

Science-Technology Linkages in an Emerging Research Platform: The case of Combinatorial Chemistry and Biology

Paper prepared for submission to the special Benelux issue of *Scientometrics*.
Comments are most welcome

Stéphane Malo^{*}
MERIT, University of Maastricht
&
Aldo Geuna⁺
SPRU, University of Sussex

First draft: January, 1999
This draft: 28, April, 1999

We are very grateful to Bart Verspagen for supplying the data on patent citation. Support from the Canadian Research Council on Human Sciences and the Commission of the European Communities, TSER Programme is also gratefully acknowledged.

* : MERIT, PO BOX 616, 6200 MD Maastricht, The Netherlands. Tel: +31 43 3883865,
Fax: +31 43 3216518, e-mail: s.malo@merit.unimaas.nl

+ : SPRU, University of Sussex, Mantell Building, Brighton, BN1 9RF, UK. Tel. +44 1273 877139,
Fax. +44 1273 685865, e-mail: a.geuna@sussex.ac.uk

Abstract

This article focuses on issues concerning science and technology relationships posed by the emergence of a new drug discovery method, namely, combinatorial chemistry and biology. We assess the scientific content of combinatorial chemistry and biology using citations in patents to scientific journals and compare this research platform with biotechnology. We also identify the institutional affiliation of all the authors of the cited papers, which leads us to an analysis of knowledge spillovers between the main participants in the research network. Finally, we examine the relevance of localisation in the process of knowledge exchange with regard to EU countries and the US. The result of the analysis provide evidence to support the view that the inventive capacity of a country is dependent upon the basic research which is carried out, especially in universities and public research centres located in the inventor's country.

1. Introduction

Recent work in the economics of science and technological change as well as in science and technology studies has revived interest in the contribution of publicly funded research to industrial innovation.¹ Such interest has led to the expansion of better measurements, a development that largely stems from the availability of new databases of science and technology indicators combined with the development of powerful desktop computers and new software. Broadly speaking, scholars attempting to gauge the contributions of publicly funded research to industrial innovation have followed three methodological paths: (1) econometric studies, (2) survey and (3) case studies. Diversified though these methods may be, they have set the stage for a better theoretical understanding of the science-technology relationship and thereby contributed to easing the task of policymakers.

Although the studies that have applied these methodological tools share the general conclusion that public research may be a prime factor behind technological change and economic growth, it does not follow that technological change will necessarily generate economic growth. Thus, the picture that emerges can in no way be taken as being a map of all inventive activities. In other words, publicly funded research may well be conducive to more industrial innovations but will not yield evenly across technological sectors.

In this context, a quantitative assessment of the science-technology linkages characterising combinatorial chemistry and biology, an emerging research platform that spans a broad spectrum of applications from drug to new materials discoveries, should be well received. Embedded in a large network of firms, universities and research centres, combinatorial synthesis methods, albeit still in their infancy, are bound to be a spawning grounds for a stream of new products. However, despite its tremendous potential, no study has ever tried to examine the issue of public research and its contribution to combinatorial innovations. The purpose of this paper is to fill this void.

The paper is organised as follows. Section 2 presents a brief technical and historical account of combinatorial synthesis methods, including parallel synthesis techniques, solution-phase methods and combinatorial biology, and discusses the improvements in

¹ For an exhaustive review of the literature on the economic benefits of publicly funded basic research see

price and performance that have been associated with their uses in drug discovery processes. Section 3 focuses on linkages with science. Responding to the drive for new insights, we began by constructing three comprehensive databases, covering all the patents, firms and publications in combinatorial chemistry and biology for the period 1980-1997. Using the CHI classification of journals from applied technology to basic research, we then analysed the scientific content of patents in combinatorial chemistry and biology to shed light on the principal fields of science cited by combinatorial patents. In Section 4, we explore the distribution of author institutions to identify the contribution of science to innovation and, finally, examine the question as to whether or not localisation affects the science-technology linkages. Section 5 provides concluding remarks and suggestions for further research.

2. An overview of combinatorial chemistry and biology methods

Combinatorial synthesis methods are here defined as “the systematic and repetitive, covalent connection of a set of different ‘building blocks’ of varying structures to each other to yield a large array of diverse molecular entities” (Gallop et al. 1994:1233). These technologies encompass a broad range of methodological schemes, the diversity of which reflects whether the building blocks are amino-acids, small chemical molecules, DNA or RNA; whether the synthesis of molecules is applied in solution, on a solid substrate or a biological system (i.e. living micro-organisms or preparations thereof such as phages, bacterial cells or DNA binding proteins); whether molecules are screened one at a time or simultaneously in a mixture; whether the biological target (i.e. molecular site of intervention) is known or unknown; whether the techniques are used for lead discovery (i.e. process of finding one or more compounds which interact with the target) or lead optimisation (i.e. process of synthesising variations of the lead compound); and so on. Yet, despite this diversity, there is little risk of over-categorising the various combinatorial synthesis methods if a clear distinction is established between the methods of solution-phase synthesis, parallel synthesis and combinatorial biology.

The first combinatorial method - solution-phase synthesis - was pioneered in 1982 by Árpád Furka at Eötvös University in Budapest (Hungary) (Furka 1995; Borman 1997). His "portioning-mixing" method, equally known as the split synthesis technique or the "divide, couple, and recombine" process, is most commonly carried out by tethering amino-acids to microparticles or polymer beads in a given number of vessels, say 3, one for each molecule. After thorough washing to remove excess reagents, these chemical units are then combined and mixed in a single dish and again separated in equal portions into 3 additional vessels. At this stage, different sets of new building blocks are added to the vessels to produce 9 different novel compounds. The protocol can be repeated as often as needed. In this example, the combinatorial procedures are simply reproduced a third time, yielding 27 new molecular entities. However, a more realistic combinatorial project using 20 different building blocks and involving a five-fold reproduction would create 20^5 different chemical combinations, or 3.2 million new compounds, each exhibiting different chemical properties (Terrett et al. 1995; Gordon et al. 1996; Furka et al. 1991).

Next came the parallel synthesis technique, which was invented by a group of scientists led by Mario Geysen (Geysen et al. 1984)² at the Australian pharmaceutical company Commonwealth Serum Laboratories in 1984. Schematically, combinatorial chemists using this technique would perform a series of individual reactions separately but simultaneously in the different reaction vessels of (usually) a microtitre plate (i.e. a moulded plastic sheet of wells filled with a few millilitres of reagents). Most scientists would begin the process by anchoring a given set of building blocks to pin-shaped polystyrene beads and inserting the resulting chemical unit-plastic bead combination into the wells of the first row, which would then be thoroughly filtered in order to wash away unreacted chemicals, that is to say, those not connected to the beads. These operations would be repeated with different sets of molecules in the wells of the other rows. Upon completion, a further set of building blocks is added to the first column, another one to the second, and so on. Unreacted chemicals are once again filtered away and the resulting compounds screened for biological activity (Plunkett and Ellman 1997). On the basis that

² It should be stressed, however, that variants of such a technique appear in parallel in many laboratories throughout the world, including the cellulose paper procedure by Frank et al. (1983) in Germany and the "tea-bag" approach by Richard Houghten (1985) in the United States. Not surprisingly, the paternity of combinatorial chemistry remains very doubtful (Lebl 1998).

microtitre plates typically contain 96 wells (i.e. 12 columns and 8 rows), a collection of 96 different new compounds could be synthesised in this way in a matter of hours, if not minutes.

Another milestone in the history of drug discovery processes concerns combinatorial biology, which was the brainchild of George Smith of the University of Missouri (US) in 1985 (Smith 1985). Combinatorial biology allows peptide libraries to be created with the help of molecular biology and some of the principles, rules of thumbs, methodologies and instruments otherwise associated with combinatorial chemistry. In a nutshell, scientists have invented an array of methods by which different sets of nucleic acids are randomly synthesised and subsequently inserted into biological systems, which in turn transcribe and translate the encoding genetic information into a vast number of different peptides.

These three methods, which set the stage for numerous incremental process innovations including the light-directed synthesis method (Fodor et al. 1991), the Diversomer technique (DeWitt and Czarnik 1995), the 'orthogonal synthesise approach (Deprez et al. 1995), have only just begun to grab the attention of policy-makers. Yet these technologies have sparked the interests of many entrepreneurs, who have founded small start-up companies in order to exploit these technologies, and large pharmaceutical companies, which have quickly built-up in-house capabilities (Glaser 1995; Brown 1996). Such companies have embarked on the combinatorial bandwagon for a variety of reasons.

The first reason revolves around the fact that combinatorial chemistry and biology offer a solution to the innovation deficit of the industry. Thus, in contrast to the orthodox chemical approach, which was devoted to yielding one pure substrate at a time in an individual test tube, combinatorial synthesis (or molecular diversity, as it is often called) allows the creation of vast in-house collections of molecular compounds, also known as combinatorial libraries. By combining mixtures of substances in tiny amounts in a selective and precise way, drugmakers are capable of producing molecular entities in previously inconceivable quantities. Up to 1991 the chemical literature described no more than an estimated 11 million compounds, a paltry performance by today's standards considering

that a small team of combinatorial chemists may now easily exceed this number in less than a week (Lowe 1995)³.

A second reason lies with the fact that combinatorial synthesis methods have also recently moved from the pharmaceutical sector into the realm of new materials; thus, they are helping to create not only new therapeutic drugs but also new catalysts, polymers, lubricants, and agrochemical products (Service 1998; Borchardt 1998; Stix 1998). For example, combinatorial chemists at the California-based combinatorial synthesis firm, Symyx, have been able to construct and screen a library of more than 25,000 different new light-emitting materials. Having discovered a blue phosphore that is entirely different from other light-emitting materials, their concerted efforts are directed towards optimising this discovery, an endeavour that could lead to applications in television and computer screens (Service 1998).

Finally, and perhaps more importantly, the new "shotgun" method of discovering therapeutic drugs (and new materials) has been a boon as far as the prospects of improving R&D efficiency are concerned. An illuminating illustration of this is presented by Thomke et al. (1998) in a field study about a combinatorial drug discovery program that aimed at treating glaucoma. Using data provided by the combinatorial synthesis firm, Pharmacoopia, a leader in the industry, the authors found that 4 combinatorial chemists had created 9000 drug-like molecules for less than \$US 19 each, a radical reduction considering that 15 chemists using traditional synthesising methods would have synthesised about 3,500 compounds at a price of \$US 5,000 per unit. Similarly, combinatorial methods have reduced development time to some three and half months, as compared to the estimated five years it would have typically taken using conventional medicinal chemistry.

The account of the evolution of combinatorial synthesis and description of its impact on the drug discovery process presented above, clearly underscores the growing importance of combinatorial chemistry and biology as a new emerging research platform in the pharmaceutical industry. To our knowledge, a detailed analysis of the scientific and technological network underlying combinatorial chemistry and biology has not yet been

³ While combinatorial chemistry and biology have yet to yield a single product, dozens of combinatorial drugs

undertaken. The following two sections are devoted to this exercise. Central to our concern will be the scientific and technological relationships that characterise research activities in universities, research institutions, and small, medium and large sized firms.

3. The scientific content of combinatorial chemistry and biology patents

Although there is wide agreement among scholars and practitioners on the contribution of public research -i.e. scientific research performed at public research institutes and universities- to the process of industrial innovation, different and sometimes conflicting measures of the relevance of its contribution have been proposed (Martin et al. 1996; Smith and Barfield, 1996). Another methodological difficulty that has come to the fore of this debate is that generalisations about the effects of fundamental science on applied research and other stages of the innovation process are virtually impossible to support. One simple but powerful reason for this is that linkages between science and technology vary greatly across industrial sectors. This is amply confirmed by a series of scientometric studies, wherein patent citations have been examined to reveal signs of science dependence (Carpenter et al. 1980; Narin 1985; Collins and Wyatt 1987; Narin and Olivastro 1991; Godin 1995; Anderson et al. 1996; Narin, Hamilton and Olivastro 1997; Verspagen 1998).

Such studies have drawn attention to how innovations in firms depend on science and thereby public research in different ways. Most of them were carried out by CHI Research Inc. or were at least largely inspired by its methodology. To understand how CHI's analytical apparatus can be employed in accounting for science and technology linkages, it is important to recall that patent data contain references to existing patents and mainstream literature. This means that the citations in patents to journal papers of specific technologies or industrial sectors can be extracted from the patent database for further analysis. In particular, one can learn more about the underlying research science base of a particular technology or industry by categorising citations in patents to literature according to a classification produced by CHI, which breaks down several thousand journals covered by the *Science Citation Index* into four different levels:

are currently awaiting approval by the regulatory agencies (Hughes 1996).

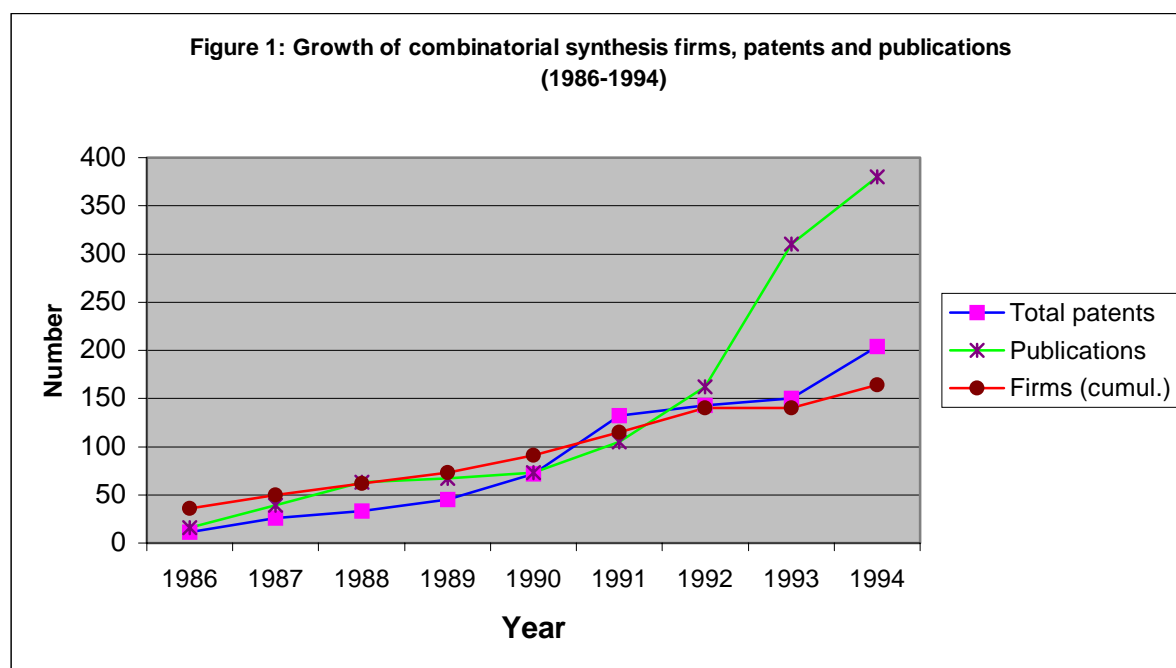
- Level 1: Applied technology (*e.g. Journal of Urology*)
- Level 2: Engineering and technological sciences (*e.g. Chemical & Engineering News*)
- Level 3: Applied research (*e.g. Journal of Chromatography*)
- Level 4: Basic research (*e.g. Nature, Journal of the Chemical Society*)

Once citations are classified along these lines, science and technology linkages may be traced by looking at the extent to which basic research journals are being cited as prior art for a given technology/industry. Underlying this reasoning is that patents granted in science-based industries ought to cite basic research journals more frequently than patents in other industrial sectors. In this context, the legitimate question that arises is whether combinatorial patents include a high level of scientific content, i.e. a high proportion of citations to patents in basic research journals.

The patent, publication, citation in patent and firm databases

To answer the above question, the exact procedures of the CHI methodology had to be reproduced using: i) a set of European Patent Office (EPO) patents dealing with combinatorial innovations and; ii) CHI's classification of journals. As data on combinatorial chemistry and biology have been noticeably scarce, a patent database had to be constructed from scratch. Such an enterprise was fraught with difficulties. Therefore, it was not until 1992 that the term "combinatorial" began spreading into the patent system, the year a patent entitled "Combinatorial strategies for polymer synthesis" (EP 0624059 - A1) was granted to the California-based combinatorial synthesis firm, Affymax. To circumvent this problem, a series of keywords such as "parallel synthesis", "split synthesis", "chemical library" and "oligonucleotide library", in addition to the term "combinatorial", had to be used. However, even then the use of such combinatorial jargon was not sufficient. Hence the database had to be solidified by the building of another two databases. The first was on firm formation and the second one on scientific publications. Moreover, a publication database has the merit of providing a point of comparison with the citation database, a point that will be treated more extensively in Section 4. Complex though this exercise is, retrieving information on combinatorial chemistry and biology both from the Internet and the CD-ROMs of the Institute for Scientific Information proved to be helpful as newly-founded firms and scientific articles often revealed new combinatorial patents. More

precisely, an iterative process was used to consolidate all three databases. An Internet discovery of a new company would lead to a new patent and publication search, whereas patent and bibliometric data often exposed yet more corporate information. In this connection, a netsite entirely dedicated to combinatorial chemistry (www.5z.com) turned out to be a particularly useful source of information, listing combinatorial synthesis companies and scientists as well as papers and patents.



Having said this, our endeavours uncovered 220 small combinatorial chemistry and biology start-ups, 1165 patent applications⁴, which spanned the period 1985-1997, as well as 2,570 scientific publications that were released between 1984 and 1996 (see Figure 1). Not surprisingly, the Commonwealth Serum Laboratories, the Australian company where Mario Geysen originally worked, was one of the first to enter the field 1) by establishing in 1988 the first combinatorial chemistry and biology company, a subsidiary named Coselco Mimotopes^{1e 1} It is worth remarking that Commonwealth Serum Laboratories resolved to sell Coselco Mimotopes (now Chiron Mimotopes) to Chiron in 1996. The decision was taken on the grounds that combinatorial chemistry and biology did not fit its strategic plan, CSL being essentially a biotechnology company seeking to develop biological products (Personal communication, Ian Gust, Director of CSL, September 7, 1998).; 2) by patenting Geysen's new synthesis method in 1986; and 3) by

publishing a paper in the well-established *Proceedings of the National Academy of Sciences in the USA* in 1984. Interestingly enough, this initial paper was rejected by the first scientific journal it was submitted to. The reviewers' decision was grounded on the apparent discrepancy between theory and empirical data; thus, it was argued that combinatorial chemists could not rigorously predict the number and amount of newly produced molecular compounds (Baum 1994).

That the emergence of combinatorial chemistry and biology start-ups is a recent trend can be captured by the fact that 127 combinatorial synthesis firms, or 58% of the total, were founded from 1992 onwards. Also worthy of note is the observation that the US leads the pack in terms of firm formation, with 170 firms offering their services in combinatorial synthesis, followed by Britain (18 firms), Germany (8), Canada (6), France (3), Australia (6), France (3), Denmark (2) and Sweden (2). This is instructive in that the relatively smaller number of innovators in Europe could be the clue as to why Europe is falling behind the United States in terms of both publication and patents share. Indeed, the number of scientific papers⁵ produced by American players had reached 1,805 by 1996, of which 760 (42%) were of academic origin, 501 (27%) from public and private non-profit research centres and 285 (16%) and 259 (14%) from small and large firms respectively. In addition, a patents count reveals that 878 (74%) patents were held by American organisations⁶, mostly small and large firms and, to a much lesser extent, by universities and public institutions.

Meanwhile, the EU was placed a distant second with 913 scientific publications, only slightly more than half the US contribution. Of these papers, 427 (46%) were from universities, 373 (40%) from public and private non-profit organisations, 29 (3%) from small firms and 84 (9%) from large, established companies. It is worth remarking, though, that the publication growth rate attributed to European universities, estimated at 66% between 1994 and 1996, significantly outpaced the overall growth rate of publications

⁴ Patents granted represent about 95% of patent applications.

⁵ While our scientific publication database comprised 2570 papers, 755 papers had been authored by 2 or more parties. Hence the total number of identified authors is higher than 2,570, totalling 3,327 different organisations.

⁶ Our EPO patent database includes 1165 patents, of which 21 were held by organisations from two or more countries.

produced by American universities, estimated at 49% for the same period. On the other hand, European organisations did not fare better in terms of patents count, accounting for a mere 234 (19% of total) patents, three time less than their American counterparts. Finally, the rest of the world, which encompasses Australia, Canada, Hungary and Japan, ranked third with 619 articles. Academia was responsible for 216 (34%) scientific publications, public and private institutions 181 (29%), small firms 165 (26%) and large firms 57 (9%). Research investigation also revealed that these countries, with 74 (6%) patents in all, lag behind the US and the European Union in terms of patents count.

Turning to citations in patents in scientific publications, we only considered patent applications up to 1994, a decision that was justified on the grounds that the data from EPO is usually a few years behind. Therefore, citations to 816 combinatorial patents, which covered the period 1986-1994, had to be extracted, and the data divided into CHI's classification whenever possible. To do this, we used the Science Citation Index (SCI) database of the Institute for Scientific Information. Of the 816 patents identified, 621 had non-patent citations, citing 2,521 publications. Unfortunately, 798 citations in patents to journals were not covered by the SCI or belonged to older journals (1980 and earlier) that were unavailable for analysis; thus, in this latter case, we only had the year of publication. Finally, the year of publication was not available for 78 of the citations, preventing us from finding any further information. As a result, we were only able to find the full bibliographical information on the SCI database for 1,645 citations in 445 patents.

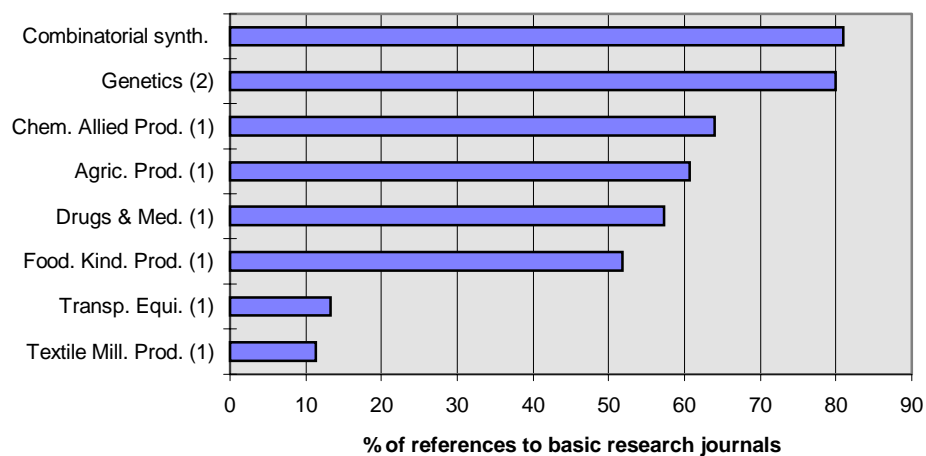
Empirical findings

Attesting to the rich variety of patterns across science-technology relationships, combinatorial chemistry and biology innovations, akin to genetics, but less so than chemistry and allied products, agricultural products, drugs and medicine and food kindred products and even less so than transport equipment and textile products (see figure 2), are shown to rely strongly on the scientific literature. Thus the results of our analysis reveal an extremely high scientific content: of the cited papers, 80.7% belong to level 4 (i.e. basic research), 17.6 % were in level 3 (i.e. applied research/clinical investigation) and the rest (1.7%) in levels 1 and 2 journals (i.e.. applied technology and engineering and technological sciences/clinical mix). Breaking down the patent data and the citations in

scientific journals by type of innovations, the role played by science in the creation of process innovations (86.0%) appears to be even more predominant than product innovation (77.8%).

Whereas our findings may come as a surprise, it should be pointed out that combinatorial synthesis approaches, which depend largely on the chemistry and application of biological and chemical oligomers as well as decorated and modified biological monomers, can be stifled by the inadequacies of existing building blocks, scaffolds, reagents and solid support substrates. To put this another way, current knowledge in medicinal chemistry cannot sustain hope for the combinatorial synthesis of *all* possible chemical combinations. These technical difficulties are further compounded with regard to process innovations as combinatorial synthesis takes place in miniaturised settings. Indeed, small-scale equipment has become the single most important issue for combinatorial applications because miniaturisation makes it possible to lower reagent costs, improve analytical sensibilities and integrate chemical synthesis with other drug discovery steps such as high-throughput screening and genomics (Borman 1998). One needs only to look at the size of combinatorial libraries, which in exceptional cases can be made up of 10^{12} compounds, to appreciate the difficulties facing scientists with regard to the isolation and identification of drug candidates. Attempts to circumvent this problem have been successful, for instance, the discoveries of deconvolution methods and tag-encoding technologies, but research along these avenue and many others is still proceeding apace (Wells 1998). To sum up, these compelling needs have been a major challenge for combinatorial chemists, who have responded by performing basic research activities to an extent and at a level that has been rarely seen in the history of drug discovery.

Figure 2: Linkage to science by product field



Sources: (1) derived from Narin and Olivastro (1992); (2) derived from Collins and Wyatt (1988).

Of equal interest is that citations to biomedical research journals represented 63% of total citations to papers, whereas the chemistry and clinical medicine journals were responsible for only 19.9% and 16.1% respectively of total patent references. It is also worth pointing out that citations in patents to chemical literature increased by 22 percent between 1991 and 1992, whereas biochemical citations decreased by 19 percent, reflecting a shift away from the biochemistry knowledge base towards the chemistry discipline. A major reason for this is that peptide libraries began to lose some of their appeal in favour of libraries made of smaller chemical molecules (i.e. with molecular weight of about 500 daltons or less). This is because small chemical compounds, albeit more difficult to assemble, give higher hit rates with more reactions to a particular target and, perhaps more importantly, can be delivered orally (i.e. through tablet or other means) rather than parenterally (i.e. by intravenous or intramuscular injections, through implantable pumps, etc.)⁷. The trend which began in 1992 is generally attributed to a team of researchers led by Bunin and Ellman of the University of California (Berkeley) which, after developing a general combinatorial method, set up a library of compounds that had previously created a broad spectrum of psychotropic drugs such as Hoffman LaRoche's Valium® (Bunin and

⁷ Peptides, when taken orally, become seriously vulnerable to digestive enzymes and often end up being destroyed in the stomach and intestines before having the chance to take effect. In this context, arguing that small chemical compounds born out of combinatorial chemistry offer a technical and commercial advantage makes perfect economic sense. The argument is further reinforced by noting that orally active medications, by causing less discomfort and less inconvenience, can increase patient compliance and wider use, a human reflex that may lead to larger markets and greater profits (Damms and Bains 1995).

Ellman 1985). This group has since been emulated by competing researchers and the method extended slowly to other small molecules such as carbohydrate molecules, λ -lactams, biphenyls, acylpiperidines and pyrrolidines (Eckker and Crooke 1995).

Two important and related propositions come out of these overall results. The first is that the fields of combinatorial chemistry and biology owe a great deal to fundamental research. Since combinatorial technologies tap intensively into the pool of generic knowledge rather than merely embedding knowledge derived from applied research programs, investments in fundamental research represent a sensible strategy for fostering technical advances in the field. The second proposition is that patents related to combinatorial synthesis methods and products heavily cite the biomedical, chemical and clinical medicine literature journals. Therefore, if policy makers are resolved to support combinatorial synthesis research activities in academic research systems, research efforts should be directed towards the pharmacology, chemistry and biomedical chemistry departments as they will offer the best potentials for generating knowledge spillovers in the industry.

4. Science-technology linkages in combinatorial synthesis: A first look

In the previous section we focused our analysis on the scientific content of the citations to the literature. Now, following the large body of literature that analyses the number of citations in a patent as an indicator of the contribution of science to innovation, we examine the relationships between the patent holder organisation and the organisation of affiliation of the author of the paper.

On average, each patent cites 3 journal articles. Throughout the period under consideration, patent citations show a static profile with a maximum number of citations in 1987 of 4.1 per patent, and a minimum in 1994 of 2.⁸ The study by Grupp et al. (1995), using comparable EPO data, identifies biotechnology as the technical area with the highest number of citations per patent, i.e., between 2 and 2.5 in the period 1985-1992.⁹ Thus,

⁸ Narin and Olivastro (1998) emphasised that the occurrence of non-patent references appears to have been relatively constant in the EPO system over the last decade.

⁹ Cited from Schmoch (1997).

patents in combinatorial chemistry and biology show extremely high linkages with science. Furthermore, the peak cited year for papers is 2 years prior to patent application which represents an extremely short time-lag. Using US patents, Narin et al. (1997) calculated that the peak time for papers cited in drugs and medicine patents is 4-6 years prior to patent grant. Therefore, even taking into account the granting period, the time lag in combinatorial synthesis between scientific discovery and technological innovation tends to be very short.

The 1,645 publications for which institutional information was available account for 3,125 organisation addresses (about 1.9 authors per paper). Of these, 66% are from the US, 20.6% from the EU and 13.4% from other countries (including Japan 5.3%, Canada 3.0%, and Australia and Switzerland about 1% each). The author institutions are classified as: (1) universities, (2) research and other institutions, (3) small firms and (4) large firms (Table 1). Universities are the most often cited type of organisation; they account for 48% of total citations. Research and other institutions, which includes, among others, hospitals, are the author institutions in 33% of the citations. Finally, slightly less than 20% of citations are shared equally between small and large firms. The 3,125 organisations addressed refer to 676 different institutions, 261 universities, 233 research and other institutions, 93 small firms and 89 large firms respectively.

Table 1: Number of organisations and citations, by institution type

Institution Type	Number of Organisations	Number of Citations
Universities	261	1,510
Research and other institutions	233	1,034
Small Firms	93	283
Large Firms	89	298
All Organisations	676	3,125

Table 2 shows the top 30 institutions in terms of number of citations. They account for about 36% of the citations, with more than one-third being attributable to 4.4% of the institutions. All are located in the US with the exception of the British Medical Research Centre, the French CNRS and the University of Tokyo. Universities (with 18 institutions among the first 30) are more than proportionally represented in the top group; not only do they have the largest share of citations, but also they are the most often cited institutions.

Cited research papers are authored principally at prestigious US universities. The top university in the EU in terms of citations is the State University of Leiden in The Netherlands with only 15 citations. Public research centres, with five institutions among the top 30 (2 European and 3 from the US), are the second most represented type of institution.

The conclusions that can be drawn from the institutional analysis of citations to literature are threefold. First, technological invention in the combinatorial synthesis research platform relies heavily on the scientific discoveries made by non-profit institutions, especially universities, and public research centres. In particular, US universities are the most important source of reference for patents in combinatorial synthesis. Second, scientific papers authored by researchers in small firms account for a relatively relevant share of citations indicating the importance of the research developed by these institutions (for instance, the case of Stratagene Cloning, Affymax and ISIS Pharmaceuticals). Finally, the citation distribution is highly skewed with 60 institutions (about 9% of the total) accounting for more than half of the citations.

Table 2: Top 30 author institutions for number of citations on all patents

Organisation	Organisation type	Country	Frequency
HARVARD UNIVERSITY	University	US	125
SCRIPPS CLINICAL & RESEARCH FOUNDATION	Private non-profit research centre	US	123
NATIONAL CANCER INSTITUTION	Public research centre	US	122
UNIVERSITY OF WASHINGTON	University	US	75
GENENTECH	Large Firm	US	66
MIT	University	US	48
PENN UNIVERSITY	University	US	42
MEDICAL RESEARCH COUNCIL	Public research centre	GB	42
UNIVERSITY OF CALIFORNIA AT SAN FRANCISCO	University	US	32
UNIVERSITY OF CALIFORNIA AT LOS ANGELES	University	US	27
STANFORD UNIVERSITY	University	US	26
UNIVERSITY OF TOKYO	University	J	26
UNIVERSITY OF MICHIGAN	University	US	26
UNIVERSITY OF ALBERTA	University	US	24
US FOOD AND DRUG AGENCY	Government agency	US	24
UNIVERSITY OF CALIFORNIA AT SAN DIEGO	University	US	24
UNIVERSITY OF MISSOURI	University	US	23
UNIVERSITY OF CALIFORNIA AT BERKELEY	University	US	23
NATIONAL INSTITUTE OF ALLERGY AND	Public research centre	US	22

INFECTIOUS DISEASES			
WORCESTER FOUNDATION FOR EXPERIMENTAL BIOLOGY	Public research centre	US	22
BRIGHAM & WOMEN'S HOSPITAL	Hospital	US	21
MERCK SHARP & DOHME LTD	Large Firm	US	21
UNIVERSITY OF TEXAS	University	US	21
STRATAGENE CLONING SYSTEM	Small Firm	US	21
CALTECH	University	US	20
UNIVERSITY COLORADO	University	US	19
MASSACHUSETTS GENERAL HOSPITAL	Hospital	US	19
CNRS	1 Public research centre	F	19
YALE UNIVERSITY	University	US	19
UNIVERSITY OF BAYLOR	University	US	18

One might wonder whether these results are affected by localisation in the process of knowledge spillover (Jaffe, Trajtenber, and Henderson, 1993; Mansfield and Lee, 1996). Narin et al. (1997) clearly showed that inventors cite their own country's papers much more than might be expected from their publication share in the SCI. First the citations to papers from the country of the inventors compared to the total distribution of citation by countries are analysed. For all countries, the number of citations to their own country's papers is higher than for all countries taken together. For the US this difference is small, while for the EU countries the difference is much more significant. About 39% of the citations to the literature in patents of the EU countries are to papers from institutions of the EU countries. This share is almost double that for total patents. In the case of France, Germany, Japan and the UK the 'own country' citations are three to four times higher than for all countries combined. The reliance on 'own country' literature becomes extreme for countries such as Italy and Canada for which more than 50% of the cited literature originates in the author's own country.

To confirm the relevance of localisation we have compared the share of 'own country' citation to science papers with the share of publications by each country in combinatorial synthesis (see Section 3 for the description of the combinatorial synthesis publication database). As in the previous case, generally there is evidence of a higher citation rate for 'own country' papers. For the US, however, the difference is marginal. As in the previous case, inventors from the other countries considered cite local literature at least twice as often. For example, the UK ranks second in terms of publications on

combinatorial synthesis with about 7.4% of the total; however, approximately 23% of all the science papers cited in UK patents are authored by employees of UK organisations.

Finally, we compared the country share of publications in the SCI with own country citations. The results confirm a bias in favour of 'local' literature. In this case, also, US inventors (with about 69% of own country citations) more often cite US literature than would be expected from the share of US papers in the SCI (about 35-45% in the scientific areas relevant to combinatorial synthesis).

The current analysis confirms that science-technology linkages in combinatorial synthesis, as measured by the citations to the literature in patents, are affected by the location of the organisations. The technological inventions of a country rely heavily on the scientific discoveries made in that same country, especially in the universities and public research centres.

5. Conclusions

This paper has provided a first exploratory analysis of science-technology linkages in combinatorial chemistry and biology. To begin with, we presented a technical and historical description of the development of combinatorial synthesis methods, showing that they have created a new research platform for drug and new material discovery. Subsequently, using three original databases on firms, patents and publications, we examined the main characteristics of science-technology linkages in this research platform.

The results of the empirical analysis provide evidence that supports the following conclusions. First, the predominance of patents citations to basic research journals supports the view that technological invention in combinatorial synthesis is characterised by an extremely high scientific content. In particular, biomedical research is the most significant knowledge base of reference for combinatorial chemistry and biology patents, although chemical literature has been increasing in relevance since the early 1990s. Second, as in the case of biotechnology and computational chemistry (Orsenigo, 1989, Mahdi and Pavitt, 1997), the US lead in firm formation is paralleled by the dominant role played by US

universities and research centres. We found some evidence that the EU countries are catching up in terms of university publishing, while the number of new combinatorial synthesis firms in Europe remain very small. Third, the inventive capacity of a country heavily depends upon the strength of the underlying universities and public research institutes. The innovation process of firms relies to a great deal upon research carried out by universities and by public research centres of their own country. Finally, the analysis of patent applications and patent citations underlines the importance of small firms in the development of the combinatorial synthesis research platform. Not only are they patenting in a significant way, but they also are producing important publications that are cited in other patents.

The preliminary results of this study emphasise the significance of the contribution of academic research to industrial innovation. This overall conclusion must be qualified given the shortcomings inherent in the methodology adopted, such as the fact that citations are insert by the inventor but also by the examiner of the patents. Nonetheless, the results of innovation surveys, such as the Yale and PACE, confirm the fact that, in the pharmaceutical and chemical industries, the research carried out by public institutes and universities is considered by research managers to be extremely important for the innovative activity of firms (Arundel et al., 1995; Geuna, 1999; Klevorick et al., 1996). Further detailed analysis of the institutional network in combinatorial synthesis is required to develop a better understanding of the micro mechanisms by which knowledge created in universities and public research centres spills over into the knowledge creation processes of firms and other institutions.

Reference

- Anderson, J., N. Williams, F. Seemungal, F. Narin, and D. Olivastro. 1996. Human Genetic Technology: Exploring the Links Between Science and Innovation. *Technology Analysis and Strategic Management* 8: 135-156.
- Arundel, A., G. Van de Pall, and L. Soete. 1995. *Innovation Strategies of Europe's Largest Industrial Firms*. PACE Report prepared for the SPRINT Programme of the European Commission.
- Borchardt, J. 1998. Combinatorial Chemistry: Not Just for Pharmaceuticals, *Today's Chemist at Work* 7: 34-35, 37, 39, 41.
- Borman, S. 1997. Combinatorial Chemistry. *Chemical & Engineering News* February 24, pp. 43-62.
- Borman, S. 1997. Combinatorial Chemistry. *Chemical & Engineering News* April 8, pp.1-24, <http://pubs.acs.org/hotarcl/cenear/980406/comb.html>
- Brown, K.S. 1996. Drug, Biotech Firms Beginning to Embrace Combinatorial Chemistry *The Scientist* 10: 1, 5.
- Bunin, B. A.; Ellman, J. A. 1992. A General and Expedient Method for the Solid-Phase Synthesis of 1,4-Benzodiazepine Derivatives. *Journal of American Chemical Society*, 114, 10997-10998.
- Carpenter, P., M.Cooper, and F. Narin. 1980. Linkage Between Research Literature and Patents. *Research Management* March, pp.30-35.
- Collins, P. Wyatt, S. 1988. Citations in Patents to the Basic Research Literature. *Research Policy* 17: 65:74.
- Damms,B. And W. Bains. 1995. The Cost of Delivering Drugs Without Needles. *Bio/Technology* 13:1438-1440.
- Deprez,B., X. Williard, L. Bourel, H. Coste, F. Hyafil, and A. Tartar. 1995. Orthogonal Combinatorial Chemical Libraries. *Journal of American Chemical Society*. 117: 5404-5406.
- De Witt, SH, and A.W. Czarnik 1995. Combinatorial Organic Synthesis Using Parke-Davies's DIVERSOMER Method. *Association of Chemical Research* 29:114-122.
- Ecker, D., and S. Crooke. 1995. Combinatorial Drug Discovery: Which Methods will Produce the Greatest Value. *Bio/Technology* 13:351-359.
- Fodor, S.P.A., J., Read, M.C. Pirrung, L. Stryer, A.T. Lu, and D. Solas. 1991 Light Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* 251: 767-
- Furka, A. 1995 The History of Combinatorial Chemistry. *Drug Development Research* 36:1-12.
- Furka, A, F. Sebestyen, M. Asgedom, G. Dibo. 1991. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *International Journal of Protein Research* 37:487-493.
- Gallop, M., R. Barrett, W. Dower, S. Fodor, and E. Gordon. 1994. Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries *Journal of Medicinal Chemistry* 37:1233-1251.
- Geysen, H.M., R.H. Meloen, and S.J. Barteling 1984. Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino acid *Proceedings in National Academy of Science USA* 81: 3998-4002.
- Geuna, A. 1999, Contribution to Innovation and the Evolution of Scientific Specialisation in the Chemical and Pharmaceutical Industries. Paper presented at the "European Meeting on Applied Evolutionary Economics", Grenoble 7-9th June.

- Glaser, V. 1995. Drug Companies Like Combinatorial Chemistry. *Bio/Technology* 13: 311-312.
- Godin, B. 1995 Research and the Practice of Publication in Industries. *Research Policy* 25:587-606.
- Gordon, E., M.Gallop, and D. Patel. 1996. Strategy and Tactics in Combinatorial Organic Synthesis. Applications to Drug Discovery. *Account Chemical Research* 2: 144-154.
- Grupp, H., G. Münt and U. Schmoch. 1995. *Wissensintensive Wirtschaft und ressourcenschonende Technik*. Report to the German Ministry of Education and Research. Karlsruhe.
- Houghten, R.A. (1985) General Method for the Rapid Solid-Phase Synthesis of Large Number of Peptides: Specificity of Antigen-Antibody Interaction at the Level of Individual Amino-Acids. *Proceedings in National Academy of Science USA* 82:5131-5135.
- Jaffe, A. 1989. Real Effects of Academic Research. *American Economic Review* 79: 957-970.
- Jaffe, A., M. Trajtenberg, and R. Henderson. 1993. Geographic Localisation of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 63: 577:598.
- Klevorick, A., R. Levin, R. Nelson, and S. Winter. 1996. On the Sources and Significance of Interindustry Differences in Technological Opportunities. *Research Policy* 24:185-205.
- Lebl, M. 1999. Parallel Personal Comments on 'Classical' Papers in Combinatorial Chemistry. *Journal of Combinatorial Chemistry* 1: 3:24.
- Lowe, G. 1995. Combinatorial Chemistry. *Chemical Society Review* 29: 309-382.
- Mahadi, S., and K. Pavitt. 1997. Key National Factors in the Emergence of Computational Chemistry Firms. *International Journal of Innovation Management* 1:355-386.
- Martin, B., A. Salter, and et al. 1996. *The Relationship Between Publicly Funded Basic Research and Economic Performance*. Brighton: SPRU, University of Sussex.
- Narin, F., and D. Olivastro. 1998. Linkage Between Patents and Papers: An Interim EPO/US Comparison. *Scientometrics* 41: 51-59.
- Narin, F., and D. Olivastro. 1992. Status Report: Linkage Between Technology and Science. *Research Policy* 21: 237:249.
- Narin F., K.S. Hamilton, and D. Olivastro. 1997. The Increasing Linkage between US Technology and Public Science. *Research Policy* 26: 317: 330.
- Orsenigo, L. 1989. *The Emergence of Biotechnology: Institutions and Markets in Industrial Innovation*. London: Pinter.
- Plunkett, M.J. and J.A. Ellman 1997. Combinatorial Chemistry and New Drugs. *Scientific American* April, pp.54-59.
- Rosenberg, N. 1990. Why do Firms do Basic Research with their Own Money? *Research Policy* 19: 169-175.
- Schmoch, U. 1997. Indicators and Relations Between Science and Technology. *Scientometrics* 38: 103-116
- Service, R. 1998. Winning Combination. *Research Policy Technology Review* May/June, pp. 34-41.
- Smith, B.L.R. and C.E. Barfield. 1996. Contributions of Research and Technological Advance. In *Technology, R&D, and the Economy*, ed. B.L.R. Smith and C.E. Barfield. Washington, D.C.: The Brookings Institution and American Enterprise Institute.

- Smith, G.P. 1985. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* 228:1315-1317.
- Terrett, N., M. Gardner, D. Gordon, R. Kobylecki, and J. Steele. 1995. Combinatorial Synthesis- The Design of Compound Libraries and their Application to Drug Discovery. *Tetrahedron* 51:8135-8173.
- Thomke, S., E. Von Hippel, and R. Franke. 1998. Modes of Experimentation: An Innovation Process -and Competitive- Variable. *Research Policy* 27:315-332.
- Van Vianen, B. G., H. F. Moed, and A. F. J. Van Raan. 1990. An Exploration of the Science Base of Recent Technology. *Research Policy* 19: 61:81.
- Verspagen, B. 1998. Large Firms and Knowledge Flows in the Dutch R&D System. A Case Study of Philips Electronics. *Technology Analysis and Strategic Management* 11:
- Wells, W. 1999. Combinatorial Chemistry. Making More Drugs.
<http://outcast.gene.com/ae/AB/BA/combiChem/>