance was made in these figures for recurring papers since each individual appearance, because they are selected by the current literature papers in that year, is important in this case. It can be seen that, analysed in this fashion, there are almost twice as many papers taken from core journal sources as from other sources. Thus although the current literature theoretically has complete freedom of choice concerning the literature that it cites, it appears to be fairly conservative in this trend. Nevertheless it was

concluded that the information given by this trend could be improved if a model was built specifically covering the type of research that it contains, with its own appropriate core journal set. Also from the data it was observed that only 3 journals survived through all three time periods, one of which the Journal of Bacteriology, the most frequently occurring journal, contained papers which mainly occurred in 1978. It was speculated that perhaps the patterns of survival and mortality of journals and the years in which they are most prominent, could act as an indicator to changing cognitive interests. This assumes that journals publish papers mainly from specialised areas of interest. However the testing of this particular hypothesis is beyond the scope of this thesis.

(8): Analysis of Acknowledgements in the Base Literature Papers of Trend 44 to First Named Trend Authors

Details of the pattern of acknowledgements are shown in Table 4.37. Only acknowledgements to first named authors were analysed since not all the co-authors were known when the papers were examined.

Links through acknowledgements were searched for because it was thought that they would show close personal links between first named trend authors and the authors of other base literature papers in the trend. It was observed that out of a total of 34 papers in the base literature of Trend 44, only 15 showed acknowledgements of assistance given by other first named trend authors. Out of these one showed affiliations to 4 other trend authors, 2 showed debts to 3, 5 showed affiliations to 2, and 7 to one. So the links did not appear to be very extensive. The acknowledgements tended to show gratitude for technical or theoretical assistance. Thus Table 4.37 shows a web of relationships, e.g. Berg is acknowledged by Johnsrud, who also acknowledges Gottesman, and Berg is also acknowledged by Kleckner. Thus Berg is directly linked to Johnsrud and Kleckner and indirectly linked to Gottesman. If we should also include the names of the co-authors on the acknowledging papers we would be

able to see how the first named authors relate to the groups who produced the papers. A web of relationships was derived manually from Table 4.37 and displayed in Fig. 4.2.

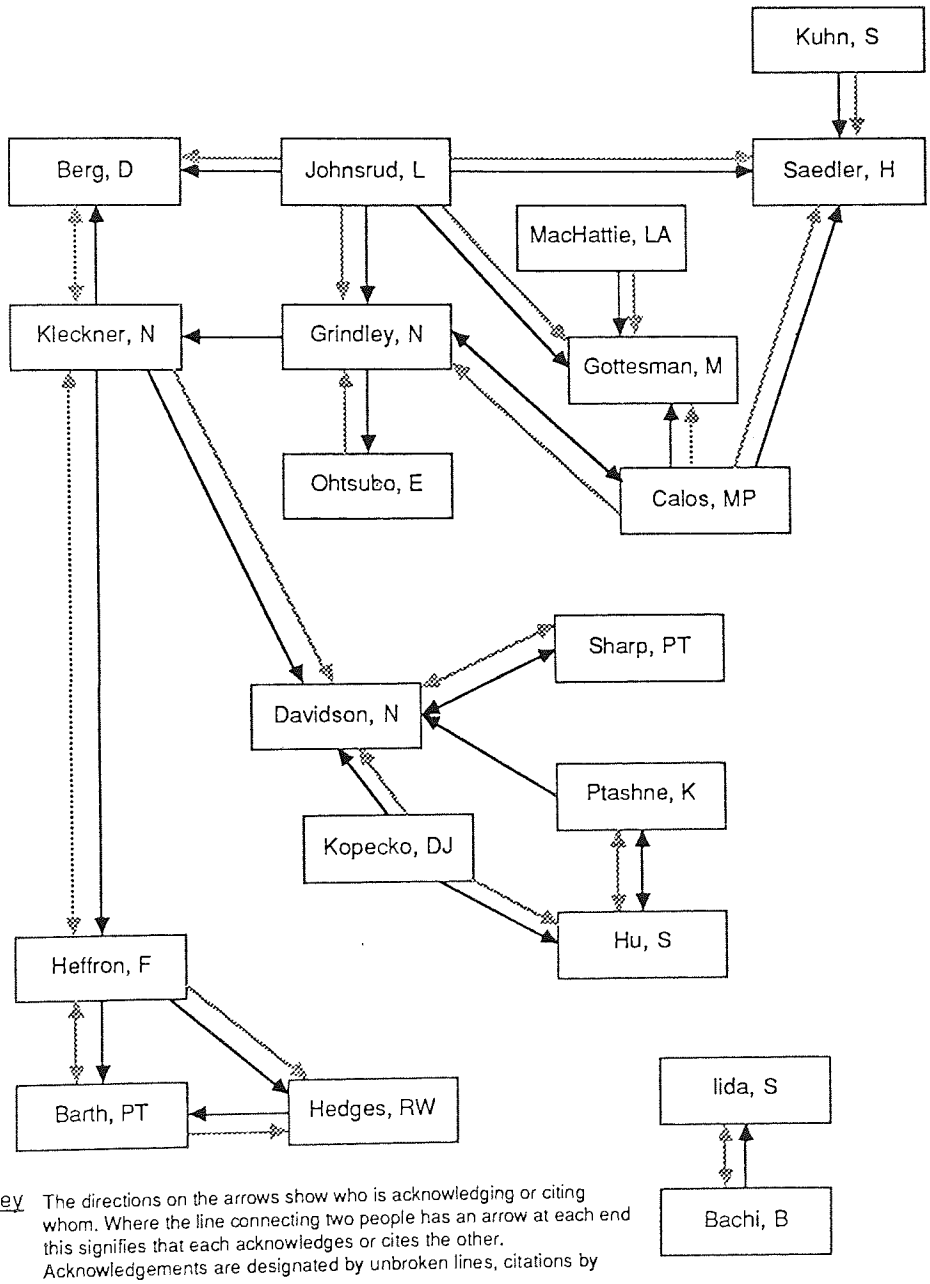
It was hypothesized that if the ackowledgements data was compared with Table 4.33 which shows how the first named trend authors cite each other, where the acknowledgements and citations coincide could indicate a close relationship. Citation links from Table 4.33 were included in Fig. 4.2 to show where this has occurred. It can be seen from Fig. 4.2 that the authors mostly cite the people that they acknowledge, most of the citations occurring in this direction, but in 7 cases the person that was acknowledged cited the acknowledger indicating a very close relationship. Only in 1 case, Ptashne and Davidson, was there no citation. Fig. 4.2 is interesting because it shows how 18 first named trend authors are linked together by a network of indebtedness. Two authors, Iida and Bachi appear to be outside this network. However looking at Table 4.33 shows that Iida cites some of the authors in the main network, but Bachi cites only Iida. This appears to provide further evidence that the authors displayed in Fig. 4.2 belong to an invisible college. The relationships displayed in Fig. 4.2 were later compared with data derived from the questionnaire and interview results in Chapters 5 and 6.

(9): Analysis of Industrial Involvement in the Current Literature

All of the papers in the base literature were found to originate from academic or medical institutions, no industrial concerns were involved. Using data from an M.Sc thesis by E. Bell, 1984, (120), which detailed the occurrence of industrial corporations in all the specialties of the Protein Determination Models for 1978, 1980 and 1982, the current literatures of the specialties in Trend 44 were searched for the presence of industrial concerns.

From Table 4.38 it was observed that only 6 companies were involved in the current literature. Of these City New York Inc. seems the most active in the field, being represented in all

Figure 4.2 Showing the Relationships Between First Named Authors in Trend 44 as Shown by their Acknowledgements and Citations to Each Other.



Key The directions on the arrows show who is acknowledging or citing whom. Where the line connecting two people has an arrow at each end this signifies that each acknowledges or cites the other. Acknowledgements are designated by unbroken lines, citations by dotted lines.

4 specialties in 1978 and in one specialty in 1980. The only other company to appear more than once was GE Corporate R and D which appeared in 2 specialties in 1978. Interestingly enough City N.Y. Inc links all 4 specialties in 1978 to the continuing specialty 222 in 1980, reflecting the linking of these specialties in the trend diagram in Fig. 4.1. However industrial involvement starts to decline in 1980 and disappears completely in 1982.

Thus there does not seem to be much industrial interest in Trend 44, a trend apparently with a base of pure scientific research. This was somewhat surprising since it impinges on the area of genetic engineering, so a high proportion of biotechnology companies would be expected to occur in the current literature. However it is possible that in the highly competitive area of biotechnology that the companies do not publish much, their activities perhaps being better recorded in the form of patents. How this compares to Trend 11 is examined in Chapter 5.

Section 5: General Conclusions for the Analysis of the Literature of Trend 44

Overall the results of the analysis of the literature were promising with respect to corroborating the hypothetical relationships between specialties over time shown by the model builders. Nothing was found to contradict these relationships, either at the level of individual specialties or of the overall trend. The results so far have demonstrated that the relationships shown in the trend appear to be verified by other unmodelled aspects of the literature that it represents. The co-cited pairs of papers were found to have links with respect to their scientific (or cognitive) content and cognitive and personal links between the base literature authors were shown to exist, (demonstrated by citation, co-authorship, institutional and acknowledgement connections between them.). Re-occurrence of authors in different time periods were found to corroborate links shown between specialties in the trend diagram in Fig. 4.1., and in some cases to show extra links. Research in the field was dominated by the US in both the base and current literatures. The

Specialty Summary indicator data given by the model builders seemed useful in explaining the specialties behaviour in the trend. However the results showed that the trend modelling procedures used by the model builders need further development in that the attachment of some specialties to the trend was found to be rather tenuous.

Thus the analysis was successful in its aim of going beyond what was modelled to see if the supposed relationships between specialties over time stood up under a close scrutiny of the literature. The problem with this so far is that it is only a partial indicator of the trends validity, in that it can only show that the trend is consistent in itself. In order to track down errors of omission, see how accurately it follows the development of the field of transposon research, test the assumptions in CCA concerning the high quality of the base literature papers and the ability of CCA to model the activities of invisible colleges, the analysis of Trend 44 was extended using questionnaire and interview data, the results of which are reported in Chapters 6 and 7.

However prior to this, the utility of the techniques of trend analysis developed in this chapter were tested on another complex trend in the Protein Determination models. The analysis of this trend, designated in the models by the number 11, is presented in Chapter 5.

Table 4.1 Details of Base Literature Papers in Specialty 320.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers in own specialty
1.	Davidson, N., Deonier, R.C., Hu, S., Ohtsubo, E.	Microbiol.1974 p56 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.coli. DNA Sequence Organization of F and F-primes, and the Sequences Involved in Hfr Formation.	Dept. Chem. Calif. Inst. of Technology, Pasadena, California, 91109, USA.	Public Health Service grant from Nat. Inst. of General Medical Sciences.	2,3	3
2.	Hu, S., Ohtsubo, E., Davidson, N.	J.Bact v.122 p749 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.coli. Structure of F13 and related F-primes.	Dept. Chem. Calif. Inst. of Technology, Pasadena, California, 91109, USA.	Grant from the Nat. Inst. of General Medical Sciences.	1,4	1,3,4
3.	Ohtsubo, E., Deonier, R.C., Lee, H.J., Davidson, N.	J.Mol.Biology v.39 p565 1974	Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli.	Dept. Chem. Calif. Inst. of Technology, Pasadena, California, 91109, USA.	Grant from US Public Health Service	1,4	4
4.	Sharp, P.A., Hsu, M.T., Ohtsubo, E., Davidson, N.	J.Mol.Biology v.71 p471 1972	Electron Microscopes Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli.	Dept. Chem. Calif. Inst. of Technology, Pasadena, California, 91109, USA.	Grant from Nat. Inst. of Health	2,3	–

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.2 Papers in Which Base Literature Authors Appeared in Specialty 320.

Author	Papers in which they Appeared (denoted by number)
1. Norman Davidson	1, 2, 3, 4
2. Richard C. Deonier	1, 3
3. Ming-Ta Hsu	4
4. Sylvia Hu	1, 2
5. Hung Jung Lee	3
6. Eiichi Ohtoubo	1, 2, 3, 4
7. Phillip A. Sharp	4

N.B. The identifying numbers for the papers are in Table 4.1.

Table 4.3 Comparison of National Rankings for Base and Current Literature Papers in Specialty 320.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. U.S.A.	4	100	1. U.S.A.	22	50
			2. W. Germany	12	27.3
			3. U.K.	4	9.1

Table 4.4 Details of Base Literature Papers in Specialty 101

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers in own specialty
1.	Barth, P.T., Datta, N., Hedges, R.W., Grinter, N.J.	J.Bact v.125 p800 1976	Transposition of a DNA Sequence Encoding Trimethoprin and Streptomycin Resistance from R483 to other Relicons.	Dept. Bacteriology, Royal Postgraduate Medical School, London. UK.	Medical Research Council, UK.	5	3, 5, 8, 10
2.	Bennett, P.M., Richmond, M.H.	J.Bact v.126 p1 1976	Translocation of a Discrete Piece of DNA Carrying an amp Gene between Replicons in E.coli.	Dept. Bacteriology, Medical School, University of Bristol, Bristol, UK.	Medical Research Council, UK.	7, 10, 12	5, 6
3.	Berg, D.E., Davies, J., Allet, B., Rochaix, J.D.	PNASUSA v.72 p3628 1975	Transposition of R-factor Genes to Bacteriophage Lambda.	Dept. Le Biologie Moleculaire, University de Geneve, Geneva, Switzerland.	Fonds Nat. Suisse de la Recherche Sci., Ligue Suisse Centre de Cancer, Suiss Amer. FDN, Sci. Exch.	4, 7, 8	5, 7, 8, 9
4.	Gottseman, M.M., Rosner, J.L.	PNASUSA v.72 p5041 1975	Acquisition of a Determinant for Chloramphenical Resistance by Celiphage Lambda.	Lab. Mol. Biol., Nat. Inst. Arithritis, Metab & Dig. Dis, NIH, Bethesda, Maryland, USA.	No grant information	3, 9, 11	3, 7, 8, 9, 10
5.	Hedges, R.W., Jacob, A.E.	Mol.Gen.Gen. v.132 p31 1974	Transposition of Ampicillin Resistance from RP4 to Other Replicons.	Dept. Bacteriology, Royal Postgraduate Medical School, London, UK.	Medical Research Council, UK.	1, 7, 10	–

Table 4.4, continued.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers in own specialty
6.	Heffron, F.H., Sublett, Hedges, R.W., Jacob, A., Falkow, S.	J.Bact v.122 p250 1975	Origin of the TEM Beta Lactamase Gene Found on Plasmids.	Dept. Microbiology, School of Medicine, University of Washington, Seattle, Washington, USA; & Dept. Bact., Royal Postgrad Med. Sch., UK.	National Science Fdn. & US Army R&D Command.	7	5
7.	Heffron, F., Rubens, C., Falkow, S.	PNASUSA v.72 p3623 1975	Translocation of a Plasmid DNA Sequence which Mediates Ampilcillin Resistance; Molecular Nature and Specificity of Insertion.	Dept. Microbiology, School of Medicine, University of Washington, Seattle, Washington, USA.	No grant information.	2, 3, 5, 6, 10	6, 8, 10
8.	Kleckner, N., Chan, R.K., Tye, B.K., Botstein, D.	J.Mol.Biology v.97 p561 1975	Mutagenesis by Insertion of a Drug Resistance Element Carrying an Inverted Repetition.	Dept. Microbiology, Massachusettes Inst. of Tech., Cambridge, Mass., USA.	Amer. Cancer Society & Nat. Inst. of Health.	3	3, 4, 6, 7, 10
9.	Kondo, E., Mitsubashi, S.	J.Bact v.88 p1266 1964	Drug Resistance of Enteric Bacteria. IV Active Transducing Bacteriophage PICM Produced by the Combination of R Factor with Bacteriophage P1.	Dept. Microbiology, School of Medicine, Gunma University, Maebahi, Japan.	Grant from Waksman FDN (Japan).	4	–

Table 4.4, continued.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers in own specialty
10.	Kopecko, D.J., Cohen, S.N.	PNASUSA v.72 p1373 1975	Site Specific Rec A-Independent Recombination between Bacterial Plasmids: Involvement of Palindromes at the Recombinant Loci.	Dept. Medicine, Stanford University, School of Medicine, Stanford, California, USA.	No grant information	2,5,7	–
11.	MacHattie, L.A., Jackowski, J.B.	DNA Insertion Elements p219 1977	Physical Structure and Deletion Effects of the Chloramphenicol Resistance Element Tn9 in Phage Lambda.	Dept. Medical Genetics, University of Toronto, Toronto, Canada.	Medical Research Coucil of Canada.	4	4, 9
12.	Rubens, G., Heffron, F., Falkow, S.	J.Bact v.128 p425 1976	Transposition of a Plasmid DNA Sequence that Mediates Ampicillin Resistance: Independence from Host rec Functions and Orientation of Insertion.	Dept. Microbiology & Immunology, School of Medicine, University of washington, Seattle, Washington, USA.	Grant from National Science Fdn., & US Army R&D Command.	2	2, 3, 5, 6, 7, 8, 10

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.5 Papers in Which Base Literature Authors Appeared in Specialty 101.

Author	Papers in which they Appeared (denoted by number)
1. Bernard Allet	3
2. Peter Barth	1
3. Peter Bennett	2
4. Douglas Berg	3
5. David Botstein	8
6. Russell Chan	8
7. Stanley Cohen	10
8. Naomi Datta	1
9. Julian Davies	3
10. Stanley Falkow	6,12,7
11. Michael Gottesman	4
12. Nigel Grinter	1
13. Robert Hedges	1,5,6
14. Fred Heffron	6,12,7
15. J.B. Jackowski	11
16. Alan Jacob	5,6
17. Nancy Kleckner	8
18. Eiko Kondo	9
19. Dennis Kopecko	10
20. L.A. MacHattie	11
21. Susumu Mitsuhashi	9
22. Mark Richmond	2
23. Jean-David Rochaix	3
24. J.L. Rosner	4
25. Craig Rubens	12,7
26. Richard Sublett	6
27. Bik Kwoon Tye	8

NB: The identifying numbers for the papers are given in Table 4.4.

Table 4.6 Comparison of National Rankings for Base and Current Literature Papers in Specialty 101.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. U.S.A.	6	50.0	1. U.S.A.	79	52.70
2. U.K.	4	33.33	2. U.K.	23	15.30
3. Japan	1	8.3			
3. Switzerland	1	8.3			
3. Canada	1	8.3			

Table 4.7 Details of Base Literature Papers in Specialty 388.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Fiandt, M., Szybalski, W., Malamy, M.H.	Mol.Gen.Genetics v119 p223 1972	Polar mutations in lac, gal and phage Lambda Consist of a few IS-DNA Sequences Inserted with Either Orientation.	McArdle Lab, Univ of Wisconsin, Madison, Wisconsin. & Dept of Molec. Biology & Microbiology, Tufts Univ, Boston, Mass. USA.	Nat. Cancer Inst. & National Science FDN & US Public Health Service, NIH, Division of General Medical Sciences.	2	–
2.	Saedler, H. & Heiß, B.	Mol.Gen.Genetics v.122 p267 1973	Multiple Copies of the Insertion DNA Sequences IS1 and IS2 in the Chromasome of E.Coli K-12.	Institute fur Genetic der Universitat zu Koln, Cologne, W. Germany.	Deutsche Forschungsgemeinschaft.	1	1

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.8 Comparison of National Rankings for Base and Current Literature Papers in Specialty 388.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. U.S.A.	1	50	1. U.S.A.	15	75
1. U.S.A.	1	50	2. W. Germany	3	15

Table 4.9 Details of Base Literature Papers in Specialty 525.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Hu, S., Ohtsubo, E. & Davidson, N.	J.Bact v122 p764 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Bacterial Plasmids: Identification and Mapping of the Insertion Sequences IS1 and IS2 in F and R Plasmids.	Dept. Chem., Calif. Inst. Tech., Pasadena, Calif, USA. & Institute fur Genetic, Univ zu Koln, Cologne, W. Germany.	Deutsche Forschungsgemeinschaft & Nat. Inst. General Medical Sciences.	2	2
2.	Ptashne, K. & Cohen, S.N.	J.Bact v.122 p776 1975	Occurrence of Insertion Sequence (IS) Regions on Plasmid DNA as Direct and Inverted Nucleotide Sequence Duplications.	Dept. Medicine, Stanford Univ, School of Medicine, Stanford, California, USA.	Nat. Inst. Allergy and Infectious Diseases; Nat. Sci FDN & American Cancer Society.	1	1

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.10 Comparison of National Rankings for Base and Current Literature Papers In Specialty 525.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. U.S.A.	2	100	1. U.S.A.	23	90.0
2. W. Germany	1	50	2. W. Germany	10	27.8
			3. U.K.	2	5.6

Table 4.11 Details of Base Literature Papers in Specialty 222.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Calos, M.P., Johnsrud, L. & Miller, J.H.	Cell v.13 p411 1978	DNA Sequence at the Integration Sites of the Insertion Element IS1.	Dept. Molec. Biol., Univ. of Geneva, Switzerland.	Nat. Inst. Health General Med. Sci., Nat. Sci. Fdn., Swiss National Fund.	4, 6, 7, 8, 9, 10	3, 4, 9, 10, 12
2.	Davidson, N., Deonier, R.C., Hu, S. & Ohtsubo, E.	Microbiology v.1974 p56 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.coli: X. DNA Sequence Organization of F and F-Primes, and the Sequences Involved in Hfr Formation.	Dept. Chem., Calif. Inst. of Technology, Pasadena, California, 91109, USA	National Institute of General Medical Sciences.	5	12
3.	Gottesman, M.M. & Rosner, J.L.	PNASUSA v.72 p5041 1975	Acquisition of a Determinant for Chloramphenicol Resistance by Celiphage Lambda.	Lab. Mol. Biol., Nat. Inst. Arth. Met & Dig. Dis, N.I.H., Bethesda, Maryland, USA.	No Grant Information.	6, 9,10	5, 11, 12
4.	Grindley, N.D.F.	Cell v.13 p419 1978	IS Insertion Generates Duplication of a Nine Base Pair Sequence at its Target Site.	Dept. Molec. Biophysics & Biochem., Yale Univ., New Haven, Connecticut, USA.	National Institutes of Health.	1, 6, 7, 8, 9, 10	1, 9, 10, 12

Table 4.11, continued.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
5.	Hu, S., Ohtsubo, E., Davidson, N. & Saedler, H.	J.Bact v.122 p764 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Bacterial Plasmids: Identification and Mapping of the Insertion Sequence IS1 and IS2 in F and R Plasmids.	Dept. Chem., Calif. Inst. of Technology, Pasadena, California, USA. and Institut fur Genetik, Univ. zu Koln, Cologne, W. Germany.	Deutsche Forschungs Gemeinschaft & Nat. Inst. General Medical Sciences.	2, 11, 12	2, 11, 12
6.	Johnsrud, L., Calos, M.P. & Miller, J.H.	Cell v.15 p1209 1978	The Transposon Tn9 Generates a Nine Base Pair Repeated Sequence During Integration.	Biology Labs, Harvard University, Cambridge, Mass., USA.	Nat. Inst. Health Gen. Med. Sci. & Swiss National Fund.	1, 3, 4, 7, 8, 9, 10	1, 4, 8, 9, 10, 12
7.	Kleckner, N.	Cell v.16 p711 1979	DNA Sequence Analysis of Tn10 Insertions. Origin and Role of Nine Base Pair Flanking Repetitions During Tn10 Translocation.	The Biology Labs, Harvard University, Cambridge, Mass., USA.	National Science Foundation & National Institutes of Health.	1, 4, 6	1, 4, 6
8.	Kuhn, S., Fritz, H.J., & Starlinger, P.	Mol.Gen. Genetics v.167 p235 1979	Close Vicinity of IS1 Integration Sites in the Leader Sequence of the gal operon of E.Coli.	Institute fur Genetik, Univ. zu. Koln, Cologne, W. Germany.	Deutsche Forschungsgemeinschaft.	1, 4, 6, 10, 12	1, 4, 9, 10

Table 4.11, continued.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
9.	MacHattie, L.A. & Jackowski, J.B.	DNA Insertion Elements p219 1977	Physical Structure and Deletion Effects of the Chloramphenical Resistance Element Tn9 in Phage Lambda.	Dept. Medical Genetics, Univ. of Toronto, Toronto, Canada.	Medical Research Council of Canada.	1, 3, 4, 6	3, 5
10.	Ohtsubo, H. & Ohtsubo, E.	PNASUSA v.75 p615 1978	Nucleotide Sequence of an Insertion Element IS1.	Dept. Microbiology, Health Science Centre, State Univ, Stony Brook, N.Y., USA.	US Public Health Service.	1, 3, 4, 6, 8, 12	5, 9
11.	Ptashne, K. & Cohen, S.N.	J.Bact v.122 p776 1975	Occurrence of Insertion Sequence (IS) Regions on Plasmid DNA as Direct and Inverted Nucleotide Sequence Duplications.	Dept. Medicine, Stanford Univ., School of Medicine, Stanford, Calif., USA.	Nat. Inst. Allergy & Infectious Diseases, Nat. Sci. Fdn & American Cancer Society.	5	5, 12
12.	Saedler, H. & Heiß, B.	Mol.Gen. Genetics v.122 p267 1973	Multiple Copies of the Insertion-DNA Sequences IS1 and IS2 in the Chromasone of E.Coli K-12.	Institut fur Genetik, Univ. zu Koln, Cologne, W. Germany.	Deutsche Forschungsgemeinschaft.	5, 8, 10	–

NB: The numbers given in the "co-cited with" and "cited columns" refer to the papers' identifying numbers.

Table 4.12 Papers in Which Base Literature Authors Appeared in Specialty 222.

	Authors	Paper in which they appeared (donted by number).
1.	Calos, Michele, P.	1,6
2.	Cohen, Stanley, N.	11
3.	Davidson, Norman.	2,5
4.	Deonier, Richard.	2
5.	Fritz, H.J.	8
6.	Gottesman, Michael.	3
7.	Grindley, Nigel.	4
8.	Heiβ, Brigitte.	12
9.	Hu, Sylvia.	2,5
10.	Jackowski, J.B.	9
11.	Johnsrud, Lorraine.	1,6
12.	Kleckner, Nancy.	7
13.	Kuhn, Siegfried.	8
14.	MacHattie, L.A.	9
15.	Miller, Jeffrey.	1,6
16.	Ohtsubo, Eiichi.	2,5,10
17.	Ohtsubo, Hisako.	10
18.	Ptashne, K.	11
19.	Rosner, J.L.	3
20.	Saedler, Heinz.	5,12
21.	Starlinger, Peter.	8

Table 4.13 Comparison of National Rankings for Base and Current Literature Papers in Specialty 222.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. USA	7	58.33	1. USA	58	56.30
2. W. Germany	3	25.00	2. Switzerland	14	13.60
3. Canada	1	8.33	4. UK	6	5.80
3. Switzerland	1	8.33			

Table 4.14 Details of Base Literature Papers in Specialty 543.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Kopecko, D.J. & Cohen, S.N.	PNASUSA v.72 p1373 1975	Site Specific Rec. A-Independent Recombination Between Bacterial Plasmids: Involvement of Palindromes at the Recombinant Loci.	Dept. Medicine, Stanford Univ., School of Medicine, Stanford, California, USA.	No grant data	2	–
2.	Kopecko, D.J., Brevet, J. & Cohen, S.N.	J.Mol.Biol v.108 p333 1976	Involvement of Multiple Translocating DNA Segments and Recombinational Hot Spots in the Structural Evolution of Bacterial Plasmids.	Dept. Medicine, Stanford Univ., School of Medicine, Stanford, California, USA.	Nat. Inst. Health, National Science FDN & American Cancer Society.	1	1

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.15 Comparison of National Rankings for Base and Current Literature Papers in Specialty 543.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. USA	2	100	1. USA	12	41.40
			2. Switzerland	3	10.30
			2. UK	3	10.30

Table 4.16 Details of Base Literature Papers in Specialty 134.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Berg, D.E.	DNA Insertion Elements p205 1977	Insertion and Excision of the Transposable Kanamycin Resistance Determinant Tn5.	Dept. Biochem, Univ of Wisconsin, Madison, Wisconsin. USA.	Fonds Nat. Suisse de la Rech. Sci.; Nat. Sci. FDN; Nat. Cancer Inst.	2	–
2.	Botstein, D., Kleckner, N.	DNA Insertion Elements p185 1977	Translocation and Illegitimate Recombination by the Tetracycline Resistance Element Tn10.	Dept. Biology, Mass. Inst. Tech., Cambridge, Massachusetts, USA.	Nat. Inst. Health; American Cancer Society.	1,3	3
3.	Kleckner, N., Chan, R.K., Tye, B.K., Botstein, D.	J.Molec.Biology v97 p561 1975	Mutagenesis by Insertion of a Drug Resistance Element Carrying an Inverted Repetition.	Dept. Biology, Mass. Inst. Tech., Cambridge, Massachusetts, USA.	Nat. Inst. of Health; American Cancer Society.	2	–

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.17 Comparison of National Rankings for Base and Current Literature Papers in Specialty 134.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. USA	3	100	1. USA	44	69.80
			2. Japan	4	6.30
			2. UK	4	6.30

Table 4.18 Details of Base Literature Papers in Specialty 82.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
1.	Arber, W., Iida, S., Jutte, H. Casporo, P., Meyer, J. & Hanni, C.	Cold Spring Harbour Symposium v.43 p1197 1978	Rearrangement of Genetic Material in E.Coli as Observed on the Bacteriophage P1 Plasmid.	Dept. Microbiology, Univ. of Basel, Biozentrum, Switzerland.	Swiss National Science Foundation.	3, 4, 5	2, 3, 6
2.	Bachi, B. & Arber, W.	Mol.General Genetics v.153 p311 1977	Physical Mapping of BglII, Bam HI, Eco R1, Hind III, PstI Restriction Fragments of Bacteriophage P1 DNA.	Biozentrum der Universitat Basel, Basel, Switzerland.	Swiss National Science Foundation.	6	–
3.	Hu, S., Ohtsubo, E., Davidson, N. & Saedler, H.	J.Bact. v.122 p764 1975	Electron Microscope Heteroduplex Studies of Sequence Relations Among Bacterial Plasmids: Identification and Mapping of the Insertion Sequences IS1 and IS2 in the F and R Plasmids.	Dept. Chem., Calif. Inst. of Technology, Pasadena, Calif, USA; Institut fur Genetik, Univ. zu Koln, Cologne, W. Germany.	Deutsche Forschungsgemeinschaft; Nat. Inst. of General Medical Sciences.	1	–

Table 4.18, continued.

No.	Authors	Reference	Title	Institution	Financial Support	Co-cited with	Papers cited in own specialty
4.	Iida, S. & Arber, W.	Molec.General Genetics v.177 p261 1980	On the Role of IS1 in the Formation of Hybrids Between the Bacteriophage P1 and the R Plasmid NR1.	Dept. Microbiology, Univ. Basel, Biozentrum, Basel, Switzerland.	Swiss National Science Foundation.	1, 6	1, 2, 3
5.	Ohtsubo, E., Zenilman, M. & Ohtsubo, H.	PNASUSA v.77 p750 1980	Plasmids Containing Insertion Elements are Potential Transposons.	Dept. Microbiology, Health Sciences Centre, State Univ., Stony Brook, N. York, USA.	US Public Health Service; N.I.H.	1	–
6.	Yun, T. & Vapnek, D.	Virology v.77 p376 1977	Electron Microscope Analysis of Bacteriophages P1, P1Cm and P7: Determination of Genome Sizes, Sequence Homology and Location of Antibiotic Resistance Determinants.	Program in Genetics, Dept. Microbiology & Dept. Biochemistry, Univ. Georgia, Athens, Georgia, USA.	Nat. Inst. of General Medical Sciences, US Public Health Service.	2, 4	3

NB: The numbers given in the "co-cited with" and "cited" columns refer to the papers' identifying numbers.

Table 4.19 Papers in which Base Literature Authors Appeared in Specialty 82.

Author	Papers in which they appeared (denoted by number)
1. Werner Arber	1, 2, 4
2. Brigitte Bachi	2
3. P. Caspars	1
4. Norman Davidson	3
5. Christine Hanni	1
6. Sylvia Hu	3
7. Shgeru Iida	1,4
8. H. Jutte	1
9. J. Meyer	1
10. Eichi Ohtsubo	3,5
11. Hisako Ohtsubo	5
12. Heinz Saedler	3
13. Daniel Vapnek	6
14. Thomas Yun	6
15. Michael Zenilman	5

NB: The identifying numbers for the papers are given in Table 4.18.

Table 4.20 Comparison of National Rankings for Base and Current Literature Papers in Specialty 82.

Base Literature			Current Literature		
Country	No. of Papers	% of Total	Country	No. of Papers	% of Total
1. U.S.A.	3	50	1. U.S.A.	15	35.7
1. Switzerland	3	50	2. Switzerland	11	26.2
3. W. Germany	1	16.66	4. U.K.	3	7.1

Table 4.21 Specialties in which the Base Literature Papers of Trend 44 Occur.

No.	First Author and Reference of Paper	Time Periods and Specialty Numbers		
		1978	1980	1982
1.	Arber, W. CSHS 43 p1197 1978			82
2.	Bachi, B. MGG 153 p311 1977			82
3.	Barth, P.T. J.Bact 125 p800 1976	101		
4.	Bennett, P.M. J.Bact 126 p7 1976	101		
5.	Berg, D.E. DIE p205 1977		134	
6.	Berg, D.E. PNASUSA 72 p3628 1975	101		
7.	Botstein, D. DIE p185 1977		134	
8.	Calos, M.P. Cell 13 p411 1978		222	
9.	Davidson, N. Microbiology p56 1974	320	222	
10.	Fiandt, M. MGG 119 p223 1972	388		
11.	Gottesman, M. PNASUSA 72 p5041 1975	101	222	
12.	Grindley, N. Cell 13 p419 1978		222	
13.	Hedges, R.W. MGG 132 p31 1974	101		
14.	Heffron, F. J.Bact 122 p250 1975	101		
15.	Heffron, F. PNASUSA 72 p3623 1975	101		
16.	Hu, S. J.Bact 122 p764 1975	525	222	82
17.	Hu, S. J.Bact 122 p749 1975	320		
18.	Iida, S. MGG 177 p261 1980			82
19.	Johnsrud, L. Cell 15 p1209 1978		222	
20.	Kleckner, N. Cell 16 p711 1979		222	
21.	Kleckner, N. JMB 97 p561 1975	101	134	
22.	Kondo, E. J.Bact 88 p1266 1964	101		
23.	Kopecko, D. JMB 108 p333 1976		543	
24.	Kopecko, D. PNASUSA 72 p1373 1975	101	543	
25.	Kuhn, S. MGG 167 p235 1979		222	
26.	MacHattie, L.A. DIE p219 1977	101	222	
27.	Ohtsubo, E. PNASUSA 77 p750 1980			82
28.	Ohtsubo, E. JMB 89 p565 1974	320		
29.	Ohtsubo, H. PNASUSA 75 p615 1978		222	

Table 4.21, continued.

No.	First Author and Reference of Paper	Time Periods and Specialty Numbers		
		1978	1980	1982
30.	Ptashne, K. J.Bact 122 p776 1975	525	222	
31.	Rubens, C. J.Bact 128 p425 1976	101		
32.	Saedler, H. MGG 122 p267 1973	388	222	
33.	Sharp, P.A. JMB 71 p471 1972	320		
34.	Yun, T. Virology 77 p376 1977			82

Key to journal abbreviations

CSHS	Cold Spring Harbour Symposium
DIE	DNA Insertion Elements (book)
J.Bact	Journal of Bacteriology
JMB	Journal of Molecular Biology
MGG	Molecular and General Genetics
PNASUSA	Proceedings of the National Academy of Sciences, United States of America

Table 4.22 Origins of Papers in the Base Literature of Specialties in Trend 44.

Specialty	Papers from previous specialties	Specialty Origin	New Papers	Specialty Origin (current literature)
543	Kopecko, D. (PNASUSA v72 p1373)	101	Kopecko, D. (JMB v108 p333)	
134	Kleckner, N. (JMB v97 p561)	101	Berg, D. (DIS p205) Botstein, D. (DIS p185)	
222	Hu, S. (J.Bact v122 p764)	525	Calos, M. (Cell v13 p411)	101, 388
	Davidson, N. (Microbiol p56)	320	Grindley, N. (Cell v13 p419)	101, 388
	Gottesman, M. (PNASUSA v72 p5041)	101	Johnsrud, L. (Cell v15 p1209)	
	MacHattie, L. (DIS p219)	101	Kleckner, N. (Cell v16 p711)	
	Ptashne, K. (J.Bact v122 p776)	525	Kuhn, S. (MGG v167 p235)	
	Saedler, H. (MGG v122 p267)	388	Ohtsubo, H. (PNASUSA v75 p615)	101, 388, 525
82	Hu, S. (J.Bact v122 p764)	222	Arber, W. (CSHS v43 p1187) Bachi, B. (MGG v153 p311) Iida, S. (MGG v177 p261) Ohtsubo, E. (PNASUSA v77 p750) Yun, T. (Virology v77 p376)	222

NB: Data is only shown here for specialities in 1980 and 1982. As no models were built prior to 1978, the origins of papers in 1978 could not be examined.

Table 4.23 Showing the Authors of the Base Literature of Trend 44, the Specialties in Which they Appear, and the Number of Times that they Appear.

	1978				1980			1982
Speciality No.	320	101	388	525	222	543	134	82
Author								
Allet, B.		1						
Arber, W.								3
Bachi, B.								1
Barth, P.T.		1						
Bennett, P.M.		1						
Berg, D.		1					1	
Botstein, D.		1					2	
Brevet, J.						1		
Calos, M.P.					2			
Caspars, P.								1
Chan, R.K.		1					1	
Cohen, S.N.		1		1	1	2		
Datta, N.		1						
Davidson, N.	4			1	2			1
Davies, J.		1						
Deonier, R.C.	2				1			
Falkow, S.		3						
Fiandt, M.			1					
Fritz, H.J.					1			
Gottesman, M.		1			1			
Grindley, N.					1			
Grinter, N.J.		1						
Hanni, C.								1
Hedges, R.W.		3						
Heffron, F.		3						
Heiß, B.			1		1			
Hsu, M.T.	1							
Hu, S.	2			1	2			1
Iida, S.								2
Jakowski, J.B.		1			1			
Jacob, A.		2						
Johnsrud, L.					2			
Jutte, H.								1
Kleckner, N.		1			1		2	
Kondo, E.		1						
Kopecko, D.J.		1				2		
Kuhn, S.					1			
Lee, H.J.	1							
MacHattie, L.A.		1			1			
Malamy, J.			1					
Meyer, J.								1
Miller, J.H.					2			
Mitsuhashi, S.		1						
Ohtsubo, E.	4			1	3			2
Ohtsubo, H.					1			1
Ptashne, K.				1	1			
Richmond, M.H.		1						
Rochaix, J.D.		1						
Rosner, J.L.		1			1			
Rubens, C.		2						
Saedler, H.			1	1	2			1
Sharp, P.A.	1							
Starlinger, P.					1			
Sublett, R.		1						
Szybalski, W.			1					
Tye, B.K.		1					1	
Vapnek, D.								1
Yun, T.								1
Zenilman, M.								1

Tables 4.24 to 4.31 Show the Occurrence of Base Literature Authors in the Current Literature of Trend 44. Specialties in 1978.

Table 4.24 Specialty 101.

Base Literature Authors	Occurrence in Base Literature of 101	No. of Papers in Current Literature	Base Literature Authors	Occurrence in Base Literature of 101	No. of Papers in Current Literature
Arber. W.		1	Iida. S.		1
Barth. P.T.	●	2	Jacob. A.E.	●	1
Bennett. P.M.	●	8	Johnsrud. L.		1
Berg. D.E.	●	4	Kleckner. N.	●	1
Botstein. D.		2	Kopecko. D.J.	●	3
Brevet. J.		1	MacHattie. L.A.	●	1
Calos. M.P.		1	Meyer. J.		1
Cohen. S.N.	●	5	Miller. J.H.		1
Datta. N.	●	3	Ohtsubo. E.		2
Davies. J.	●	3	Ohtsubo. H.		1
Falkow. S.	●	6	Richmond. M.H.	●	8
Fiandt. M.		1	Rochaix. J.D.	●	2
Grindley. N.D.		2	Saedler. H.		1
Grinter. N.J.	●	1	Sharp. P.A.		1
Hedges. R.W.	●	2	Starlinger. P.		1
Heffron. F.	●	2			

No. of papers authored by base literature people in current literature = 44
(Total No. of individual appearances = 70).
Total No. of current literature papers = 150. Proportion of current literature papers authored by base literature people = 29.3%.

Table 4.25 Specialty 525.

Base Literature Authors	Occurrence in Base Literature of 525	No. of Papers in Current Literature	Base Literature Authors	Occurrence in Base Literature of 525	No. of Papers in Current Literature
Arber. W.		1	Jacob. A.E.		1
Berg. D.E.		2	Kopecko. D.J.		3
Brevet. J.		2	Meyer. J.		1
Cohen. S.N.	●	5	Ohtsubo. E.		6
Davies. J.		1	Ohtsubo. H.		2
Hedges. R.W.		1	Saedler. H.		2
Hsu. M.T.		2	Starlinger. P.		1
Iida. S.		1			

No. of papers authored by base literature people in current literature = 18 (Total No. of individual appearances = 31). Total No. of current literature papers = 36. Proportion of current literature papers authored by base literature people. = 50%

Table 4.26 Specialty 320.

Base Literature Authors	Occurrence in Base Literature of 320	No. of Papers in Current Literature	Base Literature Authors	Occurrence in Base Literature of 320	No. of Papers in Current Literature
Brevet. J.		1	Ohtsubo. E.	●	4
Cohen. S.N.		2	Ohtsubo. H.		1
Hsu. M.T.	●	2	Saedler. H.		2
Kopecko. D.J.		2	Starlinger. P.		1

No. of papers authored by base literature people in current literature = 8 (Total No. of individual appearances = 15).
Total No. of current literature papers = 44. Proportion of base literature authored current literature papers = 18.18%.

Table 4.27 Specialty 388.

Base Literature Authors	Occurrence in Base Literature of 388	No. of Papers in Current Literature	Base Literature Authors	Occurrence in Base Literature of 388	No. of Papers in Current Literature
Calos. M.P.		1	Miller. J.H.		1
Gottesman. M.		1	Ohtsubo. E.		2
Grindley. N.D.		1	Ohtsubo. H.		1
Hsu. M.T.		1	Saedler. H.	●	2
Johnsrud. L.		1	Starlinger. P.		1

No. of papers authored by base literature people in current literature = 8 (Total No. of individual appearances = 11).
Total No. of current literature papers = 20. Proportion of base literature authored current literature papers = 40%

Table 4.28 Specialty 134.

Base Literature Authors	Occurrence in Base Literature of 134	No. of Papers in Current Literature	Base Literature Authors	Occurrence in Base Literature of 134	No. of Papers in Current Literature
Bennett. P.M.		1	Kleckner. N.	●	1
Berg. D.E.	●	2	Miller. J.H.		2
Calos. M.P.		2	Richmond. M.H.		1
Cohen. S.N.		2	Starlinger. P.		1
Falkow. S.		1			

No. of papers authored by base literature people in current literature = 10 (Total No. of individual appearances = 13).
Total No. of current literature papers = 63. Proportion of base literature authored current literature papers = 15.87%.

Table 4.29 Specialty 222.

Base literature authors	Occurrence in Base literature of 222	No. of papers in current literature	Base literature authors	Occurrence in Base literature of 222	No. of papers in current literature
Arber. W.		2	Meyer. J.		1
Bennett. P.M.		1	Miller. J.H.	●	5
Berg. D.E.		2	Ohtoubo. E.	●	3
Calos. M.P.	●	5	Ohtoubo. H.	●	3
Cohen. S.N.	●	2	Richmond. M.H.		2
Davies. J.A.		1	Rosner. J.L.	●	1
Deonier. R.C.	●	1	Saedler. H.	●	2
Falkow. S.		1	Starlinger. P.	●	4
Grindley. N.D.	●	1	Zenilman. M.		2
Iida. S.		3			

No. of papers authored by base literature people in current literature = 34 (Total No. of individual apperances = 42).
Total No. of papers in current literature = 103. Proportion of base literature authored current literature papers =33%.

Table 4.30 Specialty 543

Base literature authors	Occurrence in Base literature of 543	No. of papers in current literature	Base literature authors	Occurrence in Base literature of 543	No. of papers in current literature
Bennett P.M.		2	Miller J.H.		2
Calos M.P.		2	Richmond M.H.		1
Cohen S.N.	●	3			

No. of papers authored by base literature people in current literature = 7 (Total No. of individual apperances =10).
Total No. of papers in current literature = 29. Proportion of base literature authored current literature papers = 24.13%.

Table 4.31 Specialty 82.

Base literature authors	Occurrence in Base literature of 82	No. of papers in current literature	Base literature authors	Occurrence in Base literature of 82	No. of papers in current literature
Arber. W.	●	5	Iida. S.	●	7
Berg. D.E.		1	Meyer. J.	●	3
Datta. N.		1	Ohtsubo. H.		1
Deonier. R.C.		1	Saedler. H.		2
Hanni. C.	●	2			

No. of papers authored by base literature people in current literature = 14 (Total No. of individual apperances =23).
Total No. of papers in current literature = 42. Proportion of base literature authored current literature papers = 33.3%.

Table 4.32 Showing the Speciality Current Literatures in which the Base Literature Authors of Trend 44 Appear.

Base Literature Authors Name	1978 Specialties Current Literatures		1980 Specialties Current Literatures		1982 Specialties Current Literatures	
	No. of Papers	No. of Specialties	No. of Papers	No. of Specialties	No. of Papers	No. of Specialties
Arber, W.	2	2	2	1	6	1
Barth, P.T.	2	1				
Bennett, P.M.	8	1	4	3		
Berg, D.	6	2	4	2	1	1
Botstein, D.	2	1				
Brevet, J.	4	3				
Calos, M.P.	2	2	9	3		
Cohen, S.N.	12	3	7	3		
Datta, N.	3	1			1	1
Davies, J.	4	2	1	1		
Deonier, R.C.			1	1	1	1
Falkow, S.	6	1	2	2		
Fiandt, M.	1	1				
Gottesman, M.	1	1				
Grindley, N.	3	2	1	1		
Grinter, N.J.	1	1				
Hanni, C.					2	1
Hedges, R.W.	3	2				
Heffron, F.	2	1				
Hsu, M.T.	5	3				
Iida, S.	2	2	3	1	7	1
Jacob, A.	2	2				
Johnsrud, L.	2	2				
Kleckner, N.	1	1	1	1		
Kopecko, D.J.	8	3				
MacHattie, L.A.	1	1				
Meyer, J.	2	2	1	1	3	1
Miller, J.H.	2	2	9	3		
Ohtsubo, E.	14	4	3	1		
Ohtsubo, H.	5	4	3	1	1	1
Richmond, M.H.	8	1	4	3		
Rochaix, J.D.	2	1				
Rosner, J.L.			1	1		
Saedler, H.	7	4	2	1	2	1
Sharp, P.A.	1	1				
Starlinger, P.	4	4	5	2		
Zenilman, M.			2	1		

NB: 1978 had a total number of 4 specialities, 1980 had 3 and 1982 only one.

Table 4.33 Patterns of Citation Between First Named Base Literature Authors in Trend 44.

Citing Author Showing which paper is concerned	Cited Author (numbered as Author column). 1	2	3	4	5	7	8	9	10	11	12	13	14	16	18	19	20	22	23	25	26	27	29	30	31	32	33	34	% First named trend author cited
1. Arber, W.		●				●								●	●			●			●							●	25.0
2. Bachi, B.															●														3.5
3. Barth, P.T.			●		●							●	●	●			●		●					●				28.6
4. Bennett, P.M.												●	●				●							●				7.1
5. Berg, D.			●		●					●		●	●				●							●				25.0
6. Berg, D.					●			●		●		●	●	●			●	●	●					●				35.7
7. Botstein, D.					●					●							●							●			●	32.1
8. Calos, M.P.									●	●		●	●	●			●				●		●		●			25.0
9. Davidson, N.									●	●												●				●	●	17.8
10. Fiandt, M.									●	●																●		10.7
11. Gottesman, M.									●	●			●				●		●					●		●		28.6
12. Grindley, N.							●				●						●		●		●		●			●		28.6
13. Hedges, R.W.									●			●																10.7
14. Heffron, F.			●		●							●	●				●		●					●				28.6
15. Heffron, F.												●																3.6
16. Hu, S.								●	●					●								●		●		●	●	25.0
17. Hu, S.								●	●	●				●								●					●	17.0
18. Iida, S.	●	●			●	●	●			●	●			●	●			●			●	●	●			●		25.0
19. Johnsrud, L.																●	●			●	●							42.85
20. Kleckner, N.					●	●	●	●		●	●					●	●						●	●		●	●	25.0
21. Kleckner, N.					●	●		●		●			●	●					●					●		●	●	35.7
22. Kondo, E.	●																	●										7.1
23. Kopecko, D.			●		●			●		●		●	●	●			●		●					●				35.7
24. Kopecko, D.																		●										17.8
25. Kuhn, S.									●		●			●							●		●			●	●	21.4
26. MacHattie, L.A.									●	●				●														17.8
27. Ohtsubo, E.														●		●					●	●	●					28.6
28. Ohtsubo, E.								●		●												●						10.7
29. Ohtsubo, H.					●				●		●		●	●			●		●		●	●	●			●	●	39.3
30. Plashne, K.														●					●							●	●	14.3
31. Rubens, C.				●	●			●		●		●	●	●			●		●					●		●	●	46.4
32. Saedler, H.				●	●				●									●				●						7.1
33. Sharp, P.A.								●		●																		10.7
34. Yun, T.					●					●		●	●	●			●	●			●						●	35.7
No. of times each Trend author cited (%)	5.9	5.9	14.7	2.9	32.3	8.8	14.7	29.4	29.4	47.1	20.6	32.3	35.3	47.1	8.8	8.8	41.2	23.5	32.3	2.9	26.5	23.5	20.6	35.3	0.0	44.1	41.2	2.9	

NB: The details of the citing papers are available from Table 4.22.

Table 4.34 Proportion of References Attributable to First Named Base Literature Authors in the Reference Lists of the Base Literature Papers in Trend 44.

Paper (denoted by First Author Name)	No. of References by FNTA	Total No. of References	Proportion by FNTA (expressed as a percentage)
1. Arber, W.	8	33	24.24
2. Bachi, B.	1	28	3.57
3. Barth, P.T.	12	32	37.50
4. Bennett, P.M.	2	21	9.52
5. Berg, D.E.	8	15	53.33
6. Berg, D.E.	9	23	39.13
7. Botstein, D.	9	24	37.50
8. Calos, M.P.	13	48	27.08
9. Davidson, N.	9	21	42.
10. Fiandt, M.	3	20	15.00
11. Gottesman, M.	10	29	34.48
12. Grindley, N.	9	38	23.68
13. Hedges, R.W.	6	25	24.00
14. Heffron, F.	6	34	17.65
15. Heffron, F.	8	26	30.77
16. Hu, S.	12	38	31.58
17. Hu, S.	5	24	20.83
18. Iida, S.	10	33	30.30
19. Johnsrud, L.	13	52	25.00
20. Kleckner, N.	14	40	35.00
21. Kleckner, N.	12	54	22.22
22. Kondo, E.	3	33	9.1
23. Kopecko, D.J.	18	67	26.87
24. Kopecko, D.J.	4	30	13.33
25. Kuhn, S.	7	27	25.92
26. MacHattie, L.A.	7	16	43.75
27. Ohtsubo, E.	12	37	32.43
28. Ohtsubo, E.	5	26	19.26
29. Ohtsubo, H.	15	31	48.39
30. Ptashne, K.	5	26	19.23
31. Rubens, C.	14	37	37.84
32. Saedler, H.	3	27	11.11
33. Sharp, P.	5	52	9.61
34. Yun, T.	11	28	39.29

NB: The details of the citing papers are available from Table 4.21. FNTA stands for First Named Trend Author.

Table 4.35 Occurrence of Institutions in the Base Literature of Specialties in Trend 44.

	1978				1980			1982
Specialty No.	320	101	388	525	222	543	134	82
Institution								
1. Dept. of Chemistry, Calif. Inst. Technology, Pasadena, Calif, 91109, USA.	4			0.5	1.5			0.5
2. Dept. Bacteriology, Royal Postgrad. Med. Sch., London, W12, UK.		2.5						
3. Dept. Bacteriology, Medical School, Bristol University, Bristol, UK.		1						
4. Dept. Microb. & Immunol., Sch Medicine, Washington Univ., Seattle, 98195, USA.		2.5						
5. Mol. Biol. Lab., Nat. Inst. Arth. Met. Dig. Dis., Bethesda, Maryland, 20014, USA.		1			1			
6. Dept. Medicine, Stanford Univ., Calif, 94305, USA.		1		1	1	2		
7. Dept. Microbiol., M.I.T., Cambridge, Mass., 02139, USA.		1					2	
8. McArdle Lab., Univ of Wisconsin, Madison, Wisconsin, USA.			1				1	

Table 4.35, continued.

	1978				1980			1982
Specialty No.	320	101	388	525	222	543	134	82
Institution								
9. Dept. Mol. Biol. & Microb., Tufts Univ., Boston, Mass., USA.			1					
10. Dept. Microb., State Univ., Stony Brook, N.Y., 11794, USA.					1			1
11. Biol. Labs., Harvard Univ., Cambridge, Mass., 02138, USA.					2			
12. Dept. Mol. Bioph. Biochem., Yale Univ., New Haven, Connecticut, 06510, USA.					1			
13. Dept. Microb., Dept. Biochem., Univ. of Georgia, Athens, Georgia, 30602, USA.								1
14. Dept. Microb., Sch. Med., Gunma Univ., Maebashi, Japan.		1						
15. Dept. Mol. Biol., Univ of Geneva, Geneva, Switzerland.		1			1			
16. Dept. Med. Gen., Univ of Toronto, M5S 7A8, Canada.		1			1			
17. Inst. Gen., Koln Univ., Cologne, W. Germany.			1	0.5	2.5			0.5
18. Dept. Micro., Basel Univ., Basel, CH405E, Switzerland.								3

Key

The number of papers originating from each institution in a specialty is shown above. Where the value is not a whole number this indicates that the paper originated from more than one institution. The value of 1 for each paper, in these cases, is divided by the number of institutions involved in its production, and the value of the fraction obtained was awarded to each component institution.

Table 4.36 Distribution of Journals in the Base Literatures of Specialties in Trend 44.

Occurrence in Core Journal Set	Year of Model	1978				1980			1982	Total number of papers
	Specialty No.	320	101	388	525	222	543	134	82	
	Journal									
●	Cell					4				4
	Cold Spring Harbor Symposium								1	1
	DNA Insertion Elements (book)		1			1		2		4
●	Journal of Bacteriology	1	5		2	2			1	11
●	Journal Of Molecular Biology	2	1				1	1		5
	Microbiology 1974 (book)	1				1				2
●	Molecular & General Genetics		1	2		2			2	7
	Proceedings National Academy Sciences, USA		4			2	1		1	8
●	Virology								1	1

Key

The numbers listed under each specialty denote the numbers of papers that the Journals concerned had in those specialities.

Table 4.37 Acknowledgements to First Named Base Literature Authors in Base Literature Papers in Trend 44.

Papers in Trend 44 (denoted by First Named Author)	First Named Trend Authors See key below for names. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
2. Bachi, B.	
8. Calos, M.P.	
9. Davidson, N.	
12. Grindley, N.	
13. Hedges, R.W.	
15. Heffron, F.	
16. Hu, S.	
19. Johnsrud, L.	
20. Kleckner, N.	
21. Kleckner, N.	
23. Kopecko, D.	
24. Kopecko, D.	
25. Kuhn, S.	
26. MacHattie, L.A.	
30. Ptashne, K.	

NB: Only the papers that had any acknowledgement to first named authors in the trend are shown here (in first column). They are denoted by their first author's name andy with them in the table correspond to their identifying numbers in Table 4.21 for full reference purposes.

Key to First Named Trend Authors

1. Arber, W.	2. Bachi, B.	3. Barth, P.T.	4. Bennett, P.M.
5. Berg, D.E.	6. Botstein, D.	7. Calos, M.P.	8. Davidson, N.
9. Fiandt, M.	10. Gottesman, M.	11. Grindley, N.	12. Hedges, R.W.
13. Heffron, F.	14. Hu, S.	15. Iida, S.	16. Johnsrud, L.
17. Kleckner, N.	18. Kondo, E.	19. Kopecko, D.	20. Kuhn, S.
21. MacHattie, L.A.	22. Ohtsubo, E.	23. Ohtsubo, H.	24. Ptashne, K.
25. Rubens, C.	26. Saedler, H.	27. Sharp, P.A.	28. Yun, T.

Table 4.38 Industrial Concerns in the Current Literatures of Specialties in Trend 44.

Year	Specialty	Company
1978	101	City New York Inc.
		Ciba Geigy AG
		F. Hoffman La Roche & Co. Ltd.
1978	320	GE Corporate R&D
		City New York Inc.
1978	388	City New York Inc.
1978	525	GE Corporate R&D
		City New York Inc.
1980	543	
1980	134	Schering AG & Corporation
1980	222	City New York Inc.
		Searle R&D
1982	82	

Chapter 5: Analysis of the Base Literature of Trend 11 "The Biochemistry and Biosynthesis of Bacterial Cell Membrane Proteins"

Section 1: Introduction

Trend 11, see Fig. 5.1, was examined in order to test whether some of the patterns of trend behaviour identified from the literature of Trend 44 can be verified by data derived from another complex trend. The aim of the exercise was not only to check some of our conclusions from the analysis of Trend 44, but also to further test the trend analysis methodology developed in Chapter 4.

Trend 11 was selected not only for its complexity but also because it included some of the work of Aston University Pharmacy Department. The intention initially was to test the trends' portrayal of their work, but as time ran out for this the criteria of their involvement was eventually only used to narrow down the number of specialties examined in detail. However any future studies concentrating on Aston Pharmacy's involvement in Trend 11 would probably find the data presented in this chapter, about some of the properties of its base literature, useful.

Section 2: Trend Analysis Methodology

Trend 11 was found to be much larger than Trend 44, being composed of 17 specialties containing 168 papers. Therefore the analysis of the characteristics of the base literature was of necessity less detailed than for Trend 44. The approach to the analysis of Trend 11 was to partially validate it on the lines of the analysis of Trend 44.

Data for the analysis of the base literature of Trend 11 was obtained from the papers themselves which were examined from library sources, (the same as for the analysis of Trend 44), in order to augment the scanty information about the base

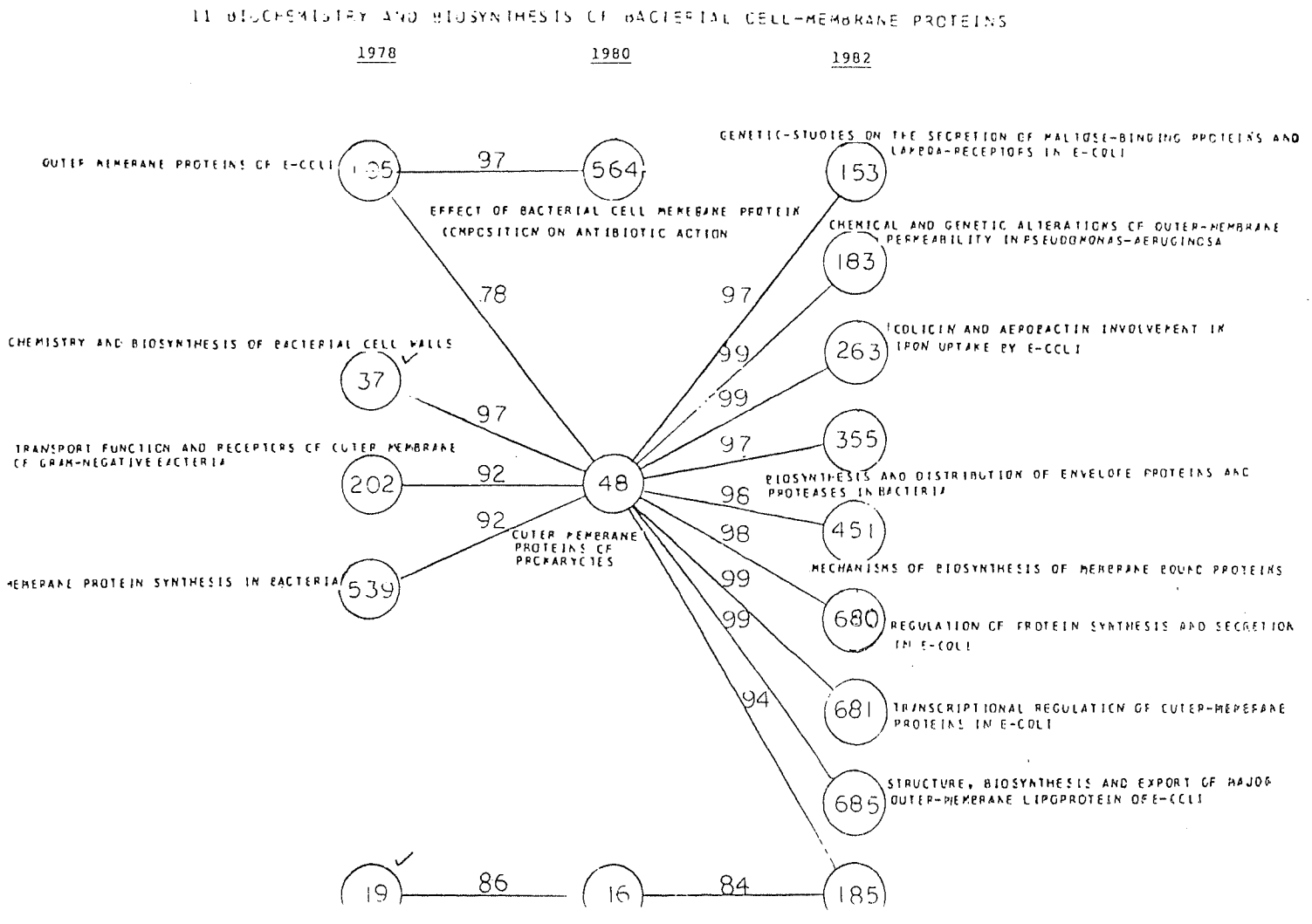


Fig. 5.1: Trend Diagram for Trend 11

literature in the models. Due to the large size of the base literature in Trend 11 a detailed analysis was done only for the specialties linked to Aston Pharmacy Department that were identified by H. Rothman in an unpublished seminar to Aston Pharmacy Department. These specialties were: no. 37, "The Biochemistry and Biosynthesis of Bacterial Cell Walls" (1978); no. 564, "Effect of Bacterial Cell Membrane Protein Composition on Antibiotic Action" (1980); no. 48, "Outer Membrane Proteins of Prokaryotes" (1980) and no. 183, "Chemical and Genetic Alterations of Outer Membrane Permeability in Pseudomonas Aeruginosa" (1982). Only sufficient information was taken from the papers in the other specialties to get an idea of the overall behaviour of the trend and the relationships between the specialties within it.

The intention was firstly to analyse Trend 11 at the level of its component specialties to see whether the same patterns of citation between co-cited pairs of papers could be observed for Trend 11 as had been found for Trend 44. At the trend level of organisation the analysis attempted to see whether the links shown between the specialties in Trend 11 could be substantiated by the patterns of re-occurrence of its authors. The data contained in the Trend Summary Report was examined to see whether it supported the patterns of birth and mortality of the trend's component specialties, (see the Trend Diagram Fig 5.1).

The method of analysis was essentially the same as for Trend 44, see Chapter 4. However no cognitive details (which would have been obtained from examination of the abstracts), were noted for any specialty since this would have made the analysis too time consuming. Similarly examination of the current literatures for the specialties in Trend 11 was abandoned as there were a total of 1370 papers in the current literature and 168 papers in the base literature which would have made searching for the origins of new base literature papers in the current literatures of previous years too lengthy a process. Also the total number of base literature authors (219 including the co-authors), would have resulted in a similar problem in the analysis of the occurrence of these authors in the current literature papers.

So for specialties other than those involving Aston Pharmacy in the current literature only the names of the authors affiliated to the base literature papers were collected. For the specialties involving Aston Pharmacy only the author details and the dates of publication were extracted.

Using this basic data, the following analyses were carried out:

The Aston affiliated specialties were examined individually (at the specialty level) to see whether the pattern of citation between a paper and its co-cited partners reflected their co-citation groupings.

Then at the trend level of analysis for every specialty it was observed;

(a): Which papers linked the specialties together in the trend.

(b): The Specialty Summary Report given for Trend 11 in the model was examined to see how it explains the development of the trend. The position of the Aston affiliated specialties in the Trend Diagram (Fig. 5.1) was also examined.

(c): All of the first named named trend base literature authors (including the co-authors identified from the papers), were tabulated against the specialties in which they appeared, in order to trace links between the specialties that existed in addition to those caused by authors attached to recurring papers. Also the specialties in which authors occurred more than once were noted, since this could indicate ties between the base literature papers.

Section 3: Data from the analysis at the Specialty Level of Trend 11

Analysis of Citation of Papers by their Co-cited Partners for Specialties Involving Aston Pharmacy Department

In the analysis of Trend 44 (Chapter 4, Section 3) it was discovered that the base literature papers tended to cite their co-cited partners in apparent support of the relationships

perceived between them by the current literature. In order to see whether this also applied to the base literature of Trend 11 the Aston affiliated specialties were examined to see whether the base literature papers cited their co-cited partners. In a manner similar to the analysis of Trend 44 the date of publication was taken into account as obviously a paper cannot cite one that was published after it. Where both papers concerned were published in the same year, the dates that they were received for publication (when given) were taken into consideration.

So for Specialty 37, see Table 5.1, which had only 2 papers and thus only 1 co-cited pair, neither of the papers were observed to cite each other. However they were received for publication within 2 months of each other and published in the same year, which would have left little opportunity for citation to occur. In Specialty 564, see Table 5.2, again only 2 papers were involved. However one preceded the other by 2 years and so although citation was possible, the other did not cite it.

The same process of analysis was applied to Specialty 48, data for which is listed in Table 5.3. However as it contains 64 papers, for the sake of brevity it was decided to discuss only those failures to cite co-cited partners that could not be explained away by the date of publication.

Bavoil.P. (number 5 in Table 5.3) failed to cite 3 of its co-cited partners. Bassford.P (numbers 2 and 3) and Foulds.J. (number 14) all failed to cite 2 of their co-cited partners. Bassford.P. (number 4), Datta.D.B. (number 11), Hasegawa.Y. (number 17), Henning.U. (number 19), Ichihara. S. (number 21), Palva.E.T. (number 37), Pugsley.A.P. (number 38), Pugsley.A.P. (number 39), Schindler. H. (number 45), Schmitges.C.J. (number 46), Schweizer.M. (number 48), Sekizawa.J. (number 49), Steven.A.C. (number 56), VanAlphen.L. (number 61), VanAlphen.W. (number 62) all failed to cite 1 of their co-cited partner papers.

Another paper that failed to cite one of its co-cited partners was Smith.W.P. (number 55) which did not cite Randall.L.L. (number 41). This was interesting since Smith.W.P. cited this work in his earlier paper (number 54). So perhaps this is why he was paired with it later as he identified a link between

his own work and that of Randalls which was picked up by other authors in connection with his next paper. It is possible that papers clustered together but which could not cite each other because of too close dates of publication or which fail to cite each other even though a sufficient time interval between their dates of publication had elapsed to make it theoretically possible, may show relationships perceived purely by the current literature. By differentiating between these 2 types of citation links, i.e. between those drawn by the base literature papers themselves and those perceived later by the current literature it may be possible to differentiate between old associations of ideas in a field and new ones.

Looking at the pairings from each papers point of view in Specialty 48 there were 306 instances of co-citation links between papers. Looking at citations between these pairings there were only 25 instances where citation failed to occur although the date of publication would have made it possible, i.e. 8.2%. The number of instances where citation of co-cited partners occurred was 93 i.e. 30.4% and of course 61.4% were instances where citation was impossible due to the time factor. Thus there were nearly 4 times as many cases of citation of co-cited partners occurring as failure to cite without excuse. This and the other observation from this analysis would tend to support the idea that base literature authors perceive relationships between themselves that are later picked up by the current literature. In Trend 44 it was observed that the papers often cited other papers in the base literature of the specialty that they were not directly paired with. This occurred in 51 of the papers in Specialty 48, i.e. 79.7% of them. Thus there seem to be cognitive relationships perceived between papers indirectly clustered together by appearing in the base literature of the same specialty.

In Specialty 183, see Table 5.4, there were only 5 papers. Out of these only Nicas failed to cite Hancock.R.E.W. (number 2 in Table 5.4), although it would have been possible to do so. Thus out of a total of 10 co-citation pairings, there was 1 instance of failure to cite (10%), 4 instances of citation of co-cited partners (40%), and 5 instances where citation would be

impossible due to the time factor (50%). Three of the papers in the specialty showed links to other papers in the base literature that they were not directly paired with, i.e. 60%.

Section 4: Data from the Analysis at the Trend Level of Organisation of Trend 11

(1): Analysis of Papers Linking the Specialties Together in the Trend

. The occurrence of base literature papers in specialties in the trend is shown in Table 5.5. Due to the large number of papers in Trend 11, only those papers that occurred more than once are listed for examination. This was done to identify the papers that link the specialties together. Table 5.5 shows the specialties that the papers occur in, the papers being denoted by first named author only and the specialties by their identifying numbers only (these numbers being the ones allocated to them in the Protein Determination models). Table 5.6 gives the reference details of the papers in Table 5.5.

The first thing that was observed from Table 5.5 was that no paper appeared in the base literature of more than one specialty in the same time period, so the links between specialties are only longitudinal and not vertical. There were a total of 51 papers that occurred in more than 1 time period. Out of these only 9 occurred in all 3 time periods, i.e. only 17.6%. The identifying numbers of these in Table 5.5 were; 5, 6, 16, 18, 20, 27, 34, 47 and 48. There was one interesting instance of a paper appearing in two time periods that were not consecutive. This was paper number 40 in Table 5.5 by Silverman.M. that appeared in the 1978 and 1982 models but not in 1980. Thus in itself it does not link specialties together in the trend since it does not appear in the intervening year. So the specialties that it appears in were linked into the trend by other papers. It is interesting that a paper can completely disappear, (if it had appeared anywhere in the 1980 Protein Determination Models base literature it would have been picked up by the algorithm and the

specialty associated with it included in Trend 11), and re-appear 4 years later. Perhaps it became temporarily redundant and then proved useful again to later research. This finding highlights the suggestion by CRP, 1983, (19) that the trend modelling techniques could be improved by searching for papers re-appearing in other than consecutive time periods.

Since only 1 paper carried Trend 44 through all 3 time periods, Trend 11 with its 9 continuous papers seems a much more firmly linked trend then Trend 44. This is also reflected in its larger size, more specialties having been included in its structure.

(2): Occurrence of Authors in Specialties in Trend 11

The names of all the authors of the base literature papers in Trend 11 were obtained from the papers themselves. The authors names, the specialties in which they appeared, if they appeared as first named and / or co-authors, and the number of their appearances are laid out in Table 5.7.

The total number of authors in the trend was 218. Out of this 114 were the first named authors identified by the model builders and 104 were those people identified only as co-authors. The number of first named trend authors who also appeared as co-authors in their specialties was 33. Where this occurs the authors have a higher influence in the base literature of a specialty than would otherwise be suspected from the first named author listings. The number of first named trend authors appearing as co-authors in specialties other than those in which they are listed, was 31, which show previously unsuspected links between specialties in the trend. For those people only occurring as co-authors 33 were found to occur in more than one specialty in the trend (which also shows previously unsuspected links), and 18 appeared more than once in any particular specialty.

The links between specialties in the trend are shown in Fig.5.1. The data in Table 5.7 was used to see whether any links could be found between specialties by the sharing of base literature authors. Since the specialties are linked together by

re-occurrence of certain papers, the authors associated with those papers were deleted from the analysis, so that only extra author links were noted. This was done to see whether any links could be found between specialties over and above those found by the trend modelling methodology. This was done to check the validity of these links as other connections should be found if the linking of specialties together in this way in any way reflects an ongoing research field.

Firstly looking at the specialties that were directly linked in the trend diagram (Fig. 5.1) it was discovered that there were a total of 34 shared author links over and above the shared paper links. When the papers were examined to see whether any extra author links could be found between specialties not directly linked in the trend, a total of 58 were discovered. Thus there is a grand total of 92 extra author links between the specialties in Trend 11, which would appear to support the links found by the trend modellers between the specialties.

(3): Analysis of the Trend Summary Data.

In a manner similar to the examination of Trend 44, the data given to analyse the properties of the component specialties of Trend 11 in the Trend Summary Report in the Protein Determination Models was analysed to see whether it could shed light on the behaviour of specialties in the trend. The links between specialties in the trend are shown in Fig.5.1. The trend summary data used to analyse them is shown in Table 5.8. The data that was used for analysis in this section was the size, the average age of all the papers and the age of the youngest paper in the base literature, together with the size of the current literature.

Using this information firstly for those specialties in 1978, it can be seen from the point of view of the sizes of the base and current literature that there are 2 very large specialties, 105 and 202. Four of the specialties had youngest paper ages of 1 year, and the average ages extended from 1 to 4 years. Only 1 specialty, 37, had an average age of 4 years and a

youngest paper age of 4 years, but since even this is not very high it can be concluded that the 5 specialties in 1978 were based on fairly new information.

Then elements of 4 of the specialties in 1978 (105, 37, 202 and 539) combined to form the large specialty 48 in 1980. From the data this specialty appears to be highly active since it has a high base literature size and current activity level and a low base literature age.

Some elements of Specialty 105 fed into Specialty 564 in 1980, a specialty that does not continue into 1982, and so appears to have stagnated and fallen away from the mainstream of research. This stagnation appears to be supported by the data, as Specialty 564 has low base literature and current literature sizes and a high average age and youngest paper age for its base literature. This is in marked contrast to its active parent specialty, 105.

So does this mean that the more active components of Specialty 105 went into 48 and not into 564? In 1980, Specialty 16 is still independent of the lines of work converging on 48 since it draws only from Specialty 19 in 1978 (which also contributes nothing to any other specialty). Specialty 16 is more active with respect to the number of base literature and current literature papers than its parent specialty, 19, although their average ages and youngest paper in the base literature are nearly the same. So it seems to be making progress.

Then in 1982 Specialty 48 seems to spawn on its own, 8 smaller but still active specialties, with low youngest paper ages and small base literatures, but considerably larger current literatures. This invites speculation on whether Specialty 48 which appears to be highly fertile has sub-divided itself into more distinct problem areas, with perhaps groups splitting off to concentrate on problems. One element of Specialty 48 combines with Specialty 16 to produce Specialty 185 in 1982, a very active specialty with a high number of base literature and current literature papers and a low age for its youngest paper. Thus it appears that 2 fruitful specialties have come together to produce another. Therefore as for Trend 44 it would appear that the summary data can shed light on the behaviour of the specialties.

Section 5: General Conclusions for the Analysis of the Literature in Trend 11

The base literature papers in the 4 specialties affiliated to Aston Pharmacy Department were observed to mostly cite their co-cited partners, the major omissions occurring in Specialty 48 which had a very large base literature of 64 papers. A high level of citation of other base literature authors in the base literature papers was found in some specialties. The trend was observed to be more firmly linked than Trend 44, having more linking papers. Examination of the occurrence of base literature authors in specialties in the trend supported the linkages in the trend diagram in Fig. 5.1. The Trend Summary data was found to apparently provide useful information about the trend. The methodology developed for the analysis of Trend 44 was also shown to be useful in the analysis of another complex trend.

The observations from the analysis of Trend 44 that the cognitive relationship between co-cited papers perceived by the current literature is also perceived by the base literature papers in their acts of citing each other, was borne out also by the analysis of the literature of Trend 11. Furthermore the links in the trend were corroborated by the continuation of base literature authors' names in subsequent time periods, although doubt was also cast on the discrete nature of the individual specialties by the sharing of authors between specialties in the same time period. The indicator data supplied by the model builders in the Trend Summary Report also seemed to give useful descriptive information about the development of specialties over the three time periods. Thus overall the conclusions from the analysis of the base literatures of Trends 11 and 44 appear to corroborate the assertions of the model builders that trend modelling can illustrate and follow the development over time of fields of scientific research.

The next stage of the analysis then moved away from the close examination of the literature of the trends to an analysis involving some of the participant scientists of the base literature of one of the trends, Trend 44. The analysis involving

the scientists consisted of two parts. The first part involved a close examination of the quality and nature of the base literature papers, the quality of the institutions, the relationships between the base literature authors, and the validity of the historical structure of the trend compared to the participants experience. This part of the analysis took the form of a questionnaire and is examined in Chapter 6. The second part of the evaluation involving participant scientists examined the history of the development of the field of mobile genetic elements research and compared it to the development of the field as portrayed by the trend, and is examined in Chapter 7.

Table 5.1 - Citation Between Base Literature Papers in Specialty 37.

No.	Authors	Reference	Co-cited with	Papers cited in own specialty
1.	Ames, G.F.L.; Spudich, E.N. & Nikaido, H.	J.Bact v117 p406 1974 (13/7/73)	2	-
2.	Koplow, J. & Goldfine, H.	J.Bact v117 p527 1974 (6/9/73)	1	-

NB: The numbers given under the "co-cited with" and "papers cited in" columns refer to the identifying numbers of the papers concerned. The numbers in brackets after the reference refer to the date that the paper was received for publication.

Table 5.2 - Citation Between Base Literature Papers in Specialty 564.

No.	Authors	Reference	Co-cited with	Papers cited in own specialty
1.	Leive, L.	Annals NY Acad.Sci v235 p109 1974	2	-
2.	Nikaido, H.	Biochem.BioPhys.Acta v433 p118 1976 (21/10/75)	1	-

NB: The numbers given under the "co-cited with" and "papers cited in" columns refer to the identifying numbers of the papers concerned. The numbers in brackets after the reference refer to the date that the paper was received for publication.

Table 5.3 Citation Between Base Literature Papers in Specialty 48.

No.	Authors	Reference	Co-cited with	Papers cited in own speciality
1.	Ames, G.F.L; Spudich, E.N; Nikaido, H.	J.Bacteriology v.117 p406 1974 (18/7/73)	18, 27, 34, 37	0
2.	Bassford, P; Beckwith, J.	Nature v.277 p538 1979	4,12, 51	8, 20, 22, 26, 41, 42, 50, 54
3.	Bassford, P; Diedrich, D L; Schnaitman, C.L; Reeves, P.	J.Bacteriology v.131 p608 1977 (26/5/77)	10, 14, 19, 21, 30, 31, 38, 39, 46, 60, 62, 63	46,52
4.	Bassford, P.J; Silhavy, T.J; Beckwith, J.R.	J.Bact v.139 p19 1979 (6/3/79)	2, 50, 51	2, 12, 22, 26, 41, 42, 49, 50, 54, 58, 59
5.	Bavoil, P; Nikaido, H; von Meyenberg, K.	Mol.Gen.Genetics v158 p23 1977 (9/9/77)	6, 14, 19, 21, 31, 33, 34, 38, 39, 43, 58, 59, 62, 63	1, 20, 31, 33, 35, 44, 46, 49, 60
6.	Braun, V; Krieger-Brauer, H.J.	Biochem.Biophys.Acta v.469 p89 1977 (28/1/77)	5, 59	7, 18, 34, 43, 44, 58, 59
7.	Braun, V; Hancock, R.E.W; Hantke, K; Hartmann, A.	J.Supramol.Structure v.5 p37 1976	14	1, 27, 34
8.	Chang, C.N; Blobel, G; Model, P.	P.Natl.Acad.Sci.U.S.A. v.75 p361 1978 (9/11/77)	22, 25, 32, 64	22, 23, 49, 54, 57
9.	Chang, C.N; Model, P; Blobel, G.	P.Natl.Acad.Sci.U.S.A. v.76 p1251 1979 (30/11/78)	25	8, 12, 50, 51, 54, 57, 64
10.	Datta, D.B; Kramer, C; Henning, U.	J.Bact v.128 p834 1976 (14/6/76)	3, 14, 21, 38, 46, 48, 49, 52, 61	1, 18, 27, 43, 46, 52
11.	Datta, D.B; Arden, B; Henning, U.	J.Bact v.131 p821 1977 (4/5/77)	33, 44, 52, 61	1, 10, 18, 27, 35, 36, 44, 46, 47, 52, 58, 60, 61
12.	Emr, S.D; Schwartz, M; Silhavy, T.J.	P.Natl.Acad.Sci.U.S.A. v.75 p5802 1978 (28/8/78)	2, 28, 42, 50, 51, 54	20, 22, 23, 28, 41, 42, 49, 50, 51, 54
13.	Endermann, R; Hindennach, I; Henning, U.	FEBS Letters v.88 p71 1978 (23/1/78)	43	3, 6, 11, 18, 19, 20, 23, 31, 33, 35, 36, 43, 44, 48, 49, 46, 58
14.	Foulds, J; Chai, T.J.	J.Bact v.133 p1478 1978 (26/10/77)	3, 5, 7, 10, 19, 21, 30, 38, 39, 62, 63	3, 11, 19, 30, 31, 33, 34, 44, 46, 49, 58
15.	Halegoua, S; Inouye, M.	J.Mol.Biology v.130 p39 1979 (18/9/78)	49	24, 41, 49, 53, 55, 54
16.	Hancock, R.E.W; Nikaido, H.	J.Bact v.136 p381 1978 (26/4/78)	35	17, 33, 34, 35, 36
17.	Hasegawa, Y; Yamada, H; Mizushima, S.	J.Biochemistry v.80 p1401 1976 (17/5/76)	18, 21, 44	44, 36, 60
18.	Henning, U; Hohn, B; Sonntag, I.	Eur.J.Biochem v.39 p27 1973 (16/6/73)	1, 17, 37, 48, 56	0
19.	Henning, U; Schmidmayr, W; Hindennach, I.	Mol.Gen.Genetics v.154 p293 1977 (4/6/77)	3, 5, 14, 21, 30, 31, 38, 39, 46, 49, 54	18, 44, 46, 47, 52, 60

Table 5.3, continued.

No.	Authors	Reference	Co-cited with	Papers cited in own speciality
20.	Hofnung, M.	Genetics v.76 p169 1974 (20/10/73)	58	43
21.	Ichihara, S; Mizushima, S.	J.Biochem.Tokyo v.83 p1095 1978 (18/11/77)	3, 5, 10, 14, 17, 19, 38, 39, 46, 62	3, 17, 19, 33, 34, 36, 44, 46, 47
22.	Inouye, H; Beckwith, J.	P.Natl.Acad.Sci.U.S.A. v.74 p1440 1977 (14/1/77)	8, 23, 42, 54, 55	0
23.	Inouye, S; Wang, S; Sekizawa, J; Halegoua, S; Inouye, M.	P.Natl.Acad.Sci.U.S.A. v.74 p1004 1977 (6/1/77)	22,49	0
24.	Ito, K; Sato, T; Yura, T.	Cell v.11 p551 1977 (14/4/77)	42	17, 30, 33, 44, 46, 50, 59
25.	Ito, K; Mandel, G; Wickner, W.	P.Natl.Acad.Sci.U.S.A. v.76 p1199 1979 (26/12/78)	8, 9, 32, 57, 64	8, 23, 24, 32, 41, 42, 49, 54, 57, 64
26.	Kellermann, O; Szmelcman, S.	Eur.J.Biochem v.47 p139 1974 (Feb/74)	43, 58, 59	20
27.	Koplow, J; Goldfine, H.	J.Bact v.117 p527 1974 (6/9/73)	1	0
28.	Lin, J.J.C; Kanazawa, H; Ozals, J; Wu, H.C.	P.Natl.Acad.Sci.U.S.A. v.75 p4891 1978 (9/8/78)	12	23
29.	Lugtenberg, B; Bronstein, H; van Selm, N; Peters, R.	Biochem.Biophys.Acta v.465 p571 1977 (28/10/76)	44, 46, 63	1, 18, 30, 35, 44, 46
30.	Lugtenberg, B; Peters, R; Bernheimer, H; Berendsen, W.	Mol.Gen.Genetics v.147 p251 1976 (29/12/75)	3, 14, 19, 39, 48, 52	1, 27, 44, 47, 52
31.	Lutkenhaus, J.	J.Bacteriology v.131 p631 1977 (13/5/77)	3, 5, 19, 33, 34, 46, 62, 63	30, 33, 34, 35, 44, 46, 52, 59
32.	Mandel, G; Wickner, W.	P.Natl.Acad.Sci.U.S.A. v.76 p236 1979 (2/11/78)	8, 25, 57	8, 23, 42, 49, 54, 57, 64
33.	Nakae, T.	Biochem.Biophys.Res.Co v.71 p877 1976 (14/6/76)	5, 11, 31, 34, 35, 63	18, 34, 35, 44, 60
34.	Nakae, T; Nikaido, H.	J.Biological.Chem v.250 p7359 1975 (13/1/75)	1, 5, 31, 33, 35, 45	1
35.	Nakae, T.	J.Biological.Chem v.251 p2176 1976 (14/10/75)	16, 33, 34, 36	1, 44
36.	Nakamura, K; Mizushima, S.	J.Biochemistry v.80 p1411 1976 (17/5/76)	35, 44, 62	17, 44, 47, 60
37.	Palva, E.T; Randall, L.L.	J.Bacteriology v.133 p279 1978 (11/7/77)	1, 18, 44, 56	17, 18, 33, 34, 44, 46, 56
38.	Pugsley, A.P; Schnaitman, C.A.	J.Bacteriology v.133 p1181 1978 (1/11/78)	3, 5, 10, 14, 19, 21, 39, 44, 53, 62, 63	3, 11, 14, 19, 31, 33, 35, 59, 62
39.	Pugsley, A.P; Schnaitman, C.A.	J.Bacteriology v.135 p1118 1978 (15/5/78)	3, 5, 14, 19, 21, 30, 38, 62, 63	3, 5, 14, 19, 31, 38, 47, 62, 63

Table 5.3, continued.

No.	Authors	Reference	Co-cited with	Papers cited in own specialty
40.	Raibaud, O; Roa, M; Braun-Breton, C; Schwartz, M.	Mol.General Genetics v.174 p241 1979 (12/1/79)	43, 58, 59	2, 6, 12, 20, 26, 42, 43, 50, 51, 58, 59
41.	Randall, L.L; Hardy, S.J.S.	Eur.J.Biochem v.75 p43 1977 (Dec/76)	42, 55	26, 34
42.	Randall, L.L; Hardy, S.J.S; Josefson, L.G.	P.Natl.Acad.Sci.U.S.A. v.75 p1209 1978 (29/12/77)	12, 22, 24, 41, 49, 51, 54	22, 23, 41, 49, 54
43.	Randall-Hazelbauer, L; Schwartz, M.	J.Bacteriology v.116 p1436 1973 (9/8/73)	5, 13, 26, 40, 58, 59	0
44.	Rosenbusch, J.	J.Biological.Chem v.249 p8019 1974 (10/4/74)	11, 17, 29, 36, 37, 38, 56	18
45.	Schindler, H; Rosenbusch, J.P.	P.Natl.Acad.Sci.U.S.A. v.75 p3751 1978 (22/5/78)	34	5, 33, 44, 56
46.	Schaitges, C; Henning, U.	Eur.J.Biochem v.63 p47 1976 (30/9/75)	3, 10, 19, 21, 29, 31, 60, 62, 63	18, 44, 47
47.	Schnaitman, C.A.	J.Bacteriology v.118 p442 1974 (12/12/73)	60	0
48.	Schweizer, M; Hindennach, I; Garten, W; Henning, U.	Eur.J.Biochem v.82 p211 1978 (July/77)	10, 18, 30, 49, 52, 61	1, 10, 11, 18, 27, 35, 44, 47, 52, 60
49.	Sekizawa, J; Inouye, S; Inouye, M.	Biochem.Biophys.Res.Co v.77 p1126 1977 (5/7/77)	10, 15, 23, 42, 48, 61	22, 23, 44
50.	Silhavy, T.J; Casadaban, M.J; Shuman, H.A; Beckwith, J.R.	P.Natl.Acad.Sci.U.S.A. v.73 p3423 1976 (15/7/76)	4, 12, 51	20, 26, 43, 58, 59
51.	Silhavy, T.J; Shuman, H.A; Beckwith, J.R; Schwartz, M.	P.Natl.Acad.Sci.U.S.A. v.74 p5411 1977 (22/9/77)	2, 4, 12, 42, 50	20, 22, 23, 43, 50, 58, 59
52.	Skurray, R.A; Hancock, R.E.W; Reeves, P.	J.Bacteriology v.119 p726 1974 (4/3/74)	10, 11, 30, 48, 61	1, 27, 47
53	Smit, J; Nikaido, H.	J.Bacteriology v.135 p687 1978 (12/1/78)	38	1, 33, 35, 36, 44
54.	Smith, W.P; Tai, P.C; Thompson, R.C; Davis, B.D.	P.Natl.Acad.Sci.U.S.A. v.74 p2830 1977 (9/5/77)	12, 22, 42, 55	22, 23, 41
55.	Smith, W.P; Tai, P.C; Davies, B.D.	P.Natl.Acad.Sci.U.S.A. v.75 p814 1978 (8/12/77)	22, 41, 54	22, 33, 54
56.	Stevens, A.C; ten Heggeler, B; Muller, R; Kistler, J; Rosenbusch, J.P.	J.Cell Biology v.72 p292 1977 (1/7/76)	18, 37, 44	35, 44
57.	Sugimoto, K; Sugisaki, H; Okamato, T; Takanami, M.	J.Mol.Biology v.111 p487 1977 (24/8/76)	25, 32	0
58.	Szmelcman, S; Schwartz, M; Silhavy, T.J; Boos, W.	Eur.J.Biochem v.65 p13 1976 (Nov/75)	5, 20, 26, 40, 43, 59	20, 26, 43, 59
59.	Szmelcman, S; Hofnung, M.	J.Bacteriology v.124 p112 1975 (25/3/75)	5, 6, 26, 40, 43, 58	26, 43
60.	Uemura, J; Mizushima, S.	Biochem.Biophys.Acta v.413 p163 1975	3, 46, 47	44, 47
61.	Van Alphen, L; Havekes, L; Lugtenberg, B.	FEBS Letters v.75 p285 1977 (4/1/77)	10, 11, 48, 49, 52	30, 47, 52
62.	Van Alphen, W; Lugtenberg, B.	J.Bacteriology v.131 p623 1977 (28/3/77)	3, 5, 14, 19, 21, 31, 36, 38, 39, 46, 63	1, 17, 27, 29, 30, 33, 35, 44, 46, 47, 61
63.	Van Alphen, W; Van Selm, N; Lugtenberg, B.	Mol.General Genetics v.159 p75 1978 (10/11/77)	3, 5, 14, 19, 29, 31, 33, 38, 39, 46, 62	7, 19, 29, 30, 31, 33, 35, 43, 44, 46, 52, 59, 62
64.	Wickner, W; Mandel, G; Zwizinski, C; Bates, M; Killick, T.	P.Natl.Acad.Sci.U.S.A. v.75 p1754 1978 (6/2/78)	8, 25	8, 57

NB: The numbers given under the "co-cited with" and "papers cited in" columns refer to the identifying numbers of the papers concerned. The numbers in brackets after the reference refer to the date that the paper was received for publication.

Table 5.4 - Citation Between Base Literature Papers in Specialty 183.

No.	Authors	Reference	Co-cited with	Papers cited in own specialty
1.	Benz, R. & Hancock, R.E.W.	Biochem.BioPhys.Acta v646 p298 1981 (24/2/81)	2	2,3
2.	Hancock, R.E.W. Decad, G.M. Nikaido, H.	Biochem.Biophys.Acta v554 p323 1979 (9/1/79)	1,4,5	3
3.	Hancock, R.E.W. Nikaido, H.	J.Bacteriology v136 p381 1978 (26/4/78)	4	-
4.	Hancock, R.E.W. Carey, A.M.	J.Bacteriology v140 p902 1979 (13/9/79)	2,3,5	2,3
5.	Nicas, T.I. Hancock, R.E.W.	J.Bacteriology v143 p872 1980	2,4	3,4

NB: The numbers given under the "co-cited with" and "papers cited in" columns refer to the identifying numbers of the papers concerned. The numbers in brackets after the reference refer to the date that the paper was received for publication.

Table 5.5 Papers Occurring More than Once in the Base Literature of Trend 11.

	Year	1978					1980			1982								
	Paper	105	37	202	539	19	564	48	16	153	183	263	355	451	680	681	685	185
1.	Adler, J.					●		●										
2.	Ames, G.F.		●					●										
3.	Bassford, P.							●		●								
3.	Bassford, P.	●						●										
5.	Berg, H.C.					●			●									●
6.	Braun, V.			●				●				●						
7.	Chang, C.N.							●					●					
8.	Chang, C.N.							●					●					
9.	Datta, D.B.	●						●										
10.	Goy, M.F.								●									●
11.	Halegoua, S.							●							●			
12.	Hancock, R.							●								●		
13.	Hasegawa, Y.	●						●										
14.	Henning, U.	●						●										
15.	Henning, U.	●						●										
16.	Hofnung, M.			●				●										●
17.	Inouye, H.				●			●										
18.	Inouye, S.				●			●									●	
19.	Ito, K.							●						●				
20	Kellermann, O.			●				●										●
21.	Koplow, J.		●					●										
22.	Kashland, D.					●			●									
23.	Kort, E.N.								●									●
24.	Lugtenberg, B.	●						●										
25.	Lugtenberg, B.	●						●										
26.	Lutkenhaus, J.	●						●										
27.	MacNab, R.M.					●			●									●
28	Nakae, T.	●						●										
29.	Nakae, T.	●						●										
30.	Nakamura, K.	●						●										
31.	Nikaido, H.	●					●											
32	Raibaud, O.							●										●
33.	Randall, L.				●			●										
34.	Randall-Hazelbauer, L.			●				●										●
35.	Rosenbusch, J.	●						●										
36.	Schmitges, C.	●						●										
37.	Schnaitman, C.	●						●										
38.	Sekizawa, J.				●			●										
39.	Silhavy, T.J.							●		●								
40.	Silverman, M.					●												●
41.	Silverman, M.								●									●
42.	Skurray, R.A.	●						●										
43.	Smith, W.P.				●			●										
44.	Springer, M.S.								●									●
45.	Springer, W.R.								●									●
46.	Stock, J.B.								●									●
47.	Szmelcman, S.			●				●										●
48.	Szmelcman, S.			●				●										●
49.	Uemura, J.	●						●										
50.	Van Alphen, L.	●						●										
51.	Van Alphen, W.							●								●		

NB. The numbers listed under the years refer to the specialties that the papers occur in.

Table 5.6 Reference Details of Papers in Table 5.5

No.	Author (first named only)	Reference
1.	Adler, J.	Science v.166 p1588 1969
2.	Ames, G.F.	J.Bacteriology v.117 p406 1974
3.	Bassford, P.	Nature v.277 p538 1979
4.	Bassford, P.	J.Bacteriology v.131 p608 1977
5.	Berg, H.C.	Nature v.239 p500 1972
6.	Braun, V.	J.Supramol.Struc v.5 p37 1976
7.	Chang, C.N.	P.Natl.Acad.Sci.U.S.A. v.75 p381 1978
8.	Chang, C.N.	P.Natl.Acad.Sci.U.S.A. v.76 p1251 1979
9.	Datta, D.B.	J.Bacteriology v.131 p821 1977
10.	Coy, M.F.	P.Natl.Acad.Sci.U.S.A. v.74 p4964 1977
11.	Halegoua, S.	J.Mol.Biology v.130 p39 1979
12.	Hancock, R.E.W.	J.Bacteriology v.136 p381 1978
13.	Hasegawa, Y.	J.Biochemistry v.80 p1401 1976
14.	Henning, U.	Eur.J.Biochem v.39 p27 1973
15.	Henning, U.	Mol.Gen.Genetics v.154 p293 1977
16.	Hofnung, M.	Genetics v.76 p169 1974
17.	Inouye, H.	P.Natl.Acad.Sci.U.S.A. v.74 p1440 1977
18.	Inouye, S.	P.Natl.Acad.Sci.U.S.A. v.74 p1004 1977
19.	Ito, K.	P.Natl.Acad.Sci.U.S.A. v.76 p1199 1979
20.	Kellermann, O.	Eur.J.Biochem v.47 p139 1974
21.	Koplow, J.	J.Bacteriology v.117 p527 1974
22.	Kashland, D.	Science v.196 p1055 1977
23.	Kort, E.N.	P.Natl.Acad.Sci.U.S.A. v.72 p3939 1975
24.	Lugtenberg, B.	Mol.Gen.Genetics v.147 p251 1976
25.	Lugtenberg, B.	Biochem.Biophys.Acta v.465 p571 1977
26.	Lutkenhaus, J.	J.Bacteriology v.131 p631 1977
27.	MacNab, R.M.	P.Natl.Acad.Sci.U.S.A. v.69 p2509 1972
28.	Nakae, T.	J.Biological.Chem v.251 p2176 1976
29.	Nakae, T.	Biochem.Biophys.Res.Co v.71 p877 1976
30.	Nakamura, K.	J.Biochemistry v.80 p1411 1976
31.	Nikaido, H.	Biochem.Biophys.Acta v.483 p118 1976
32.	Raibaud, O.	Mol.Gen.Genetics v.174 p241 1979
33.	Randall, L.L.	Eur.J.Biochem v.75 p43 1977
34.	Randall-Hazelbauer, L.	J.Bacteriology v.116 p1436 1973
35.	Rosenbusch, J.	J.Biological.Chem v.249 p8019 1974
36.	Schmitges, C.	Eur.J.Biochem v.63 p47 1976
37.	Schnaitman, C.A.	J.Bacteriology v.118 p442 1974
38.	Sekizawa, J.	Biochem.Biophys.Res.Co v.77 p1126 1977
39.	Silhavy, T.J.	P.Natl.Acad.Sci.U.S.A. v.74 p5411 1977
40.	Silverman, M.	Nature v.249 p73 1974
41.	Silverman, M.	P.Natl.Acad.Sci.U.S.A. v.74 p3317 1977
42.	Skurray, R.A.	J.Bacteriology v.119 p726 1974
43.	Smith, W.P.	P.Natl.Acad.Sci.U.S.A. v.74 p2830 1977
44.	Springer, M.S.	P.Natl.Acad.Sci.U.S.A. v.74 p3312 1977
45.	Springer, W.R.	P.Natl.Acad.Sci.U.S.A. v.74 p533 1977
46.	Stock, J.B.	P.Natl.Acad.Sci.U.S.A. v.75 3659 1978
47.	Szmelcman, S.	J.Bacteriology v.124 p112 1975
48.	Szmelcman, S.	Eur.J.Biochem v.65 p13 1976
49.	Uemura, J.	Biochem.Biophys.Acta v.413 p163 1975
50.	Van Alpen, L.	Febs Letters v.75 p285 1977
51.	Van Alpen, W.	J.Bacteriology v.131 p623 1977

Table 5.7 Occurrence of Base Literature Authors in Specialties in Trend 11.

Author	Occurrence as First Named Author	Occurrence as co-Author
Adler, J.	19(2), 16(2)	19(1), 16(7), 185(8)
Ames, G.F.	37(1), 48(1)	
Arden, B.		105(1), 48(1)
Aswanikumar, S.		16(1)
Axebod, J.		16(2)
Ball, C.B.		16(1)
Bassford, P.J.	105(1), 48(3), 153(1)	153(2), 185(1)
Bates, M.		48(1)
Baty, D.		680(1)
Bavoil, P.	48(1)	
Beckwith, John.		539(1), 48(4), 153(5), 681(1), 185(2)
Beckwith, Jonathan.		48(1)
Bedouelle, H.	153(1)	
Benz, R.	183(1)	
Berenden, W.		105(1), 48(1)
Berg, H.C.	19(2), 16(2), 185(1)	16(1)
Bernheimer, H.		105(1), 48(1)
Blobel, G.		48(2), 355(2)
Boos, W.		202(1), 48(1), 185(1)
Boyd, A.	185(1)	
Bradbeer, C.	202(1)	202(1)
Bragg, P.D.	105(1)	
Braun, V.	105(1), 202(2), 48(2), 263(2)	202(2), 685(1)
Braun-Breton, C.		48(1), 185(1)
Brickman, E.		185(1)
Bronstein, H.		105(1), 48(1)
Brown, D.A.		19(1), 16(1), 185(1)
Carey, A.M.		183(1)
Casadaban, M.J.		48(1), 185(1)
Chai, T.J.	105(2)	48(1)
Chang, C.N.	48(2), 335(2)	
Chelsky, D.	185(1)	
Clement, J.M.		153(1)
Corcoran, B.A.		16(1)
Dahlquist, F.W.		185(1)
Date, T.	451(1)	
Datta, D.B.	105(1), 48(2)	
Davies, J.K.	105(1)	
Davis, B.D.		539(1), 48(2)
Debarbouille, M.	185(1)	
Decad, G.M.		183(1)
Defranco, A.L.	185(1)	
de Graaff, P.J.		105(1)
Diedrich, D.L.		105(1), 48(1)
Diliberto, E.J.	16(1)	
Dimasi, D.R.	202(1)	
Earhart, C.F.		263(1)
Ear, S.D.	48(1), 153(3)	
Enderman, R.	48(1)	
Enequist, H.G.	451(1)	

Table 5.7, continued.

Author	Occurrence as First Named Author	Occurrence as co-Author
Engstrom, P.	185(1)	
Ferengi, T.	185(1)	185(1)
Forn de Salsas, M.		105(1)
Foulds, J.	48(1)	105(2)
Fowler, A.V.		153(2)
Gargus, J.J.		16(1)
Garten, W.	105(1)	48(1)
Goldfine, H.		37(1), 48(1)
Goodman, J.M.		451(1)
Goy, M.F.	16(1), 185(2)	16(2), 185(3)
Greenwood, J.K.T.		263(1)
Guarente, L.		185(1)
Halegoua, S.	48(1), 680(1)	539(1), 48(1), 685(1)
Hall, M.N.	681(3)	153(1)
Haller, I.		105(1)
Hancock, R.E.W.	202(1), 48(1), 183(3)	105(1), 202(1), 48(2), 183(2), 263(1)
Hanley-Way, S.		153(1)
Hantke, K.	202(2), 685(1)	202(2), 48(1), 263(1)
Harayama, S.		451(1)
Hardy, F.M.		16(1)
Hardy, S.J.S.		539(1), 48(2), 451(1)
Hartmann, A.		202(1), 48(1), 263(1)
Hasegawa, Y.	105(1) 48(1)	
Havekes, L.		105(1), 48(1)
Hayashi, H.		185(1)
Hazelbauer, G.L.	202(1), 185(2)	185(1)
Hedblom, M.L.	185(1)	
Hedgpeth, J.		153(1)
Heggeler, B.T.		48(1)
Henning, U.	105(3), 48(2)	105(4), 48(5)
Hindennach, I.	105(1)	105(2), 48(3)
Hirst, T.R.		451(1)
Hofnung, M.	202(1), 48(1), 185(1)	202(1), 48(1), 153(2), 185(1)
Hogg, R.W.		16(1)
Hohn, B.		105(1), 48(1)
Hsu, C.		105(1)
Hussain, M.	685(1)	685(1)
Ichihara, S.	48(1), 685(1)	685(1)
Inouye, H.	539(1), 48(1)	
Inouye, M.		539(2), 48(3), 680(1), 685(1)
Inouye, S.	539(1), 48(1), 685(1)	539(1), 48(1)
Ito, K.	48(2), 451(1)	
Joseffson, L.G.		48(1)
Kanazawa, H.		48(1)
Kellermann, O.	202(1), 48(1), 185(1)	
Khalifah, L.I.		202(1)
Kihara, M.	185(1)	
Killick, T.		48(1)
Kistler, J.		48(1)
Kleene, S.J.	16(1)	

Table 5.7, continued.

Author	Occurrence as First Named Author	Occurrence as co-Author
Koiwai, O.	185(1)	
Kondoh, H.	16(1)	
Koplow, J.	37(1), 48(1)	
Kort, E.N.	16(1), 185(1)	19(1)
Koshland, D.E.	19(1), 16(1)	19(1), 16(6), 185(5)
Kramer, C.		48(1)
Krieger-Brauer, H.		48(1)
Larsen, S.H.	19(1), 16(1)	16(1), 185(1)
Lazjunski, C.		680(1)
Leive, L.		564(1)
Lin, J.J.C.	48(1)	
Luckey, M.	185(1)	
Lugtenberg, B.	105(1), 48(1)	105(1), 48(3), 681(1)
Lugtenberg, E.J.		105(1)
Luke, R.K.J.		263(1)
Lutkenhaus, J.F.	105(1), 48(1)	
Mandel, G.	48(1), 451(1)	48(2)
Manson, M.D.	16(1)	
Meijers, J.		105(1)
Meyenberg, K.V.		48(1)
Miller, J.B.	16(1)	
Mizushima, S.	105(1)	105(4), 48(4), 685(2)
Model, P.		48(2), 355(2)
Moreno, F.	153(1)	
Muller, R.		48(1)
MacNab, R.M.	19(1), 16(1), 185(1)	16(1), 185(1)
McIntosh, M.	263(1)	
Nakae, T.	105(2), 48(3)	
Nakamura, K.	105(1), 48(1), 685(1)	
Neilands, J.B.		202(1)
Nicas, T.I.	183(1)	
Nikaido, H.	105(2), 564(1)	37(1), 48(5), 183(2), 185(1)
Nurminen, M.		105(1)
Odea, R.F.	16(1)	
Okamoto, K.		48(1)
Oliver, D.B.	153(1)	
Ozols, J.		48(1)
Pages, J.M.		680(1)
Palva, E.T.	48(1)	
Parkinson, J.S.	16(1)	
Peters, R.		105(3), 48(2)
Pugsley, A.P.	202(3), 48(2)	
Raibaud, O.	48(1), 185(1)	
Randall, L.L.	539(1), 48(2)	48(1), 451(1)
Randall-Hazelbauer, L.	202(1), 48(1)	
Reader, R.W.		19(1)
Reeves, P.		105(3), 202(3), 48(2)
Repaske, D.R.	185(1)	
Roa, M.		48(1), 185(1)
Rosenberg, H.	263(1)	

Table 5.7, continued.

Author	Occurrence as First Named Author	Occurrence as co-Author
Rosenbusch, J.P.	105(1), 48(1)	48(2)
Sabet, S.F.	202(1)	
Sarthy, A.		681(1)
Sato, T.		48(1)
Schaller, K.		202(1)
Schiffmann, E.		16(1)
Schindler, H.	48(1)	
Schidmayr, W.		105(1), 48(1)
Schmitges, C.J.	105(1), 48(1)	
Schnaitman, C.A.	105(2), 48(1)	105(1), 202(2), 48(3)
Schwartz, M.		202(2), 48(5), 153(2), 185(6)
Schwartz, V.		185(1)
Schweizer, M.	48(1)	
Sekizawa, J.	539(1), 48(1)	539(1), 48(1), 685(1)
Shaltiel, L.		105(1)
Shuman, H.A.	185(2)	48(2), 153(1), 185(2)
Silhavy, T.J.	48(2), 153(2), 185(1)	202(1), 48(3), 153(4), 681(3), 185(3)
Silverman, M.	19(1), 16(2), 185(2)	
Simon, M.I.		19(1), 16(2), 185(3)
Skurray, R.A.	105(1), 48(1)	
Smit, J.	48(1)	
Smith, D.		105(1)
Smith, W.P.	539(1), 48(2)	
Song, S.A.		105(1)
Sonntag, I.		105(1), 48(1)
Springer, M.	16(1), 185(2)	16(1), 185(1)
Springer, W.R.	16(1), 185(1)	
Spudich, E.N.		37(1), 48(1)
Steven, A.C.	48(1)	
Stock, J.B.	16(1), 185(1)	
Stuart, S.J.	263(1)	
Sugimoto, K.	48(1)	
Sugisaki, H.		48(1)
Szmelcman, S.	202(2), 48(2), 185(2)	202(1), 48(1), 185(1)
Tai, P.C.		539(1), 48(1)
Takahami, M.		48(1)
Tedesco, P.		16(1)
Thompson, R.C.		539(1), 48(1)
Toews, M.L.		16(1)
Tsang, N.	16(1)	
Tso, W.W.	16(1)	19(1)
Uemura, J.	105(1), 48(1)	
Van Alphen, L.	105(1), 48(1)	105(1)
Van Alphen, W.	48(2), 681(1)	
Van der Drift, C.		16(1)
Van der Hoek, P.		105(1)
Van der Werf, P.	16(1)	
Van Selm, N.		105(1), 48(2)
Verhoef, C.	105(1)	
Viveros, O.		16(2)
Wandersman, C.	185(1)	

Table 5.7, continued.

Author	Occurrence as First Named Author	Occurrence as co-Author
Wang, F.A.	185(1)	
Wang, S.		539(1), 48(1), 685(1)
Wanner, B.L.	681(1)	
Warner, P.J.		263(1)
Wayne, R.	202(1)	
White, J.C.		202(1)
Wickner, W.	48(1)	48(2), 451(2)
Williams, P.H.	263(2)	
Wolff, H.		202(1)
Woodrow, M.L.		202(1)
Wu, H.C.		48(1)
Yamaja, H.		105(2), 48(1)
Young, I.G.		263(1)
Yu, F.	105(1)	
Yura, T.		48(1)
Zabin, I.		153(2)
Zwizinski, C.		48(1)

Key

The first number in the occurrence columns refers to the speciality number. The number in brackets (after it) refers to the number of times the author appeared as a first named author or co-author respectively.

Table 5.8 Trend Summary Data for Trend 11.

Specialty No. & Year	Size of B.L. (No. of papers)	Average age of B.L. (years)	Youngest papers in B.L. (years)	Size of C.L. (No. of papers)
19 (1978)	8	4.7	1	62
37 (1978)	2	4.0	4	38
105 (1978)	32	2.4	1	168
202 (1978)	18	3.0	1	93
539 (1978)	5	1.0	1	53
16 (1980)	24	4.2	1	136
48 (1980)	64	3.2	1	325
564 (1980)	2	5.0	4	33
153 (1982)	9	2.3	1	63
183 (1982)	5	2.6	1	35
185 (1982)	35	4.5	1	171
263 (1982)	6	4.3	2	38
355 (1982)	2	3.5	3	28
451 (1982)	3	2.0	1	27
680 (1982)	2	3.0	3	24
681 (1982)	5	2.6	1	30
685 (1982)	5	4.0	1	46

NB: B.L. means base literature.

Chapter 6: The validation of Trend 44 by Scientists in the Base Literature:

Part I: Data Obtained from Questionnaires

Section 1: Aims, Methodology and Initial Results of the Questionnaire Analysis

The analysis of Trends 11 and 44 in Chapters 4 and 5 showed that on the whole trend modelling seemed capable of grouping related papers over time and so able to follow the development of a field of research. Evidence was also found for the existence of an invisible college of practising scientists at base literature level, and for the validity of grouping base literature papers into co-cited pairs. Therefore some of the basic premises of CCA analysis concerning the validity of the way that it organizes the scientific literature, and its ability to follow the work of invisible colleges and the dynamics of science (see Chapters 2 and 3), seem to have been substantiated.

Thus the purpose of Chapters 6 and 7 is to extend the validation of Trend 44 in order to further test these findings. This was carried out by consulting scientific personnel, with some knowledge of the field of mobile genetic elements, through questionnaires and interviews. This was essential because the methodology used to analyse and validate trends in Chapters 4 and 5 is only capable of evaluating trends as they stand. It is not capable of locating errors of omission in a trends coverage of the development of a scientific field. Such errors are just as important when considering the utility of trend modelling as a tool for policy making as errors of commission.

That such errors of omission may have occurred in Trend 44 was demonstrated in Chapter 4, where the analysis of the literature of the trend indicated possible methodological problems associated with the criteria of continuity currently used to generate trends. Presumably these problems might result in a failure to adequately represent the scientific fields concerned. It was considered that the only way to discover if errors had

occurred, and if so to what extent, was to approach scientists with a working knowledge of the field. A reasonable starting point for this was considered to be the list of base literature authors supplied in Trend 44. A list of their names can be seen in Table 1 from the accompanying booklet to the questionnaire (see Appendix D.3). This list provided a manageable number for the analysis and also provided the opportunity to continue the in-depth study of the base literature of Trend 44 that was begun in Chapter 4. This extended analysis was only carried out for Trend 44 since the size of Trend 11, and the time available for the analysis did not permit it to be examined in any detail.

It was decided to use a questionnaire approach in order to gain the data needed to test these previous findings and locate any errors of omission. The questionnaire was followed up where possible by in-depth interviews with the scientists (see Chapter 7). It was felt that alternative methods of canvassing the scientists opinions, for example the exclusive use of telephone interviews, apart from the expense of the many international calls that would have been required would also not have produced sufficient reliable data for such an analysis. This was because the highly complex nature of the data obtained in Chapter 4 would have meant that the questions concerning it would have been confusing to the scientists if only interviews had been used. Thus the questionnaire aimed to present some of the data obtained from the analysis of the base literature papers of Trend 44 in a coherent format to the scientists who had produced the papers for their validation and comment. The data presented included the relationships shown between specialties in Trend 44 by the model builders, the details of the papers in the base literatures of the specialties and lists of the authors and their associated institutions. The data and associated questions presented in the questionnaire were designed to test the view of the development of the field of transposable elements presented by the modellers in the Specialty Summary Report and Trend Diagram, (combined in Fig.4.1); to gather further evidence for the existence of an invisible college, and to test the assumption in CCA that papers

linked by co-citation are the important documents providing the intellectual foundations for a specialty.

In order to achieve these aims the questionnaire was designed in 5 sections (see Appendix C), each section examining a different aspect of Trend 44. The information concerning the trend that the scientists needed in order to answer the questions was supplied in a separate booklet, which can be seen in Appendix D. In Section A of the questionnaire the scientists were asked to comment on the validity of the relationships shown between specialties over time by the model builders. A copy of the trend diagram supplied for their perusal is in Appendix D.1. In Section B they were asked to examine the papers in the base literature of each specialty (see Appendix D.2) and answer questions relating to their quality and importance. Section C was designed to gain information relating to the structure of any invisible college that might be present, by requesting the respondent to show their relationships to other scientists in the base literature on the table supplied for the purpose in Appendix D.3. Section D provided a list of the institutions associated with the base literature (see Appendix D.4), and was designed to gain information relating to their perceived importance and to uncover institutional links between the scientists that might provide further information relating to an invisible college. Section E simply aimed to explore the possibility that errors of omission in the trend had resulted from the omission of relevant journals in the core journal sets used to construct the models, by asking which journals the scientists normally published in. A covering letter requesting the scientists co-operation was also enclosed, together with a letter explaining some of the ideas behind scientometrics, see Appendices A and B. Thus the questionnaire was different from previous ones employed by e.g. Small, 1977, (81), in that it was quite open about the source of the data and did not ask general questions about the nature and evolution of the field (which were saved for subsequent interviews). Many of the other questionnaires sent out by other researchers to test citation data have been essentially covert in nature.

The resulting questionnaire was piloted on helpful scientists in the Pharmacy Department at the University of Aston. The consensus of opinion was that it was overlong and that it was probable that most scientists would not have the time to fill it in. However it was decided to take a gamble and send the questionnaire out in its original format on the basis that if at least some of the scientists could be persuaded to do it, some interesting data would become available. This was considered to be a worthwhile risk because the base literature scientists would be more familiar with the subject matter and possibly would be more motivated to fill it in. In the initial mailshot, because some of the base literature papers were 10 or more years old, it was considered possible that some of the scientists involved would have moved to other institutions. In an attempt to overcome this problem questionnaires were initially sent to the institutions connected with the scientists names in the base literature, with a covering note asking the departmental secretaries to forward questionnaires as appropriate. Replies were encouraged by enclosing addressed envelopes with pre-paid postage in the form of international reply coupons.

A few completed questionnaires were received from this mail shot. However after a couple of months, which was deemed to be adequate time for them to have reached their various, mostly foreign destinations, it was decided to enquire after them via the telephone. This required many international calls because the scientists had often moved on from the original institution. It was not uncommon for them to have moved several times, necessitating calls to a series of addresses until the scientist was located. As a result questionnaires had often failed to reach the scientists concerned, thus necessitating a second mail shot. Thereafter the number of completed questionnaires received greatly improved by reminding scientists at intervals over the telephone. A letter of thanks was sent to each of the respondents on receipt of the completed questionnaires.

Questionnaires were initially sent to all of the authors that were identified as being associated with the base literature papers of Trend 44 (See Chapter 4), a total of 59. Out of these

one scientist was found to have died, one to be seriously ill, 3 were technicians, one had only been a masters student in the field and could not be traced, and one who had been a graduate student but had since left science completely to become an interpreter at the United Nations. Together with one scientist who could not be traced by any of the methods used to find the others such as forwarding addresses, information from other scientists in the mail shot or addresses obtained from library sources such as the SCI, meant that the maximum number of replies that could be expected was 51.

One interesting fact that emerged from the tracing of these scientists was that 10 (20%) of them had left traditional academia for industrial concerns, one of whom had subsequently moved between 2 companies. The companies involved were, firstly in the USA; Amgen, Pioneer Hi-Bred Inc., Promega Biotechnology Co. Ltd, and Johnson and Johnson Biotechnology. In the UK there were ICI and Searle, and in Switzerland Ciba Geigy, Hoffmann La Roche and Biogen SA. The analysis in Chapter 4 showed that all the base literature papers were produced in academic institutions. So the movement of scientists out of academia into industry is interesting as it could imply increasing interest in the research area typified by Trend 44 by industrial concerns. Since the research area contains material closely allied to genetic engineering techniques, the interest of biotechnology and pharmaceutical companies is not surprising.

The high mobility of the group of 51 people was shown in that only 14 (27%) had stayed in the same institution. The other 37 (72%) had all moved. Ten of the scientists movements showed international co-operation. Out of these 2 had moved back to Japan from the USA, 4 showed links between the USA and Switzerland, 1 between the USA and France, 1 between Canada and the USA, 1 between the UK and Belgium and 1 between Switzerland and France. The deceased scientist, who was not included in the sample had also worked in Switzerland and the USA. A good proportion of the scientists, 25 (49%), were found to be presently working outside the USA, although the USA is still obviously influential in the field. Some very distinguished people were identified in the

sample, 1 Nobel prize winner, 2 fellows of the Royal Society, and 2 scientists held in great esteem by their colleagues (see Chapter 7). The most notable of the 2 scientists in this latter category was Stanley Cohen of Stanford University (the virtual founder of genetic engineering) who unfortunately, but not surprisingly, was too busy to respond.

An interesting range of responses were received. These varied from enthusiastic replies where the scientists not only completed the whole questionnaire, but carried on to write virtual essays on the subject on separate pieces of paper, enclose review papers (in one case a paper on the subject of proper nomenclature in the research area), to extreme hostility, involving in one case an abusive letter. All of the scientists displayed some professional cynicism towards the data presented. Replies to the questionnaire, and some of the off the questionnaire comments, showed that very few felt that it accurately represented the development of the field as they saw it.

The completed questionnaires were analysed section by section since not all of the respondants felt competent enough to answer all of the questions. Thus the percentage response was calculated separately for each question. The detailed aims of the questions in each section and the analysis of results obtained are examined below.

Section 2: Questionnaire Results

(1): Section A

In this section the trend diagram supplied by the model builders for Trend 44, showing the linkages between the specialties over the different time periods, was presented for perusal. The scientists were asked whether the links made sense to them when compared with their own knowledge of the field and were supplied with a "skeleton diagram" showing only the specialties, thus allowing them to draw their own links. Thus this question was

designed to test the ability of Trend 44 to show the development of the field of mobile genetic elements

A total of 23 people responded to this section. Out of these 6 people (26%) agreed with the links in the diagram, 1 person entirely disagreed (4%) and 16 (70%) thought that the links were only partially correct. From consultation with statisticians it was found that no inferential statistics would be suitable for endeavouring to decide how representative these results would be for the population of 51 base literature authors. This is because such techniques are designed to draw inferences from a small, randomly selected sample, to a much larger population. In this case the sample was not random because only people featured in the base literature of Trend 44 were surveyed, and the sample represented a sizeable proportion of the total population. The opinions of non-replying scientists in the base literature would not be expected to be very different from those replying because from telephone calls it was deduced that the main reason for non-response was lack of time to fill in the lengthy questionnaire. This factor concerning the inapplicability of inferential statistics was found to apply to all questions of this type in the questionnaire. Thus it was concluded that the so-called "eyeball" test (see F. Clegg, 1982, (10)) would be sufficient for analysing the data.

A variety of modifications were made to the trend diagram by the 14 respondants who had answered "no" or "partial" to the first question, resulting in some rather confusing results. All of the replies showed different linkages, adding or subtracting from those in the original diagram, e.g. 1 respondent agreed with all the linkages between 1978, 1980 and 1982 apart from the link between Specialties 101 in 1978 and 222 in 1980, and added a link between Specialty 543 in 1980 and Specialty 82 in 1982.

Of especial interest were those dealing with Specialties 543 and 134 in the 1980 time period, which according to the trend diagram had no links to the 1982 time period. Three respondants linked only Specialty 543 to Specialty 82 in the 1982 time period, 3 linked only Specialty 134 to 82, and 3 linked both of them to

82. So 9 respondants (39%) thought that the "extinct" specialties of 134 and 543 continued in some way into the 1982 timeframe.

Some of the scientists disagreed with the timeframes in which specialties occurred. One respondent (4%) moved Specialty 388 from 1978 to an earlier, unspecified time period, and showed it as the intellectual precursor of the other 3 specialties in 1978. Two other respondants moved specialties to 1976. One moved Specialty 222 from 1980 and showed it as the precursor of all the specialties in 1978 and even of Specialty 82 in 1982, and the other moved specialties 101 and 525 and then linked them to the remaining specialties in 1978.

The specialties were not seen as distinct entities, one respondent (4%) commented that the distinctions between some specialties seemed very fine. Three respondants (13%) perceived Specialties 134 and 222 in 1980 as being about the same mobile genetics elements concepts, or as one respondent put it some transposons have insertion sequences. Three respondants (13%) also saw Specialties 101 and 525 in 1978 as being the same. One respondent saw Specialties 543 and 82 as being the same. One respondent also pointed out a relevant area of research that had been entirely omitted, saying that bacteriophage Mu is a transposon that is very important in this system and that the whole branch of this subsystem was overlooked. It is possible that papers in this area occur in the current literature of the trend, however the time factor did not permit the exploring of this possibility.

Thus although relatively few scientists openly disagreed with properties of the trend such as the time frames used, in the context of the majority who considered the trend to be not or only partially correct these opinions would appear to be significant. The different opinions concerning the linkages and the nature of the specialties themselves, support the comments of 1 scientist who said that all the specialties in Trend 44 were related and difficult to distinguish between. These findings thus cast doubt on the ability of the trend diagram to convey much useful information.

The relationship of the trend diagram to the historical development of the field is explored in more detail in Chapter 7 where the oral accounts given by scientists in interviews are examined.

(2): Section B

The questions in this section of the questionnaire were designed to test the quality of the base literature papers in each specialty. Details of these papers were supplied in the booklet accompanying the questionnaire (see Appendix D.2) which gave summaries of the content of each of the base literature papers and its reference. The scientists were then asked for their opinions on whether or not any of the papers could be described as key paradigm, and if they thought that only some were, to state which ones. Key paradigm in this case was taken to mean papers that contained data, methodology or theory of importance to the field. Whether they selected their own papers in this section was checked in the analysis afterwards. This was done to test the assumption in CCA that the co-cited papers embody important concepts, methods or data, (See H. Small, 1973, (84)).

In order to discover whether the field had a mostly theoretical or experimental base, the assumption was made that scientific papers fall into 2 major groups, experimental or theoretical. It was also assumed that a gradation exists between these 2 sorts of papers which was represented on the questionnaire as a line between 2 poles. The scientists were asked to place a cross on the line showing which category they thought that the majority of the base literature papers in each specialty belonged to. They were also asked to do the same for how good or bad and how unknown or very well known they considered the majority of the papers in each specialty to be. This was done to assess the quality of the papers in each specialty base literature to test Small's assumption in CCA that co-cited pairs of papers are the important papers for a particular research field (see Chapter 2).

In order to gain some measure of how experimental / theoretical, bad / good or unknown / very well

known the respondents thought the papers to be, the results were analysed by measuring the distance of each cross from the left hand pole. The absolute measurements are given in Tables 6.2 to 6.4. In the subsequent analysis the ranges and medians were expressed as percentages of the total length of the line (which was 8.5cms long). The median was used as a measure of central tendency instead of the mean because it is not as sensitive to changes in the extreme values of the scores, which is important considering the subjective nature of the replies. The operation for calculating the median can be found in F. Clegg, 1982, (10). Thus if all the papers in a specialty were entirely experimental, or bad or unknown, measurements of 0% in each case would be obtained. Conversely if they were all theoretical, or good, or very well known, values of 100% would occur.

Results for Specialties in Section B

(a): Specialty 388

There were a total of 22 respondants to the analysis of this specialty. Out of these 72% thought that all the papers were key paradigm, 22% thought that only some were and 4% that none were. Two of the respondants commented that all the papers were very important but felt the list to be limited. All the papers in the base literature were thought to be key paradigm, see Table 6.1. One respondent went so far as to add that the paper by Fiandt et al opened up the area since it showed that insertion sequences are specific sequences, and first named them IS1, IS2, IS3 and IS4. It was a fundamental paper for this area as prior to it the insertions were believed to be unspecific. One respondent said that the 2 papers were produced by Szybalski's group and Starlinger's group who knew about each others work and decided to publish at the same time.

Then looking at the characteristics of the papers the range for the experimental / theoretical nature of the papers was found to be 91% and the median was 21%, so it can be deduced that the papers were mostly considered to be experimental, see Table 6.2. Similarly for how bad / good the papers were, the range was

37% and the median 85% so the 2 papers were considered to be good overall, see Table 6.3. For unknown / very well known the range was 68% and the median 85%, so they were generally considered to be well known, see Table 6.4. The ranges obtained showed more agreement concerning the bad / good characteristics of the papers than for the other options. An additional comment was made by 1 respondent concerning the title of the specialty which was deemed inappropriate since neither paper reported DNA sequencing.

Thus it was concluded that for this specialty CCA had succeeded in picking up the important papers, one of which was shown to be vital for the area concerned.

(b): Specialty 320

There were a total of 22 respondants to this section. Out of these 36% thought that all of the 4 papers in the base literature of this specialty were key paradigm (1 actually commenting that they were all good), 50% thought that only some were and 9% that none were. The papers were all selected as key paradigm, see Table 6.1. The 2 instances of self citation that occurred can be discounted because the papers involved were also selected by other people. Two of the respondants commented that all of the papers were derived from the same laboratory (N. Davidson's), and that they were all important contributions well summarised in N. Davidson et al (which was a review article covering the work).

For the experimental / theoretical nature of the papers, see Table 6.2, the range was 88% and the median 21%, so the papers were considered to be predominantly experimental. As to their good / bad nature, see Table 6.3, the range was 41% and the median 88%, so they were considered to be good. For unknown / very well known, see Table 6.4, the range was 65% and the median 81%, so they were considered to be well known. The ranges showed more agreement between the respondents concerning the bad / good nature of the papers than for the other options.

This specialty was slightly different from the others in that a further question was posed where scientists were asked to comment on the correspondence between the base literature paper

titles and the title of the specialty that had been deduced from the titles of the current papers by the model builders. This question was restricted to this specialty since this was the only specialty identified where the connection was not immediately obvious (See Section 3, Chapter 4). The base literature paper titles were listed and the scientists were asked whether they considered that the given specialty title was appropriate for the subject matter of the base literature papers as indicated by their titles. In answer to this 45% said yes and 55% no, although the difference was too close to be significant. Thus the majority felt that the correspondence was not good. This group were then asked to write down a name that they felt to be more appropriate. Thirteen titles were proposed, the 2 most typical being "Specialized and General Recombination of IS and Tn elements on F and the Bacterial chromosome" and "The F - Factor and the Role of IS elements in F - Prime and Hfr Formation". Interestingly enough 2 of the respondants selected the title from the paper by N. Davidson et al "Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E. Coli: DNA Sequence Organization of F and of F - Primes, and the Sequences Involved in HFr formation." The titles suggested were observed to be similar to the specialty title "F - Sex Factor Genes and Bacterial Recombination ", but were more detailed. One respondent provided an explanation for the difference between the titles of the base literature papers and the title of the specialty. Apparently Norman Davidson defined the field and set the standard for appropriate titles, but because of his caution you would not guess from the titles alone that these papers dealt with transposable elements.

Thus it was concluded that all of the papers were key paradigm, good, very well known and highly experimental. Even the apparent incongruity of the specialty title was of no great importance and had a rational explanation.

(C): Specialty 525

This had 21 respondants. Out of these 57% felt that all of the papers were key paradigm, 29% felt that only some were, and

14% felt that none were. All the papers in the base literature were selected as being key paradigm, see Table 6.1.

For the experimental / theoretical nature of the papers, the range was 88% and the median 26%, see Table 6.2, indicating that they were mostly experimental. With respect to how bad / good the papers were the range was 96% and the median 78%, see Table 6.3. Thus although the papers were found overall to be quite good, the great differences of opinion were found to have arisen over the dual nature of the Ptashne and Cohen paper. This was considered by the scientists to have both good and bad characteristics, with the scientists often entering 2 crosses on the line (with one at 0%) to show this. The consensus of opinion was that this paper reported incorrect results that were later corrected by them, but that the theoretical principle was correct. The paper apparently contained the important idea that the inverted repetitions in Tn10 were each transposons but it was shown later by Kleckner that the IS element associated with tet r was not related to IS3. For unknown / very well known, the range was 47% and the median 75%, see Table 6.4, so they were generally considered to be well known. The ranges showed a diversity of opinion concerning all the characteristics.

Two of the respondants commented on the specialty title saying that it was not really a new concept and that it was misleading. They said that there were probably better examples of title that could be used. One of the respondants regarded the whole specialty as being redundant, and said that the papers from this specialty should be added to the paradigm papers of Specialty 101.

Thus it was concluded that the subject matter of this specialty was highly controversial, particularly relating to the subject matter of one of the papers, but extending even to its title and in one case doubts about its viability. Its papers were generally considered to be experimental and good, even the controversial paper was independently selected as key paradigm, but there was less dispute about both papers being well known.

(d): Specialty 101

This had 20 respondants. Out of these 30% felt that all the papers were key paradigm, 70% felt that only some were, and 0% thought that none were. All the papers in the base literature were selected by various people to be key paradigm, see Table 6.1.

For the experimental / theoretical nature of the papers the range was 82% and the median 29%, see Table 6.2. For bad / good the range was 46% and the median 82%, see Table 6.3. Similarly for unknown / very well known, the range was 69% and the median 79%, see Table 6.4. The ranges showed more agreement over the bad / good natures of the papers than for the other options. So the papers were considered to be predominantly key paradigm, experimental, good and very well known.

(e): Specialty 222

There was a total of 19 respondants. Out of these 21% felt that all were key paradigm papers, 79% felt that only some were and 0% felt that none were. All 12 papers in the base literature were deemed to be key paradigm, see Table 6.1.

For the experimental / theoretical nature of the papers, the range was 93% and the median 38%, see Table 6.2. For bad / good the range was 41% and the median 83%, see Table 6.3. For unknown / very well known the range was 63% and the median 78%, see Table 6.4. Therefore all the papers had been selected as key paradigm, and generally experimental (although there was the most diversity of opinion over this), good and well known.

(f): Specialty 543

There were 19 respondants. Out of these 11% thought that all the papers were key paradigm, 16% thought that only some were, and 73% that none were. Both of the papers in the base literature were cited as being key paradigm, see Table 6.1.

For the experimental / theoretical nature of the papers, the range was 82% and the median 38%, see Table 6.2. For bad / good the range was 82% and the median 58%, see Table 6.3. For unknown / very well known the range was 79% and the median 53%,

see Table 6.4. Thus the papers were considered to be mainly experimental, average in quality and only averagely known.

The respondants generally criticised the specialty. Three respondants observed that the base literature only dealt with prokaryotic elements, not yeast, so that it did not match the specialty title, and that the specialty category made no sense. One respondent provided an explanation for this by saying that workers in yeast have observed transposition, but as this aspect of their field developed after the investigation of mobile elements in prokaryotes, they would be expected to cite the prokaryote work. Two respondants went so far as to say that it did not appear to be a separate specialty, one said that the prokaryote observations were covered in specialties 222 and 101, and the other that Specialty 543 meant something close to Specialty 82.

The poor results obtained in the validation of the base literature of this specialty are interesting considering that this was one of the "extinct" specialties, i.e. one that failed to carry on into the 1982 time period.

(g): Specialty 134

There were 19 respondants. Out of these 37% thought that all the papers were key paradigm, 42% thought that only some were, and 21% thought that none were. All of the papers were said to be key paradigm, see Table 6.1.

For the experimental / theoretical nature of the papers, the range was 83% and the median 26%, see Table 6.2. For bad / good the range was 46% and the median 81%, see Table 6.3. For unknown / very well known the range was 82% and the median 71%, see Table 6.4. So generally the papers were considered to be experimental (although there was some divergence of opinion), good, and well known (although again there was a divergence of opinion).

Two of the respondants felt that there was no difference between this specialty and 222, an interesting comment considering that Specialty 134 was an "extinct" specialty. Two of the papers in the base literature of this specialty were published in a book

called DNA Insertion Elements. One of the respondants thought that the other papers in this book were just as important as the two in the base literature. Another respondent thought that the paper by D. Berg et al from Specialty 101 should have been included as a key paradigm paper here. One comment by a respondent showed a conceptual link with Specialty 134's ancestral specialty in 1978, 101, in that the D. Botstein paper in 134 was apparently an extension of the P.M. Bennett paper in 101 with some overlap.

Thus it can be seen that the scientists had a higher opinion of the quality of the papers in Specialty 134 than for its companion extinct specialty, 543.

(h): Specialty 82

There were 19 respondants. Out of these 10% thought that all the papers were key paradigm, 79% thought that only some were, and 10% that none were. Out of the 6 papers in the base literature, 5 were selected separately as key paradigm, the one that everyone excluded being by B. Bachi et al, see Table 6.1.

For the experimental / theoretical nature of the papers, the range was 91% and the median 27%, see Table 6.2. For bad / good the range was 69% and the median 79%, see Table 6.3. Similarly for unknown / very well known, the range was 92% and the median 67%, see Table 6.4. So generally the papers were thought to be experimental, good and moderately well known, although there was much divergence of opinion on all 3 counts.

Two of the respondants thought that the specialty title was not descriptive of the content of the cited papers. The example of this that was quoted was that the papers by T. Yun et al and B. Bachi et al concerned genome structure of P lambda rather than IS mediated recombination.

Also as part of Section B, there was a section of general questions where the scientists were asked whether any key papers were missing from the base literature of specialties in the trend, and if so, what they were. They were then asked whether, considering the entire base literature of the trend, there were

any outstandingly good or useful papers, and if so what they were. The purpose of these questions was firstly to ascertain the gaps in the coverage of key paradigm papers by the model builders (and so expose possible errors of omission), and secondly to identify papers of general importance to the field.

There were 20 respondants to this part. In answer to the first question, 65% of these thought that key papers were missing, 25% thought that none were and 10% were not sure. Many missing papers were listed, sometimes accompanied by an explanation as to why they were important. In order to avoid problems with self citation, missing papers proposed only by their authors were excluded. Firstly, those that were considered to be missing from the trend as a whole are listed in Table 6.5. N.B. full references were not always supplied by the respondents.

Thus a total of 23 papers were considered to be missing generally from the base literature of Trend 44. In addition to these, one respondent listed the papers numbered 2, 21 and 23 in Table 6.5 as being representative of a whole area missing from the trend which dealt with bacteriophage integration into the chromosome of a host.

The respondents also proposed a number of papers that were said to be missing from the base literatures of specific specialties (papers proposed only by their authors were again excluded). These are listed in Table 6.6.

Table 6.6 shows that a total of 9 papers were believed to be missing from 3 specialties. The papers by Malamy and Shapiro that were listed as missing from Specialty 388 also appear in Table 6.5 as numbers 11 and 17. In Specialty 320 the paper by Sharp et al was interesting as its title "Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E. Coli. II. Structure of Drug Resistance (R) Factors and F Factors." was very similar to the titles of the other base literature papers in the specialty. The paper by N. Davidson missing from Specialty 82 occurs in the base literatures of Specialties 320 and 222 in Trend 44.

Therefore a total of 32 papers were deemed to be missing from the base literature of Trend 44. However specific gaps in the coverage of areas relevant to mobile genetic elements research were also suggested, the following papers being proposed by scientists to cover the perceived gaps. One comment was that papers dealing with models of transposition that serve as the base of much present research activity had been completely missed, the most important of these in the time scale 1978 to 1982 being numbers 1, 5 and 20 listed in Table 6.5. It was also said that the trend should include papers reporting the nucleotide sequence determination of transposable elements e.g. Ohtsubo et al Proc. Natl. Acad. Sci. USA 75:615 (1978) and Heffron et al J. Bacteriology 122:250 (1975). Papers dealing with bacteriophage lambda and phage Mu and their insertion / properties (the work of Bukhari, Campbell, Toussaint and Taylor), and papers on proposed models for the mechanism of insertion and excision (Shapiro, Sherratt, Galas and Reznikoff) were missing. The omission of papers dealing with Mu research was considered to be particularly serious by 2 of the respondants who commented that since 1977 (See DNA Insertion Elements, Plasmids and Episomes, Cold Spring Harbor Lab) Mu has been considered to be a complex mobile genetic element. This work apparently contributed significantly to the field by furthering the understanding of transposable elements, the trend missed the most important theoretical papers. One very important paper by B. McClintock which demonstrated the possibility of transposition and for which she later got the Nobel prize was missing. Two respondants summarised the feelings of many of the other scientists in saying that there were very many inconsistencies and gaps in the base literature.

Thus evidence had been obtained in the questionnaire for the existence of errors of omission in Trend 44. In order to test whether the trend modelling methodology had omitted to include these papers from the Protein Determination models, the specialty index: current literature data supplied in the models was examined. The search aimed to see whether the missing papers occurred in the current literatures of specialties in Trend 44 or specialties located elsewhere in the models. Only papers published

in 1978, 1980 and 1982 were searched for since papers published in other years would not have been included in the current literature. Of the papers that the respondants said had been omitted, 3 were found to be in 1978 and 1 in 1980. Only 2 of these were located in the current literature of specialties in Trend 44; Iida.S. et al, Plasmid 1: 357 to 365 (1978) in Specialties 525 and 101, and Ohtsubo et al Proc. Natl. Acad. Sci. USA 75: 615 (1978) in Specialties 388 and 101. Thus 50% of the papers were located. It is possible that the remaining 2 occur in the base literatures of other specialties in the Protein Determination models, however time was not available to examine this.

The question concerning whether the base literature contained any outstandingly good or useful papers was asked in order to test the ability of trend modelling to locate papers of especial importance to a field and so test claims made early on in the development of CCA by Small (see Chapters 2 and 3). In answer to this 65% thought that it did, 5% thought that it did not and 30% were not sure. The one person who thought that it did not, still thought well of the papers, the reason being that he thought it was hard to single out any one paper since he regarded all of them as having been useful and having contributed to the overall picture. These sentiments were echoed by 2 other respondants who thought that they were all important but very similar. After those papers that only appeared as a result of self citation were excluded from the analysis, the particularly outstanding papers proposed by the respondents were listed in Table 6.7. In this table the papers are identified by the authors names and the specialties in which they appear, references are only given where extra clarification is required. In Specialties 101, 222 and 134 the book DNA Insertion Elements which was said to be the first book on insertions, was selected instead of its constituent papers. It was concluded from this that CCA trend modelling is capable of locating papers of significance to an on-going area of research.

(3): Section C

The purpose of the questions in this section was mainly to clarify and extend knowledge of the professional relationships between these people in order to prove the existence and show the workings of any invisible college that may be present. The questionnaire respondants were asked to examine the list of base literature authors (which contained all of the authors identified from library sources) presented in Table 1 in the questionnaire (See Appendix D.3) and to indicate on the table their relationships to them. The possible relationships covered were: Read work of, co-authored with, co-worked with, teacher of, student of, same institution as, provided / received technical assistance, first year of knowing, know well, do not know and other. Positive answers to these questions were considered to indicate the presence of an invisible college and negative answers the converse. A total of 24 people filled in the table. The results are presented in Tables 6.8 to 6.15.

Table 6.8 shows base literature authors whose work the respondants have read. The table shows that, except in a few cases, the trend base literature authors work has been read by their colleagues among the respondants to the questionnaire. Thus they appear to be generally aware of each others work.

Table 6.9 shows the networks of more personal relationships between the authors, as it deals with who they have co-worked with. The authors were found to divide roughly into 10 working groups, the distinctions between which were found to be vague due to the high mobility of the scientists concerned. Only 2 groups were found to exist outside the main network. Rosner and Gottesman and Vapnek and Yun showed no patterns of co-working with the others, although Table 6.8 shows that they have an extensive knowledge of the others work through the literature. Another group that was on the periphery was the Swiss group composed of Hanni, Arber, Bachi, Caspars, Iida, Jutte and Meyer. Here a co-working link with members of the main network occurred only through Arber, a Nobel prize winning scientist. Of the originally British group of Datta, Barth, Grinter, Hedges, Jacob, Richmond and Bennett,

links to the others came through Datta and Bennett who each had one link to the main network, and Hedges who had 2. The other groups identified were: Kopecko, Brevet, Cohen and Ptashne (where the latter 2 scientists were the ones that connected with the network); Botstein, Chan, Johnsrud, Kleckner, Miller, Sharp, Tye and Calos; Kuhn, Fritz, Saedler and Starlinger (representing the Germans, with again the last 2 being the main interactants); Berg, Allet, Davidson, Davies and Rochaix; Deonier, Hsu, Hu, Lee and Ohtsubo.E.; Rubens, Falkow, Heffron and Sublett; Szybalski, Fiandt, MacHattie, Malamy and Grindley. These groups thus had many links to each other.

Table 6.10 shows who the respondants have co-authored with. Comparison with Table 6.9 shows that this does not necessarily correspond with whom they have worked. Examples of this include Hu, who authored a paper with one person and Jacob and Sublett who had each authored papers with 2 people that they had not worked with. Also it was found to be quite common to work with other scientists, but not publish papers with them. This is interesting since it gives support to some of the critics of such techniques in the sociology of science (see Chapters 2 and 3) by showing that literature models of science cannot follow all of the interactions between scientists, and that they need to be supplemented with questionnaire and interview data to get a fuller picture of what goes on.

The next item examined was mentor relationships between the scientists. This data was not presented in a table since it was simple enough to describe verbally. These were found to be more clear-cut than working relationships, as they were observed to fall into 12 distinct groups, with no apparent interactions. This could perhaps be an artefact of the questionnaire replies received since some of the non-respondants might have demonstrated links between these groups. Richmond, Grindley and Heffron were found to have been the mentors of Bennett. Similarly Datta was the mentor of Grinter and Jacob, Miller of Calos, Cohen of Kopecko, Vapnek of Yun, Davies of Berg and Szybalski of Fiandt. Falkow was the mentor of Rubens and Sublett, and Saedler and Starlinger were

the mentors of Kuhn. Botstein was the mentor of 4 other scientists, Chan, Johnsrud, Kleckner and Tye.

Slightly more complex mentor relationships were shown by the remaining 2 groups. Davidson was the mentor of Sharp, Deonier and Hu, however Sharp was also the mentor of Deonier and of another scientist, Hsu. An even more interesting set of relationships which were the reason why traditional teacher / student relationships were abandoned in this analysis in favour of the more general mentor relationship, was that exhibited by scientists at the University of Basel in Switzerland. The relationships were observed to be circular in that Iida was the mentor of Arber, who was the mentor of Caspars, who was the mentor of Iida. Caspars also had one other mentor, Meyer.

Then the occurrence of scientists in the same institution was examined. This was not just analysed with respect to current affiliations, if they had ever been in the same institution was considered to be equally important. Again it was observed that the scientists generally did not fall into clearly defined groups, with respect to who had worked in which institution. There were many institutional links which reflected the high mobility of the scientists concerned. Only 3 groups were apparently institutionally isolated. These were the group at the University of Basel, Caspars, Arber, Bachi, Hanni, Iida, Jutte and Meyer and 2 small groups (each composed of only 2 people) of Rosner and Gottesman, and Vapnek and Yun, the same groups that were excluded from the co-working network. The remaining scientists for which co-institutional data was supplied by the respondants (See Table 6.11) showed quite dense networks of interaction. The only respondent who showed no institutional link with any other author was Kuhn.

The patterns of exchange of technical assistance were then examined because scientists often have contact through the exchange of technical information and advice and gifts of substances or even living organisms such as bacterial strains needed for experiments. This sort of information has been called tacit knowledge (e.g. See H.M.Collins, 1982, (11).), which some sociologists of science regard as being vital for the production

of scientific knowledge as a whole. This sort of assistance tends not to be recorded in the traditional format of citations, but sometimes appears in the acknowledgements in scientific papers. This question was intended to elucidate professional relationships not shown by co-working in the published literature.

The data shown in Table 6.12 shows that 2 main groups were identified, one being very much larger than the other. The main group contained the scientists Fiandt, Davies, Allet, Berg, Sublett, Grindley, Heffron, Bennett, Chan, Saedler, Davidson, Tye, Malamy, Botstein, Johnsrud, Kleckner, Rosner, Gottesman, MacHattie, Ptashne, Sharp, Vapnek, Miller, Szybalski, Jackowski, Kopecko, Brevet, Cohen, Calos, Grinter, Datta, Barth, Hedges, Jacob, Deonier, Hsu, Hu, Ohtsubo.E. and Falkow. The second group were the scientists from the University of Basel, Arber, Iida, Caspars, Bachi, Jutte, Meyer and Hanni. These scientists exchanged much technical assistance between themselves, but their only link with the main group was through Arber, who collaborated in this fashion with Malamy and Szybalski. The people who exchanged no technical information with the respondants were: Fritz, Heib, Hu, Kondo, Kuhn, Lee, Mitsuhashi, Richmond, Rochaix, Rubens, Yun and Zenilman. These links would probably have been more extensive and possibly included those authors omitted, had more questionnaires been received.

The data presented in Table 6.13 showing the first year that the respondants came to know of the other authors was examined in a cursory fashion. This question was designed to test whether the time period modelled by the trend provided a representative view of the field. A possible indicator for this was considered to be the length of time that scientists have been active in the field. Most of the scientists appear to have been known from the early 1970's, although a handful do go back to the late 1960's. Only a very few go back to the early 1960's, or even late 1950's. These were Szybalski, Arber, Cohen, Datta, Davidson, Davies, Falkow, MacHattie, Malamy, Miller and Starlinger. It would be interesting to see whether these are the "old hands" of the field who got it started. Stanley Cohen would certainly be in this category. Certainly the field appears to have gained momentum in

the 1970's when most of the authors appear to have entered the field. This idea was further explored through the interviews in Chapter 7.

Then Tables 6.14 and 6.15, which respectively contain data on who the respondants know well and who they do not know were examined. Table 6.14 showed that only 3 people, Kondo, Yun and Zenilman were not well known by any of the respondants. Otherwise all of the other authors showed densely interconnected patterns of knowledge of each other. There were no peripheral groups of scientists. The best known were Arber, Berg, Botstein, Cohen, Davidson, Davies, Falkow, Grindley, Heffron, Kleckner, Miller, Ohtsubo.E., Saedler, Starlinger and Vapnek who each had only up to 3 people who did not know them. Table 6.15 showed that none of the authors were known by all of the respondants, but otherwise it complemented Table 6.14.

The column headed "other" in the table in the questionnaire was used to allow the respondants to indicate other types of relationships to those listed. Only 7 respondants replied to this. Out of these Hu and Deonier used it to show that they were married. Iida said that he had once listened to Sharp giving a seminar. Calos and Datta used it to show acquaintances: Calos with Allet, Arber, Berg, Davies, Kleckner, MacHattie, E and H. Ohtsubo (who are a married couple), Rochaix, Saedler, Starlinger and Szybalski; and Datta with Arber, Bennett, Berg, Kleckner, Kondo, MacHattie, Mitsuhashi, E and H. Ohtsubo, Rubens, Saedler, Starlinger, Szybalski and Vapnek. Bennett used it to show people that he had met at international conferences in Europe and the USA, i.e. Barth, Berg, Cohen, Datta, Falkow, Grindley, Hedges, Heffron, Kleckner, Kopecko, Miller, Saedler and Starlinger. Jacob ticked the names of Bennett, Cohen, Grindley, Heffron, Kopecko, Saedler and Starlinger without specifying the relationship. Thus this section did not provide much further information, but did show the presence of more distant professional relationships, such as seminars, conferences etc.

In order to locate any errors of omission in the trend with respect to missing members of an invisible college, the respondants were asked to state whether any people that they knew

in the field were missing from the list. Out of a total of 20 respondants to this question 30% felt that all the people they knew were present and 70% felt that some were missing. It was interesting that even a small number of people felt the list to be complete since it had been thought likely that all would feel it to be incomplete.

The respondants that felt that some were missing were then asked to list the names of the absent scientists, how they came to know of them and the year that they first came across them. The names of the following scientists were presented, see Table 6.16. Several scientists were mentioned in connection with research on Mu, an area that apparently omitted in the trend. These are listed in Table 6.17. One respondent said that the omission of Bukhari, Shapiro, Sherratt, Galas, Chandler, Kahman and Toussaint was particularly serious as they had provided theoretical models for transposition.

The names of the missing scientists were then searched for in the base and current literatures of all the specialties in the 3 Protein Determination models to see if they could be located, using the author indexes supplied by the model builders. The list of names also included those that were proposed along with the missing papers (Section B)

A total of 38 researchers names were searched for. Only Craig, Morse, Murphy and Roth were excluded as their initials had not been supplied by the respondents, which meant that they could not be identified with any degree of certainty from the database. A total of 57 specialties in 1978, 52 in 1980 and 83 in 1982 were found to be affiliated to them. Due to this large number of specialties, only those that occur in Trend 44, or that appear to be close to the subject matter of the trend or to areas that the respondants claim to have been overlooked, will be discussed. Details of these specialties are presented in Table 6.18.

Table 6.18 shows that all of the specialties in Trend 44 contained people that the respondants said had been missed. This encouraging result was supported by the fact that out of all the missing researchers searched for in the models only 5 could not be located. So the models on the whole have not failed to capture the

work of the missing scientists. A consequence of this search was that a total of 51 specialties over all 3 time periods were discovered that appeared to contain material relevant to the trend. These were particularly interesting in that they covered some of the areas of research that the respondants said had been omitted from the trend (See Section B). The only area that still appears to be missing is that concerned with models of transposition, although a close inspection of the literature of these specialties might show work on this.

(4): Section D

In this section the institutions that were found to be associated with the base literature papers were listed in the questionnaire (See Appendix D.4). The scientists were then asked which institutions in the list they considered to be important to the research area, which ones they had worked at, and which they had collaborated with in their work. These questions were asked in order to test the hypothesis that institutions affiliated to base literature papers are prestigious institutions in the research area of mobile genetic elements. Scientists were asked which institutions they had worked at in order to eliminate those proposing their own institutions as being prestigious and to obtain further information relating to institutional links between authors. Apart for some overall observations, their answers are dealt with for each institution in Table 6.19. There were 22 respondants to this section. Out of these only 3 (14%) felt that all of the institutions were important, and only 2 (9%) said that they had had no collaboration with other institutions. The institutions below are in the same order as the list in the questionnaire, where their full addresses may be obtained. One respondent thought that institutions numbers 1 to 8, 10 and 13 to 16 in Table 6.19 were only important up to 1980 and that institutions 9, 11 and 12 were currently important. That respondents opinion was supported in the case of institution number 2 where 3 of the other respondents (14%) also said that it

was no longer important. The explanation given was that the research group disbanded when Datta retired.

Table 6.19 shows that all of the institutions were claimed to be important, all having gained support in this connection from respondants who had never worked at the institutions concerned. Apparently the most important were (having received support from over 50% of the respondants), the California Institute of Technology, Stanford University (presumably because of its connection with Cohen), Massachusetts Institute of Technology, Harvard University and the Institut fur Genetik. The lowest in importance was apparently the University of Georgia with 18%.

The respondants were were then asked whether any important institutions were missing, and if so what they were. In reply to this 18% thought that none were missing, 50% thought that some were and 32% did not know. Four of the respondants pointed out the difficulty in assessing the importance of institutions due to the high mobility of the scientists concerned. The leading scientists in the field have apparently changed their affiliations once or twice (one respondent said that most of the people listed in the questionnaire were no longer at the same institutions), and that the reputation of an institution in a particular field was often dependant on only a few people. This explains why some institutions were only considered to be important up to 1980. It was thus concluded that the institution itself was of little importance.

When the respondents were asked to list important institutions that they thought had been omitted from the list, it was found that the institutions they proposed tended to be mentioned either as the new addresses of base literature authors or in connection with other eminent people that were put forward by the respondants in answer to Section C. The missing institutions are listed in Table 6.20. This table shows that institutions numbers 4, 11, 22 and 24 were selected by respondents who had worked in them. Out of these only 22 and 24 were proposed only by their workers.

Because it was established that the institutions are only considered important because of their association with the base literature authors or other eminent people it was not considered necessary to search for their occurrence in the current literatures of specialties in the Protein Determination Models.

(5): Section E

In this section the scientists were asked to list the journals that they publish in. This was an attempt to locate possible errors of omission that might be attributable to the exclusion of relevant journals from the core journal set used to generate the Protein Determination Models. Twenty respondants answered this section, publishing in a total of 24 journals. These are listed in Table 6.21.

Table 6.21 shows that the journals that the scientists most frequently published in were contained in the core journal set. Three journals were not included, and of these Nature and the Proceedings of the National Academy of Science USA were deliberately excluded by the model builders because (See Coward et al, 1984, (14)) they were considered to be too broad in their intellectual focus to be considered part of the core of any single research field. Thus since the journals used mostly by the respondants appeared for the most part in the core journal lists, and as the model builders also included papers from non core journal sets that cited any of the same references cited by the core papers in the models, it was concluded that any gaps in the coverage of the field by the trend did not originate in the journals selected by the model builders.

Section 3: General Conclusions for the Questionnaire Analysis

The majority of respondants thought that the links shown in the trend diagram were only partially correct. There was little consistency in the linkages that they proposed themselves but, interestingly enough, a sizeable proportion of them thought that

the "extinct" specialties of 1980, 543 and 134, were connected somehow with Specialty 82 in 1982. There was also some disagreement with the timeframes in which specialties occur, the respondants often moving specialties to an earlier period, and showing them as the precursors of the other specialties in Trend 44. This could be explained by specialties in the model representing ongoing areas of research, whose origins go back to a few years earlier. This hypothesis was tested in Chapter 7.

The respondants also tended not to see the specialties as distinct entities, saying that the distinctions between them seemed to be very fine. This gives credence to the idea proposed in Chapter 4 that the trend may be the most important representation of an area of research, not its component specialties. Some of the most interesting observations of this type related to Specialties 543 and 134. Specialty 134 was said to be the same as Specialty 222 and Specialty 543 the same as Specialties 222 and 82. Thus it would appear that the "extinct" specialties, 543 and 134 are still very much a part of the subject matter of the trend and do survive into the 1982 time period. All of the observations in this part of the questionnaire tended to cast doubts on the ability of the trend diagram to convey much useful information. A serious omission found in the trend was that it had overlooked bacteriophage Mu research, an area that was considered to be very important in transposon research.

Then from the answers to Section B it was discovered that all of the papers in the base literatures of every specialty (except one in Specialty 82), were selected by the respondants as being key paradigm. So CCA does seem to pick up the good papers, especially as in all 3 time periods the majority of respondants felt that the trend contained outstandingly good papers.

All of the papers in the trend were judged to be experimental in nature, and were also generally regarded as being good and well known, apart from those in Specialty 543 where the papers were only considered to be average in quality and only averagely known. For Specialty 320 the majority of respondants thought that the specialty title was not appropriate for the content of the base literature papers. However a rational

explanation for this incompatability was supplied by 1 respondent and the alternative titles that were suggested were found to be similar to the specialty title, only more detailed. Interestingly, some of the respondants went on to cast similar doubts on the titles of Specialties 525, 543 and 82.

The majority of respondants thought that key papers were missing. A total of 23 papers were judged to be missing from the entire trend and 9 from individual specialties. It would be interesting to see whether these papers appear in the base literatures of other specialties in the Protein Determination models. Although time was not available for this, 50% of the papers published in the years covered by the models were found to be present in the current literatures of specialties in Trend 44.

In Section C the ability of the trend to list the members of an invisible college was tested. The respondants were found to have an extensive knowledge of each others work. Patterns of co-working showed ill-defined groups, the distinctions between them being vague because of the high mobility of the scientists concerned. The groups were found to generally have many links to each other. Study of the patterns of co-authorship proved interesting since it was expected to find that it reflected the co-working relationships. However, although this was generally found to be the case, it was not always so. It was found to be quite common to work with other scientists but not to publish with them. Thus it was concluded that literature models of science cannot show all the interactions between scientists.

Mentor relationships were found to form more distinct groups, with few apparent interactions. This was as expected, since these relationships naturally tend to involve few people. Institutional affiliations showed that the scientists again generally did not fall into clearly defined groups, and there were many international links which again reflected the high mobility of the scientists concerned. Examination of patterns of exchanges of technical assistance showed the presence of one very large network of exchanges, and one isolated group.

Thus it was concluded that most of the scientists belonged to one main network. The density of the interactions

found were impressive considering the relatively few replies received. So it can be concluded that the base literature authors seem to belong to one large invisible college. The only exceptions to this were the group of scientists at the University of Basel who were peripheral to this main network (but had links via Arber), and 2 other small groups. The group at Basel were on the whole institutionally isolated and restricted co-working and provision of technical assistance between themselves. The 2 other groups, each composed of only 2 people, appeared to be very isolated. These groups contained the scientists Rosner and Gottesman and Vapnek and Yun. They were isolated institutionally and through co-working. However Rosner, Gottesman and Vapnek were found to interact with the others via the literature and the provision of technical assistance. Yun can probably be discounted since he was a masters student and probably only in the field for a short time. Thus the majority of the authors belong to an invisible college, a view which is further supported by the fact that when asked who they knew they showed dense patterns of interconnections. It is possible that as the invisible college in the base literature was shown to be very close knit in its various manifestations that it was vital for the fields development. This hypothesis is examined further in Chapter 7.

Examination of data showing the years in which scientists first came to be known by their colleagues, shows that very few appeared before the 1970's, when the field appears to have gained momentum. Many scientists were said to be missing from the trend. When their names were searched for in the base and current literatures of all the specialties of the current literature models, all bar 5 were located. The specialties that they were affiliated to were found to include specialties which appeared to be on similar subjects to those contained within Trend 44, and more interestingly on subjects that the trend was said to have omitted. In Section B these missing areas of research were said by respondants to be; Bacteriophage lambda and Mu, the nucleotide sequence determination of transposable elements and proposed mechanisms of insertion and excision.

Section D showed that very few of the respondants felt that all of the institutions were currently important, although all had been considered to be important at some stage. The explanation provided was that the reputation of a given institution in this field was dependant on a few scientists, who tended to be highly mobile and change their institutional affiliations often. This was highlighted by the difficulty in tracking many of the respondants down. This emphasizes the need to constantly update bibliometric data in this respect if ranking of institutions is to be useful for public policy purposes. In support of this, important institutions that were listed as having been omitted were mentioned in connection with either base literature authors or other eminent scientists.

It was concluded from Section E that the scientists tend to publish in the core journals used to build the Protein Determination models. So omissions in the coverage of the field of mobile genetic elements research by Trend 44 did not arise through this. Thus overall it was concluded that the trend suffers from errors of omission rather than commission in that it omitted important papers, authors and research areas. However as this omitted information was generally found to occur elsewhere in the models, the problem may by soluble and the trend generating algorithm may be altered accordingly.

It was generally concluded that the main methodological problem associated with the questionnaire was the subjectivity of the replies to Sections A, B and parts of Section D. This was to be expected because these sections asked the scientists to express opinions about the data presented, and personal prejudices could be expected to affect their judgement. However the subsequent analysis attempted to reduce this effect firstly by canvassing the opinions of many scientists and secondly by eliminating obviously suspect data (for example uncorroborated high opinions of a scientists own work). The unsuitability of statistical techniques in the analysis of the data prevents overt generalisation of some of the findings in order to make predictions about the properties of trends in general. However the results obtained should be able

to act as a guide to the problems that may occur with data derived from other trends.

One further word of warning must be issued with respect to using trend data to analyse developments in a given scientific field. One of the respondants emphasized the importance of small important meetings in the development of the field where individuals who did not loom large in the publications had a crucial impact, the example cited being that of Peter Starlinger's influence on Heinz Saedler.

Table 6.1 Showing Base Literature Papers in Trend 44 that were Deemed to be Key Paradigm by the Respondents.

Specialty	Paper (FNA)	No. of times Selected as Key Paradigm	No. of Self Citations
388	Saedler, H.	2	0
	Fiandt, M.	4	0
320	Hu, S.	6	0
	Ohtsubo, E.	5	0
	Sharp, P.A.	9	1
	Davidson, N.	3	1
525	Hu, S.	4	0
	Ptashne, K.	2	0
101	Barth, P.T.	6	1
	Bennett, P.M.	3	0
	Heffron, F.	7	1
	Kondo, E.	3	0
	Rubens, C.	5	0
	Berg, D.	8	1
	Gottesman, M.	6	0
	Heffron, F.	12	0
	Kopecko, D.J.	3	1
	Kleckner, N.	10	1
	MacHattie, L.A.	5	0
	Hedges, R.W.	5	1
222	Hu, S.	7	0
	Ptashne, K.	4	0
	Gottesman, M.	6	1
	Ohtsubo, H.	10	0
	Johnsrud, L.	12	0
	Kleckner, N.	9	0
	MacHattie, L.A.	7	0
	Calos, M.P.	11	1
	Grindley, N.	12	0
	Kuhn, S.	3	0
	Saedler, H.	6	0
	Davidson, N.	2	0
543	Kopecko, D.J. (J.Mol.Biol)	1	0
	Kopecko, D.J. (PNAS)	2	0
134	Berg, D.E.	4	0
	Botstein, D.	3	0
	Kleckner, N.	6	1
82	Bachi, B.	0	0
	Hu, S.	4	0
	Ohtsubo, E.	10	0
	Yun, T.	5	1
	Arber, W.	9	1
	Iida, S.	6	0

NB: FNA indicates that the papers are identified by their first named authors.

Table 6.2 Ranked Absolute Scores for Experimental/ Theoretical Characteristics of Base Literature Papers in Trend 44.

Specialty	388	320	525	101	222	543	134	82
Scores	7.9	7.8	8.0	7.9	8.4	7.9	7.6	8.0
(Distance	7.75	7.3	7.9	7.5	8.2	7.8	6.8	7.9
in cms	7.4	7.2	7.7	7.5	8.0	7.7	5.3	7.5
along line)	7.3	4.6	4.4	4.3	4.7	4.7	4.6	4.9
	4.5	4.3	4.3	3.6	4.5	4.7	3.7	3.5
	4.2	3.8	4.2	3.6	4.5	4.6	3.5	3.0
	3.7	3.7	3.7	3.5	4.1	4.6	3.4	2.5
	2.7	2.5	2.8	3.4	3.8	3.2	2.7	2.5
	2.6	2.5	2.3	2.6	3.5	3.2	2.6	2.3
	2.1	2.5	2.3	2.5	3.0	2.5	1.8	2.3
	2.0	1.8	2.2	2.2	2.1	2.5	1.5	2.3
	1.6	1.7	2.1	1.8	1.9	2.0	1.5	2.1
	1.6	1.5	2.1	1.8	1.6	1.8	1.5	2.1
	1.4	1.3	1.6	1.7	1.3	1.1	1.2	1.4
	1.3	1.0	1.6	1.3	1.1	0.9	0.9	1.1
	1.2	0.7	1.5	1.2	1.0		0.8	0.9
	1.0	0.7	1.3	1.2	0.6		0.6	0.8
	0.6	0.6	1.3	1.2	0.5		0.5	0.5
	0.6	0.5	1.2	0.9				0.3
	0.5	0.4	1.1					
	0.5	0.3	0.5					
	0.1							
Median	1.8	1.8	2.2	2.5	3.25	3.2	2.2	2.3
% Median	21.2	21.2	25.9	29.4	38.2	37.6	25.9	27.1

Table 6.3 Ranked Absolute Scores for Bad / Good Characteristics of Base Literature Papers in Trend 44.

Specialty	388	320	525	101	222	543	134	82
Scores	8.2	8.2	8.2	7.9	8.5	7.9	8.1	8.2
	8.1	8.0	8.2	7.8	8.2	7.7	7.8	8.2
	8.1	8.0	8.2	7.8	8.2	6.7	7.7	8.0
	8.0	7.9	8.1	7.8	8.1	6.5	7.7	7.7
	8.0	7.9	7.8	7.7	8.0	5.6	7.7	7.7
	7.9	7.75	7.7	7.7	8.0	5.1	7.6	7.7
	7.9	7.7	7.6	7.5	7.3	5.0	7.6	7.2
	7.75	7.7	7.5	7.5	7.2	4.9	7.1	7.0
	7.7	7.7	7.3	7.0	7.2	4.8	6.9	6.8
	7.6	7.6	7.3	7.0	6.9	4.7	6.9	6.7
	7.3	7.5	7.1	6.9	6.8	4.2	6.7	6.2
	7.3	7.4	7.0	6.6	6.7	4.1	6.7	6.1
	7.3	7.3	6.3	6.5	6.6	3.5	6.5	5.9
	7.3	7.3	6.3	6.5	6.5	0.9	6.3	5.5
	7.1	7.2	6.2	5.9	6.3	0.9	6.1	5.2
	7.0	7.0	6.1	5.5	6.2		5.4	5.1
	6.9	6.8	5.7	5.3	5.2		5.2	4.6
	6.9	6.6	5.3	5.0	5.0		4.2	2.3
	6.9	5.7	4.4	4.0				
	6.5	4.8	4.3					
	5.5	4.7	4.1					
	5.2		1.9					
	5.0		0.25					
Median	7.3	7.5	6.65	7.0	7.05	4.9	6.9	6.75
% Median	85.9	88.2	78.2	82.3	82.9	57.6	81.2	79.4

Table 6.4 Ranked Absolute Scores for Unknown / Very Well Known Characteristics of Base Literature Papers in Trend **44.**

Specialty	388	320	525	101	222	543	134	82
Scores	8.5	8.4	8.1	7.9	8.3	8.0	7.8	8.1
	8.2	8.0	8.1	7.9	8.0	7.5	7.8	8.0
	8.0	8.0	8.0	7.9	8.0	7.2	7.8	7.7
	7.75	7.8	7.8	7.8	7.9	5.5	7.8	7.6
	7.7	7.5	7.7	7.7	7.5	5.5	7.7	7.4
	7.7	7.5	7.7	7.3	7.2	5.0	7.7	6.9
	7.6	7.5	7.3	7.0	7.1	4.5	7.0	6.7
	7.4	7.1	7.3	6.9	7.0	3.5	6.8	6.3
	7.3	7.1	6.8	6.7	6.9	3.3	6.6	5.7
	7.3	7.1	6.5	6.7	6.6	2.9	6.5	5.1
	7.2	6.7	6.4	6.6	6.5	2.8	6.4	4.9
	7.2	6.7	6.1	6.5	6.3	2.7	6.3	4.7
	7.2	6.6	5.0	6.2	6.3	2.3	6.1	4.5
	7.2	6.6	5.0	6.0	6.2	1.3	4.9	4.2
	7.2	6.3	4.8	5.6	6.0		4.9	4.1
	6.8	6.0	4.8	5.5	6.0		4.2	2.5
	6.7	4.6	4.8	5.5	6.0		3.4	2.5
	5.7	4.5	4.5	4.1	4.2		0.8	
	5.4	3.8	4.3	2.0	2.9			
	5.3	2.9	4.3					
	2.7		4.1					
Median	7.2	6.9	6.4	6.7	6.6	4.5	6.55	5.7
% Median	84.7	81.2	75.3	78.8	77.6	52.9	77.1	67.1

Table 6.5 Papers Generally Missing from the Base Literature of Trend 44.

Authors	Reference	No. of Proposals	Reason for Importance
1. Arthur & Sherratt	Mol.Gen.Gen. v175 p267 1979	2	
2. Bertani et.al	Virology v6 p357 1958	1	
3. Blottner	Virology v62 p458 1974	1	Showed isolation of new IS (IS5) by selective system.
4. Brachet	Mol.Gen.Gen v108 p266 1970	1	Isolated IS insertions that block replication.
5. Bukhari	Trends in Bloch SCI v6 p56 1981	1	
6. Campbell et.al.	Gene v5 p197 1978	1	IS and transposon nomenclature
7. Casadaban & Cohen	No Ref.	1	Use of gene fusions to analyse Tn3
8. Fiandt et.al	Gene v2 p55 1977	1	Showed subject of Morse et.al paper to be IS2
9. Heffron	Several Papers, no Ref.	1	Said to be outstanding
10. Jordan, E; Saedler, H. & Starlinger, P.	Mol.Gen.Gen v102 p353-363 1968	2	
11. Malamy et.al	Mol.Gen.Gen v119 p207-222 1972	1	
12. Morse et.al	Genetics v41 p158 1956	1	First paper on first IS insertion gal 3 (=IS2)
13. Pilacwski et.al	Gene v2 p61 1977	1	Showed IS2 could acquire promoter (INT mutants)
14. Reif & Saedler	Mol.Gen.Gen v137 p17-28 1975	1	
15. Roth & Botstein	J.Mol.Biol 1978?	1	Demonstrated the broad use of transposable elements.
16. Saedler et.al	Mol.Gen.Gen v115 p258-265 1972	1	
17. Shapiro	J.Molec.Biol v40 p93 1969	1	
18. Shapiro	J.Molec.Biol v40 p249 1969	1	
19. Shapiro et.al	Mobile Genetic Elements Academic Press NY	1	
20. Shapiro	PNASUSA v76 p1933 1979	4	
21. Shimada et.al	J.Mol.Biol v63 p483 1971	1	
22. Szybalski	Cold Spring Harbor NY p583-590 1977	1	First listing of all IS elements
23. Taylor	PNASUSA v50 p1043 1963	1	

Table 6.6 Papers Missing from Specific Specialties in Trend 44.

Specialty	Authors	Reference	No. of Proposals
388	Hirsch, Saedler & Starlinger	Mol.Gen.Gen v115 p266 1972	2
388	Hirsch et.al	Mol.Gen.Gen v119 p191 (19?)	1
388	Malamy, M.	Cold Spring Harbor Lab p359 1967	1
388	Malamy et.al	Mol.Gen.Gen v119 p207 1972	1
388	Shapiro, J.	J.Mol.Biol v40 p93 1969	1
320	Sharp, P.A; Cohen, S.N; Davidson, N.	J.Mol.Biol v75 p235-255 1973	1
320	Westmoreland et.al	Science v163 p1343 1969	1
82	Arber et.al	Cold Spr.Harb.Symp. Quan.Biol. v45 p38 (1980?)	1
82	Davidson et.al	Microbiology v1474 p56 1974	1

Table 6.7 Base Literature Papers Selected as Being Outstandingly Good or Useful.

Specialty	Paper	No. of Proposals
388	Saedler et.al	3
320	Hu, S. et.al	1
	Ohtsubo, E. et.al	1
	Sharp, P.A. et.al	4
	Davidson, N. et.al	3
525	Hu, S. et.al	3
	Ptashne, K. et.al	1
101	Bennett, P.M.	1
	Berg, D. et.al	1
	MacHattie, L.A. et.al	1
	Rubens, C. et.al	2
	Hedges, R.W. et.al	2
	Gottesman, M. et.al	3
	Heffron, F. et.al (J.Bact)	4
	Heffron, F. et.al (PNASUSA)	5
	Kleckner, N. et.al	5
222	Ptashne, K. et.al	1
	Gottesman, M. et.al	1
	Ohtsubo, H. et.al	1
	Davidson, N. et.al	1
	MacHattie, L.A. et.al	2
	Calos, M.P. et.al	2
	Hu, S et.al	3
	Saedler, H. et.al	3
	Johnsrud, L. et.al	4
	Kleckner, N.	5
	Grindley, N.D.F.	5
543	Kopecko, D.J. et.al (J.Mol.Biol)	1
134	Berg, D.E.	1
	Boststein, D. et.al	2
	Kleckner, N. et.al	5
82	Hu, S. et.al	1
	Arber, W. et.al	1
	Iida, S. et.al	1
	Ohtsubo, E. et.al	2

Table 6.8 Showing Base Literature Authors of Trend 44 Whose Work the Respondents have Read.

Authors	Respondents (numbered as author column.)																							
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
1. Allet, B.	●		●		●	●	●		●	●			●		●		●	●	●				●	●
2. Arber, W.	●	●	●		●	●	●	●	●	●	●		●	●	●	●	●	●	●			●	●	●
3. Bachi, B.						●					●		●					●	●					
4. Barth, P.T.	●	●		●	●			●	●	●			●	●	●		●	●	●	●		●		●
5. Bennett, P.M.		●		●	●			●	●	●			●	●	●			●	●	●				●
6. Berg, D.	●		●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●		●	●	●
7. Botstein, D.	●	●		●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●		●	●	●
8. Brevet, J.	●	●						●					●		●		●	●	●					
9. Calos, M.P.	●	●	●	●		●	●		●	●			●	●	●	●	●	●	●			●	●	●
10. Caspars, P.													●					●						
11. Chan, R.K.		●							●				●		●		●	●	●	●				
12. Cohen, S.N.	●	●	●		●	●	●	●	●			●	●	●	●		●	●	●	●	●	●	●	●
13. Datta, N.	●	●	●	●	●		●		●	●	●		●	●	●		●	●	●	●		●	●	●
14. Davidson, N.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●	●	●	●	●	●
15. Davies, J.	●	●	●	●	●	●	●	●		●			●	●	●		●	●	●	●		●	●	●
16. Deonier, R.C.		●	●		●		●	●				●	●	●	●		●	●	●	●			●	●
17. Falkow, S.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●	●	●	●	●	●
18. Fiandt, M.	●	●	●		●		●	●	●	●			●	●	●		●	●	●	●	●	●	●	●
19. Fritz, H.J.													●					●						
20. Gottesman, M.	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●			●	●
21. Grindley, N.	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●			●	●
22. Gunter, N.	●	●		●	●			●	●	●			●	●	●			●		●				●
23. Hanni, C.	●					●							●					●		●				●
24. Hedges, R.W.	●	●	●	●	●		●	●	●	●			●	●	●		●	●	●	●		●		●
25. Heffron, F.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●	●		●	●	●
26. Heiß, B.	●									●			●		●			●	●					
27. Hsu, M.T.	●		●	●			●	●	●	●		●	●		●		●	●	●	●				
28. Hu, S.	●	●	●	●	●	●		●	●	●	●		●	●	●	●	●	●	●	●				●
29. Iida, S.	●	●	●	●	●	●		●	●	●	●		●		●		●	●	●					●
30. Jakowski, J.B.	●		●							●			●		●			●	●	●				
31. Jacob, A.	●	●	●					●	●				●		●		●	●	●	●		●		●
32. Johnsrud, L.	●	●	●		●	●	●		●	●	●		●		●	●	●	●	●					●
33. Jutte, H.						●							●					●						
34. Kleckner, N	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●
35. Kopecko, D.J.	●	●	●		●	●	●	●	●	●	●	●	●	●			●	●	●	●		●	●	●
36. Kondo, E.	●		●					●	●				●		●			●	●	●				●
37. Kuhn, S.								●					●		●			●						
38. Lee, H.J	●									●		●	●		●			●	●					
39. MacHattie, L.A.	●	●	●		●	●	●	●	●	●			●	●	●		●	●	●			●	●	●
40. Malamy, J	●	●	●	●	●		●	●	●	●			●	●	●			●	●	●	●	●	●	●
41. Meyer, J.	●	●			●	●		●	●		●		●		●			●	●					
42. Miller, J.H.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●			●	●	●
43. Mitsuhashi, S.	●	●	●					●	●				●	●	●			●	●	●				
44. Ohtsubo, E.	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
45. Ohtsubo, H.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●	●		●	●	●
46. Ptashne, K.	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●	●		●		●
47. Richmond, M.H.	●	●	●	●	●			●	●	●			●	●	●		●	●	●	●			●	●
48. Rochaix, J.D.	●	●	●		●		●	●	●				●		●			●	●					
49. Rosner, J.L.	●	●	●		●	●	●	●	●	●			●		●		●	●		●			●	●
50. Rubens, C.	●	●	●	●			●	●	●	●			●		●			●	●					
51. Saedler, A.	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●
52. Sharp, P.A.	●	●	●	●	●	●	●	●	●	●	●		●	●	●		●	●	●	●			●	●
53. Starlinger, P.	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	●	●
54. Sublett, R.	●	●	●												●			●		●				
55. Szybalski, W.	●	●	●		●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●			●
56. Tye, B.K.		●	●		●		●		●	●			●		●		●	●	●	●			●	●
57. Vapnek, D.	●	●	●		●	●		●	●	●	●	●	●	●	●		●	●	●	●	●		●	
58. Yun, T.													●		●			●	●					
59. Zenilman, M.		●								●			●					●						

Table 6.9 Showing Base Literature Authors of Trend 44 that the Respondents have Co-worked with.

Authors	Respondents (numbered as the author column in Table 6.8)																						
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.		●							●														●	
Arber, W.						●					●		●				●	●					●	
Bachi, B.											●		●					●						
Barth, P.T.								●						●										
Bennett, P.M.																								
Berg, D.									●														●	
Botstein, D.							●																	
Brevet, J.															●									
Calos, M.P.																								
Caspars, P.											●		●					●						
Chan, R.K.			●						●															
Cohen, S.N.				●								●			●		●				●		●	
Datta, N.									●					●										
Davidson, N.		●								●											●		●	
Davies, J.		●															●						●	
Deonier, R.C.												●												
Falkow, S.	●																			●		●		
Fiandt, M.									●								●						●	
Fritz, H.J.																●								
Gottesman, M.																			●					
Grindley, N.																	●							
Grinter, N.								●						●										
Hanni, C.						●							●					●						
Hedges, R.W.								●	●					●			●							
Heffron, F.	●																			●		●		
Heiß, B.																								
Hsu, M.T.										●		●									●			
Hu, S.										●														
Iida, S.						●					●							●						
Jakowski, J.B.																								
Jacob, A.								●																
Johnsrud, L.		●	●		●																			
Jutte, H.						●							●					●						
Kleckner, N.			●				●																	
Kopecko, D.J.																								
Kondo, E.																								
Kuhn, S.																								
Lee, H.J.										●		●												
MacHattie, L.A																	●						●	
Malamy, J.																							●	
Meyer, J.						●							●											
Miller, J.H.			●		●												●						●	
Mitsuhashi, S.																								
Ohtsubo, E.										●		●									●			
Ohtsubo, H.																								
Ptashne, K.				●											●		●							
Richmond, M.H.	●							●																
Rochaix, J.D.		●							●															
Rosner, J.L.																								
Rubens, C.																								
Saedler, A.												●					●							
Sharp, P.A.			●																					
Starlinger, P.																●	●						●	
Sublett, R.																								
Szybalski, W.																	●							
Tye, B.K.			●				●																	
Vapnek, D.																								
Yun, T.																								●
Zenilman, M.																								

Table 6.10 Showing Base Literature Authors of Trend 44 the Respondents have Co-Authored with.

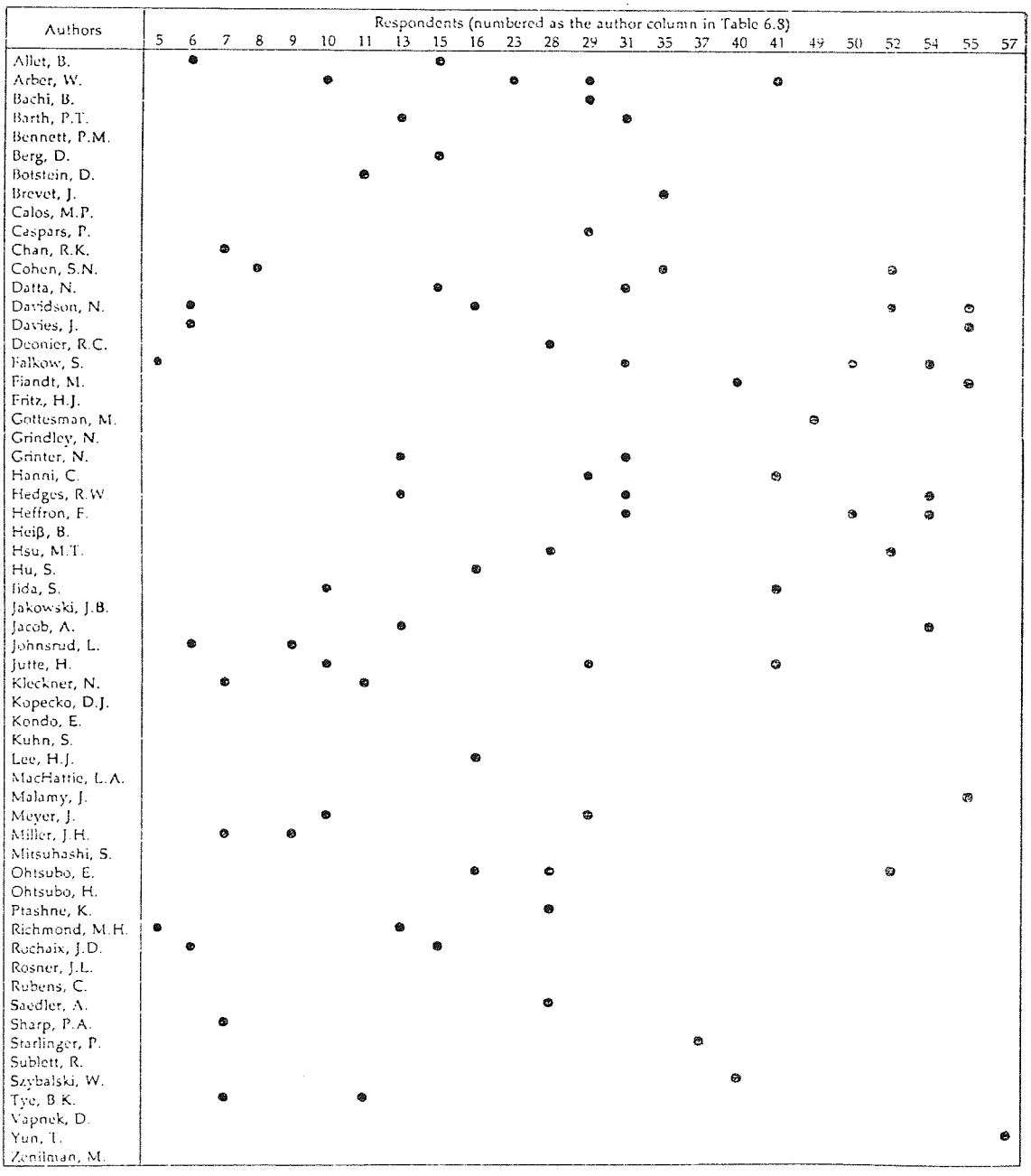


Table 6.11 Showing Base Literature Authors of Trend 44 that the Respondents have Been in the Same Institution as.

Authors	Respondents (numbered as the author column in Table 6.8)																						
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.		●			●		●																	
Arber, W.						●					●		●					●						
Bachi, B.						●							●					●						
Barth, P.T.								●	●					●										
Bennett, P.M.																								
Berg, D.									●															
Botstein, D.							●																	
Brevet, J.															●		●							
Calos, M.P.																								
Caspars, P.													●					●						
Chan, R.K.									●															
Cohen, S.N.		●			●										●									
Datta, N.									●					●										
Davidson, N.												●									●			
Davies, J.		●					●																●	
Deonier, R.C.												●									●			
Falkow, S.					●		●													●		●		
Fiandt, M.									●														●	
Fritz, H.J.																								
Gottesman, M.																			●					
Grindley, N.																								
Grinter, N.								●						●										
Hanni, C.						●							●					●						
Hedges, R.W.								●	●					●			●							
Heffron, F.							●													●		●		
Heiß, B.																								
Hsu, M.T.										●		●									●			
Hu, S.										●														
Iida, S.						●					●							●						
Jakowski, J.B.																								
Jacob, A.								●	●															
Johnsrud, L.		●			●																			
Jutte, H.						●							●					●						
Kleckner, N.			●		●		●																	
Kopecko, D.J.																								
Kondo, E.																								
Kuhn, S.																								
Lee, H.J.										●		●												
MacHattie, L.A.																								
Malamy, J.																								
Meyer, J.						●							●											
Miller, J.H.		●			●				●															
Mitsuhashi, S.																								
Ohtsubo, E.												●									●			
Ohtsubo, H.																								
Ptashne, K.					●										●									
Richmond, M.H.	●																							
Rochaix, J.D.		●			●				●															
Rosner, J.L.																								
Rubens, C.	●						●																	
Saedler, A.		●																						
Sharp, P.A.			●						●	●														
Starlinger, P.																								
Sublett, R.																								
Szybalski, W.		●							●															
Tye, B.K.							●																	
Vapnek, D.																								
Yun, T.																								●
Zemilman, M.																								

Table 6.12 Exchange of Technical Assistance Between Respondents and Base Literature Authors of Trend **14.**

Authors	Respondents (numbered as the author column in Table 6.8)																						
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.									●														◉	
Arber, W.						●											●						◉	
Bachi, B.						●																		
Barth, P.T.								●						●										
Bennett, P.M.																								
Berg, D.									●								●		●			●	●	
Botstein, D.																	●							
Brevet, J.															●									
Calos, M.P.																								
Caspars, P.													●											
Chan, R.K.			●						●															
Cohen, S.N.															●		●				●		●	
Datta, N.									●					●			●							
Davidson, N.			●												●						●		●	
Davies, J.														●			●						●	
Deonier, R.C.												●			●		●				●			
Falkow, S.														●	●		●					●		
Fiandt, M.									●								●						●	
Fritz, H.J.																								
Gottesman, M.			●														●							
Grindley, N.	●		●														●							●
Grinter, N.								●						●										
Hanni, C.						●							●											
Hedges, R.W.								●						●			●							
Heffron, F.	●		●					●						●	●		●					●		●
Heiß, B.																								
Hsu, M.T.										●		●			●						●			
Hu, S.																								
Iida, S.						●					●													
Jakowski, J.B.															●									
Jacob, A.																								
Johnsrud, L.			●																					
Jutte, H.						●							●											
Kleckner, N.			●														●							
Kupecko, D.J.																								
Kondo, E.																								
Kuhn, S.																								
Lee, H.J.																								
MacHattie, L.A.			●												●		●						●	
Malamy, J.			●																				●	
Meyer, J.						●							●											
Miller, J.H.			●		●												●						●	
Mitsuhashi, S.																								
Ohtsubo, E.												●					●				●			
Ohtsubo, H.																	●							
Ptashne, K.			●												●		●							
Richmond, M.H.																								
Rochaix, J.D.																								
Rosner, J.L.			●				●								●									
Rubens, C.																								
Saedler, A.			●									●			●		●							
Sharp, P.A.			●							●					●		●							●
Starlinger, P.																	●							
Sublett, R.																								
Szybalski, W.			●														●							
Tye, B.K.			●																					
Vapnek, D																	●				●			
Yun, T.																								
Zenilman, M.																								

Table 6.13 First Year that the Respondents Came to Know of Base Literature Authors of Trend 44.

Authors	Respondents (numbered as the author column in Table 6.8)																							
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.		74			77		70		74	73	67		78					79	73				70	76
Arber, W.		71			77	74	70	68	71	78	75		74	72	81	75	81	76	63				58	74
Bachi, B.						72							75					76						77
Barth, P.T.	75	76		76	78			70	74	75				70						76				68
Bennett, P.M.		75			78			74	72	80			79	74	75					77				76
Berg, D.	75			75	78		74	75	74	73	70		78	75	75		74		70	75		79	68	74
Botstein, D.	75	68		75	77		68		70	73	72		80	74	75		72		68	74		72		72
Brevet, J.		75													75									
Calos, M.P.	78	78		75			78		79	80			80	80	78	78			75				78	78
Caspars, P.											75		78											
Chan, R.K.									72										68	74				
Cohen, S.N.	75	69		75	78		72	75	71	72			77	70	72			84	75	73	69	75	65	70
Datta, N.	72	74		75	78		74		71	80	74		77	69	72					73		72	65	72
Davidson, N.	75	69			78		70	74	73	71	73	70		74	72			74		73	69	75	64	70
Davies, J.	75	74			78		70			78			77	72	72				75	73			60	72
Deonier, R.C.		71					74					71	79	72	75				78	73	71			74
Falkow, S.	75	73			77		74	68	72	78			77	71	69			83	65	73	69	72	62	69
Fiandt, M.		70					70		72					72						73	70		63	
Fritz, H.J.																77								
Gottesman, M.		75					70		74	77	70			72	75			74	73	74				72
Grindley, N.	73	74			77		78	73	76	78	76		79	74	75	78			77	76				74
Grinter, N.		74						75	72	75			84	71						75				74
Hanni, C.						77							75					76						
Hedges, R.W.	73	74					76	70	72	75				70	77					73		72		76
Heffron, F.	75	73			77		74	73	74	75	70		79	74	75				75	73		72		74
Heiß, B.										73														
Hsu, M.T.							70			71		71			75					75	70			

Table 6.13, continued.

Authors	Respondents (numbered as the author column in Table 6.8)																							
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Hu, S.		75								72	75			74		77				74				74
Iida, S.		73				77				79	75							76	73					74
Jakowski, J.B.										77					75									
Jacob, A.	74	74							73						77					73		72		76
Johnsrud, L.	78	73			75		78			78	74		79			78								76
Jutte, H.						77							75											
Kleckner, N.	75	74		74	77		68		76	78	75		78	74	75	79			68	75		77		79
Kondo, E.	76																			74				
Kopecko, D.J.	75	74					70		74	77			79	72					75	75				75
Kuhn, S.																								
Lee, H.J.										71		71												
MacHattie, L.A.	77	75			78		70			80				74	75				70				60	76
Malamy, J.	73	73					70		67	78	75		84	72							70		60	77
Meyer, J.						77			76		70		76											
Miller, J.H.	73	71			77		70		73	78			78	72	78			77	70				61	69
Mitsuhashi, S.		75							65					69	79					73				
Ohtsubo, E.	74	71			77		70		77	72	73	71	79	78	74	78			75	74	70		67	77
Ohtsubo, H.		71			77		70		77	78	73		79	78	74				75	74				76
Ptashne, K.		75		74						75	73			69	73					76				76
Richmond, M.H.	72	75							72	77			77	69	72					73				75
Rochaix, J.D.		74			78		74		74				77											
Rosner, J.L.		73					70		76	77			79		75					73				70
Rubens, C.	75	75					74			78														73
Saedler, H.	73	69			77	78	70		76	73	75	73	77	74	75	73		77	72	73	70	77		74
Sharp, P.A.	73	70					73		79		74			72	80					74				70
Starlinger, P.	73	69		75	77	78	70		76	80	75		76	74	75	72		78	72	74	70		58	74
Sublett, R.																				75				
Szybalski, W.	73	69			77		69		64	74	72		76	75	84			80	68	73	70			
Tye, B.K.		75					70		78					84					72	75				72
Vapnek, D.		75							77	78	75	80	80	70	79				75	74	69			
Yun, T.																								72
Zenilman, M.										78														

Table 6.14 Base Literature Authors of Trend 44 that the Respondents Claim to Know Well.

Authors	Respondents (numbered as the author column in Table 6.8)																						
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.			●				●		●				●					●					●	
Arber, W.		●	●			●			●		●		●					●	●				●	●
Bachi, B.						●						●												
Barth, P.T.								●	●					●										●
Bennett, P.M.									●						●									
Berg, D.			●				●		●	●			●		●		●		●			●	●	●
Botstein, D.		●			●		●		●						●		●		●			●	●	●
Brevet, J.			●												●		●							
Calos, M.P.													●										●	●
Caspars, P.														●										
Chan, R.K.			●						●										●					
Cohen, S.N.			●	●	●			●	●	●			●		●		●		●		●	●	●	●
Datta, N.									●				●	●	●							●	●	●
Davidson, N.		●	●					●		●		●						●			●	●	●	●
Davies, J.		●	●				●						●	●	●		●		●			●	●	●
Deonier, R.C.		●	●									●	●				●				●			●
Falkow, S.		●	●				●	●	●				●	●	●		●	●	●	●	●	●	●	●
Fiandt, M.									●								●						●	
Fritz, H.J.																●								
Gottesman, M.		●	●				●		●						●		●	●	●				●	●
Grindley, N.		●	●		●			●	●				●		●		●		●				●	●
Grinter, N.								●	●				●	●										
Hanni, C.						●							●					●						●
Hedges, R.W.								●	●					●			●					●	●	●
Heffron, F.		●	●		●		●	●	●	●			●		●		●			●		●	●	●
Heiß, B.																								●
Hsu, M.T										●		●			●						●			●
Hu, S.																								●
Iida, S.		●				●					●							●	●				●	●
Jakowski, J.B.																								●
Jacob, A.									●														●	●
Johnsrud, L.		●	●		●								●											●
Jutte, H.						●							●					●						●
Kleckner, N.		●	●	●			●		●	●			●		●		●		●			●	●	●
Kopecko, D.J.		●							●				●						●				●	●
Kondo, E.																								
Kuhn, S.																								
Lee, H.J.										●														
MacHattie, L.A.		●													●		●		●				●	●
Malamy, J.		●							●	●			●										●	●
Meyer, J.						●			●		●		●											
Miller, J.H.			●		●		●		●				●		●		●	●	●				●	●
Mitsuhashi, S.			●						●						●									
Ohtsubo, E.		●	●						●			●	●		●		●		●		●		●	●
Ohtsubo, H.		●	●						●				●		●		●		●				●	●
Ptashne, K.				●	●										●									●
Richmond, M.H.								●	●				●	●			●						●	●
Rochaix, J.D.		●							●				●					●						
Rosner, J.L.		●	●				●		●				●		●								●	●
Rubens, C.							●																	
Saedler, A.		●	●			●			●	●		●	●		●	●	●	●	●			●	●	
Sharp, P.A.		●	●				●		●														●	
Starlinger, P.		●	●			●			●				●		●	●	●	●	●				●	
Sublett, R.																								
Szybalski, W.		●							●	●			●		●		●	●	●					
Tye, B K.		●	●				●								●								●	●
Vapnek, D.		●	●						●	●		●	●		●		●		●				●	
Yun, T.																								
Zemilman, M.																								

Table 6.15 Base Literature Authors of Trend 44 that the Respondents Claim Not to Know.

Authors	Respondents (numbered as the author column in Table 6.8)																					
	5	6	7	8	9	10	11	13	15	16	23	28	29	31	35	37	40	41	49	50	52	54	55	57
Allet, B.						●		●							●		●		●			●		●
Arber, W.														●										
Bachi, B.	●	●	●				●	●		●					●		●		●			●	●	●
Barth, P.T.		●	●			●	●						●		●		●		●				●	
Bennett, P.M.			●			●	●										●		●			●	●	
Berg, D.						●								●										
Botstein, D.						●		●						●										
Brevet, J.		●				●	●	●		●			●						●			●	●	●
Calos, M.P.						●		●	●					●			●							
Caspars, P.	●	●	●				●	●		●					●		●		●			●	●	●
Chan, R.K.	●	●				●		●		●			●		●		●					●	●	●
Cohen, S.N.						●																		
Datta, N.		●	●			●	●												●					
Davidson, N.						●			●				●	●					●					
Davies, J.						●																		
Deonier, R.C	●					●		●	●					●								●		
Falkow, S.						●																		
Fiandt, M.		●	●			●		●		●			●	●	●				●		●			●
Fritz, H.J.	●	●				●	●	●		●			●		●				●			●	●	●
Gottesman, M.		●				●		●					●	●								●		
Grindley, N.						●				●												●		
Grinter, N.		●	●			●	●								●				●			●	●	
Hanni, C.		●	●				●	●	●	●					●				●			●	●	
Hedges, R.W.		●	●			●				●			●						●					
Heffron, F.						●																		
Heiß, B.		●	●			●	●	●	●	●			●						●			●	●	
Hsu, M.T.		●	●			●		●	●				●						●			●	●	
Hu, S		●	●			●	●	●	●				●	●	●				●			●	●	
Iida, S.			●				●		●						●							●		
Jakowski, J.B.		●	●			●	●		●	●			●						●			●	●	
Jacob, A.		●	●			●	●			●			●						●					
Johnsrud, L.						●			●						●				●			●	●	
Jutte, H.	●	●	●				●		●	●					●				●			●		
Kleckner, N.						●								●										
Kopecko, D.J.			●			●																		
Kondo, E.		●	●			●	●		●	●			●		●				●			●	●	●
Kuhn, S		●	●			●	●		●	●			●		●				●			●	●	●
Lee, H.J.		●	●			●	●		●				●		●							●	●	●
MacHattie, L.A.			●			●			●				●	●										
Malamy, J.			●			●								●	●				●		●			
Meyer, J.		●	●				●			●					●				●			●	●	●
Miller, J.H.					●								●											
Mitsuhashi, S.		●				●	●			●			●	●					●				●	●
Ohtsubo, E.						●								●										
Ohtsubo, H.						●								●										
Ptashne, K.		●	●			●	●		●	●			●	●					●				●	
Richmond, M.H.		●	●			●	●												●			●		
Rochaix, J.D.			●			●				●					●				●			●	●	●
Rosner, J.L.					●																	●		
Rubens, C.		●	●			●			●				●		●				●			●	●	
Saedler, A.																					●			
Sharp, P.A.						●								●					●			●		
Starlinger, P.																					●			
Sublett, R.			●			●	●		●	●					●				●				●	●
Szybalski, W.			●			●								●										
Tye, B.K.	●					●			●				●									●		
Vapnek, D.						●	●							●								●		
Yun, T.	●	●	●			●	●		●	●			●		●				●			●	●	
Zenilman, M.	●	●	●			●	●		●				●		●				●			●	●	●

Table 6.16 Details of Scientists Missing from the List of Base Literature Authors in Trend 44.

Name	Connection with Field	Year of entry to field	No. of proposals
1. Altenbuchner, J.		Late 1970s	1
2. Brown, N.		Late 1970s	1
3. Bukhari & Adhya	First book on DNA insertions		1
4. Campbell, A.	Stanford Univ.		1
5. Caro, L	F & R Factor integration	1960s	2
6. Chandler, M.		1970s & 80s	4
7. Clewell, D.		1973	1
8. Courvalin, P.	Co-worker of Davies & Brevet	Mid 1970s	2
9. Cruz, F. de la			1
10. Davis, R.	Electron microscopy pioneer	1970s	1
11. Fink, G.	Work on yeast transposons	1978	1
12. Galas, D.		1970s & 80s	3
13. Grinsted, J.		1972	1
14. Kretschmer, P.		1974	1
15. McClintock, B.	Nobel laureat (1983) - deduced existence of mobile genetic elements from genetic research with maize.	1950s	1
16. Morse & Lederberg	Isolated first IS (IS2 = gal 3)		1
17. Muger, M.	Transposable sequences		1
18. Novick, R.		1973	1
19. Reid, R.		1980	1
20. Reznikoff, W.		1972	1
21. Richards, H.	Royal P/G Med.	1974	1
22. Schmitt, R.	Worked in Bennett's Bristol lab	Late 1970s	2
23. Scott, J.	Isolated Lambda Cm	mid 1970s	1
24. Shapiro, J.	First to demonstrate IS as an insertion of DNA, also IS2 & development of a transposition model.	1970s	5
25. Sherratt, D.	Mechanisms/ models of transposition	1970s	4
26. Weidemann, B.		1972	1

Table 6.17 Details of Missing Scientists Mentioned Specifically in Connection with Bacteriophage Mu Research.

Name	Year of entry to field	No. of proposals
1. Bukhari, A.I.	1960s & 70s	3
2. Bukhari & Taylor	1960s & 70s	1
3. Casadaban, M.		1
4. Hasskey, R.		1
5. Howe, M.		1
6. Kahman, R.	1979	1
7. Mizuuchi, K.		3
8. Pato, M.		1
9. Putte, P. van de		1
10. Symonds, N.		1
11. Toussaint, A.		2

Table 6.18 Specialties in the Protein Determination Models of 1978, 1980 and 1982 that Contained Missing Scientists from Trend 44 in their Current Literatures.

Model Year	Specialty No.	Specialty Name	No. Missing authors in current literature
1978	12	Molecular Biology of Bacterial Plasmids.	8
	18	Aspects of Genetic Mapping & DNA Replication in Bacteriophage Lambda.	2
	32	Characterisation of Bacteriophage Mu DNA & its Role in Chromosomal Re-arrangements.	7
	101	Transposons Associated with Bacterial Antibiotic Resistance.	10
	110	Plasmid Replication in Bacteria.	1
	150	Killer Plasmid of Yeast.	1
	186	Sequences of Agrobacterium TI Plasmids.	1
	231	Gene Insertion, Replication & Gene Products of Bacteriophage Mu.	6
	245	Bacteriophage Lambda Integration Sites in E.Coli.	7
	320	F-Sex factor Genes & Bacterial Recombination.	1
	351	Genes & Gene Products of Bacteriophage Lambda.	1
	388	DNA Sequence of Insertion Element IS1.	1
	506	Plasmids in Gene Cloning.	2
	525	Transposable DNA Segments in Plasmids.	3
	542	Plasmids as Vehicles for Gene Cloning.	1
	673	Plasmid Vectors and Expression of Cloned Genes.	2
1980	12	Surface Plasmid Exclusion Factor of E.Coli.	1
	33	Relation of Retrovirus to Bacterial Transposons.	8
	65	Transposons and its Elements in Gene Function.	14
	81	Repressor and Cro Protein Interactions with Operator DNA of Bacteriophage Lambda.	2
	87	Use of Plasmids as Cloning Vectors.	2
	134	Genetic Elements & Transposons.	6
	213	Using Translocatable Genetic Elements such as Phage Mu to Determine Bacterial Gene Structure.	7
	222	Insertion Sequences & Transposons.	10
	224	Transposable Elements & Analysis of Repeated DNA Sequences	2
	290	Plasmids Microbiology & their Transfer of Antibiotic Resistance & Enterotoxins.	3
	330	Streptococcus Pneumoniae. Plasmid Molecular Biology.	3
	430	Sequencing DNA Cloned in Single Stranded Phage Vectors.	2
	479	Characterisation of Genetic Recombination Using Bacteriophage Lambda Cloning Vehicles.	1
	526	Bacteriophage Lambda Integration.	2
	538	Transposable Genetic Elements in Eukaryotes & Prokaryotes.	5
	543	Nucleotide Sequence & Activity of Transposable Genetic Elements in Prokaryotes & Yeast.	6
	577	Use of E.Coli Plasmid Cloning Vectors in Recombinant DNA Research.	2
	619	Antibiotic Resistance Conferred by Transposable Elements.	3
	637	Control of Plasmid Replication.	2
	648	Transcriptional Termination in Bacteriophage Lambda.	3

Table 6.18, continued.

Model Year	Specialty No.	Specialty Name	No. Missing authors in current literature
1982	16	Recognising Termination of Transcriptional Controls in Lambda Phage.	1
	63	Repeated Sequences & Transposable Elements in Drosophila.	2
	82	Characterising transposable Elements of Bacterial Genomes.	4
	90	Correlation Between Insertion Site, Frequency and Activity of Transposons in Eukaryotes.	21
	150	Genetic Analysis of Several Plasmids and Construction of RP4 Prime Plasmids.	3
	165	Transfer of DNA Segments and Plasmid Cloning Vectors in Yeast.	1
	203	Cloning & Transfer of Antibiotic Resistance Genes in Streptomyces.	1
	228	Plasmid Studies in Crown Gall Tumorigenesis.	1
	255	Comparison of Lipid Containing Bacteria & Plasmid Characterisation.	1
	256	Comparison of Plasmids & Plasmids Dependent Phages.	2
	327	Initiation of DNA Replication in E.Coli & Bacteriophage Lambda.	1
	349	Characterisation of Transposable Elements in Yeast.	2
	453	Antibiotic Resistance Genes in Pathogenic Animal Bacteria.	1
	486	Role of Integration Site, Long Terminal Repeat Sequences & Transposable Elements in Viral Transposition.	2
	531	Structure & Transformation of Plasmids Specifying Antibiotic Resistance.	2
	581	Transferable Antibiotic Resistance in Bacteroides & Streptococcus.	3
	646	Transfer RNA Structure & DNA Transfer by Plasmid Vectors in Yeast.	1
	676	Genetic Mapping & Plasmid Cloning Vectors in Pseudomonas Aeruginosa.	1
	762	Control of DNA Replication in Cole 1 & Other Plasmids.	1
	858	Use of Plasmids as Cloning Vectors.	1
	895	Repair of Damaged DNA by Polymerases, Transposition, Recombination & Replication of IS1 Flanked Transposons & Plasmids.	3
	946	Plasmid Copy Number Control Genes.	3

Table 6.19 Comparative Importance of Institutions in Trend 44.

Institution	No. of respondents believing it to be important	No. of respondents who have worked there	Names of respondents who have worked there	Collaborating respondents
1. Dept. Chem. Calif. Inst. Tech.	12 (54.6%)	3 (13.6%)	Deonier, Sharp, Hu	Szybalski, Berg
2. Dept. Bact. Royal Postgrad. Med. Sch.	10 (45.5%)	3 (13.6%)	Datta, Davies, Jacob	Rubens
3. Dept. Bact. Univ. Bristol	8 (36.4%)	1 (4.5%)	Bennett	Datta
4. Dept. Microbiol. Immun. Univ. Washington, Seattle	9 (40.9%)	3 (13.6%)	Rubens, Sublett, Bennett	Jacob
5. Mol. Biol. Lab, NIH, Bethesda	10 (45.5%)	1 (4.5%)	Rosner	Deonier
6. Dept. Medicine, Stanfoed Univ.	15 (68.2%)	4 (18.2%)	Kopecko, Brevet, Calos, Berg	Malamy, Datta, Rubens, Sharp, Hu
7. Dept. Microbiol., MIT	13 (59.1%)	2 (9.1%)	Sharp, Chan	Malamy
8. McArdle Lab, Univ. Wisconsin	10 (45.6%)	2 (9.1%)	Szybalski, Berg	Davies, Malamy
9. Dept. Molec. Biol. Microbiol., Tufts Univ.	6 (27.3%)	1 (4.5%)	Malamy	–
10. Dept. Microbiol. State Univ.	8 (36.4%)	0	–	Malamy
11. Biol. Labs., Harvard Univ.	14 (63.7%)	1 (4.5%)	Calos	Malamy
12. Dept. Molec. Bioph. Bioch. Yale Univ.	10 (45.5%)	0	–	Malamy
13. Prog. Genet. Univ. Georgia	4 (18.2%)	0	–	–
14. Dept. Microbiol. Gunma Univ.	5 (22.7%)	0	–	–
15. Dept. Mol. Biol. Univ. Geneva	10 (45.5%)	4 (18.2%)	Davies, Berg, Vapnek, Calos	Iida, Malamy
16. Dept. Med. Gen. Univ. Toronto	5 (22.7%)	0	–	–
17. Inst. Gen. Univ. Koln	15 (68.2%)	0	–	Szybawski, Deonier, Iida, Hu, Malamy
18. Dept. Microbiol. Univ. Basel	10 (45.5%)	4 (18.2%)	Caspars, Meyers, Hanni, Iida	–

Table 6.20 Missing Important Institutions

Institution	Scientists working at institution	No. of Proposals
1. Dept. Biology, Stanford Univ.	Campbell, A.M.	1
2. Dept. Microbiology, Univ. of Chicago	MacHattie, L.A., Shapiro, J.A.	1
3. Dept. Genetics & Dept. Biochem., Univ. Wisconsin	Lederberg, Reznikoff, Morse	2
4. Dept. Microbiology, Washington Univ., St. Louis	Berg, D., Hartl	2
5. Cold Spring Harbor Labs.	Bukhari, A.I.	3
6. Nat. Cancer Inst., Bethesda, Maryland	Adhya, S.L.	1
7. Univ. Libre, Bruxelles	Toussaint	1
8. CNRS, Toulouse	Chandler	2
9. Genetics, Univ. Regensberg	Schmitt	3
10. Univ. California (LA)	Miller, Galas	2
11. Inst. Pasteur	Davies, Courvalin	4
12. Microbiology, Univ. of Michigan	Clewell, D.	1
13. Microbiology, Howard Medical School	Syvanen	1
14. Biology, Univ. of Utah	Roth	1
15. App. Microbiology, Univ. of Tokyo	Ohtsubo, E.	1
16. Genetics, Univ. of Glasgow	Sherratt	2
17. Hopkins Univ.	Reed, R.	1
18. Microbiology, Univ. Calif. (San Francisco)	Craig	1
19. Public Health Research Inst., NY	Murphy, E. Novick, R.	2
20. Univ. Freiburg, Germany	Saedler, H.	1
21. Cornell Univ., NY	Fink, G.	1
22. Inst. de Microbiologie, Univ. Paris (Sud)	Brevet	1
23. Program in Microbiology, Univ. of Texas, Dallas	Clowes, R.C.	1
24. Univ. Manchester Inst. of Science & Tech.	Jacob	1

Table 6.21 Journals that the Respondents Publish In

Journal	Percent of respondents publishing in it	Inclusion in core journal set
1. Journal of Bacteriology	65	●
2. Proc. Nat. Acad. Sci. USA	60	
3. Molecular & General Genetics	55	●
4. Journal of Molecular Biology	45	●
5. Gene	30	●
6. Plasmid	25	●
7. Nature	30	
8. Virology	25	●
9. Journal General Microbiology	20	
10. Antimicrobial Agents Chemotherapy	15	
11. Cold Spring Harbor Symposium for Quantitative Biology	10	
12. Nucleic Acids Research	10	●
13. Journal Infectious Diseases	10	
14. Infection & Immunity	10	
15. Cell	10	●
16. Journal of Virology	10	●
17. Molecular & Cellular Biochemistry	10	●
18. European Molecular Biology Organisation	10	
19. Genetics	5	
20. Journal of Clinical Microbiology	5	
21. Science	5	
22. European Journal of Cell Biology	5	
23. Journal of Cell Science	5	

Chapter 7: The Validation of Trend 44 by Scientists in the Base Literature:
Part II: Data Obtained from Interviews.

Section 1: Aims and Methodology

In Chapter 6 the properties of the base literature of Trend 44 with respect to the quality of the papers involved and their associated authors and institutions were quantitatively examined. The results from the analysis of the answers to Section A of the questionnaire (see Chapter 6, Section 2) showed that the trend diagram supplied by the model builders gave a misleading picture of the development of mobile genetic elements research and of the relationships between the specialties. The scientists considered the timeframes in which the specialties occurred in the trend to be incorrect, with many of the specialties being considered to more appropriately belong to earlier time periods. As a result of this it was decided that the only way to elucidate the source of these errors in the trend was to obtain a cognitive history of the development of the field. This was considered to be important since the main attraction of trend modelling techniques to policy makers and academics interested in the history and sociology of science is their apparent ability to follow the development of scientific fields over time.

The answers to Section C (Chapter 6, Section 2) indicated the presence of a closely knit invisible college among the authors of the base literature. It was surmised that the invisible college network may have played an important role in the early development of the field. So it was decided to examine the history of the field to see to what extent the base literature authors had been vital to its development and to establish what role the invisible college had played. The information needed to answer these questions was derived from literature and oral sources.

In order to do this scientists in the base literature of Trend 44 were interviewed. Interviews were restricted to firstly those scientists who had replied to the questionnaire and secondly

to those who presently worked in the UK, France and Switzerland. The first consideration came about as a result of the need to locate scientists who had already showed that they might be prepared to allocate the necessary time, and the second because the Science and Engineering Research Council (SERC) had only allocated sufficient funds to visit scientists in these locations. The only scientist who did not fit into the latter category was Douglas Berg from Washington University, St. Louis in the USA, who was kind enough to make a detour from a conference in Manchester to be interviewed in Birmingham.

The interviews were essentially unstructured in that the scientists were allowed to talk about their work with minimal prompting. The areas that the scientists were asked to cover were; The general historical development of the field of mobile genetic elements, their own work in relation to it, expansions on any comments they had made on the questionnaire and their relationships with the scientists that they had indicated they knew well in Table 1 in the Questionnaire, (see Appendix D.3). The interviews were tape recorded with the scientists permission and later transcribed. This was found to be an efficient way of recording the copious amounts of information supplied by them. Thus face to face interviews were considered to be more effective than for example telephone interviews, not only because it would have been difficult to link a tape recorder up to the telephone system, but also because holding the interviewee's attention might have posed a problem.

A total of 9 interviews were carried out. In addition to Berg the scientists interviewed were; Naomi Datta (retired) from London, Julian Davies at the Institut Pasteur in Paris, Jean Brevet at the Universite de Paris Sud in Orsay, Peter Bennett from the University of Bristol and Christine Hanni, Jurg Meyer, Patrick Caspars and Shigeru Iida from the University of Basel. This interview data was supplemented by 3 review papers supplied by one of the respondents.

The information gained from these sources is examined in the following 4 sections. Section 2 covers the main points of the development of the field of mobile genetic elements by examining

its central ideas in a historical context. Thus this section also provides an account of the fields scientific content. Section 3 examines the role that the base literature authors of Trend 44 played in order to assess the importance of the invisible college in the field's development. Section 4 examines the additional comments made by interviewees concerning their replies to the questionnaire. Section 5 draws on information from these previous sections in order to analyse what Trend 44 has succeeded in showing of the field's structure and development. Thus it is hoped to evaluate the trend's strengths and weaknesses in this respect.

Section 2: Analysis of the General Development and Content of the Field of Mobile Genetic Elements

The 3 review papers used were C. Berg and D. Berg, 1984, (2), C. Berg and D. Berg, 1986, (3) and D. Berg and C. Berg, 1983, (4). In these papers the authors provided a general explanation as to the content of the field and its historical development, an account which was later found to be supported by the other scientists interviewed. Thus these review papers provided a general framework for the analysis with much extra detail being supplied in the interviews. These two sources of information were combined to give the account of mobile genetic elements research below.

Transposable elements can best be described as jumping genes in that they are segments of DNA which do not follow the usual rules of behaviour for DNA and so are capable of moving from one part of a genome to another in the absence of homology. They were first discovered by Barbara McClintock in the 1940's and 50's. Her pioneering studies on transposable elements in maize led her to hypothesize that their movements could be responsible for the observed changes in gene expression during development. She was awarded the Nobel prize for this work in 1983. Her work was remarkable in that it provided an alternative to the conventional view that all mutations (which are heritable changes in the DNA sequence) arise from the the addition, deletion or substitution of a few base pairs in the DNA sequence or even from the deletion of

a gene cluster. The significance of her work then lay dormant for about two decades until in the late 1960's a new class of mutations was discovered during analysis of galactose nonfermenting mutants of E.Coli. At the molecular level it was discovered that the mutations were due to insertions of DNA segments, which were called insertion sequences (IS), and it was found that the movement of these segments was a major cause of spontaneous mutation in the genome.

Then while geneticists were studying this type of mutation, another area of research arose looking for the reasons for the rapid and alarming epidemic spread of antibiotic resistance in bacteria. Many pathogenic bacteria had developed resistance to common antibiotics such as penicillin, tetracycline, streptomycin and kanamycin, and it became necessary to develop new antibiotics that often proved to be more expensive and sometimes less effective. Since the bacteria could often then develop resistance to these new drugs research into this phenomenon became a major medical concern.

Then in the mid 1970's the connection between these lines of research was made when it was generally recognised that many antibiotic resistance genes were associated with insertion elements, the resulting structures being called transposons. The discovery occurred when resistance genes were found on small DNA molecules called plasmids or R (Resistance) factors. Plasmids are pieces of DNA separate from the main bacterial chromosome which are transferable from one cell to another irrespective of species or generic boundaries, enabling the spread of antibiotic resistance to pathogenic bacteria. As well as plasmids, bacteria can contain symbiotic viruses called temperate bacteriophages or phages, the best known of which is phage lambda of E.Coli. It was the finding that resistance determinants could move from a plasmid to the lambda DNA molecule that led to the discovery that some of the resistance genes of bacterial plasmids were parts of mobile elements more complex than IS elements, or as they came to be known, transposons.

The conceptual contributions outlined above would not have been possible without the technical advances that occurred

about the same time. The technical background consisted of electron microscope heteroduplex mapping (which became generally available in the early 1970's), restriction endonuclease digestion and electropheresis in the mid 1970's and DNA sequencing in the late 1970's. In the mid 1970's the role of mobile elements as a cause of spontaneous mutations became clear, with the approximately simultaneous and independent discoveries of Tn3, Tn5, Tn9 and Tn10. The discovery of these transposons resulted in great excitement in the scientific community when their potential as research tools was realised. The use of these techniques resulted in the discovery that when a pair of IS elements move in unison they carry with them the intervening DNA. Thus it was concluded that transposons probably arise by a chance bracketing of a drug resistance gene by a pair of IS elements.

Apart from their importance in understanding the origin of certain mutations, transposons have been found to have many applications in science. Thus research into their application as tools became an important part of the later work of the field. These techniques have been described as a transposon based molecular genetics by Berg, 1984, (2), which facilitated the genetic analysis of bacterial species that had been difficult to work with using traditional genetic approaches. Lack of species specificity makes it possible to move well characterised transposons into other species, a useful tool for transposon mutagenesis studies in other genera. They have been found to be useful in genetic analysis where they are used to tag genes. They have been used in genetic engineering to mark the ends of a gene, to introduce specific genes into new hosts and to provide portable restriction sites in or next to a gene for use in *in vitro* cloning, to investigate the genetic control of metabolic pathways in bacteria and facilitate analyses of complex traits such as bacterial pathogenicity and the analysis of foreign genes cloned in bacteria.

At present about a hundred different transposable elements have been identified, which have been divided into several different families depending on their structure. For example one family is represented by Tn5 which contains a gene

encoding kanamycin resistance flanked by 2 IS*50* elements. Another family where the genes are flanked by IS*1* elements, for example contain transposons such as Tn*1681* which encodes for a toxin involved in a common diarrheal disease, and Tn*9* which encodes chloramphenicol resistance. Then there is a large family of transposons, the Tn*3* family, whose different members carry resistance to the antibiotics penicillin, tetracycline, streptomycin or sulphonamides, for fermentation of lactose or resistance to mercury. They differ from the previous 2 families in that they do not contain a pair of complete IS elements. Transposons also exist which act as vectors for characteristics other than antibiotic resistance. An example of this is phage Mu which carries genes for phage morphogenesis, and the transposable elements of yeast. As one interviewee pointed out, although yeast transposition was not one of the earliest historical events in the field it was an important development in the understanding of the mechanism of transposition. Phage Mu was worked on in the early 1970's, its similarities to mobile genetic elements only later became obvious in the late 1970's. "Super transposons" have also been found to occur in which a pair of transposable elements such as Tn*1*, Tn*5* or Mu can act in an analogous fashion to the pair of IS elements in a transposon and cause the transposition of interstitial segments of bacterial DNA.

There are 2 main models for the mechanism of transposition, conservative and replicative. In conservative transposition breaks are made at each junction of the donor DNA molecule and the transposable element that it contains, and also in the target DNA molecule. The element then moves from the donor molecule and inserts into the target molecule. In this process transposition is completed without replication of the element and the remainder of the donor molecule is lost. Then a new copy of the donor molecule containing a copy of the transposon is replicated. Tn*5* and Tn*10* appear to use this mechanism. Replicative models were developed subsequently to explain certain findings not covered by the conservative model. In replicative transposition the junctions between the donor and transposable element are nicked with one strand of DNA at each end of the transposable

element joining to the target molecule where it has previously been cut. The transposable element then replicates in situ forming a single joined molecule where the donor and target replicons are joined together via the elements (which is known as a co-integrate). The co-integrate then resolves into 2 separate replicons, the donor plus the target containing the transposable element. The Tn3 family of transposons appear to follow this model. However it was later considered that the conservative and replicative models may represent opposite ends of a continuum with the finding that transposition is often more complex than the 2 mechanisms outlined above. For example phage lambda uses a different conservative mechanism involving site specific recombination, and Tn9 and phage Mu use mechanisms which are conservative in some events and replicative in others. Insertion into the genome is often found to be non random since there may be repeated insertions of a given element at either one or a few highly preferred regions (hot spots). The complex rules by which these specific insertion sites are chosen are only just beginning to be understood.

It has now become possible to derive new transposable elements by genetic engineering, many useful derivatives of Mu, Tn5 and Tn10 having been constructed for analyses of gene structure and regulation. MudII was the first of this new generation of transposable elements.

One of the more recent developments, according to the interviewees, is in the area of *in vitro* transposition. Mizuuchi was one of the first to demonstrate it and published in about 1984 or 1985. He succeeded in isolating transposase from the bacteriophage Mu.

At present in the field there is still much work to be done on the mechanism of transposition and in refining techniques for using transposons for other purposes. Also medically, apart from continuing research into their functions as transmitters of drug resistance, there is also much interest in their possible role as carcinogens.

Thus in conclusion the whole area of the study of transposons has greatly enhanced knowledge of DNA joining

mechanisms and of the roles played by DNA rearrangements in normal and abnormal development. One interviewee said this was the reason why transposons have remained the subject of intensive scientific investigation in the last decade.

Section 3: The Contribution of the Base Literature Authors to the Development of the Field.

The purpose of this section is to show the part that the interviewees among the base literature authors and the colleagues that they were prepared to talk about, had played in the discoveries outlined above. Also included in this are some of the comments made by scientists (who were not interviewed) in letters.

One of the first major advances in the field came from a group of Germans in Cologne led by Starlinger and Saedler, who analysed IS1 and IS2. The second set of contributions to the field came from Norman Davidson's group, who made electron microscopical observation of the elements possible, and the third was that contributed by the British group led by Datta who were interested in drug resistance. These 3 different lines of research fused together about the time of the Cold Spring Harbor meeting in 1975. Also of early importance to the field were Cohen and Falkow. The various contributions of these people are discussed in more detail below.

Peter Starlinger had worked since the 1950's on the genes of E.Coli controlling galactose fermentation. As a part of this work he examined the basis of spontaneous mutations in this system with his student Heinz Saedler. In order to explain their unusual results they proposed 2 ideas, firstly that they resulted from insertion of new DNA disrupting the gene, or secondly from inversion of the DNA. On testing these ideas physically they were the first to prove the presence of extra DNA. Their contributions spanned the late 60's, the first paper on the subject appearing about 1965. Another of the base literature authors, Berg came in contact with Saedler when Saedler came to Stanford to work with Dow Woodward as a post doc around 1970, on a subject unrelated to mobile genetic elements because he was then convinced that

insertion sequences were too difficult to work with. However after a couple of years in Woodward's lab he realised that the insertion sequences that he had worked on as a student in Starlinger's lab were important and amenable to further study. Another scientist working on the area at the same time was J. Shapiro who worked on IS1 at the University of Cambridge. Berg followed his work and visited his laboratory. The combined work of Saedler, Starlinger and Shapiro led to the discovery of 3 elements, IS1, IS2 and IS3 which laid the foundation of all the transposon work done in bacteria. Naomi Datta first met Saedler at a meeting in Cologne and Starlinger on a visit to the Medical Research Council (MRC) unit on microbial genetics in Hammersmith and kept in touch for the remainder of her working career. Saedler and Starlinger are still working in the area on mobile genetic elements in plants.

Naomi Datta worked on antibiotic resistance as a medical bacteriologist at the Royal Postgraduate Medical School in London. She said that antibiotic resistance plasmids were described in Japan in 1958 but were only published in English in 1960. D. Mitchison, her boss at that time, pointed out to her in 1960 the paper about the genetics of antibiotic resistance which at first was not believed. However Datta had a collection of Salmonella strains which she examined for antibiotic resistance and found that it was transmissible. At that time her work only included the tetracycline, streptomycin and sulphonamide resistances which had been described by the Japanese. She discovered ampicillin resistance could be carried on plasmids at the same time as Xeni Kontamikalou, a scientist from Athens, in the mid 1960's. Datta and Robert Hedges, who had come to her lab as a post doc about 1970, then worked on classifying recombination relationships between plasmids. Hedges recognised that an odd sort of recombination was occurring between them in which it always seemed to be ampicillin resistance going from one plasmid into another rather than all the different recombinants that it was possible to have. They then found that a plasmid which had acquired the resistance gene from another could pass it on to a third. It was Hedges idea that this was not a straightforward recombination but a discrete movement of a piece of genetic material from one

plasmid to another. So he got Alan Jacob, who was working in the lab at the time, to measure the plasmids by ultracentrifugation (they did not have access to gel electropheresis then) to see whether the same size piece of DNA was always added to the plasmid when it acquired ampicillin resistance. This was found to be the case and Hedges and Jacob published a paper in Molecular and General Genetics where they used the name transposon for the first time to describe it. This paper was described by their rival in Bristol, Peter Bennett, as being very important, as in it the authors had made a great conceptual leap. Bennett said that Datta's group had been aware of Starlinger and Saedler's work on IS elements, who in their turn were aware of McClintock's. Julian Davies said that although they did not characterise it by DNA studies and electron microscopy there was no doubt that Hedges and Jacob had the first real antibiotic resistance transposition, the transposon concerned being Tn3. Jacob was also the first person, along with Stanley Cohen who published independently, to show that plasmids were transmissable between Streptococci. Then in 1975 there was a meeting about plasmids in California where Datta met Nancy Kleckner for the first time. Datta talked about transposition but only in the context that it occurred, whereas Nancy Kleckner and Fred Heffron had advanced to the stage where they were doing experiments on how it worked. Kleckner had not yet published her work on tetracycline resistance transposition, but she gave Datta a preprint which proved to be useful. Datta commented on the rapid explosion of interest in transposons around this time which led to many meetings all over the world. Datta retired in 1984 and is now a fellow of the Royal Society.

Another scientist associated with Datta's group was Peter Barth who was a post doctoral researcher in her lab. Jean Brevet, a scientist originally from Stanley Cohen's group, met him in a meeting, where as they were both working with Tn7 they exchanged results. Nigel Grinter also worked in Datta's department where he had initially been taken on as a technician but later did a Ph.D.

Stanley Falkow was working on the genetics of Salmonella and Shigella at the University of Washington in Seattle. Fred Heffron was working with Falkow as his student at the same time as

Hedges and Jacob were identifying the first transposon in London. Falkow invited Datta to the very first plasmid meeting that she had ever been to in 1968 when he was at Georgetown University. He performed some elegant experiments which were more sophisticated than could be done in London. Craig Rubens was one of Falkow's collaborators. Rubens also worked with Heffron and afterwards published papers on other transposons including those carrying gentamycin resistance, and was one of the first to report that gentamycin resistance was transposable. Datta met Rubens in the USA and followed his work until her retirement. Bennett (from the research group at the University of Bristol) met Rubens when he visited Falkow's lab in 1975. This came about because the leader of the Bristol group, Mark Richmond, knew Falkow and since Heffron had developed some electron microscope techniques that Richmond wanted to use, Bennett was sent to their lab. Later on Falkow came to Bristol and published a paper with Bennett. Richard Sublett had worked as an undergraduate in Stan Falkow's lab, but as a graduate student at the University of California at Los Angeles (UCLA) he only used transposons as tools, a purpose for which he still uses them today. Berg met Falkow on one of his visits to his old university at Seattle where he talked to him about Tn3.

Stanley Cohen worked at Stanford University at the beginning of genetic manipulation when he was one of the first workers in the field. His contribution was so remarkable that many of the scientists interviewed felt that he should have received a Nobel prize. Berg knew him from meetings and visits to his lab. Jean Brevet worked as a post doc in Cohen's lab, working with the transposon Tn7. Although Brevet no longer works directly on transposition, he is working in the area of genetic engineering relating to plants, where transposons are often used as tools. K.Ptashne worked in Cohen's lab at the same time as Brevet. She also worked in the same lab as Brevet's wife and visited them when they returned to Paris where Brevet now works in the Universite de Paris Sud.

Norman Davidson was important to the field as one of the contributors to the technical background. He developed electron microscope heteroduplex mapping techniques and elucidated the

structure of many of the elements carried on plasmids. Berg, when he first started out with Dale Kaiser, had worked with him. Richard C. Deonier had been a post doc with Davidson and Sylvia Hu had also worked in his lab. Eiichi and Hisako Ohtsubo, a married couple, were also post docs with Davidson. Their group was the first to sequence the whole of IS1. Berg met the Ohtsubo's in meetings when he was collaborating with another of Davidson's students, Louise Child, and Christine Haenni of the Basel group met them when she was a Ph.D. student in New York, when Eiichi Ohtsubo visited the lab to see her supervisor. Another member of the Basel group, Shigeru Iida, knew them very well, whenever he went to the US he would visit them in their lab at Stony Brook.

Philip Sharp was initially a physicist who came to Norman Davidson's lab to work on R (resistance) factor plasmids. He was part of the team that Davidson had assembled to do electron microscope analysis. This work led to the discovery that the plasmids were partially related in DNA sequence due to the transposable elements that they carried. Sharp's work also involved the characterisation of Tn10. Berg met him at a couple of meetings.

Douglas Berg started out in the University of Washington in Seattle where he first studied the classical work of Barbara McClintock. He then did a post doc with Dale Kaiser at Stanford University where he started to work with phage lambda. Tn5 was discovered and the first studies in connection with it carried out while Douglas Berg was a research associate at the University of Geneva with Lucien Caro, after he had left Stanford. The discovery came about when Berg decided that it would be useful to have lambda phages that were marked with antibiotic resistance determinants. The coincidence was that Julian Davies, who was interested in antibiotic resistance, had come from the University of Wisconsin to Geneva with various antibiotic resistant strains of bacteria. Berg obtained a couple of R factors from him and they passed lambda through these strains to try to get lambda to pick them up. They succeeded in obtaining phages that carried kanamycin, tetracycline or mercury resistance. Then Jean-David Rochaix returned from Joe Gall's laboratory with experience in

electron microscope heteroduplex mapping and Bernard Allet came back from the Cold Spring Harbor lab with restriction enzymes. So it became possible for Berg to avoid looking at the insertion of the resistance genes using genetic mapping and instead to look at the structure of the DNA using microscopy and restriction endonuclease cleavage. It became clear that the mechanism that Berg had envisaged of the phage picking up the resistance gene by aberrant excision did not explain the observed structure of the DNA of the lambda kanamycin resistance carrying phages. Berg then realised that the explanation had to be that of McClintock's transposable elements occurring in bacteria. This was the first clear illustration that Tn5 was a transposable element. When Davies and Berg submitted their paper they found that Stanley Falkow had done the same thing with Tn3. The implications of this discovery for explaining the spread of antibiotic resistance in bacteria quickly became clear. As a result of this Berg decided to abandon his work on lambda replication and concentrate on transposable elements, which he has done since since 1975. From about 1976 to the early 1980's there was an assumption underlying transposition work that all the elements were using fundamentally the same mechanism of transposition, the replicative model. However Berg found that the model proposed for replicative transposition did not fit his Tn5 data. His critique of it led to the elaboration of the conservative model of transposition. Then followed models like Galas and Chandler's which tried to bring the 2 models together in a split flow scheme. Berg continued to work with Julian Davies when he left Geneva as he went to work in his lab in Wisconsin for a couple of years. William Reznikoff who was in the next lab to them, was working on the characterisation of Tn10 and Tn5 with a student called Jorgenssen. Here Davies and Berg became interested in searching for transposons in other organisms and devised new ways of achieving this. Berg is still working on Tn5 at Washington University, St.Louis.

Lorraine Johnsrud had worked with Berg at Washington University, St.Louis where she was working on IS1, determining its DNA sequence and its specificity of insertion. After completing her Ph.D. she left the field, before her untimely death in 1983.

Julian Davies came originally from the UK. He was at Wisconsin for about 10 years (during which time he took his sabbatical in Geneva) and then was head of research at Biogen (a large industrial biotechnology company) in Geneva in Switzerland for about 7 years, where he worked on an area unrelated to transposition. Davies said that at the time that he was working with Berg in Switzerland they were unaware that Falkow was working on a similar area at the same time. However Davies knew of the work that Hedges and Jacob had done in London because he had spent some time in Datta's lab working on restriction enzymes, before he went on sabbatical in Geneva. After his move he continued to correspond with Datta. He recently left his position at Biogen to work in the Institut Pasteur in Paris, where he returned to work on transposition to try to create transposons that can be used in gram positive organisms, and to examine the application of transposons as tools. He shares his lab at Pasteur with someone who was recently a post doc with Nancy Kleckner.

Nigel Grindley was working in Bristol when Mark Richmond was the head of the department. Berg met him at a conference in Berlin (and has stayed in contact since). There was much contact between this department and Datta's lab in London, as Datta co-worked and co-authored with Richmond. Datta had known Richmond since he was working in the mid 1960's at the National Institute at Mill Hill in London, where he assayed some bacteria carrying ampicillin resistance for her. Peter Bennett worked as a post doc under Richmond looking at drug resistance in bacterial plasmids, in the 1970's when the field was beginning to open up very rapidly. The group at Bristol were working on virtually the same thing as Datta's group, but were not so far advanced and were beaten to publication by her colleagues. Bennett's group published a paper in 1976 which showed that you did not need rec (recombinant) functions for transposition, but again this was pre-empted, this time by one from Falkow's lab in 1975, whose primary author was Fred Heffron. Bennett then published a paper on the mechanism of transposition in about 1978. He said that this may have influenced J.A. Shapiro and independently D. Sherratt. Shapiro proposed a model of replicative transposition which for

the first time explained all the phenomena, and had a big impact. At the same time David Sherratt came up with a very similar model and published in the same year. Shapiro and Sherratt were not included in the list of base literature authors, which the interviewees felt to be a great omission. Bennett is still working in the area of transposition. Richmond left the field some years ago, gained a knighthood, and is now the chairman of the Vice Chancellors of the UK.

The group composed of Botstein, Kleckner, Tye and Chan were very active in the discovery and the elucidation of bacterial transposons. They discovered that tetracycline resistance was carried by a transposable element, Tn10. One of the interviewees claimed that Russell Chan discovered Tn10 without really knowing it while working with P22. Kleckner worked on the problem of plasmids derived from phage lambda for her doctorate when Berg was working on the same subject as a post doc. She discovered Tn10 at about the same time that Berg discovered Tn5, and has been in communication with Berg since 1975. Nancy Kleckner's group whilst working on Tn10 also concluded that the replicative model of transposition did not fit their data. Bik-Kwoon Tye was a Ph.D. student of Botstein's and was a co-author of the Tn10 paper. However she had only a momentary contact with the field and is now working on other things. Roth and Botstein produced a paper in the Journal of Molecular Biology that one of the interviewees said was an excellent paper because in it they were the first to collect together the possible uses of transposons and emphasise their potential. Julian Davies visited their lab where he met Botstein and Chan, and subsequently met them several times.

There was a research group headed by Werner Arber, who obtained a Nobel prize in the early 1980's for his work on restriction endonucleases, in Basel. At first the group in Basel were rather isolated, but then formed links to groups working in the same field in Geneva, Zurich and Freiburg (when Saedler came to work there). They also knew of Starlinger's group in Cologne. In 1979 they started having regular meetings with these groups, forming a regional network. Their links with US researchers came about as a result of Arber regularly attending the Cold Spring

Harbor phage meetings. Berg knew of Werner Arber in Basel, firstly because his sister Claire Berg had worked with Arber in Geneva and secondly because Arber had been a student of Kellenberger, with whom Douglas Berg had shared a lab in Geneva. Berg met Arber on his visits to Geneva and was offered a place by him at the University of Basel, which he declined because of his intention to return to the USA. Datta knew Arber through meetings.

One of the members of this group, Shigeru Iida came to Basel in 1974 as a post doc, during which time he made some discoveries about the transduction of drug resistance markers by phage P1. He stated that in this work the turning point for him came in 1975 when he attended a conference organised by NATO in Athens in Greece, the same year that the Cold Spring Harbor meeting was held which came out as the book "Insertion Sequences, Transposons and Episomes" in 1976. Many of the prominent people that attended that meeting came to the meeting in Athens. Prior to this meeting Iida had not realised that the work on IS elements and transposition was connected to his work. At the meeting he realised that he could interpret his data in these terms. Iida believed that this meeting had a similar effect on many people and that it led to more interest generally. He attended several meetings after that on transposition which had become fashionable in Europe once it had become established at Cold Spring Harbor. It was at these meetings that he met most of the European and some of the American scientists. He and Meyer worked on the structure of this phage genome and discovered that the drug resistance markers were flanked by IS1 elements. Jurg Meyer only entered the field of mobile genetic elements when he went to the University of Basel in 1976 to join Arber's research group. Meyer is still working in the area looking at physiological parameters that affect transposition, and on a new very large transposable element. All of Meyer's work related to IS elements, and since his speciality is electron microscopy he also examined plasmids under the electron microscope for the group. Meyer and Iida examined the role of IS elements in the restructuring of the genetic material of E.Coli and so became interested in the functions of IS elements. They discovered new IS elements and showed that these

mobile elements could mediate the creation of new transposons by mobilising drug resistance markers from the chromosome onto the plasmid or from one plasmid to another. They demonstrated that although the transposition frequencies were extremely low, with selective pressure over time these rare events could be found. Their first paper on this subject came out in 1978. In the same year Arber went to the Cold Spring Harbor meeting and presented their mapping data of these transducing phages. Iida and Arber considered this contribution to have been the starting point for the group in Basel. Iida did the initial experiments and produced all the transducing phages which were then later analysed, and was the most active of all the people working in the group. He has since left the group to work on plant genetics. Whereas Werner Arber was described as being the person in the group with the comprehensive theoretical background who knew all the connections between different topics, Iida was considered by his colleagues to be the practical expert. Iida described himself as having been greatly influenced by Saedler and was responsible for introducing the other members of the Basel group to him. Berg met Shigeru Iida during one of his visits to Basel and has been in close communication ever since.

Brigitte Bachi was a doctoral student when Meyer first went to Basel, and carried on as a post doc in the department, working on restriction map analysis of P1 phage. Meyer collaborated with her on this project.

Patrick Caspars was a doctoral student of Arber's working on one of the new IS elements that they had discovered at Basel, who started work in 1978. He discovered a new insertion sequence of E.Coli, IS30 which was incorporated in the P1 genome. He worked in the field until 1983 when he went to the US for 3 years. Caspars has since returned to Basel to work at Hoffman La Roche and is no longer working in transposition.

Christine Haenni was a post doc who joined Arber's group in 1976 to look at transposition. She worked with Iida and Arber in attempting to follow the transposon carrying chloramphenicol mercury resistance from the R factor to the E.Coli chromosome and then to P7 and P1. She left the field completely for a while but

has now partially returned to it in her work for Ciba Geigy. She knew Arber before she came to work in his lab in Basel, as she had met him in 1967 in Geneva where he was a professor. She only worked there for a few months before going to the states. She commented that the work at Basel was conducted very much within the group, with little contact with outsiders. She met Jeffrey Miller at a meeting in Geneva, and was very aware of his work. Another member of Arber's group, who made an important overall contribution to the research was Helga Jutte, who worked with the research group as a technician.

MacHattie and Jackowski worked on transposons mediating chloramphenical drug resistance. Meyer said that their work was related to what the group at the University of Basel had been doing and that the 2 groups were essentially working in parallel. The groups came to the conclusion that some of the transposons were flanked by repeats of DNA sequences that had previously been recognised as IS sequences at approximately the same time, with MacHattie and Jackowski publishing in 1977 and the Basel group publishing in 1978. Berg met L.A.MacHattie, who had been working with Tn9 at a meeting organized by Bukhari at Cold Spring Harbor in 1976.

Eiko Kondo, in Japan, described the chloramphenical resistance transposon although he did not realise that that is what it was at the time. Apart from this no other information was available from the interviewees about him.

J.Rosner was known casually by Berg because of his work with bacteriophage P1 and Tn9. His co-author Michael Gottesman worked with lambda phage and he gave technical assistance to Haenni when she was in New York.

Waclaw Szybalski named the IS element and applied electron microscope heteroduplex techniques to show that they were specific sequences and not just random DNA inserts. Szybalski was described by one of the interviewees as being one of the great summarizers of the field. He did a lot of original work on the control of N gene expression during phage lambda infections and how it in turn regulates the expression of other genes of lambda. He founded the journal Gene which has since proven to be a very

useful journal to the field. Datta first met him at the Cold Spring Harbor lab, but because of his fame she had heard of him beforehand. Michael Fiandt was an electron microscopist who worked with Szybalski. He did all the heteroduplex work on transpositions in that lab and carried out heteroduplex analyses for Davies and also for Michael Malamy of Tufts University in Boston who was working on IS sequences at the time.

Daniel Vapnek was known to Berg for about a decade through meetings. Vapnek was on sabbatical in Boyer's lab from Georgia when Berg met him and talked to him about IS1 and transposon sequencing. Datta knew of him through meetings.

The area of phage Mu research which had been omitted by the trend, and hence the role played by some of the scientists that had been missed, was described by the interviewees. Mu phage was discovered by Taylor in Colorado in about 1974 or 1975, but one of the first people to do much of the work in the area was Bukhari who died in an accident in about 1984. Bukhari, who worked in the Cold Spring Harbor Lab, was one of the editors of the original Cold Spring Harbor manual of transposons and episomes, and influenced many people in this area, for example A. Toussaint. Subsequently many people worked with Mu such as Malcolm Casadaban in Chicago who did some of his work on Mu in Stan Cohen's lab. Mick Chandler started his work on transposition in Geneva with Lucien Caro and David Galas on IS sequences. He is continuing his work with IS sequences in collaboration with Galas in Toulouse. Courvalin was also reported as having done a lot of interesting work recently in this area. Don Clewell found the first conjugated transposon while doing work on Streptococcus. One of the interviewees said that many of the groups who were previously working with mobile genetic elements in prokaryotes have changed their topics, with many having now moved into working with mobile genetic elements in plants.

Section 4: The Interviewees Perceptions of the Trend Data.

During the interviews the scientists were asked for any general comments and criticisms of the trend data as presented in

the questionnaire. This was done in order to clarify any of the comments made to this effect by them in answer to it, and to draw out any extra information that would help to put the trend into perspective against the background of the field that it is supposed to represent.

The interviewees generally agreed that there were gaps in the trends coverage, but the only one who provided concrete extra information was Douglas Berg who commented that the list of the authors of the base literature was a good summary of some of the important people in the field, although it had omitted many of the new workers. The Trend diagram was also very different from his experience of the field. He provided an interesting explanation of the small size of Specialty 543 in 1980. He said that it only had a few papers in it then because DNA sequencing was very difficult at that time and not accessible to many laboratories, whereas now it is considered indispensible as a basic description of transposons with many papers having been published in that area since 1980. He said that the papers in Specialty 134 were very important because at the time of the production of the base literature in the late 1970's the field was exploding with the discoveries of many transposable elements. Since at the time only half a dozen elements had been found every new description of an element attracted a lot of interest.

Section 5: Discussion and Conclusions Concerning the Comparison of Model and Oral Data

Chapter 4 showed that Trend 44 only covers a 4 year period of research in genetic elements and transposons from 1978 to 1982. However the oral accounts of the field's development (see Section 2) show that the major advances occurred in the 1960's, examples of which were the discovery of IS elements by Saedler and Starlinger and the work on antibiotic resistance by Datta. There were also conceptual inputs into the field from work done before the 1960's, notably that of the original identification of jumping genes by McClintock in the 1940's and 50's. This early pioneering work was followed by research into transposition in the early and

mid 1970's with the discovery of many transposable elements by groups as diverse as Berg's, Kleckner's and Falkow's. Thus the trend occurs too late to model the original relationships between these research areas (designated in the trend by the specialty titles), and appears instead to cover a period of consolidation or maturity in the field. This would seem to provide an explanation for the inappropriate timeframes observed for the specialties by respondent scientists in Chapter 6. The earlier formative work was covered by the trend by virtue of it appearing in the base literature of the specialties from 1978 onwards. Thus the trend shows strong conceptual links between the later, consolidating work, and the early pioneering work of the field. Section 4 showed that the base literature appeared to provide a representative list of the early workers.

Examination of the trend using information from Section 2 shows that it has succeeded in picking up the contribution of physical techniques to the field of mobile genetic elements, (for example the work of Davidson's lab on electron microscope heteroduplex analysis in the 1970's), but again only showed their relationship to the consolidating work of the field. The trend portrays continuing interest in research into IS elements, which it displays alongside specialties dealing with transposition (see Fig. 4.1), although historically the work on IS sequences preceded it (see Section 2). The continuing interest in the properties of IS elements can be explained by the activities of groups such as the one in Basel (see Section 3).

Oral data in Section 2 showed that the period of consolidation in the late 1970's and 1980's can be explained by work on the properties of known transposons and the development of them as tools for other purposes such as genetic analysis. The trend picked up the later contribution to the field of work on yeast transposition in 1980. Sections 2 and 4 confirm that although the trend showed something of the roles of the 2 "extinct" specialties of 1980, 543 and 134, these specialties remain closely involved in the development of the field and do not die out as the trend portrays.

The interview data covered in Section 3 emphasised the important role that the base literature authors and the invisible college of relationships between them, played in the cognitive developments of the field, since important ideas and techniques followed personnel between research centres. For example Berg's work on the discovery of Tn5 was linked to the work of Datta's group by Davies who was interested in antibiotic resistance and spent time in Datta's lab in London before going to Switzerland.

The results from the analysis of Section C of the questionnaire (Chapter 6, Section 2) showed the group at the University of Basel to exist on the periphery of the invisible college network. This was borne out in the interviews (see Section 3) where the Basel group was indeed found to be relatively isolated from other workers in the US, although Iida and Arber had connections with Berg, and Iida and Haenni with the Ohtsubo's from Davidson's lab. This confirms the findings to this effect from the analysis of the questionnaire in Chapter 6. Otherwise their links were regional in nature, most notably with Saedler and Starlinger.

Section 3 also showed that apart from these close working relationships there were many other professional links between the scientists through for example exchange of information at meetings and visits to each others labs. One of the most encouraging aspects of the interviews was the extensive knowledge that the scientists showed of each others work, which shows the high level of communication between them.

Thus it can be concluded that the trend base literature has picked up the invisible college responsible for the early development of the field, with the exception of McClintock and Taylor (see Section 2). The early history of the field was partly contained in the base literature, although the original relationships between these concepts is obscured by their later use in the field. The exclusion of McClintock and Taylor probably occurred because of the age of their work, with elements of their work being included in the conceptual background of the base literature authors and no longer explicitly cited by later workers. Therefore the time periods for which Trend 44 was constructed could not provide a true test of the ability of

current trend modelling techniques to follow the historical development of scientific fields. In order to test this it would be necessary to construct extra models for the decade prior to 1978.

The interviewees also emphasised the importance of the work of Shapiro, Sherratt and the Mu researchers, (who were excluded from the base literature of Trend 44 (see Section 3)), to the field. This emphasises the need to include the specialties in the Protein Determination models in which their work was identified (in Chapter 6), in Trend 44.

Chapter 8: Discussion and Conclusions

Many claims have been made for trend or time series analyses since its inception in 1976 by Henry Small (87), such as those made by L. Simon (74) that they could provide a systematic survey of changes in research fields, (see Chapter 3). The validity of some of these claims are discussed below with respect to the characteristics of Trends numbers 11 and 44, discussed in the analysis presented in Chapters 4 to 7.

Section 1: Preliminary Problems Perceived in Trend Analysis

The ability of a trend to accurately portray the development of a given scientific field is very important considering the possible policy applications of these techniques, suggested by J. Ronayne, 1983, (70) and H. Roberts Coward, 1984, (14) (Chapter 3).

Initial results in this direction seemed promising such as in the study by Small, 1977, (81) where on examining the literature of collagen research from 1970 to 1974 it was found that the map of clusters of collagen research over time corresponded well with historical data supplied by collagen workers in questionnaires. In this study Small used annual accumulations of data from the SCI, a standard practise in CCA as this amount of data is needed to make the techniques statistically viable. However Sullivan et al, 1979, (90) in their study of the field of weak interactions showed that important theoretical changes often occur on a monthly time scale and so concluded that annual CCA plots may obscure the dynamics of theory change. This could have implications for trend studies as a whole which often use annual plots, although it did not appear to cause any problems in Small's study.

These observations had immediate implications for the validity of the trends presented in the ABRC study, (see H. Roberts Coward, 1984, (14)) from which the trends analysed in this present study were drawn, since the CCA data for this study was

only plotted every 2 years. Sullivan's observations would thus predict that much of the intellectual history of the fields that the trends portray would be obscured. Another possible problem that could add to the inaccuracy of trends was that identified by the CRP, 1983, (19) where they criticised the criteria of continuity that is used to generate trends. However in spite of the problems in a trend's accuracy that would be expected to result from these difficulties, the research council scientific experts that were consulted in the validation of the trends in the ABRC Science Policy Study concluded that the trends accurately reflected the intellectual evolution of scientific activity in the fields concerned.

Whether this conclusion was valid or not was examined in the present study where it was reasoned that if the trend modelling methodology was capable of capturing the progress of the scientific fields represented by Trends 11 and 44 then cognitive consistency and social continuity (perhaps in the form of an invisible college) should be found in the literature that they contain. The large number of papers in the current and base literatures of the 2 trends meant that an in depth analysis of the type proposed was only feasible for the base literature. The base literature was selected for this purpose in preference to the current because it is at this level that links between specialties in different time periods are drawn, and so it was reasoned that it is at this level that the criteria of continuity currently employed by the model builders needs to be tested. In order to achieve this the analysis started with an in depth examination of the base literature papers of the 2 trends in Chapters 4 and 5. This was followed up in Chapters 6 and 7 by a further check on these qualities in the case of Trend 44 with the addition of the question of the historical accuracy of the trend from interview and questionnaire data supplied by the scientists involved. The implications for trend modelling of the results obtained from this analysis are examined in detail in the sections below.

Section 2: The Cognitive Validity of Trends 11 and 44

Firstly the cognitive validity of the individual specialty base literatures of Trend 44 was examined in Chapter 4 by comparing the titles of the papers with each other and also with the titles given to the specialties in the CCA models. The aim of this was two-fold as firstly it was reasoned that if the base literature papers were cognitively related this might reasonably be expected to be reflected in their titles, and also since the specialty titles are derived from a keyword analysis of the current literature paper titles it would be interesting to see whether it was also applicable to the base literature. The first aim was designed to test one of the basic assumptions in CCA that the current literature co-cites papers that it deems to be related in some way (which is the foundation upon which the whole edifice of co-citation theory is built, see H. Small, 1973, (84)), and the second was investigated as a potential indicator of rapidly advancing research fronts in science. Presumably if the research front (which according to CCA theory is represented by the current literature) is advancing rapidly it should leave the base literature intellectually behind, which might be indicated by a change in the wording of the paper titles. It might also indicate a completely new field drawing on an older one using for example established techniques.

Marked subject similarity was found between the specialty base literature papers, and also between them and the specialty titles, in Trend 44 by this method. It was thus concluded that the co-cited papers in the base literature were cognitively related as predicted by CCA, and that there was strong cognitive continuity between the base and current literatures. Only one of the specialties examined in Chapter 4, Specialty 320, showed no obvious continuity. When this problem was presented to the scientists who had worked in the field in the questionnaire, see Chapter 6, the majority agreed that the match was not good. However one scientist provided an interesting explanation for this anomaly saying that all of the base literature papers were produced by the same group and that they were so cautious with

their paper titles that it would be impossible to guess that the papers were in fact on the same topic as the current literature. Thus it was concluded that the specialty titles were a reasonable indicator of the content of the base literature. This is in itself quite an important result since Noma, 1984, (59) said that there was a problem of an assumption in citation analysis that all citing articles view the literature from a common point of view. It would seem that in Trend 44 this is not a problem, but that they do actually share a common viewpoint, since there is obviously a shared area of interest between the two sets of papers. Since there was so much similarity between the base and current literatures it was not possible to test the idea that a difference in the paper titles might indicate a rapidly advancing research front, but it is still a possibility that could be explored in future studies. Although the finding in Specialty 320 shows that there is danger in assuming that any difference found always indicates this.

Because the base literature papers are grouped together in CCA by the citing actions of the current literature papers, it was decided to see whether the base literature authors perceive cognitive links to each other or whether they are unaware of the papers that they are grouped with. The obvious way of answering this question was to see whether they cited their co-cited partners. This was examined for Trends 11 and 44 in Chapters 4 and 5 where a high incidence of citation between papers grouped in co-cited pairs was found. Citations were also observed between papers that were not directly paired together. It was therefore concluded that the base literature papers show cognitive debts to each other that pervade the whole base literature of a specialty but are particularly seen in the co-cited pairs. If the assumption in CCA that the base literature papers are authored by prominent people is allowed, then it can be assumed that their authors are the members of the central groups described by Chubin and Studer, 1979, (9). Thus it can be argued that this finding of the high level of citation between the base literature papers supports Chubin and Studer's finding that the members of the central group cite each other at a high level. This finding also supports the

hypothesis in co-citation theory that co-citation can show the relationship between ideas in the base literature (H. Small, 1973, (84)). However the implications of this are even more far reaching because the assumption in CCA is that they are cognitively related from the point of view of the current literature, whereas these results show that they are cognitively related also from the point of view of the base literature. It thus seems possible that the relationships shown by CCA between papers are primarily perceived by the base literature authors and then picked up by the current literature authors (probably by their examining of base literature papers reference lists for other related literature). This would most likely be the case in mature scientific fields such as Trend 44 was shown to be in Chapter 7, where the connections between topics are well understood. In very new fields where the current literature has linked previously little related topics together in the base literature, the base literature papers would be expected to have little knowledge of each other. Possibly this could be combined with the previous idea of an indicator of cognitive dissimilarity to indicate very new fields.

A claim made by Henry Small, 1973, (84), that the frequently cited papers represented key concepts, methods and data was tested in Chapter 6 where this aspect was examined for the base literature papers in Trend 44. In the questionnaire the scientists that had produced the base literature were asked whether they would describe any of the papers as key paradigm and also how well known and how good they considered the papers to be. Once favourable reports by scientists about their own work had been excluded from the analysis, it was found that for each specialty, apart from one paper in Specialty 82, all of the base literature papers were selected as key paradigm. One paper was praised so far as to say that it had been vital for the area concerned. This finding was supported by the result that for every specialty the papers were mostly considered to be good and well known, and also by the conclusion in Chapter 7, where oral and review accounts of the history of the field were examined, that the papers were often milestones in the fields development and that the authors of these papers played important roles. This

substantiates the finding in Small, 1977, (81) where it was shown that the highly cited documents in the collagen trend were the significant papers from the point of view of the workers in the current literature. The results for Trend 44 show that the papers were already considered to be significant in the point of view of authors other than those that had produced them in the base literature. So it would appear that again the current literature only reflected judgements that had been previously made by earlier researchers, which could again be characteristic of mature fields. However Chapter 6 also showed that many important papers were missing from the base literature of Trend 44. These were proposed by the scientists to cover perceived gaps in the coverage of their field by the trend. This is dealt with more fully in Section 4 of this discussion.

Another aspect of the base literature papers of Trend 44 that was examined in the questionnaire was their experimental or theoretical characteristics. This issue was raised in connection with the studies by Sullivan et al, 1977, (91) and D. Hicks, 1987, (39) where both observed that experimental papers were under-represented compared to theoretical papers in the co-citation clusters that they examined. Hicks found in her study of a spin glass cluster that experimental papers were consistently under-represented in the clusters in relation to their share of highly cited papers. However in Chapter 6 when the scientists were asked how experimental or theoretical they considered the papers to be they indicated that the papers did not fit easily into one category or another, but that many had elements of both. However averaging the scores out showed that the papers were mostly considered to be experimental. This supported the observation by the author of this thesis in the initial reading of the base literature papers of Trend 44 for their analysis in Chapter 4, that the papers reported extensive experimental results which were then used in the same papers to discuss theoretical questions. So it would appear that experimental papers are not under-represented in at least the base literature of Trend 44, in contrast to these other two studies. This highlights a danger in assuming that the differences between experimental and theoretical work may always

be as clear-cut as they apparently are in the fields of spin glass (Hicks) or weak interactions research (Sullivan et al). The interesting qualities that these two categories can confer onto a paper was illustrated in Specialty 525 in Trend 44, where a paper, although it was selected as key paradigm by all of the respondents, was considered to be excellent theoretically but poor experimentally.

Thus it was concluded that the specialties comprising the trends examined in this study were at least cognitively consistent, and that many of the basic assumptions of CCA did seem to apply. The next stage was to examine the properties of trends as a whole.

Section 3: Investigation of the Coverage of the Activities of an Invisible College by Trend Modelling.

Price, 1965, (63) talked of there existing informal networks of scientists with a common interest which he called invisible colleges. He claimed that these networks were limited to about a hundred members, this being the limit at which he believed them to be sustainable, and that they were composed of important people who sent each other pre-prints, met at select conferences and collaborated as co-authors. Later in 1973, (84) Small also claimed that the highly co-cited papers in specialties are produced by members of specialties or invisible colleges. N.C. Mullins, 1977, (55) in looking at the group structures of co-citation clusters found that the authors of the clusters showed dense patterns of contact and concluded that his data supported this claim. Then Small, 1977, (81) looked at a trend analysis of collagen research and found that the research area was composed of a small group of people with very close informal communication and frequent exchanges of personnel between a small number of research centres. From this he concluded that CCA had identified a research specialty with the attributes of an invisible college.

It was thus decided to see for the present study if any group of people with the patterns of informal contact attributable to an invisible college could be identified in Trend 44. The

search for the presence of an invisible college started with an analysis of the base literature papers of Trend 44 in Chapter 4, with a search for shared co-authors among the papers, acknowledgements to other authors, and shared institutional affiliations. A high level of co-authorship was found in most of the specialties, although this varied from Specialty 388 where there was none, to Specialty 320 where all the papers were produced by the same group. Analysis of institutional affiliations supported this, for example in Specialty 320 the group that produced the papers were all at the same institution. When the patterns of acknowledgements and citations between the base literature papers were compared, a high level of citation of people that the authors acknowledged for assistance was found. These results were taken as being indicative of close working relationships.

Also in Chapter 4 the occurrence in the current literature of the authors that had been found to have been associated with the base literature papers was examined. All of the specialties were found to have base literature authors working in the current literature, which provided further evidence for continuity between the cognitive bases and research fronts of each specialty. It was subjectively observed that although there were many papers in the current literature that were authored by researchers other than those in the base literature, there was much interaction between these two groups in the form of co-authorship in the current literature.

From this data it was concluded in Chapter 4 that some of the relationships attributable to an invisible college had been observed amongst the authors of the base literature, and that it appeared to survive into the current literature, with many new members being drawn into the network. However it was also concluded that more information would have to be obtained before the existence of an invisible college was deemed to be certain.

So in Chapter 6 this question was further examined in the questionnaire that was sent to the base literature authors, in which they were requested to indicate their relationships to the other base literature authors. This also provided the opportunity

for examining the sub-structure of the invisible college in the form of its various groups. The use of questionnaire and interview data to obtain reports from the specialty participants in this study was in sympathy with techniques used by sociologists of knowledge, see Chubin, 1985, (8).

The replies showed a very high level of awareness among all the scientists of each others work in the literature, and extensive co-working links. Only four researchers were observed to exist outside this co-working network completely. One group, the researchers at Basel, was observed to be peripheral, since their only co-working link was through the activities of one scientist. An interesting finding was that the co-working links were not necessarily reflected in the patterns of co-authorship shown by the scientists. As was expected not all co-workers published together (co-working may only indicate practical assistance), however it was unexpectedly found that not all those who published together had worked together. These findings illustrated the meaning of Sir Mark Richmond's comments in his letter to the author that the published literature was but a poor reflection of activities in the field of mobile genetic elements. Many mentor relationships were found, although predictably due to the very close personal nature of this type of relationship, only isolated groups which did not fit into a network were found. Institutional links between scientists were found to form a complex network reflecting their high mobility. The Basel group were observed to be also relatively institutionally isolated, a fact that was also observed for the four scientists who were excluded from the co-working network. The patterns of exchange of technical assistance also showed that most of the scientists fell into one large network, with the Basel group again being relatively self contained with links to the main network only occurring through one scientist.

The relative isolation of the Basel group was later confirmed in interviews with them (See Chapter 7). However the four scientists that were apparently not included in the invisible college network might well have proved to be members had more questionnaire replies been obtained.

That the invisible college had been well represented by the base literature was shown by one of the scientists interviewed in Chapter 7 who regarded the base literature authors as being a good list of the early workers in the field. The interviews also showed that the members of this invisible college were responsible for many of the early developments in the field of mobile genetic elements. This would appear to confirm Small's observation, 1977, (81) that the authors of the highly cited papers are the leading researchers in a field. The interview data demonstrated how cognitive developments in the field closely followed the interactions in the invisible college network, with for example the exchange of ideas at meetings and the transmission of technical expertise with the movement of personnel. The importance of the invisible college in the field's development was emphasised by the scientists beliefs that these personal links were the driving force in the field, with the literature acting more as an archive.

From this data it was concluded that the trend base literature had identified a group of scientists with the attributes of an invisible college that were responsible for the early development of the field. This raises the interesting question of whether this is a general feature of the base literature of trends produced according to Small's methodology, (a possible subject for further study). However the data did appear to support the conclusions of Mulkay et al, 1975, (54) who said that the creation of scientific knowledge involves a complex web of social relationships. The results in Chapter 6 showed that the finding by Chubin and Studer, 1979, (9) in biomedicine that most of the highly cited authors in the network had worked together with a high degree of co-authorship and connectivity, was also true for the highly cited authors in the base literature of Trend 44. Chubin and Studer also observed circulation of researchers around elite groups in biomedicine. This was also observed in Chapter 7 for Trend 44, the elite groups here being headed by such prominent researchers as Starlinger, Datta, Falkow and Davidson.

The findings in this section seem to refute the criticisms of CCA made by Noma, 1984, (59), who said that since

all the articles do not necessarily view the literature from the same point of view, the citing articles should be limited to those written by the members of an invisible college. This could be criticised by reference to P. Lenk, 1983, (44) who concluded that there was a relationship between the invisible colleges and the researchers citatory behaviour. The findings in this thesis suggest that the scientists in both the base and current literature of Trend 44, do form an invisible college.

Section 4: Analysis of Trend 44's Portrayal of the History and Development of The Field of Mobile Genetic Elements.

Small, 1977, (81), examined the literature of collagen research from 1970 to 1974 and found that the map of the CCA clusters of collagen over time corresponded well with historical data supplied by collagen workers in questionnaires. This result supported Small's view that trend modelling or as he termed it, cluster strings, could follow the development of a field by linking specialties that share the same cognitive base. Trend 44 was produced in accordance with this hypothesis, and its ability to accurately reflect the development of a field is even more important concerning its role in a science policy study. Thus it was deemed important to know the accuracy of its portrayal of the history of the field of mobile genetic elements.

The history of the field was examined in 3 ways. Firstly in the questionnaire with respect to how well the Trend Diagram supplied in the Protein Determination Models corresponded to the scientists' perceptions of the development of the field; secondly from information contained in review papers and thirdly from oral histories obtained from the scientists themselves.

The questionnaire replies showed that the links between the specialties in the trend diagram were only deemed to be partially correct in the majority of cases. The main objection was that the timeframes in which the specialties occurred were incorrect, with many of the specialties being moved to earlier time periods by the respondents. There was also disagreement with the trends portrayal of the demise of Specialties 543 and 134 in

the 1980 time period with opinions being expressed that they continued into 1982. The historical validity of the trend was further criticised with respect to whole areas of the field that had apparently been overlooked, such as work on bacteriophage Mu which was said to be important to the field's overall development.

The applicability of the time span of 1978 to 1982 covered by the trend to the field of mobile genetic elements, was examined in the questionnaire by asking the scientists the first year that they came to know of their colleagues. Apart from the few who went back to the 1950's and 1960's, most of the scientists became known to their colleagues in the early 1970's. It was assumed in the analysis of the questionnaire that this was probably about when the invisible college developed and the field gained momemtum. It was concluded that if this was the case then since the first year that the trend models is 1978, it would appear that the field was modelled too late to portray the major developments. It was therefore postulated that the specialties shown in the trend represent ongoing areas of research, whose origins go back to the early 1970's. It was considered in Chapter 6 that if this was the case that it would account for the scientists disagreement with the historical validity of the trend.

Then in Chapter 7 these suspicions were confirmed against the background of the cognitive history of the field gained from the interviews and review papers. This data showed that apart from a few discoveries pertinent to the field in the 1940's, 1950's and 1960's, the major cognitive developments of the field occurred in the early 1970's. It was surmised from this that the trend covered a period of consolidation in the work of the field, and that the most striking historical developments occurred much earlier. Thus it was concluded that the trend diagram was not as historically inaccurate as the replies to the questionnaire would indicate, but that it models the field too late for the original relationships between the developments to be shown.

It is likely that as the correspondence of the trend to the history of the field was assessed by the base literature authors who were the early workers, the portrayal of the consolidation period of the field by the trend might be better

examined in future studies by interviews with scientists who had worked in the current literature in that time period. The correspondence of the trend to the historical data supplied in this thesis could also be better tested by building additional models to cover the 1960's and 1970's. Thus although the findings here do not refute Small's claim, further analyses of Trend 44 along the lines suggested would be needed in order to fully support it.

However the trends inaccurate portrayal of the demise of Specialties 543 and 134 could not be vindicated since the interviews showed that they remained an integral part of the fabric of the field. Similarly the involvement and importance of the areas omitted was shown. The other main feature of the trend is the apparent decline of the field in 1982, where it contracts down to one specialty. However analysis of the history of the field in Chapter 7 suggests that research into these genetic elements in their role as tools etc is still a going concern in 1988. The possible origins of these errors relating to the coverage of the field of mobile genetic elements in the trend modelling methodology, are examined below.

Section 5: The Sources of Problems in Trend Modelling, and Their Possible Solutions

The errors in Trend 44 discovered in the analysis of the questionnaires and interviews in Chapters 6 and 7 were examined in relation to inconsistencies discovered in the literature analysis of the trend in Chapter 4.

The discoverer of CCA, Henry Small in developing his trend modelling methodology observed in 1977, (81) that as the span of time increases, the number of surviving papers in the base literature decreases, since not one paper was observed by him to survive over 3 years in collagen research. A similar phenomenon was observed in the base literature of Trend 44 in Chapter 4, where it was observed that only one paper survived through all 3 time periods. Since the time periods used in the Protein Determination Models were not consecutive as in Small's study, the

survivability of this paper was at first sight encouraging since it had survived over a longer time period of 5 years. In Chapter 5 the survival of base literature papers in Trend 11 was found to be even better since a total of 9 papers survived the 5 year period. However it was considered that the effect of the decline of the number of surviving papers could have a serious effect on the ability of trend modelling techniques, as they stand, to follow the development of research areas over any appreciable time period. This phenomenon would seem to cast doubt on the currently used criteria of continuity which is totally dependant on the survival of base literature papers.

In Chapter 4 the extent of the effect that this phenomenon might have had in the production of the errors of omission observed in Trend 44 was examined by analysis of the papers linking the various time periods. It was observed that the trend specialty in 1982 was only linked to the 1980 time period by this one surviving paper. Also the 2 specialties that apparently became redundant in the 1980 time period, Specialties 543 and 134, were observed to have been included in the trend on the strength of only one paper each. The rather narrow margin of inclusion of these specialties could explain why the trend appears to show a decline of the field of mobile genetic elements as it contracts down to one specialty (perhaps many specialties were excluded) and the apparent demise of Specialties 543 and 134, which according to the scientists in the questionnaires and interviews (Chapters 6 and 7), should have continued into 1982.

It was concluded from this that the criteria for determining the continuation of specialties in trends may be too narrow. This was also the conclusion of a study conducted by the CRP, 1983, (19) who found from looking at a time series view of ceramics research that the assumption that the disappearance of the intellectual base always signals the departure of a specialty may be wrong. They believed that it could be due to transient changes of focus, and therefore recommended the broadening of the criteria of continuity to include any time slices, not just adjacent ones in order to detect a papers re-appearance. Support for this view in the present study was found in the analysis of

the base literature papers of Trend 11 (see Chapter 5), where one base literature paper was observed to appear in another time period after a gap. This was the paper by M. Silverman which appeared in 1978 and 1982.

There is another possible source of error in the trend that it was considered had to be investigated before proposals for altering the criteria of continuity could be proposed. The Protein Determination Models were ordered by the MRC in the ABRC Science Policy Study to study techniques concerned with elucidating the structure of proteins. Since this type of information is central to the understanding of biochemistry and molecular biology, the CCA models produced for it turned out instead to be general models of these research areas, see Bell, 1984, (1) and Coward, 1984, (14). Thus the models were not specifically targeted on the field of mobile genetic elements covered by Trend 44. This is in contrast to the work of Small, 1977, (81) where the CCA data was targeted on the field of collagen research. The core journal sets originally used in the construction of the models in the ABRC study, supplied the papers that formed the current literature, so although omissions of relevant journal literature might occur at this level, their citations allow the inclusion of missing journals in the base literature. Therefore in order to see whether any of the relevant literature had been excluded from the Protein Determination models, the journals that had been found to be associated with the base literature papers were searched for in the core journal sets. In Chapter 4 it was discovered that two thirds of the journals identified in the base literature were included in the core journal set. However this left a third that were not included, which by their very inclusion in the base literature could mean that they are important journals in the field. This question was further explored in Chapter 6 in the questionnaire where the scientists were asked which journals they normally published in. It was found that the journals they most commonly published in were in the core set. It was concluded from this that although the coverage of the field of mobile genetic elements in Trend 44 might be improved by the construction of CCA models including these journals in the core set, it would be

unlikely to make any significant improvement. Thus any errors in the trend were unlikely to have occurred at this level in the CCA models, and any improvement that is needed, needs to be introduced into the basic trend modelling methodology, i.e. into the criteria of continuity.

In Chapter 4 further possibilities for expanding the criteria of continuity were explored. The first possibility was investigated as a result of considering Derek de Solla Price's belief that research builds from the skin i.e. that new work draws intellectually mainly from recent papers. In the context of trends it was hypothesised that this could mean that one time periods current literature papers become the intellectual foundations of the current papers in the next. On examining the origins of the newcomers to the base literatures of specialties in Trend 44 it was discovered that some of them originated in the current literatures of previous specialties in the trend. Although time was not available for a search for the origins of all the base literature papers in Trend 44 in the current literatures of other specialties throughout the Protein Determination Models it was considered that such links would probably be found if such a search was made. Therefore it was suggested that if the present criteria of continuity was altered to include searching for this continuation of papers between models, that current literature - base literature links could be forged between specialties in different time periods, as well as the more usual base literature - base literature links. Using this technique could help overcome the problem with the criteria of continuity currently used, of the apparent disappearance of specialties that in fact remain at the research front but are disguised by a new set of base literature papers.

Another technique that was tried in Chapter 4 for linking specialties in trends together was the re-appearance of authors that had been found to be associated with the base literature in different papers in other time periods. This is rather like searching for other papers in an oeuvre, or body of writings by one person, as it is considered likely that they would be intellectually related. When this search was done among the base

literature papers of Trend 44 in Chapter 4 it was discovered that some of the authors did re-appear in different base literature papers in the separate time periods. That this technique could be potentially useful was verified in Chapter 5, where many extra such author links were found in Trend 11. Indeed here extra links were found between specialties not directly linked by the trend.

Another possible technique was discovered in searching for the important workers in the area covered by Trend 44 that the respondents to the questionnaire said had been overlooked. In Chapter 6 the names given by the respondents were searched for in the models and were found to be connected to a total of 192 specialties. Predictably some of them were found to occur in the current literatures of specialties in Trend 44. However when the keywords in the titles of the specialties in the trend were used to analyse the titles of the specialties identified by this method, it was discovered that a total of 51 specialties over the 3 time periods were closely related to the subject matter of the trend. These newly identified specialties were found to cover the subject areas that the scientists in the questionnaires and interviews of Chapters 6 and 7 said had been omitted. So it would seem to be quite important to include this in the broadening of the definition of the criteria of continuity, by identifying key words in the titles of specialties linked in the usual manner and then searching for these specialties in other parts of the models. Linking these specialties into the overall trend would then presumably result in a more complete coverage of the fields concerned. For example in Trend 44 use of this technique would have identified the contributions of Shapiro, Sherratt, Galas and Chandler who were said to have been vital to its cognitive history. This approach to placing a trend in its broader cognitive context might fit very well with an approach with a similar aim in the ABRC Science Policy Study, 1984, (14), where the trends were considered with respect to the regions that they passed through.

It was concluded from this analysis that the answer to Coward et al's, 1984, (14) question of what is the explanation for the specialties that completely pull up their roots, is that they probably do not. It is more likely that the present criteria of

continuity used to generate trends does not pick up their continuance.

Another problem with Trend 44 that was identified by the scientists in the questionnaire (Chapter 6) was that they did not perceive its component specialties as always being distinct entities, since much overlap was seen between them. For example the supposedly defunct specialties of 543 and 134 were seen as overlapping greatly with continuing specialties. This viewpoint was supported by the finding in Chapter 4 that Specialties 543 and 134 shared authors with Specialty 222, the continuing specialty in 1980. It is thus possible that elements of the supposedly defunct specialties may have been incorporated in continuing areas of research in this fashion. Certainly the subjects covered by these 2 specialties continued to play a role in the field after 1980, as was shown by the interviews in Chapter 7. The scientists observations of the lack of discreteness of the specialty areas was supported by the observation in Chapter 4 that the authors of the base literature papers were often observed to appear in the current literatures of specialties other than their own. The only positive finding for what the trend shows of these specialties was that the small size of 543 was accounted for by the inaccessibility at the time of the techniques that it represents.

It was concluded from this information about the lack of discreteness and the high interactivity between specialties that the overall trend may be a more important unit of analysis than its component specialties, providing a relatively stable framework. The authors activities in these specialties show that the distinctions between them are somewhat blurred, and that perhaps the trend represents the overall research interests of a group of people in an invisible college. This would mean that too much significance should not be attached to the mortality of specialty clusters, since it appears that they can be absorbed into the mainstream of research in ways not captured by the current criteria of continuity used in trend modelling.

The data contained in the Trend Summary Report of Trend 44 that was supplied in the models was examined in relation to the patterns of birth and mortality of specialties shown in the trend.

This data when examined in Chapter 4 was found to apparently reflect the major conclusions of the trend with respect to the discontinuation of 2 specialties and the overall contraction of the field in 1982. The Trend Summary Report showed that the 3 specialties concerned had lower current activity levels and higher base literature ages than for example the continuing specialty in 1980. However since the data obtained from the scientists concerned in the field in Chapters 6 and 7 disagreed with these conclusions, doubt must also be cast on the validity of the data supplied in the Trend Summary Report which supported it. The model builders propose using these types of indicators to develop theories about specialty development, and also for evaluating scientific research for policy purposes, since they purport to show something about the properties, development and behaviour of specialties in trends. The findings in this thesis would suggest that more work needs to be done before they can be used for these purposes.

Another problem associated with data presented with Trend 44 was found to be that concerning the importance of institutions. It was discovered in Chapter 4 that in Trend 44 no institutions survived independently from its continuing base literature papers through all 3 time periods. Since important institutions could be expected to show their continuing importance in this manner the suspicion arose that their importance may have been limited. This suspicion found supporting evidence in the questionnaire in Chapter 6, where it was discovered that all of the institutions had been important in the field, but that their importance was dependent on the activities of a few people, who were said to be highly mobile. This was supported in the study in the difficulty in tracking down the base literature authors. It can be concluded form this that not too much significance should be attached to the ranking of institutions in the models, since these lists would seem to need constant updating. One respondent pointed out that this would particularly be a problem with US institutions, where mobility of personnel is very high. Chapters 6 and 7 also showed a high level of migration of scientists between the different countries involved. So taking this into account and bearing in

mind that national affiliations are gained from the institutional locations, it can be concluded that the national rankings should also be regarded as inconstant and updated accordingly.

Section 6: General Conclusions With Respect To The Validity of Trends

It was concluded from the analysis of data pertaining to Trend 44 that the trend did not portray the development of the field of mobile genetic elements as the scientists who had been involved in its base literature perceived it. This was apparently due to the fact that the major cognitive advances had occurred before the earliest time period of the trend (1978) in the 1960's and early 1970's. Trend 44 instead appears to cover a period of consolidation and maturity in the field. So these findings do not necessarily contradict Small's, 1977, (81) conclusion that trends correspond well with historical data. It is probable that as the period covered by Trend 44 was one of consolidation, historical progress might not be so clear cut, and that in order to properly test this aspect of trends against the history of mobile genetic elements, it would be necessary to construct extra models to cover these earlier time periods.

Apart from this problem of the less than optimally placed time periods, other problems relating to the current trend modelling methodology were identified. The problems were firstly the omission of whole research areas relevant to the field from the trend, and secondly incorrect information conveyed by it about the development of the field, with respect to the apparent overall decline in 1982, and the demise of 2 specialties in 1980. The field apparently continued beyond 1982 with the 2 supposedly defunct specialties continuing to play a role.

These errors and omissions were found not to originate at the level of the journals used to build the CCA models, but appeared to lie in the techniques used to produce trends in CCA models. The problems were found to lie with the criteria of continuity used by Henry Small, 1976, (87), of searching for shared base literature papers, reflecting criticisms originally

put forward by CRP, 1983, (19). Evidence was found in the analysis of Trend 11 to support CRP's suggestion that the present methodology should be modified to search for the re-appearance of papers in any time slices, not just adjacent ones. This study also advocates 3 other modifications, which should be used in addition to the conventional criteria of continuity.

(1): To search for continuity between papers in the current literatures of one time period and the base literature of the next.

(2): To search for the names of authors of the base and current literatures of a trend generated in the conventional manner in other specialties in the models.

(3): To identify key words from the titles of the specialties of a trend generated in the conventional manner, and search for other specialties in the models containing them.

It was discovered that use of these modifications eliminated the anomalies of lack of coverage of certain areas and the false demise of Specialties 543 and 134 in Trend 44. It was considered that it was also probable that regenerating the trend along the lines suggested would show more activity in the 1982 time period, when the area supposedly declined.

It is appreciated that the use of these modifications would entail the need for considerable computer expertise. It is believed however that if this was found to be technically and economically viable, that a more realistic picture of time series development of scientific fields would result. It is likely that a very large number of specialties would be drawn into trends in this manner, some of which might be peripheral to the main interests of the field. This might suggest a function for scientific experts who could eliminate these to make the resultant trends more discrete. The trends might then be combined with regional data from the CCA models as was done in the ABRC study, see H. Roberts Coward, 1984, (14) to place them in a broader context.

It is suspected that the present view of the relatively isolated development of scientific fields as portrayed by the

conventional trend modelling methodology might disappear to be replaced by an interacting network of the advance of the research fronts of entire disciplines or subdisciplines. I feel that this would be a more realistic time series view of science.

Also there were perceived to be problems with the validity of some of the indicators presented with the trends. The institutional and national ranking data was found to be affected by the high mobility of the scientists concerned. However this was not considered to be a serious problem in these two cases, since it just emphasises the need to use the most up to date data available for any relevant policy decisions. However the indicators used to describe the development potential of specialties, such as the size of the current literature and the age of the base literature, were found to be more seriously in error. In Trend 44 information derived from these indicators had appeared to support patterns of decline and mortality that were found to be at odds with oral accounts of the fields development. Thus the claims by for example L. Simon, (74) that time series gave a systematic survey of research fields and that analysis using these types of properties might lead to the development of a general theory of specialty behaviour, will have to be re-assessed. I suggest that trends should first be generated along the lines of the modifications outlined above, and the significance of indicators such as the age of the base literature, re-assessed in the light of the new patterns of continuity found. This would hopefully clarify their meanings, and hence lead to theories of specialty development.

However there were also some positive results supporting Small's claim, 1977, (81), that trend modelling in CCA had identified a research specialty with the attributes of an invisible college. This was found to be the case in Trend 44. The results also support another claim made by Small in the same paper that the highly cited documents in the trends were the significant papers from the point of view of the current literature. This study showed that they were also significant from the point of view of base literature authors other than those that had authored them. The analysis of the history of the field of Trend 44

supported the claim by Small that the authors of the base literature were the leading researchers. Support was also found for Small's, 1973, (84) assertion that the frequently cited papers represented key concepts, methods and data, and that co-citation could show the relationships between these ideas. Support for the former came directly from the scientists involved, and support for the latter from the observed high levels of citation between papers that were co-cited together.

In conclusion it can generally be said that no faults had been found with the basic premises of CCA, or with the coverage of the relevant literature in the Protein Determination Models. The problems were found to lie with the criteria of continuity used to generate trends from CCA data. However problems with this would seem to be soluble by the inclusion of the additional criteria suggested earlier. One encouraging finding in support of CCA models was that the data that was missed from Trend 44 was located elsewhere in the models, only needing to be linked in. It is believed that by incorporating the suggested modifications that more realistic time surveys of scientific fields can be achieved. The results of this study are in sympathy with Leydesdorff's, 1987, (47) finding that co-citation time series derived from the Dutch Advisory Council for Science Policy Study in 1983 systematically underestimated the structure of the data over time. He blamed the high turnover of papers from year to year for the inadequate results and argued that some of the fundamental assumptions in CCA were in need of revision. The problem of the high turnover of papers can be related to the criteria of continuity.

From a policy point of view it was concluded that trend modelling as it stands would not provide much useful information and at worst could be quite misleading. However if the suggested modifications were employed, trends might become a useful tool in the policy makers repertoire. Another warning should be made with respect to taking the appearance and disappearance of specialties within trends too literally, since they often appear to have considerable overlap with other specialties in the trend framework. Also measures taken to indicate the health of

specialties such as the base literature age should be viewed in the same light, since Trend 44 showed that they can apparently show the decline and disappearance of specialties that in fact remain a part of the scientific field. It is thus suggested that trends should in future be generated along the lines proposed, and the significance of these measures re-assessed in the light of the new relationships shown between specialties. Only then will it be possible to develop theories of specialty behaviour from data of this type.

The results of this study also showed that the methodology developed in this thesis to evaluate trends was successful in tracking down the problems associated with them, which could prove to be useful to other researchers wishing to investigate this area. Also the benefits of being honest with the members of the scientific community approached for the validation of Trend 44 was shown. This type of approach enabled a more in-depth analysis, with the scientists concerned being able to pinpoint problem areas, than if questions had been asked of them in the more usual general and covert manner. Although using this type of approach can always be expected to elicit some hostility, the prospective researcher in this type of area might also be pleasantly surprised by the number of scientists prepared to assist.

REFERENCES

1. Bell, E.R.J; Corporate Activity in Protein Structure Research - A Preliminary Assessment of the Utility of Co-Citation Modelling Techniques in Research in This Area, M.Sc. Thesis, Technology Policy Unit, Aston University 1984

2. Berg, C.M. & Berg, D.E; "Jumping Genes: The Transposable DNAs of Bacteria" The American Biology Teacher, 46, 1984 p431-439

3. Berg, C.M. & Berg, D.E; "Uses of Transposable Elements and Maps of Known Insertions" Chapter 5, Section B, Part III: E. Coli & S. Typhimurium" Cellular & Molecular Biology. 1986

4. Berg, D.E. & Berg, C.M; "The Prokaryotic Transposable Element Tn5", Biotechnology 1(5) July 1983

5. Bernal, J.D; The Social Function of Science, 1939.

6. Cawkell, A.E; "Search Strategy, Construction and Use of Citation Networks With a Socio-scientific Example - Amorphous Semiconductors and Ovshinsky SR" J.Am.S.Info.Sci. 25(2) 1974 p130-133

7. Cawkell, A.E; "Understanding Science by Analysing its Literature" Inf. Scient. 10(1) 1976 p3-10

8. Chubin, D.E; "Beyond Invisible Colleges: Inspirations and Aspirations of Post 1972 Social Studies of Science" Scientometrics 7(3-6) 1985 p.221-254

9. Chubin, D.E & Studer, K.E; "Knowledge and Structures of Scientific Growth: Measurement of a Cancer Problem Domain" Scientometrics 1(2) 1979 p171-193

10. Clegg, Frances; Simple Statistics - A Course Book for the Social Sciences. Cambridge University Press, 1982

11. Collins, H.M; "Tacit Knowledge and Scientific Networks" In Science in Context: Readings in the Sociology of Science Eds. Barnes, B. & Edge, D; Milton Keynes: Open University Press 1982

12. Coward, H.R; Citation Data for Planning International Scientific Co-operation For National Science Foundation Division of International Programs Symposium: 8-10 Dec 1981

13. Coward, H.R; Second Workshop on the Measurement of R&D Output. 5-6 December 1979. Co-Citation Cluster Development as Science Indicators OECD Directorate for Science, Technology and Industry. Paper 4.3, Paris 1979.

14. Coward, H.R; Franklin, J.J. & Simon, L; Final Report. ABRC Science Policy Study: Co-Citation Bibliometric Models CRP, Philadelphia, USA. July 1984

15. Coward, H.R; Franklin, J.J. & Simon, L; Preliminary Report - A Bibliometric Model of Microelectronics - Solid State Devices and Processes. Analysis of 1982 Epitaxy and Microlithography Research CRP, Philadelphia, USA, 1984.

16. Crane, D; Invisible Colleges: Diffusion of Knowledge in Scientific Communities University of Chicago Press 1972 p2131

17. Crawford, J.W. & Crawford, S; "Research in Psychiatry - A Co-Citation Analysis" Am.J.Psychi. 137(1) 1980 p52-55

18. Crawford, S; "Informal Communication Among Scientists in Sleep Research" J.Am.S.Info.Sci. 22 1971 p301-310

19. CRP; Report on: Construction of A Bibliometric Model of Ceramics Research Activity from 1974-1981 Philadelphia USA, March 1983

20. CRP; "Description of Stratified Model/ Hypothesis 4" Memorandum to Stratified Model File. From LS/RC 24 April 1985. Unpublished.

21. Cronin, B. The Citation Process - The Role and Significance of Citations in Scientific Communication published by Taylor Graham, 1984

22. Cronin, B. "The Need for a Theory of Citing" J.Doc 37(1) 1981 p16-24

23. Department of Science and Technology, Australia. A Literature Based Model of Australian Science 20 August 1984

24. Edge, D; "Quantitative Measures of Communication in Science: A Critical Review", Hist. Sci. 17, p. 102 to 134. 1979

25. Edge, D; "Why I am Not a Co-citationist" Soc.Soc.St.Sci.Newsl. 2(3) 1977 p13-19

26. Franklin, J. Jeffrey; "Testing and Using Quantitative Methods in Science Policy Contexts". To be published Soc. St. Sci. (1988)

27. Franklin, J. Jeffrey; "The Importance of Methods and Assumptions in Testing Quantitative Methods for Science Policy Applications" February, 1988. Unpublished.

28. Franklin, J. Jeffrey; Coward, H. Roberts; "Identifying the Science-Technology Interface: Matching Patent Data to a Bibliometric Model" Submitted to Science and Technology Studies, December, 1987.

29. Garfield, E; "Citation Indexing for Studying Science" Nature 227 (5259) 1970 p669

30. Garfield, E; "Is Citation Analysis a Legitimate Evaluation Tool?" Scientometrics 1(4) 1979 p359-375

31. Garfield, E; "Mapping the Structure of Science" Chapter 8 Citation Indexing - Its Theory and Application in Science, Technology and Humanities John Wiley and Sons, 1979

32. Garfield, E; Sher, I; and Torpie, R.J; "The Use of Citation Data in Writing the History of Science" Philadelphia, I.S.I. 1964 p86

33. Gibbons, M; Georghiou, L; "Evaluation of Research Synthesis Report". Report for the Organisation for Economic Co-Operation and Development. University of Manchester, 1986.

34. Griffith, B.C; "The Social and Behavioural Sciences Literature" In Information Choices and Policies: Proceedings of the 1979 ASIS Annual Meeting v.16 Ed. Tally, R.D. and Deultgen, R.R. 1979

35. Griffith, B.C. and Mullins, N.C; "Coherent Social Groups in Scientific Change" Science 177 1972 p959-964

36. Griffith, B.C; Small, H.G; Stonehill, J.A; Dey, S; "The Structure of Scientific Literature II: Toward a Macro and Micro Structure for Science" Science Studies 4 1974 p339-365

37. Healey, P. and Rothman, H; ABRC Science Policy Study 1983/4. Meeting on 25 July 1984 Department of Education and Science, London

38. Healey, P; Rothman, H and Hoch, P.K. "An Experiment in Science Mapping for Research Planning" Research Policy 15 1986 p.233-251

39. Hicks, D; "Limitations of Co-Citation Analysis as a Tool for Science Policy" Soc.St.Sci. 17, p.295 to 316, 1987.

40. Irvine, J; Martin, B.R; "What Direction for Basic Scientific Research?" In Science and Technology Policy in the 1980's and Beyond Ed. Gibbons, M; Gummett, P. and Udgaonkar, B.M; Longman, London 1984 p67-98

41. Kessler, M.M; "Bibliographic Coupling Between Scientific Papers" Amer.Doc 14(1) 1963 p10-25

42. King, J; "A Review of Bibliometric and Other Science Indicators and Their Role in Research Evaluation" J.Inf.Sci. 13 1987 p. 261-276

43. Kuhn, T.S; The Structure of Scientific Revolutions 1962 2nd Edition The University of Chicago Press 1970

44. Lenk, P; "Mappings of Fields Based on Nominations" J.Am.S.Info.Sci. 34(2) 1983 p115-122

45. Lenoir, T; "Quantitative Foundations for the Sociology of Science - Linking Blockmodelling with Co-Citation Analysis" Soc.St.Sci 9(4) 1979 p235-236

46. Lester, George; The Structuring and Retrieving of Scientific Literature Based Citation Indexes. Ph.D. thesis, University of Manchester, U.K., 1977

47. Leydesdorff, L; "Various Methods for the Mapping of Science" Scientometrics 11(5-6) 1987 p.295-324

48. Marshakova, I.V; "Bibliographic Coupling System Based on Cited References" Nouchno Tekhnicheskaya Informatsiya Seriya 2(6) 1973 p3-8

49. Marshakova, I.V; "Citation Networks in Information Science" Scientometrics 3(1) 1981 p13-25

50. Martin, B.R; Irvine, J; and Crouch, D; "Bibliometric Analyses of Ocean Currents and Protein Crystallography: A Report for the ABRC Science Policy Study Science Policy Research Unit, Sussex University, 1984

51. Martyn, J; "An Examination of Citation Indexes" Aslib.Proc. 17(6) p. 184-192, 1965.

52. Martyn, J; "Citation Analysis" J.Doc 31(4) 1975 p290-297

53. McCain, K.W; "The Author Co-Citation Structure of Macroeconomics" Scientometrics 5(5) 1983 p277-289

54. Mulkay, M.J; Gilbert, G.N; Woolgar, S; "Problem Areas and Research Networks in Science" Sociology 9 May p187-203 1975

55. Mullins, N.C; Hargens, L.L; Hecht, P.K. & Kick, E.L; "Group Structures of Co-Citation Clusters - A Comparative Study" Am.Soc.R 42(4) 1977 p552-562

56. Nadel, E; "Citation and Co-Citation Indicators of a Phased Impact of the BCS Theory in the Physics of Superconductivity" Scientometrics 3(3) 1981 p203-221

57. Nadel, E; "Commitment and Co-Citation - An Indicator of Incommensurability in Patterns of Formal Communication" Soc.St.Sci 13(2) 1983 p255-283

58. Narin, F; Concordance Between Subjective and Bibliometric Indicators of the Nature and Quality of Performed Biomedical Research - A Program Evaluation Report for the Office of Program Planning and Evaluation April 1983 National Institutes of Health, USA.

59. Noma, E; "Co-Citation Analysis and the Invisible College" J.Am.S.Info.Sci. 35(1) 1984 p29-33

60. Popper, K.R; Objective Knowledge Oxford University Press, 1972

61. Porter, A.L; "Citation Analysis: Queries and Caveats" Soc.St.Sci 7 1977 p257-267

62. Price, D. de Solla; "Little Science, Big Science New York, Columbia University Press, 1963

63. Price, D. de Solla; "Networks of Scientific Papers" Science 149(3683) 1965 p510-515

64. Price, D. de Solla; "The Revolution in the Mapping of Science" In Information Choices and Policies: Proceedings of the 1979 ASIS Annual Meeting vol.16 Ed. Tally, R.D. and Deultgen, R.R; 1979

65. Price, D. de Solla; "The Scientific Foundations of Science Policy" Nature 206(4981) 1965 p233-238

66. Price, D. de Solla. and Gursey, S; "Studies in Scientometrics (1): Transience and Continuance in Scientific Authorship" Int.Forum 1(2) 1976 p17-24

67. Price, D. de Solla. and Gursey, S; "Studies in Scientometrics (2): Relation Between Source Author and Cited Author Population" Int.Forum 1(3) 1976 p19-22

68. Price, D.J.D; "Science: The Science of Scientists" Medical Opinion and Review 1(10) 1966 p88-97

69. Rip, A. and Courtial, J.P.; "Case Study F - Co-Word Analysis of a Biotechnology Journal" In: Commission for the EEC D.G.XII F.A.S.T. Programme. Sub Programme Biosociety. C.1.4. Analysis of Scientific Disciplines Germane to Biotechnology. Final Report, 1984.

70. Ronayne, J; Extracts from Final Report - Australian Science and Technology Indicators Feasibility Study. The Government and Private Non Profit Sectors. 1983.

71. Rothman, H; Insect Pest Control Research: The Analysis of Historical Trends with Special Reference to Scientometric Analysis Ph.D. thesis, Aston University 1987

72. Rothman, H; "Science Mapping for Strategic Planning" In Science and Technology Policy in the 1980's and Beyond Ed. Gibbons, M; Gummett, P. and Udgaonkar, B.M; Longman 1984 p.99-116

73. Rothman, H; "Techniques for Measuring the Outputs and Advances in R&D" F.A.S.T. Programme. Sub-Programme Biosociety. Analysis of Scientific Disciplines Germane to Biotechnology. Final Report Technology Policy Unit, Aston University, 1984

74. Simon, L; Literature Models of Science for Research Planning Centre for Research Planning, Philadelphia. No date appended.

75. Simon, L; Franklin, J.J. and Coward, H.R; A Bibliometric Evaluation of the US-Italy Cooperative Scientific Research Programme - Final Report 5 February 1985, Centre for Research Planning, Philadelphia

76. Small, H; "Knowledge Representation Via Co-Citation Clusters" In The Information Community: An Alliance for Progress - Proceedings of the 44th ASIS Annual Meeting 1981 vol.18 Ed. Lunin, L.F; Henderson, M and Wooster, H.

77. Small, H; Mapping the Specialty Structure of the Social Sciences Using the Social Science Citation Index Unpublished Report of a Study for the National Science Foundation, 1975

78. Small, H; "The Relationship of Information Science to the Social Sciences - A Co-Citation Analysis" Inf.Pr.Man 17(1) 1981 p39-50

79. Small, H; "The Synthesis of Speciality Narratives from Co-Citation - Clusters" ISI Early 1985. To be published in J.Am.S.Info.Sci.

80. Small, H. and Greenlee, E; "Citation Context Analysis of a CC Cluster - Recombinant DNA" Scientometrics 2(4) 1980 p277-301

81. Small, H.G; "A Co-Citation Model of a Scientific Specialty. A Longitudinal Study of Collagen Research" Soc.St.Sci 7(2) 1977 p139-166

82. Small, H.G; "Co-Citation Context Analysis - The Relationship Between Bibliometric Structure and Knowledge" P.Am.S.Info.Sci. 16 1979 p270-275

83. Small, H.G; "Co-Citation Context Analysis - The Relationship Between Bibliometric Structure and Knowledge" P.Am.S.Info.Sci. 16 1979 p270-275

84. Small, H.G; "Co-Citation in the Scientific Literature - A New Measure of the Relationship Between Two Documents" J.Am.S.Info.Sci. 24(4) 1973 p265-269

85. Small, H.G; "Cited Documents as Concept Symbols" Soc.St.Sci 8 1978 p327-340

86. Small, H.G; "Multiple Citation Patterns in Scientific Literature: The Circle and Hill Models" Inform.Stor.Retri 10(11-12) 1974 p393-402

87. Small, H.G; "Structural Dynamics of Scientific Literature" Internat.Class 3(2) 1976 p67-74

88. Small, H.G. and Crane, D; "Specialties and Disciplines in Science and Social Science - An Examination of Their Structures Using Citation Indexes" Scientometrics 1(5-6) 1979 p445-461

89. Small, H.G. and Griffith, B.C; "The Structure of Scientific Literature I: Identifying and Graphing Specialities" Science Studies 4 1974 p17-40

90. Sullivan, D; Koester, D; White, D.H. and Kern, R; "Understanding Rapid Theoretical Change in Particle Physics: A Month-by-Month Co-Citation Analysis" In Information Choices and Policies: Proceedings of the 1979 ASIS Annual Meeting vol.16 Ed. Tally, R.D. and Deultgen, R.R; 1979

91. Sullivan, D; White, D.H. and Barboni, E.J; "Co-Citation Analyses of Science: An Evaluation" Soc.St.Sci 7(2) 1977 p223-240

92. Sullivan, D; White, D.H. and Barboni, E.J; "The State of a Science: Indicators in the Speciality of Weak Interactions" Soc.St.Sci 7 1977 p167-200

93. Weingart, P; Sehringer, R; and Winterhanger, M; "Bibliometric Indicators for Assessing Strengths and Weaknesses of West German Science" Handbook of the Quantitative Studies of Science and Technology. Ed by Van Raan, A.F.J; Amsterdam, Elsevier. To be published 1988

94. White, H.D; "A Co-Citation Map of the Social Indicators Movement" J.Am.S.Info.Sci. 34(5) 1983 p307-312

95. White, H.D. and Griffith, B.C; "Authors as Markers of Intellectual Space - Co-Citation in Studies of Science, Technology and Society" J.Doc 38(4) 1982 p255-272

96. White, H.D. and Griffith, B.C; "Author Co-Citation: A Literature Measure of Intellectual Structure" J.Am.S.Info.Sci. 32(3) 1981 p163-171

APPENDICES

The following appendices contain standard examples of the letters, data and questionnaires sent to scientists in Chapter 6.

Appendix A: Letter

Technology Policy Unit
Room 911, South Wing,
Aston University,
Aston Triangle,
Birmingham, B4 7ET,
United Kingdom.

Date: 1 May 1986

Dear Dr.

I am a doctoral research student at Aston University in the Technology Policy Unit.

I am working on a bibliometric model of research concerned with the determination of the structure of complex molecules. This model is based on a survey of the scientific literature in the fields of biochemistry and molecular biology for the years 1978, 1980 and 1982. The study is one of a number of such models commissioned by the Advisory Board of the Research Councils of the U.K. from the Centre of Research Planning, Philadelphia, U.S.A. These models were commissioned to be part of a wider study concerned with research evaluation techniques.

The feature of the models which I am most interested in is the research trends, which attempt to model the development of research areas over time. I am writing to you about one trend in which your work appears to be prominent. The trend is called "Transposons and Insertion Sequences of Bacterial Plasmids and Chromasones." This trend is particularly interesting in that it covers a topic of interest to the Royal Society's study on "The Health of British Science", which among other things is examining the area of transposon research.

Therefore it would greatly aid my research if you would complete the enclosed questionnaire and return it to me at Aston University in the enclosed stamped addressed envelope. I have enclosed an introduction to my questionnaire explaining some of the concepts behind bibliometrics. Thanking you in advance for your help,

Yours Faithfully,

Elizabeth Bell

E.R.J. Bell.

Appendix B: Explanatory Letter Concerning Research Into Bibliometric Modelling Techniques

Introduction To Some Of The Concepts Behind Bibliometric Models Of Scientific Fields

Science is a highly complex activity, involving hundreds of thousands of scientists working worldwide. The many disciplines of science are sub-divided into countless thousands of specialties. The literary output of this scientific "machine" is enormous, involving many thousands of journals. So it becomes increasingly difficult for an individual scientist, perhaps starting out on a new research topic, to track down all the relevant literature.

A pioneering attempt to tackle this "information crisis" in science was made in 1955 when Eugene Garfield proposed the use of citation indexes. These were designed to overcome some of the limitations of traditional subject indexes with their arbitrary classification schemes. This led to the publication of the first Science Citation Index (SCI) in 1964.

However it was soon apparent that the SCI could readily identify highly cited papers i.e. those that could be considered to have made an impact on science. In 1963 the late Derek de Solla Price, Professor of History of Science at Yale University, said that the study of citation relationships among documents might enable the structure of science to be regarded in geographic terms. It then came to be realized that the SCI database might make this possible.

Then in 1974 Henry Small and Belver Griffith showed that the specialty (or research area) structure of science can be objectively defined. The citation measure they used was co-citation strength which is defined as the number of times that a given pair of documents are cited together by other documents. If highly cited items reflect the significant concepts in a field, then co-citation associations between them are the clusters of related concepts.

Various validation studies have shown that these clusters of highly co-cited papers represent the intellectual base of concepts, data and methodologies driving the work in a given research specialty. The co-cited clusters can be said to represent the base literature of the specialty. The papers that cite the base represent the current work in that specialty.

These techniques made it feasible for the Atlas of Biochemistry and Molecular Biology to be produced at the Institute of Scientific Information, U.S.A. This database was found to contain accurate, rapidly retrievable information of potential use to individual scientists.

A further development in this area is a technique called bibliometric modelling. In this the co-cited cluster papers (the base literature) and the papers that cite them (the current literature) are brought together in a computer to form an interacting set of specialties that define the research front of the modelled field. These models contain a great variety of data

e.g. the researchers, organisations and countries associated with the current papers define the demographics ("geographical" layout) of each specialty and the number of current papers is a measure of activity level. This rich collection of data can then be manipulated to gain insights into research patterns and events.

In 1983 the Advisory Board for the Research Councils of the United Kingdom commissioned an exploratory study of methods for the mapping and evaluation of research activity. As a part of this study, the Centre for Research Planning was asked to build five bibliometric models, each model depicting recent activity in a research field chosen by each research council. These models were validated in a series of workshops with technical experts from the various research councils. The models held up very well under expert scrutiny, e.g. one expert commented that decisions using model data would be "Policy decisions taken with better data than ever before.". The basic data for my research is taken from the model "Protein Determination" that was selected by the Medical Research Council.

These models provide a completely new way of looking at science. From them we can see what the specialties are in the way of subject matter, who the current researchers are and on what they are basing their work. How the specialties relate to others in the field and how specialties merge or break up over time to form new specialties. This is the area that I am particularly interested in, as it is hoped that this type of literature based modelling can give us some insights into specialty growth and development. This type of modelling is essentially dynamic in that the specialties show the emphasis in research at a particular time and can monitor changes in that emphasis over time.

It is probable that this modelling of science or scientometrics will be put to a variety of uses in the future. For example it could help in the design of degree courses and then keep them updated. The models show where published industrial research occurs in a field which could help in applied courses. An individual scientist embarking on a new project could use it as a way to get to grips quickly with research in that area. It could also be of use to policy makers as was shown in the A.B.R.C. study. Other studies have shown that the demographics contained in the model data can help a nation see where its research lies with respect to the rest of the world and so identify its areas of strengths and weaknesses. It can also help policy at the level of the laboratory where the next research project is decided upon.

Research into modelling techniques is still being carried out at the Centre of Research Planning. More comprehensive models are being developed such as stratified models. They now have the computing power to cluster the whole of the SCI (which covers the entire output of over 3000 journals) and so produce global maps of science that show everything from how the different disciplines relate to each other down to the level of individuals working in each specialty.

Your work has turned up in the model. From what we can tell it seems to be prominent. That is why I would be grateful if you would answer this questionnaire to give an insiders view of the field. I would then like to interview you at a later date to get a feel for how the field has developed and a better knowledge of your opinions.

Appendix C: The questionnaire.

Questionnaire on Trend 44 "Transposons and Insertion Sequences of Bacterial Plasmids and Chromosomes"

Name:

Address:

How To Answer.

This questionnaire should only take about 25 minutes to answer. It is in 5 sections;A, B, C, D and E.The questions in Section A refer to Diagram 1. The questions in Section B refer to the separate document Section B. The questions in Section C refer to Table 1, and those in Section D to Table 2. These tables, diagrams etc. are in the enclosed booklet.

Section A

The questions in this section pertain to Diagram 1, which is a diagram of Trend 44. We believe that some specialties are more closely linked to each other than they are to the rest of the specialties within a given field. Such links are shown in the diagram. The links purport to show relationships between specialties over time.

Q.1. Do these links make cognitive sense to you? Please delete as appropriate: Yes / No / Partial.
(i) (ii) (iii)
N.B. "Cognitive Sense" meaning that do the links between ideas in the research field (ideas being roughly represented by the specialty names) make sense to you?

Q.2. If your answer to question 1 was (ii) or (iii), how would you link them? Please show on the diagram below:

1978	1980	1982
101	543	
320	134	
388	222	82
525		

Please delete or add to the diagram as you see fit.

Section B

In this section I would like you to take a closer look at the base literature or highly co-cited documents in each specialty. I have enclosed summaries of the abstracts of these papers with the questionnaire. In these summaries each paragraph represents a different paper. N.B. that the list of references at the end of each specialty is in the same order as the summaries i.e. the first reference is the paper summarized in the first paragraph.
The questions that follow are in the following categories:
1. Assuming that these papers in the base literature of a specialty represent its "paradigm" i.e. key concepts, methods, data etc. are any of the papers shown key paradigm?.
2. Assuming that scientific papers fall into two major groups, theory and experimental, and assuming that a gradation between these sorts of papers exists which can be represented by two poles, where would you place the majority of papers in specialty x?. Assuming the same for bad/good and unknown/very well known papers please enter a cross on the line accordingly for each specialty.

Specialty 388:DNA Sequence of Insertion Element IS1 (1978).

Q.1. Are any of the papers key paradigm? Please delete as appropriate:

All/ None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1. is "Some" please state which ones below:

Q.3. Please place a cross on the spectrum where you think most of these papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (or unimportant) ——————————— Very well known (Very important)

Specialty 320:F-Sex Factor Genes and Bacterial Recombination (1978)

Q.1. Are any of these papers key paradigm?, please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1. is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (unimportant) ——————————— Very well known (Very important)

Q.4. For all of the specialties the specialty titles were synthesized from the titles of the current literature papers. I have listed the titles of the base literature papers for this specialty below:

(a) Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli:Structure of F 13 and related F-primes.

(b) Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli.

(c) Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli.

(d) Electron Microscope Heteroduplex Studies of Sequence Relations Among Plasmids of E.Coli: DNA Sequence Organisation of F and of F-primes, and the sequences involved in Hfr formation.

So for this specialty only do you consider that the given specialty title is appropriate for the subject matter of the base literature papers as indicated by their titles? Yes/No

If the answer is no please write down a name that you would consider more appropriate:

Specialty 525:Transposable DNA Segments in Plasmids. (1978).

Q.1. Are any of the papers key paradigm?, please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1 is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (unimportant) ——————————— Very well known (Very important)

Specialty 101:Transposons Associated with Bacterial Antibiotic Resistance (1978)

Q.1. Are any of the papers key paradigm? please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1 is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (unimportant) ——————————— Very well known (Very important)

Specialty 222:Insertion Sequences and Transposons (1980)

Q.1. Are any of the papers key paradigm?, please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1. is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (Unimportant) ——————————— Very well known (very important)

Specialty 543:Nucleotide Sequence and Activity of Transposable Genetic Elements in Prokaryotes and Yeast (1980)

Q.1. Are any of the papers key paradigm? please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If the answer to Q.1. is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (Unimportant) ——————————— Very well known (very important)

Specialty 134:Genetic Elements and Transposons (1980).

Q.1. Are any of the papers key paradigm? Please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1. is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (Unimportant) ——————————— Very well known (Very important)

Specialty 82:Characterizing Transposable Elements of Bacterial Genomes (1980)

Q.1. Are any of the papers key paradigm?, Please delete as appropriate:

All / None / Some
(i) (ii) (iii)

Q.2. If your answer to Q.1. is "Some" please state which ones:

Q.3. Please place a cross on the spectrum where you think most papers lie:

Experimental ——————————— Theoretical

Bad ——————————— Good

Unknown (Unimportant) ——————————— Very well known (Very important)

General Questions

Q.1. Now that you have looked at the base literature or "paradigm" papers for each specialty are any key papers missing? Please delete as appropriate: Yes / No

If the answer is Yes please give full details of the papers (author, reference etc). If you feel that any of these papers should only apply to certain specialties please indicate which specialty(s) in your answer.

Q.2. In the papers outlined in the base literature of the trend are there any outstandingly good or useful papers? Please delete as appropriate:

Yes / No

If the answer is Yes please state which ones:

Section C

The authors involved in the creation of the base literature are listed in Table 1.

Q.1. Please indicate on Table 1 by placing ticks in the relevant boxes, your relationship to the authors and/or their work.

N.B. Under "First year of knowing" please put down the year you first came across them or their work. If you select "Other" for any of the authors please explain your relationship below:

Q.2. Are any people that you know working in this field missing? Please delete as appropriate:

Yes / No

Q.3. If your answer to Q.2. is Yes please give their name, how you know of them and the year that you first came across them in the space below:

Section D

The institutions involved in the base literature of the trend are listed in Table 2. Each organisation has been given a number in the table. Please list the identifying numbers of the institutions in answer to the questions below:

Q.1. Which of the listed organisations (if any) do you consider to be important in this research area? Please list:

Q.2. Which of these institutions have you worked at? Please list:

Q.3. Which of these institutions have you collaborated with in your work? Please list:

Q.4. Are any important institutions in the field missing?
Yes / No

If the answer is Yes please list them by name below:

Section E

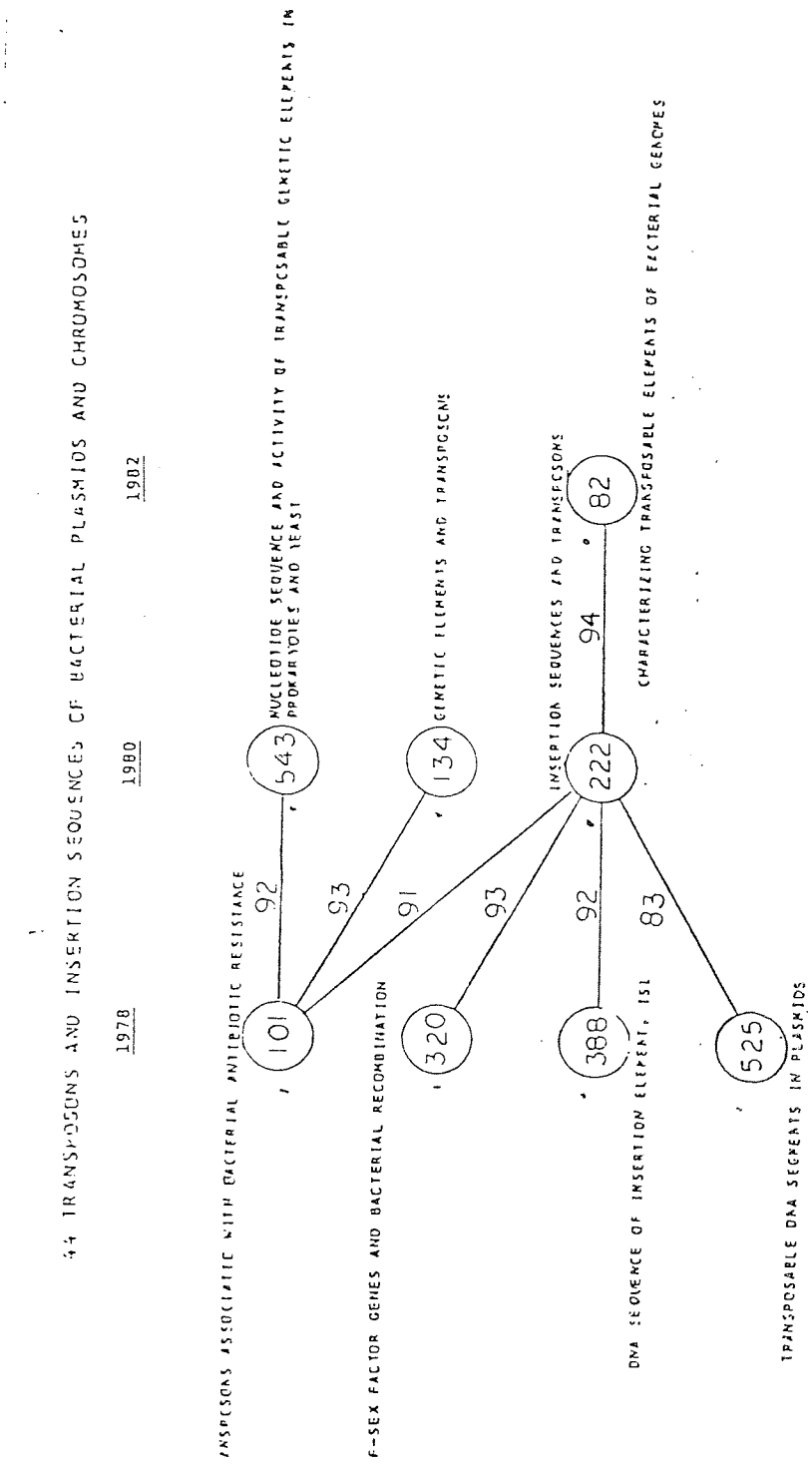
Please list the names of the journals that you normally publish your work in:

Lastly would you be prepared to see me later in the year?
Yes / No

Thank you for your time spent filling in this questionnaire.
Please return it to me in the envelope provided.

Appendix D: Booklet of Relevant Data Needed to Answer Questionnaire

Appendix D.1: Trend Diagram for Trend 44



Appendix D.2

DATA FOR SECTION B

Specialty 388:DNA Sequence of Insertion Element IS1 (1978).

1. One study found multiple copies of the insertion DNA sequences IS1 and IS2 in the chromosome of E.Coli K-12 (about 8 copies of DNA sequence IS1 and about 5 copies of IS2), detected by DNA-DNA hybridization. Both IS1 and IS2 are found in the DNA of the F plasmid, but no homology was observed between them. No small plasmids exclusively carrying insertion sequences could be isolated.

2. Another study found that polar mutations in lac, gal and phage lambda consist of a few IS-DNA sequences inserted with either orientation."Short insertions" (80-750 nucleotides) were the most commonly observed and were integrated into the lac and gal operons of E.Coli with either orientation. "Long" insertions (1170-1490 nucleotide pairs) were detected in the lac (IS3) and gal (IS2, IS4) operons and in the Y and and PQ regions (IS2) of the lambda genome.

The papers involved were;

1.H.Saedler et al.Molecular General Genetics V.122 p.267 1973
2.M.Fiandt et al.Molecular General Genetics V.119 p.223 1972

Specialty 320:F-Sex Factor Genes and Bacterial Recombination (1978).

1. All the papers in the base literature of this specialty used electron microscope heteroduplex techniques to examine the sequence relations among the plasmids of E.Coli. One paper proposed that Hfr 13 (the parent of F 13) was formed by reciprocal recombination between IS2 or F and an IS2 resident at a point between lac and proC on the chromosome of the F parent of Hfr 13. It was proposed that this IS2 and several alpha-beta sequences on the chromosonal part of F 13 are hot spots for recombination with F, i.e. for Hfr formation and that the point of origin and direction of transfer of many Hfrs can be explained by this.

2. In another study the structure of F14 was examined and heterogeneity found in closed circular plasmid molecules extracted from it. It was proposed that this heterogeneity was due to intramolecular recombination events occurring with F14 between the duplicated 2.8 to 8.5 F sequence which can account for the previously observed genetic instability of F14. It was found that another F prime plasmid, F186, was identical to F14.

3. Another study found that all those F prime factors studied had a short piece of F missing. This missing piece was found to be the same in F lac and in its close relatives F450 and FIS. Thus there is evidence for hot spots on F for insertion at different places in the E.Coli K.12 chromosome to give Hfrs and for hot spots for excision to get F-primes. A general model was proposed for the structure of F.

4. The last paper was a review of the studies from one laboratory on the sequence arrangement of F and of F-prime factors

which led to a tentative map of the fundamental organisation of the genetic sequences of F. Emphasis was placed on recent results bearing on the nature of the recombination sites involved in Hfr formation in an F cell. Details were given of the genetic sequence organisation of F;F sequences involved in Hfr insertion and other recombination events;F integration and hot spots in F recombination.

The papers involved were:

1.S.Hu et al. J.Bacteriology V.122 p.749 1975.
2.E.Ohtsubo et al. J.Molecular Biology V.89 p.565 1974.
3.P.A.Sharp et al J.Molecular Biology V71 p.471 1972.
4.N.Davidson et al Microbiology 1974 V56 1975.

Specialty 525:Transposable DNA segments In Plasmids (1978)

1. Electron microscope heteroduplex studies were used to identify and map insertion sequences IS1 and IS2 in F and R plasmids. IS1 was found to occur at R6 at the two previously mapped junctions of resistance transfer factor (RTF) DNA with R determinant DNA. Correlation of these results with previous studies show that insertion sequences play a role in a variety of F and R related intra-and-inter molecular recombination phenomenon.

2. As insertion sequences were previously identified as a cause of strongly polar mutations in E.Coli and several bacteriophages, another study did experiments to show that genetically characterised IS regions occur on bacterial plasmid DNA as both direct and inverted DNA sequence duplications. The DNA insertion shown previously to control expression of tetracycline resistance in the R6-5 plasmid and which occurs as directly and inversely repeated DNA sequences adjacent to the region believed to contain the tetracycline resistance gene was identified as IS3. A second genetically characterised insertion sequence (IS1) was identified as a direct DNA duplication occurring at both junctions of the resistance transfer factor and R-determinant components of R6-5 and related plasmids. A model was presented for the reversible dissociation of resistance transfer factor and R-determinant components of co-integrated R plasmids at the sites of DNA sequence homology provided by the repeated IS regions.

The papers involved were:

1.Hu.S et al J.Bacteriology.V.122 p.764.1975.
2.Ptashne.K et al J.Bacteriology.V.122 p.776.1975.

Specialty 101:Transposons Associated With Bacterial Antibiotic Resistance (1978)

1. The transposition of a DNA sequence encoding trimethoprim and streptomycin resistance from R483 to other replicons was studied. From this it was postulated that the Tp Sm segment of R483 is a transposon (Tn9) with specific boundary sequences.

2. In the translocation of DNA carrying an amp gene between replicons in E.Coli some R plasmids will mobilize an integrated amp gene from the chromosome of E.Coli K-12. However it was found

that R391 does not act like this. Instead the process involves recombination. Each mobilizing plasmid acquires an additional piece of DNA (MW=4X10).

3. The origin of the TEM beta lactamase gene found on plasmids was examined. It was found to be present on the small plasmid RSF 1030 and also on the naturally occurring plasmids of the F1, F11, N, X, O, I, C and W incompatibility groups that specify the TEM beta lactamase.

4. In studying the drug resistance of enteric bacteria, an active transducing bacteriophage P1CM was produced by the combination of R factor with bacteriophage P1. A phage lysate was thus obtained that was capable of transducing the characteristic of chloramphenicol resistance in extremely high frequency.

5. Transposition of a plasmid DNA sequence that mediates ampicillin resistance, its orientation of insertion and independance from host recombination functions was studied. It was found that insertion of a transposable DNA sequence that specifies the TEM beta lactamase occurred in 19 sites on plasmid RSF1010. Heteroduplex studies were used to determine the site and orientation of TnA, and found that insertions in orientation P were strongly polar on distal gene expression, whereas insertions in orientation M were mutagenic but not polar. TnA elements from different R plasmids show fine structure heterogeneity. TnA insertion at a site adjacent to the origin of replication causes an increase in plasmid copy number.

6. In the study of the transposition of R-factor genes to bacteriophage lambda (R factors being antibiotic resistance plasmids), E.Coli cells harbouring R factors that determine kanamycin resistance were infected with phage lambda and got lambda-kan transducing lines. The properties of the lambda-kan phages suggest that R factors contain systems capable of mediating genetic exchange in the absence of extensive DNA homology. It was then postulated that such systems of exchange may have had an important part in R factor evolution.

7. The acquisition of a determinant for chloramphenicol resistance by celiphage lambda was studied. This determinant was initially detected on a resistance transfer factor (RTF), then transferred to phage P1 and can be acquired from P1 by celiphage lambda. Lambda-pcam was obtained when a lambda prophage is induced in bacteria which also harbour P1cam prophage. Using electron microscope heteroduplex analysis it was concluded that the determinant for chloramphenicol resistance is contained on a unique piece of DNA which facilitates its insertion into a number of unrelated genomes.

8. In order to study the translocation of a plasmid DNA sequence mediating ampicillin resistance, recombinant plasmids were generated in E.Coli in which the TEM beta lactamase translocon (TnA) was inserted into the small plasmid RSF1010. Recombinant plasmids were classified into three clearly defined phenotypic groups. Using electron microscope heteroduplex analysis it was found that insertions of TnA occur at 12 distinct sites in a region corresponding to a third of RSF1010 DNA molecules. These give rise to the particular phenotypes.

9. Involvement of palindromes at the recombinant loci in site specific recA independant recombination between bacterial plasmids was studied. A recA independant recombination event was

shown which results in insertion of an entire plasmid genome at a unique site of another plasmid and coincident excision of a DNA segment present at the point of insertion. The resulting recombinant molecules can then undergo site specific translocation of their component segments or inversion of their original DNA sequence orientation. The events entail non reciprocal exchange of genetic material and involve a discrete nucleotide sequence that is duplicated in rotationally symmetrical reverse orientation on plasmid DNA (i.e.inverted repeat or palindrome).

10. In the study of mutagenesis by insertion of a drug resistance element carrying an inverted repetition it was found that a genetic element carrying genes for tetracycline resistance was capable of translocation from one DNA molecule to another. The tet element was acquired by bacteriophage P22 and inserts into a large number of different sites on the Salmonella chromosome. Insertion of a tet is mutagenic when it occurs within a structural gene and polar when it occurs within an operon. Insertion is usually precise, excision usually not, but both processes are independant of the recA function.

11. The DNA structures of a lambda-cam phage and a "cam excision" phenomenon by which Tn9 is spontaneously lost during phage growth were investigated. Tn9 (a chloramphenicol resistance element) was shown to promote deletions of a lambda sequence adjacent to the insertion site. The structural gene for chloramphenicol acetyl transferase (CAT), the enzyme responsible for the drug resistant phenotype was found to be contained in a 890 base pair segment at the centre of Tn9 bracketed on either side by a copy of the insertion sequence IS1. Excision of Tn9 normally occurs by reciprocal recombination between two copies of IS1, promoted by both the red and rec generalized recombination systems.

12. In the transposition of ampicillin resistance (beta-lactamase) from RP4 to other replicons, plasmids acquiring this resistance show an increase of molecular weight, and can also transpose the ampicillin resistance to other replicons. A former study (1974) suggested that those plasmids whose presence leads to the formation of large quantities of beta-lactamase might replicate under relaxed control. However this study now shows that some R factors replicate under stringent control, so the notion that production of large amounts of beta-lactamase implies multiple plasmid copies cannot be universally applied.

The papers involved were:

1.P.T.Barth et al J.Bacteriology V.125 p.800 1976
2.P.M.Bennett J.Bacteriology V.126 p.1 1976
3.F.Heffron et al J.Bacteriology V.122 p.250 1975
4.E.Kondo et al J.Bacteriology V.88 p.1266 1964
5.C.Rubens et al J.Bacteriology V.128 p.425 1976
6.D.Berg et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.3628 1975
7.M.Gottesman et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.5041 1975
8.F.Heffron et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.3623 1975
9.D.J.Kopecko et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.1373 1975
10.N.Kleckner et al J.Molecular Biology V.97 p.561 1975
11.L.A.MacHattie et al D.N.A.Insertion Elements p.219.1977
12.R.W.Hedges et al Molecular.General Genetics V.132 p.31.1974.

1. In the study of the sequence relations among bacterial plasmids, the insertion sequences IS1 and IS2 in F and R plasmids were identified and mapped using electron microscope heteroduplex analysis. It was shown that IS1 occurs on R6 at the two previously mapped junctions of resistance transfer factor (RTF) DNA with R determinant DNA. Correlation of these results with previous studies show that insertion sequences play a role in a variety of F and R related intra and inter-molecular recombination phenomenon.

2. Another study showed that genetically characterised IS regions occur on bacterial plasmid DNA as both direct and inverted DNA sequence duplications. The DNA insertion shown previously to control expression of tetracycline resistance in the R6-5 plasmid was identified as IS3. IS1 was identified as a direct DNA duplication occurring at both junctions of the resistance transfer factor and R-determinant components of R6-5 and related plasmids. A model was then proposed for the reversible dissociation of resistance transfer factor and R-determinant components of co-integrate R plasmids at the sites of DNA sequence homology provided by the repeated IS regions.

3. A determinant for chloramphenicol resistance, cam, was initially detected on a resistance transfer factor (RTF) and since transferred to phage P1. It was acquired from P1 by celiphage lambda. After electron microscope heteroduplex analysis it was concluded that the determinant for chloramphenicol resistance is contained on a unique piece of DNA which facilitates its insertion into a number of unrelated genomes.

4. Small plasmids, PSM2, PSM1 and PSM15 were used to identify regions neighbouring IS1 as well as the IS1 DNA itself. It was shown that IS1 contains 768 bases, about 30 bases at the ends of which were found to be repeated in an inverted order. The deletions occurring at the ends of IS1 were found to be due to illegitimate recombination. The hypothesis that RNA polymerase could play an important role in such recombination phenomenon is discussed based on the nucleotide sequences surrounding recombinant hot spots.

5. A genetic and sequencing analysis was carried out on the insertions of the transposon Tn9 into the lac operon of E.Coli. Although these are preferred regions for insertion these consist of multiple integration points within a small area. Sequence analysis of Tn9 insertions shows that Tn9 integration is associated with a direct repeat of 9 base pairs of host sequence. They show that these extra 9 nucleotide pairs are generated upon insertion and not brought in with the element.

6. The sequence of Tn10 insertions and the origin and role of 9 base pair flanking repetitions during Tn10 translocation were analysed (Tn10 being a tetracycline resistance element). The flanking repetitions are generated by duplication of information present only in the target DNA molecule. The repetitions do not contain genetic or structural information important for translocation. Tn10 insertions cluster at preferred positions along a target DNA. It was then speculated that homology between Tn10 and its target, at some distance from the site of the actual

recombination event, could be relevant to the preference of Tn10 for particular insertion sites.

7. In a study on the physical structure and deletion effects of the chloramphenicol resistance element Tn9 in phage lambda, a "cam excision" phenomenon by which Tn9 is spontaneously lost during phage growth was examined. The structure of Tn9 is given, bracketed on either side by a copy of IS1. Tn9 insertion promotes deletions of lambda sequences adjacent to the insertion site. Tn9 excision occurs by reciprocal recombination between two copies of IS1 and is promoted by both the red and rec generalized recombination systems.

8. The DNA sequences at the integration sites of the insertion element IS1 were studied. Two independant occurrences of insertion mutations were detected in the lacI gene of E.Coli which were both found to involve IS1. It was found that integration of IS1 into the one gene DNA is associated with a directly repeated sequence of nine nucleotides appearing at each end of the insertion element. As one of these sequences was present in the wild type gene, the second sequence either pre-existed in the IS1 before integration or else was generated by the process of insertion itself. The relevance of these findings to the mechanism of integration of transposable elements was discussed.

9. An IS1 insertion generates duplication of a 9 base pair sequence at its target site. Three independant integrations of the E.Coli insertion sequence into the gal operon were analysed. Two features were noteworthy; Similar sequences appeared in inverted orientation and in all three insertions a 9 base pair segment found once in the wild type sequence at the site of insertion is duplicated and appears in the same orientation at each end of the inserted element. No homology was observed between the inverted repeat sequences at the ends of IS1 and the sequences of the target sites. Models were proposed for the mechanism of IS1 insertion. The mechanism of IS1 translocation is in contrast to that of generalized recombination which requires the host recA function and extensive homology between the DNA species. It was observed that it remains to be seen whether the specific properties of IS1 insertion observed here apply to other translocatable elements.

10. The close vicinity of IS1 integration sites in the leader sequence of the gal operon of E.Coli were investigated. Four insertions of IS1 were analysed for their different positions and orientations. In each case 9 base pairs of the leader sequence of the gal operon are duplicated directly and found flanking the termini of IS1 at its junction with the gal operon.

11. Multiple copies of the insertion DNA sequences IS1 and IS2 in the chromosome of E.Coli K-12 were analysed. Eight copies were obtained of IS1 and five of IS2. No homology was observed between IS1 and IS2. Both IS1 and IS2 are also found in the DNA of the F plasmid. So it can be shown that no small plasmids exclusively carrying insertion sequences can be isolated from E.Coli K-12 cells and there is evidence for multiple copies of insertion DNA sequences located in its chromosome.

12. A review was given of the heteroduplex studies from one laboratory on the DNA sequence organisation of F and of F primes, and the sequences involved in Hfr formation. Recent results were emphasized which bear on the nature of the recombination sites

involved in Hfr formation in an F cell. Details were given of the general sequence organisation of F, F sequences involved in Hfr insertion and other recombination events, sequences involved in F integration and hot spots in F recombination.

The papers involved were:

1.S.Hu et al J.Bacteriology V.122 p.764 1975
2.K.Ptashne et al J.Bacteriology V.122 p.776
3.M.Gottesman et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.5041 1975
4.H.Ohtsubo et al Proc.Natl.Acad.Sci.U.S.A. V.75 p.615 1978
5.L.Johnsrud et al Cell V.15 p.1209 1978
6.N.Kleckner Cell V.16 p.711 1979
7.L.A.Machattie et al DNA Insertion Elements p.219 1977
8.M.P.Calos et al Cell V.13 p.411 1978
9.N.D.F.Grindley Cell V.13 p.419 1978
10.S.Kuhn et al Molecular.General Genetics. V.167 p.235 1979
11.H.Saedler et al Molecular.General Genetics. V.122 p.267 1973
12.N.Davidson et al Microbiology 1974 p.56

Specialty 543:Nucleotide Sequence and Activity of Transposable Genetic Elements in Prokaryotes and Yeast (1980)

1. Site specific recA independant recombination between bacterial plasmids and the involvement of palindromes at the recombinant loci were studied. It was found that a recA independant recombination event results in the insertion of an entire plasmid genome at a unique site of another plasmid with coincident excision of a DNA segment present at the point of insertion. The resultant recombinant molecules can subsequently undergo site specific translocation of their component segments or inversion of their original DNA sequence orientation. The events entail non reciprocal exchange of genetic material and involve a discrete nucleotide sequence that is duplicated in a rotationally symmetrical reverse orientation on plasmid DNA (i.e. inverted repeat or palindrome).

2. Heteroduplex studies indicate that one genetic (transposable) element (TnA) is contained within the length of the other (TnS) on the parental and recombinant plasmids. Movement of TnA and TnS among plasmids as discrete structural units implies that specific sites at their termini are involved in the translocation process. The non random distribution of recipient loci suggest that insertion is site specific and may involve recognition of a frequently recurring nucleotide sequence. In addition to their role in the intergenomic transfer of DNA sequences, the termini of transposable genetic elements can serve as hotspots for other types of recombination events. Findings suggest that evolutionary divergence of several groups of related co-integrated antibiotic resistance plasmids has occurred segmentally by a series of site specific recombination events involving the translocation, insertion and/or deletion of structurally defined segments of DNA.

The papers involved were:

1.D.J.Kopecko et al Proc.Natl.Acad.Sci.U.S.A. V.72 p.1373 1975
2.D.J.Kopecko et al J.Molecular Biology V.108 p.333 1976.

Specialty 134:Genetic Elements and Transposons (1980)

1. The insertion and excision of the transposable kanamycin resistance determinant Tn5 was examined. The frequency of transposition of the Tn5 element was found to be 10 10 . Tn5 can insert into many different sites. The results implied that the excision of Tn5, detectable as reversion is not associated with the transposition of Tn5 to new sites. The proposed models would account for the lack of specificity in the choice of a Tn5 insertion site, the apparent loss of the DNA molecule that had contained Tn5 during its transposition to new sites and the capacity of Tn5 insertion mutants to revert to the wild type.
2. The translocation and illegitimate recombination by the tetracycline resistance element Tn10 was studied. Tn10 behaviour is analogous to that observed for the IS sequences IS2 and IS1; deletions are generated, translocation occurs , precise excision is independant of temperature, events are independant of recA function, insertions are polar and show some degree of specificity. It is highly probable that the properties of Tn10 as a translocatable element derive from the properties of the IS3 sequences at its ends. It was stated that even though the mechanism involved in any illegitimate recombination events is not yet understood, the list of Tn10 properties should limit the number of models.
3. The mutagenesis by the insertion of a drug resistance element carrying an inverted repetition was studied. Tet element (which carries the genes for tetracycline resistance) was acquired by bacteriophage P22 and observed to insert into a large number of different sites on the Salmonella chromosome. Tet element insertion is mutagenic when it occurs within a structural gene and polar when it occurs within an operon. Insertion is usually precise, occurring without loss of information on the recipient DNA molecule. Excision is usually not precise. Both insertion and excision processes are independant of the recA function.

The papers involved were:

1.D.E.Berg DNA Insertion Elements p.205 1977.
2.D.Botstein et al DNA Insertion Elements p.185 1977.
3.N.Kleckner et al J.Molecular Biology V.97 p.561 1975.

Specialty 82:Characterizing Transposable Elements of Bacterial Genomes (1982)

1. Electron microscope heteroduplex analysis was used to identify and map the insertion sequences IS1 and IS2 in the F and R plasmids. It was shown that IS1 occurs on plasmid R6 at the two previously mapped junctions of resistance transfer factor (RTF) DNA with R-determinant DNA. Correlation of these results with previous studies show that insertion sequences play a role in a variety of F and R-related intra and inter-molecular recombination phenomenon.
2. Plasmids containing insertion elements are potential transposons. An in vivo recombination between plasmid PHS1 (a temperature sensitive replication mutant carrying tetracycline resistance) and PSM1 - a small plasmid carrying one copy of IS1

was studied. It was shown that PSM1 integrated at its IS1 into numerous sites on PHS1 giving rise to a duplication of IS1 in the same orientation at both junctions. Analysis of the recombinant plasmids and their parental plasmid DNA revealed that nine nucleotides at a target site were duplicated at the junction of each IS1. This phenomenon implies that plasmids containing a translocatable DNA element can be potential transposons.

3. An electron microscope analysis of bacteriophage P1, P1Cm and P7 was carried out to determine the genome sizes, sequence homology and the location of antibiotic resistance determinants. Analysis of hybrids formed in vitro between P1 and P1Cm DNA showed one region of non homology which was presumed to be the Cm determinant. Similar hybrids between P1 and P7 DNA showed six regions of non homology. Further analysis showed that the insertion of a 5.5 kb segment contained the determinant for ampicillin resistance. Prophage DNA molecules of P1Cm and P7 were larger than the P1 prophage by an amount close to the size of the antibiotic determinant which they carried.

4. The re-arrangement of genetic material in E.Coli as observed on the bacteriophage P1 plasmid was studied. It was pointed out that IS elements have an important role in the structural re-arrangements of genetic materials, yet most of the IS and Tn elements characterised so far were found by chance. So it would be helpful to design a more systematic approach. A method was suggested to achieve this, involving introducing a large plasmid as a trap into the bacterial strain. IS elements and the larger Tn elements would then be expected to transpose onto the plasmid often inactivating some of its genes. A screening of subclones for functional deficiencies would allow a sorting out of such altered plasmid genomes. Then subject plasmids to genetic characterisation (cleavage analysis and heteroduplex studies) to get the nature and site of the alteration causing the mutation. It was expected that this method would have wider applications.

5. The physical mapping of BglII, Bam HI, Eco R1, Hind III and PstI restriction fragments of bacteriophage P1 DNA was carried out. Cleavage maps of the P1 DNA were prepared by reciprocal double digestion with various restriction endonucleases. The relative orders of the fragments were obtained. The P1 genome was divided into 100 map units and the PstI site arbitrarily set as a reference point at map unit 20. It was found that DNA packaging into phage heads starts preferentially at map unit 92 and proceeds towards higher map units. The two inverted repeat sequences of P1 DNA map at about units 30 or 40. The cleavage maps obtained should complement research towards the establishment of the genetic map of P1.

6. A study was done on the role of IS1 in the formation of hybrids between the bacteriophage P1 and the R plasmid NR1. Two representatives were found of a class of oversized P1CmSmSu phages which were identified as P1 carrying the entire r-determinant of NR1 together with its two flanking, directly repeated IS1. In one case the r-determinant insertion is carried at the site of the residential IS1 of P1, in the other case it transposed into another region of the P1 genome. It was postulated that the first type resulted from reciprocal recombination within IS1 elements and that the formation of the second type of P1-R hybrid depended on IS1 mediated transposition and reciprocal recombination. The

implication of this, that any genetic material carried adjacent to an IS1 element may undergo passive transposition is discussed.

The papers involved were:

1.S.Hu et al J.Bacteriology V.122 p.764 1975
2.E.Ohtsubo et al Proc.Natl.Acad.Sci.U.S.A. V.77 p.750 1980.
3.T.Yun et al Virology V.77 p.376 1977.
4.W.Arber et al Cold Spring Harbour Symposium V.43 p.1197 1978.
5.B.Bachi et al Molecular.General Genetics V.153 p.311 1977.
6.S.Iida et al Molecular.General Genetics V.177 p.261 1980.

Appendix D.3

Table 1

	Read Work Of	Co-Worked With	Co-Authored With	Teacher of	Student of	Same Institution as	Provided/ Received Technical Assistance	First Year of Knowing	Know Well	Don't Know	Other
Allet, B.											
Arber, W.											
Bacni, B.											
Barth, P.T.											
Bennett, P.M.											
Berg, D.											
Botstein, D.											
Brevet, J.											
Calos, M.P.											
Caspers, P.											
Chan, R.K											
Cohen, S.N.											
Datta, N.											
Davidson, N.											
Davies, J.											
Deonier, R.C.											
Falkow, S.											
Fiandt, M.											
Fritz, H.J.											
Gottesman, M.											
Grindley, N.											
Grinter, N.J.											
Hanni, C.											
Hedges, R.W.											
Helffron, F.											
Heiß, B.											
Hsu, M.T.											
Hu, S.											
Iida, S.											
Jackowski, J.B.											
Jacob, A.											
Johnsrud, L.											
Jutte, H.											
Kleckner, N.											
Kopecko, D.J.											
Kondo, E.											
Kuhn, S.											
Lee, H.J.											
MacHattie, L.A.											
Malamy, M.H.											
Meyer, J.											
Miller, J.H.											
Mitsuhashi, S.											
Ohtsubo, E.											
Ohtsubo, H.											
Ptashne, K.											
Richmond, M.H.											
Rochaix, J.D.											
Rosner, J.L.											
Rubens, C.											
Saedler, H.											
Sharp, P.A.											
Starlinger, P.											
Sublett, R.											
Szybalski, W.											
Tye, B.K.											
Vapnek, D.											
Yun, T.											
Zenilman, M.											

Appendix D.4

TABLE 2

1. Dept. of Chemistry, California Institute of Technology, Pasadena, California, 91109, U.S.A.
2. Dept. of Bacteriology, Royal Postgraduate Medical School, London, W.12, England.
3. Dept. of Bacteriology, The Medical School, University of Bristol, University Walk, Bristol, BS8, 7TD, England.
4. Dept. of Microbiology and Immunology, School of Medicine, University of Washington, Seattle, Washington, 98195.
5. Lab. of Molecular Biology, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institute of Health, Bethesda, Maryland, 20014.
6. Dept. of Medicine, Stanford University, School of Medicine, Stanford, California, 94305.
7. Dept. of Microbiology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, U.S.A.
8. Mc.Ardle Lab., University of Wisconsin, Madison, Wisconsin.
9. Dept. of Molecular Biology and Microbiology, Tufts University, Boston, Massachusetts.
10. Dept. of Microbiology, Health Science Centre, State University, New York at Stony Brook, New York, 11794.
11. Biology Labs, Harvard University, Cambridge, Massachusetts, 02138.
12. Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, 06510.
13. Program in Genetics, Dept. of Microbiology and Dept. of Biochemistry, University of Georgia, Athens, Georgia, 30602.
14. Dept. of Microbiology, School of Medicine, Gunma University, Maebashi, Japan.
15. Dept. of Molecular Biology, University of Geneva, Geneva, Switzerland.
16. Dept. of Medical Genetics, University of Toronto, Toronto, Canada, M5S 7A8.
17. Institut fur Genetik der Universitat zu Koln, Cologne, W.Germany.
18. Dept. of Microbiology, University of Basel, Biozentrum CH4056, Basel, Switzerland.