Electronic Working Papers Series

Paper No. 32

An Examination of Technology Strategies for the Integration of Bioinformatics in Pharmaceutical R&D Processes

Michael M. Hopkins

Science Policy Research Unit
Mantell Building
University of Sussex
Falmer, Brighton
BN1 9RF, UK

Tel: +44 (0) 1273 686758
Fax: +44 (0) 1273 685865
Email:
M.E.Winder@sussex.ac.uk
http://www.sussex.ac.uk/spru/

An Examination of Technology Strategies for the Integration of Bioinformatics in Pharmaceutical R&D Processes

Michael M Hopkins

Submitted in partial fulfillment of the requirements for the degree of MSc in Technology and Innovation Management

SPRU, Science and Technology Policy Research
University of Sussex
September 1998

Abstract

Bioinformatics is the use of computers for the storage, recall and analysis of data derived from scientific research aimed at providing answers to biological questions. Increasingly pharmaceutical firms are choosing to incorporate bioinformatics into their drug research and development (R&D) programs.

Based on examination of recent literature and case studies of a cross section of the pharmaceutical industry and research community, this research concludes the following key points:

- Bioinformatics has been widely accepted as a new core competency for pharmaceutical research.
- Pharmaceutical firms are broadening their competencies to different extents to access new technologies.
- Integrated bioinformatics systems are systemic innovations and require strong in-house capabilities, however bioinformatics tools may also be used in an autonomous manner. Thus two groups of industrial users are emerging: those who use bioinformatics tools piecemeal and those who are attempting to integrate these tools into systems.
- Integration poses a complex set of technical and organisational problems, but may enable high throughput programs for drug discovery that optimise resources more effectively and thus provide competitive advantages.
- Such integrated systems must be tailor-made and are possible only through gaining competencies that are difficult to replicate.

Additionally in this study theoretical frameworks are used to map technological change in the pharmaceutical process, showing how barriers to innovation and limiting steps within the R&D process are changing as a result of bioinformatics.

Chapter 1: Introduction

List of Figures

Page

Chapter 1

Chapter 2

Chapter 3

Chapter 1: Introduction

1.1 Purpose

The early 1990's saw the establishment of bioinformatics departments in the largest firms of the pharmaceutical industry. This move was prompted primarily by the need to store and analysis large amounts of DNA sequence data that had become available through the activities of academics and small specialised genomics[1] firms such as Human Genome Sciences (HGS) and Incyte Pharmaceuticals in the United States.

Bioinformatics is the:

"storage, retrieval and analysis of data for scientific research aimed at providing answers to biological questions."

This definition encompasses computerised tools and techniques used for the study of sequences and structures relating to DNA, proteins and other molecules in biological systems. The term Computational Biology may also be used by some sources to describe this field. At present the most commonly used of these in pharmaceutical research are of the following types:

Software tools[2] exist to create and compare 3-D structural models of proteins and other molecules, to aid prediction of their interactions. Models may be drawn from local or remote databases. Additionally, tools also exist to find and compare DNA and protein sequences drawn from local and remote databases. Both these groups of tools are generally run on workstations with intra/ internet links. These computers have more powerful processing and graphics abilities than standard PCs, but are less powerful and expensive than mainframe computers.

It was hoped the ability to read DNA sequences would allow deeper understanding of the susceptibility of individuals to disease, as well as giving greater insight into the

[1] See glossary for definition

structure and assembly of proteins. Knowledge of protein structure would enhance the ability of pharmaceutical firms to find or design molecules to "dock" with these proteins and thus alter their performance producing a therapeutic effect. Thus bioinformatics was needed as a set of enabling tools to store and interpret the newly available data.

Given the growing importance of bioinformatics to the pharmaceutical industry, the key questions addressed in this project are:

- **How has bioinformatics been integrated into the drug research and development processes of the firm?**
- **What strategies have pharmaceutical firms used to acquire capabilities in bioinformatics?**

1.2 Argument

The argument presented here is bioinformatics constitutes a new core competency for pharmaceutical firms. In an industry with increasing regulatory requirements, rising R&D costs and increasing competition for market share, bioinformatics offers the means of acquiring a greater insight into the causes of disease and the structure and mechanisms of lead compounds[3] for therapeutic agents. Increasing such knowledge guides strategic decisions, and thus may reduce the likelihood of poor drug candidates[4] entering the clinical trials process and resulting in expensive failures. Additionally by automating many stages of research, higher throughput of lead candidates may be achieved, thus enhancing firm's development pipelines. Therefore firms optimising bioinformatics systems within their R&D process will

[2] See glossary for definition
[3] See glossary for definition
[4] See glossary for definition

derive a competitive advantage[5] over firms failing to effectively utilise the new tools. Firms' strategies for accessing bioinformatics are varied, as are their internal organisational capabilities, and this may result in a new source of competitive advantage in the industry. The extent to which firms can develop bioinformatics is dependent on available resources. Organisational and technological problems arise due to the complexity of integrating software systems and these are expensive and time consuming to solve. Only the largest firms with great investments of time and resources can achieve fully integrated bioinformatics systems and derive maximum competitive advantage. Whilst smaller organisations can access basic bioinformatics tools at relatively low cost, they cannot afford the numbers of staff and other resources necessary to integrate these tools into a tailor-made system allowing optimum results.

1.3 Frameworks Adopted

Two bodies of literature are used in pursuing the argument above: literature describing pharmaceutical R&D processes specifically, and literature debating the management of technological innovation within corporate strategy.

In mapping technological change in the drug R&D process it is necessary to examine frameworks describing this process, as presented by Nightingale[6], Lyall[7], and Jean[8]. These frameworks show that the classical drug R&D process has been changed dramatically, with new stages replacing old, and a broader choice of techniques

[5] A relative advantage over competitors in that lowering overall costs of failed products allows more resources to be spent on developing greater numbers of viable drug candidates before competing firms can gain market share in the therapeutic market for a given condition.

[6] Nightingale, P (1997) Knowledge and Technical Change: Computer Simulations and the changing Innovation Process, Sussex University

[7] Lyall, A, (1996) Bioinformatics in the Pharmaceutical Industry, Trends In Biotechnology, vol.14 p.308-312 & Lyall, A, (1998) "Informatics – The Key to redesigning Drug Discovery" European Biopharmaceutical Review, March.

[8] Jean, T, (1997) Drug Discovery CROs, European Pharmaceutical Review, September, p.36-41

available for use. Limiting factors that determine time and cost of drug discovery are also revealed.

In the management of technological innovation literature, it is suggested the nature of technological change in the pharmaceutical industry has consisted of dramatic paradigm shifts, as defined by Dosi[9&10], where by the heuristics of R&D have been changed, first by biotechnology[11] and more recently by bioinformatics[12]. In theories of technological innovation authors have asserted firms faced with change are constrained in their choice of action by their technological trajectory[13]. To adapt to changing technological circumstances, firms must access new capabilities. The question of whether to develop these capabilities in-house or to access them through external linkages may be answered by considering their application and whether it relates to autonomous or systemic innovation[14] as suggested by Chesbourgh & Teece[15]. This concept is important for firms considering how best to access and use bioinformatics capabilities. Granstrand, Patel & Pavitt[16] demonstrate that large multi-technology firms have a broad spectrum of technological competencies, wider than indicated by examining their product range. Firms must maintain this position in order to explore new technological options available to them, as well as to manage

[9] Dosi, G (1982) Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technological change, Research Policy, Vol. 11 No.3 p.147-62
[10] Dosi, G (1988) The Nature of Innovative Process, Chapter 10, in Dosi, G, Freeman, C, Nelson R, Silverberg, G, and Soete, L, Technical Change and economic Theory, Frances Pinter, London
[11] Sharp, M (1989) Collaboration and The Pharmaceutical Industry Is It The Way Forward, SPRU, Sussex university
[12] Lander, E, Langridge, R, Saccocio, D (1991) Mapping and Integrating Biological Information, Communications of the ACM Vo1. 34, No.11
[13] Nelson, R & Winter, S (1977) In search of a useful theory of Innovation, Research Policy, vol. 5, p.36-76
[14] The benefit from Autonomous innovations can be realised independently of other systems, for example a turbocharger can be introduced to car without having to design the engine. However systemic innovations require changes in supporting systems, for example Polaroid technology required not only a new film, but a new camera too (Chesbourgh & Teece, 1996).
[15] Chesbourgh, H, & Teece, D (1996) When is Virtual Virtuous? Harvard Business Review, January – February, p.65-74
[16] Granstrand O, Patel P, Pavitt K (1997) Multi-Technology Corporations: Why They Have Distributed Rather Than Distinctive Core Competencies, California Management Review Vol.39 No.4 p.8-25

and co-ordinate technical change. Thus Granstrand Patel & Pavitt imply firms should develop bioinformatics capabilities in-house to fully manage and benefit from technological changes in bioinformatics and, because it is an enabling technology, other related fields. Teece and Pisano[17] argue the co-ordination of technological change and the ability to gain competitive advantage from it is derived from firms' dynamic capabilities. Like Granstrand Patel & Pavitt, Teece and Pisano also emphasise it is not just capabilities that are important, but the ability to change and adapt capabilities as well. They emphasise the importance of firms' paths, positions and processes in contributing to the formation of dynamic capabilities and the importance of apropriability regimes in determining whether or not a firm will be successful in creating competitive advantage from new capabilities such as bioinformatics.

The frameworks presented by Chesbrough and Teece (1996) ,Granstrand, Patel, and Pavitt (1997), and Teece and Pisano (1994) are further considered in explaining the significance of this study's findings.

1.4 An Introduction to Biotechnology – A New paradigm for Pharmaceutical Development

Biotechnology is a broad term for a group of enabling technologies that allows "the application of scientific and engineering principles to the processing of material by biological agents"[18]The development of biotechnology can be divided into three generations[19]. 3rd generation biotechnology (or "modern" biotechnology) has its

[17] Teece D, Pisano, G (1994) "The Dynamic Capabilities of Firms", Industrial & Corporate Change, Vol.3 No. 3, p537-556.
[18] OECD (1992) Biotechnology: International Trends and Perspectives
[19] Sharp, M (1990) Pharmaceuticals and Biotechnology: Perspectives for the European Industry, SPRU Sussex University:

genesis three years after the first software computer. In 1952, Watson and Crick determined the structure of Deoxyribose Nucleic Acid (DNA), the molecule that acts as a blueprint for virtually all organisms. In 1973, Cohen and Boyer published an account of the first successful recombination or splicing of DNA from two species. For the first time characteristics could be transferred between unrelated species, and existing characteristics could be altered. Cohen and Boyer went on to form Genentech in the US, the first modern "biotech firm". The ability to sequence DNA, and "read" the instructions to make the building blocks of life came from the work of Fredrick Sanger at Cambridge University in 1975. His technique proved it possible to predict certain characteristics of individuals from studying their DNA. Over recent years, DNA sequencing[20] has become more efficient and less expensive, leading to the routine sequencing of genes and even entire genomes in laboratories across the world. The establishment of the Human Genome Project (supported by the US government and the Wellcome Trust) has brought hopes that all 3 billion bases and 100,000 genes of human DNA will be available to the public sector by 2005. Meanwhile Human Geneome Sciences (HGS) and Perkin Elmer working in a joint venture hope to complete their own similar program some years earlier to provide novel targets[21] for pharmaceutical research[22]. Research into the genomes of numerous other species is being undertaken. These include animals such as the mouse, and *Drosphilla* (fruit fly), plants including *Arabidopsis*, rice, wheat, and soya, micro-

---

1st generation: The use of bacteriological fermentation in processes for producing goods such as wine and cheese; cross breeding of plants and animals to produce offspring with favoured characteristics.
2nd generation: The microbiological "revolution" of the late 19th and early 20th centuries brought advances in fermentation techniques, the first vaccines and antibiotic treatments. Additionally enzymes and applied microbiology because increasingly used.
3rd generation: The discovery of DNA and Genetic engineering, allows transfer of characteristics between species by use of various vectors, and the design of drugs, proteins and therapeutic tools.

[20] See glossary for definition

[21] See glossary for definition

[22] Knight, J & Kleiner, K, Genome Project Goes Into Overdrive, New Scientist, No.2134 16th May 1998

organisms such as yeast (the first completed eukaryotic genome), *E.coli*, as well as a host of extremeophiles (bacteria with remarkable adaptations to difficult living environments such as volcanic springs or deep ocean trenches). The research from such projects provides novel techniques and products not just for the pharmaceutical industry, but the healthcare, agriculture, environmental management, mining, chemical industries too, and it may well find applications in other areas.

1.5 The Changing face of the Pharmaceutical Industry

Over recent years the sources of innovation for the pharmaceutical industry have changed dramatically from the traditional pharmacological techniques which sufficed for several decades[23]. New developments resulting from modern biotechnology (referred to from now on as simply biotechnology) offered revolutionary technological change, indeed some speculators saw biotechnology as a Schumpeterian competence destroying advance which would bring down the incumbent pharmaceutical firms[24]. Entrepreneurs sparked an explosion in new Dedicated Biotechnology Firms[25] (DBFs) which started in the US and spread globally. There are around 1300 such firms in the US and 1000 in Europe[26]. DBFs faced numerous problems both in becoming integrated pharmaceutical firms and in producing marketable products. They lacked down stream capabilities such as production, sales and distribution, had no experience of running clinical trials, and more fundamentally many of the DNA or protein based therapies they developed had inadequate delivery mechanisms to ensure they reached the cause of the patient's illness. As a result a

[23] Galambos, L & Sturchio, J (1997) Science in the Twentieth Century, Chapter 13
[24] Kenney, M, (1986) "Schumpeterian Innovation and entrepreneurs in Capitalism: A Case Study of the U.S. Biotechnology Industry" Research policy 15, p.21-31.
[25] See glossary for definition
[26] Ernst & Young (1998) European Life Sciences 98 - Continental Shift

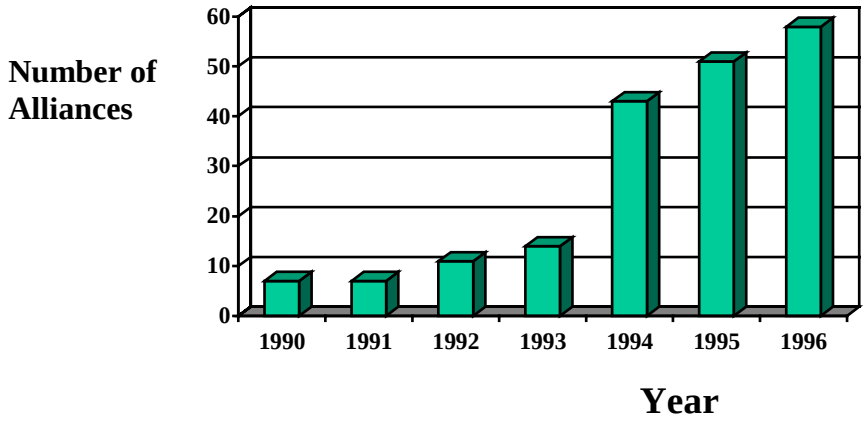
mere 40 or so therapeutic compounds derived from biotechnology have been FDA[27] licensed, and only a handful of profitable DBFs exist[28]. Faced with high barriers to entry into the pharmaceutical industry as fully integrated firms, DBFs are increasingly contenting themselves in becoming "drug discovery" firms, or Contract Research Organisations[29] (CRO) undertaking niche activities for other firms[30]. Thus it appears biotechnology has not proved competence destroying and the incumbent pharmaceutical firms' competitive advantage has so far been maintained[31].

Until the development of biotechnology, pharmaceutical firms had generally concentrated their research in-house. However, many of the techniques that biotechnology made possible were attractive to pharmaceutical firms, but only available from DBFs and universities. Pharmaceutical firms found these new techniques attractive because traditional methods of drug discovery were producing diminishing rates of return. These techniques, used since the "golden age" of pharmaceuticals in the 1950's and 1960's, had uncovered the obvious drug candidates, and finding new therapeutics was becoming increasingly expensive and time consuming[32]. To gain access to the new technologies offered by biotechnology, firms had to form alliances either with DBF's or universities with these capabilities. Indeed one reason for the growth of DBF's is seen as entrepreneurial academics facilitating the requirements of large firms[33]. The initial skill shortage was such that large firms might have had difficulty in establishing internal departments in these

[27] See glossary for definition
[28] Aharonian, G, Patent News, 21 August 1997 (available from srctran@world.std.com)
[29] See glossary for definition
[30] Rau, N (1997) Biopharmaceuticals in Europe - A Rich ground for Partnering, European Biopharmaceutical Review, September. p.20-29
[31] Pisano, P (1990) "The R&D Boundaries of The Firm: An Emperical Analysis", Administrative Science Quarterly, 35 p. 153-176
[32] Sharp, M (1989) Collaboration and The Pharmaceutical Industry Is It The Way Forward, SPRU, Sussex University, p.8.
[33] Sharp, M (1989) Collaboration and The Pharmaceutical Industry Is It The Way Forward, SPRU, Sussex University, p.17.

areas, and this was one reason behind the growth of an alliance culture. However, the large firms also benefited from alliance formation with DBF's because they could maintain a window of opportunity on biotechnology developments without the expense of investing too heavily to bring risky projects in-house. In return for investment large firms would receive licensing rights and could share knowledge created within the alliance. Only later when new trajectories in the development of biotechnology became clearer did firms choose to develop internal capabilities. Figure 1.1 below shows the growth of alliances between European pharmaceutical firms and DBFs in the 1990's.

Figure 1.1 The Growth of Alliances Between European Pharmaceutical Firms and DBFs



Source: p.12, Thomas S, Birtwistle N, Hopkins M & Simmonds N, (1998) The Impact of Intellectual Property on The Development of European Biotechnology, SPRU, Sussex University.

As Sharp predicted[34] alliance formation has occurred mainly between large pharmaceutical firms and DBFs. There is an absence of significant numbers of alliances between two large firms (other than mergers)[35] as may be seen in other industries such as consumer electronics (e.g. Sony and Philips in the case of CD development). However, large firms have been involved in significant consolidation activity, which has contributed towards an increased concentration in market share in an industry that has been historically relatively un-concentrated. As the data in figures 1.2, 1.3, and 1.4 below show, the concentration of market share held by the top ten pharmaceutical firms has increased by over 20% in the past nine years. Furthermore, the top ten firms in 1995/6 consists of 13 top 25 firms from 1989. More appetite for consolidation was shown by American Home Products, Glaxo-Wellcome, and Smithkline Beecham in abortive merger talks earlier this year. Although the trend of consolidation is slow relative to other industries (such as Aerospace or Electronics) over recent years, it is important to note increasing financial pressures on some of the world's largest firms.

[34] Sharp, M (1989) Collaboration and The Pharmaceutical Industry Is It The Way Forward, SPRU, Sussex university, p.27

[35] Thomas, Birtwistle, Hopkins & Simmonds, (1998) The Impact of Intellectual Property on The Development of European Biotechnology, SPRU, Sussex University.

Figure 1.2: Top Ten Pharmaceutical Firms in 1987 (by Global Market Share)

Position	Pharmaceutical Firm	Market Share (%)
1	**Merck**	**3.8**
2	**Hoecht**	**3.2**
3	**Glaxo**	**3.1**
4	**CibaGeigy**	**2.9**
5	**Bayer**	**2.7**
6	**American Home Products**	**2.6**
7	**Takeda**	**2.5**
8	**Sandoz**	**2.5**
9	**Eli Lilly**	**2.2**
10	**Abbott**	**2.1**
		Total
		27.6 %

Source: p.7 Sharp, M (1989) Collaboration and the Pharmaceutical Industry- Is It The Way Forward? SPRU, Sussex University

Figure 1.3: Top Ten Pharmaceutical Firms in 1989 (by Global Market Share)

Position	Pharmaceutical Firm	Market Share (%)
1	**Merck**	**4.5**
2	**Bristol Squibb**	**3.5**
3	**Glaxo**	**3.5**
4	**Hoechst**	**3.3**
5	**SmithKline Beecham**	**3.0**
6	**Bayer**	**2.8**
7	American Home Products	**2.6**
8	**Ciba Geigy**	**2.5**
9	**Eli Lilly**	**2.4**
10	**Sandoz**	**2.3**
		Total
		30.4 %

Source: adapted from p.3 Sharp(1990) Pharmaceuticals and Biotechnology: Perspectives for the European Industry, SPRU, Sussex University

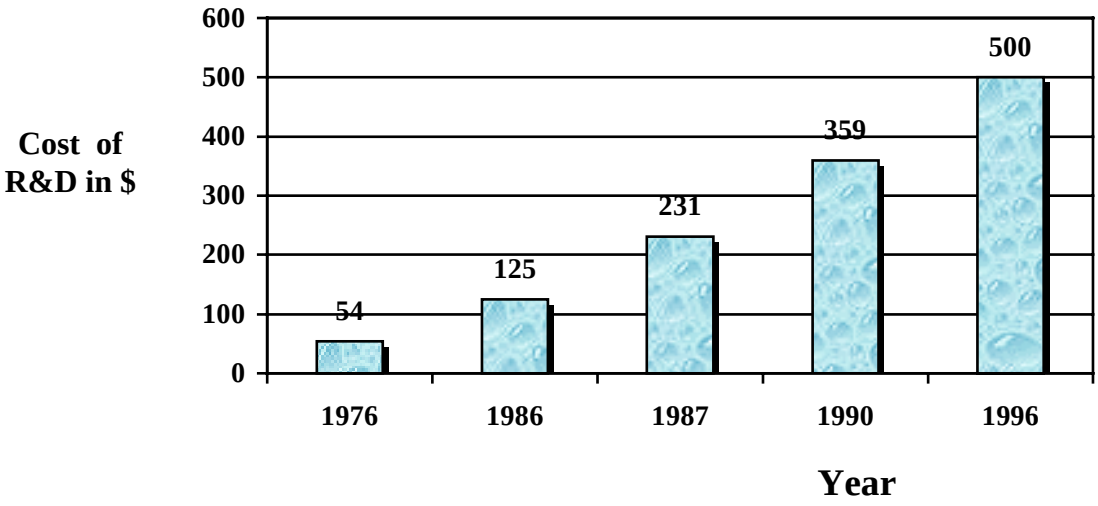
Figure 1.4: Top Ten Pharmaceutical Firms in 1995/96 (by Global Market Share)

Position	Pharmaceutical Firm	Market Share (%)
1	**Glaxo Wellcome**	**4.7**
2	**Novartis (Sandoz + Ciba Geigy)**	**4.5**
3	**Merck**	**3.5**
4	**Hoechst/ MMD**	**3.5**
5	**Bristol Myers Squibb**	**3.1**
6	**American Home Products**	**3.0**
7	**Pfizer**	**2.9**
8	**Johnson & Johnson**	**2.9**
9	**Roche**	**2.6**
10	**SmithKline Beecham**	**2.5**
		Total
		33.2 %

Source: p.9 Sharp, Patel & Pavitt (1996) Europe's Pharmaceutical Industry: An Innovation Profile SPRU, Sussex University

One of the factors driving this trend towards consolidation is the increasing cost of R&D. Although estimates of the true cost of bringing a drug to market are not always clear,[36] estimates vary between $100-400 million dollars per drug, and even more if the cost of paying for failed projects is factored into the R&D cost of successful products (See figure 1.5 below).

Figure 1.5 The Rising R&D Cost of Bringing A New Drug To Market



Source: p.1318, Drews & Ryser (1997) The role of Innovation in Drug Development, Nature Biotechnology Vol.15 No.13

This trend is likely to continue, given FDA plans to make drug licensing more stringent[37].

[36] Love, J, Call for More Reliable Costs Data on Clinical Trials, Marketletter, January 13th 1997, p.24 (also available at http://www.cptech.org/pharm/marketletter.html)
[37] Brower, Vicki, Biotech and Pharma Face More Costly Clinical Trials, Nature Biotechnology, vol.16 No. 8 p.714

Further financial pressure is placed on pharmaceutical firms through the cost cutting activities of healthcare organisations globally. These organisations seek to cut pharmaceutical expenditure as they become more stretched by aging populations and funding shortages. One cost saving strategy is to purchase cheaper generic copies of novel drugs, once these become available as patents expire.

Although consolidation in the industry may save a degree of expense, and combining product portfolios ensures protected revenue streams, this is horizontal integration, and large firms still require vertical integration to access new research techniques generated by biotechnology[38]. Acquisitions of DBFs by large pharmaceutical firms have occurred (for example Hoffman La Roche and Genentech in 1990, Rhone Poulenc Rorer and Applied Immune Sciences in 1993, Ciba-Geigy and Chiron in 1994, Glaxo and Affymax in 1995 all for sums between $100million and $2000 million[39]). However, alliance formation provides cheaper access to research, and so pharmaceutical firms are able to form portfolios of alliances.

This strategy for gaining access to new technologies is now widespread. Gambardella notes:

"In view of increasing complexity and multi-disciplinarity of knowledge, external information is critical to the development of innovations. What is nowadays more and more important is not the production of information but the dynamics and transformation of a larger pool of information. Information exchange, rather than retaining it within one's own organisational boundaries, is a major determinant of successful innovation. But this requires that one is prepared to diffuse research findings in exchange for the knowledge produced by others: To be part of a network, and to be able to effectively exploit the information that circulates in the network, has

[38] Sharp, M (1989) Collaboration and the Pharmaceutical Industry – Is It The Way Forward? SPRU, Sussex University.

become even more valuable than being able to generate new knowledge autonomously."

Source: p.32 Gambardella, A (1991) Competitive Advantages From In-House Scientific Research: The US Pharmaceutical Industry in the 1990's, IEFE, Universities of Bocconi

Additionally, competitive advantage may also be derived from the ability to follow external events and select the right alliance partners to benefit from technological advances, given there may be many potential DBF partners seeking alliance with any one large pharmaceutical firm. This is an ability firms might not be able to optimise without maintaining internal competencies in the field[40].

The rise of information intensive technologies such as genomics, proteomics[41], structural modeling and combinatorial chemistry[42], cause alliances to increasingly center on the exchange of vast quantities of data. These data are often difficult or even impossible to interpret without computer programs. Thus to effectively utilize data across large (sometimes global) R&D departments requires a complex computer infrastructure. Within this infrastructure, bioinformatics facilitates the use of new technologies that would otherwise overwhelm researchers with raw information. Like biotechnology before it, bioinformatics has been termed a paradigm shift[43]. It changes the fundamental way in which drug discovery is undertaken in almost every stage of the process, and without it the use of a host of other technologies cannot be optimised. Indeed as Gelbert and Gregg of Bristol Myers Squibb noted:

"No one technology has been or is anticipated to be, the major driving force behind the revolution in the discovery and development of new drugs.....The major scientific

[39] as listed on the biotech alliance database available at www.RECAP.COM.
[40] Granstrand O, Patel P, Pavitt K (1997) Multi-Technology Corporations: Why They Have Distributed Rather Than Distinctive Core Competencies, California Management Review Vol.39 No.4 p.8-25
[41] See glossary for definition
[42] See glossary for definition
[43] Lander, E, Langridge, R, Saccocio, D (1991) Mapping and Intergrating Biological Information, Communications of the ACM Vo1. 34, No.11

disciplines and technologies that are revolutionizing drug discovery are advances in cellular and molecular biology, the advent of combinatorial chemistry, the development of high-throughput screening[44], high resolution structural biology and data acquisition and processing systems to handle the ever increasing stream of information from biology, screening and chemistry."

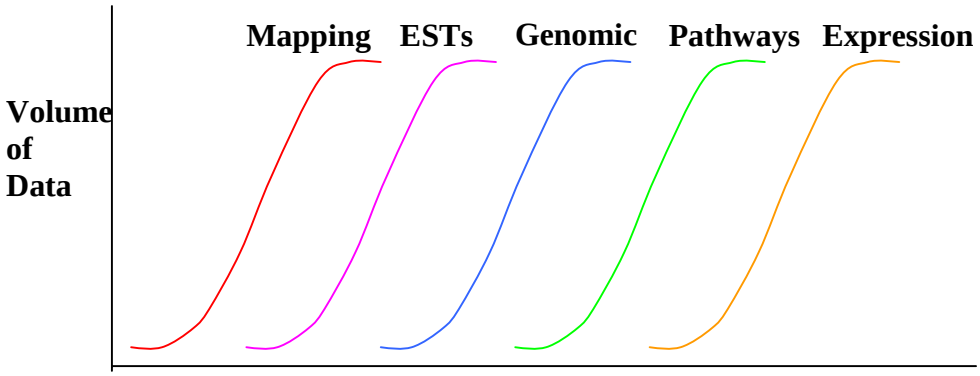
Source: p.669 Gelbert, L, and Gregg, R, (1997) Will genetics really revolutionize the drug discovery process? Current Opinion in Biotechnology, Vol.8

The benefit of these new technologies to pharmaceutical and drug discovery firms relies on developing a capability in bioinformatics. Again, as with the development of biotechnology, there is a shortage of trained bioinformaticians (those who can create, integrate and use specialist software tools). The demands firms place on their bioinformatics infrastructures are also complex, as the rate of progress in these fields is very rapid[45]. The nature of research questions posed changes quickly, and so the nature of data generated also changes quickly. This results in successive waves of data that must be fed into the systems databases for analysis as Lyall demonstrates in figure 1.6. Shown as several successive waves of data, new fields of interest may hold a focus of attention for a short period, rapidly generating large data sets. However the next fields focused on may generate very different types of data, which must be related to the previous set. For example to find the genes responsible for controlling a metabolic pathway, it might be necessary to draw on data from mapping, ESTs, and genomic data sets, but mapping information will be positional, where as

[44] See glossary for definition
[45] Gelbert, L, and Gregg, R, (1997) Will genetics really revolutionize the drug discovery process? Current Opinion in Biotechnolgy, Vol.8 p.669-674

genomic information will be related to gene sequence. The information can be related in complex databases, but failure to do so results in creation of islands of data.

Figure 1.6: Waves of New Data To Be Integrated Into Bioinformatics Systems



Source: p.57 Lyall, A, "Informatics – The Key to redesigning Drug Discovery" European Biopharmaceutical Review, March 1998.

Additionally challenges such as integrating new tools with each other and with existing processes, as well as training staff must all be taken into account. Although he does not mention bioinformatics or genomics in his book on process innovation in the pharmaceutical industry, Pisano notes there is little discussion in current strategy literature of how a strong process development capability in the pharmaceutical industry may strengthen a firm's competitive ability[46]. Certainly such abilities would be important in solving the problems bioinformatics poses:

- The need to restructure existing routines to incorporate new techniques.
- Complex nature of data produced and their interpretation.
- Rapid pace of change, requiring new software tools and robust data storage methodologies.

[46] Pisano, G (1997) The Development Factory, Harvard Business School Press, p.3

- Complexity of integrating software tools together to form suites of tools and databases.
- Cultural problems surrounding re-training of old staff and recruitment of new staff.

Firms that fail to demonstrate an ability to manage these changes may find the integration of bioinformatics difficult, and thus will not realize the full potential of the new technologies associated with it. Given the increasingly competitive environment they inhabit, this may prove to be a very important factor in determining their future success.

1.6 Methodology

To answer the questions posed in section 1.1, data were gathered from both primary and secondary sources. The primary sources for interview were nine organisations involved with bioinformatics who provided a total of 12 interviewees (Appendix 1). Secondary data from numerous publications and web sites were also used.

1.6.1 Interviews and Their Limitations:

Given the finance and time constraints of this project only a small number of UK based interviewees were selected. Due to these limitations, interviews were arranged with a representative cross section of the bioinformatics user community in order to establish current opinions in the field.

Five firms were interviewed. These included three large fully integrated pharmaceutical firms referred to as A, B and C one medium sized fully integrated pharmaceutical firm, D, and a small DBF specializing in drug discovery, E. Additionally a representative from one small firm, F, specializing in bioinformatics

development and consultancy was interviewed. Given the importance of public research organizations in the growth of bioinformatics, the cross section also included two representatives of the academic research community which were the biochemistry departments at UCL and Sussex University. Finally, researchers from a government research council, the Biological and Biotechnological Science Research Council (BBSRC), responsible for reviewing the funding of bioinformatics were also consulted.

Interviewees were contacted primarily by letter and then by telephone and E-mail to confirm meeting details. Interviews were scheduled at 45 minutes, however the majority of interviewees generously gave additional time, exceeding 2 hours or more in some cases. Structured questionnaires were used as a basis for all interviews, although interviewees were encouraged to speak more broadly in areas they saw to be important. Two slightly different questionnaires were used, one for pharmaceutical firms and one for all others (see Appendix 2 & 3).

It is accepted that the views of those sampled may not reflect the views of the wide range of organizations involved with bioinformatics. Thus, every care has been taken in the presentation of the data to avoid generalizations and assumptions based on the opinions of a single individual or organization. Additionally it must be noted some questions asked could not be fully answered for reasons of commercial sensitivity.

1.6.2 Secondary Data and Its Limitations:

Some firms involved in bioinformatics, including some of those interviewed have published information about their bioinformatics programs or strategies. This information has been used to supplement interview findings. In particular DBFs involved in genomics and combinatorial chemistry have been keen to publish accounts of their new technologies. However, it must be accepted that a number of articles in this field are written more as advertisements rather than critical analyses of the changing nature of technology in the drug discovery process. In particular, care must be taken to separate the facts from over enthusiastic hopes for which biotechnology has become notorious.

The US Patent and Trademark Office database was also used to triangulate patterns of firm activity in computational biology and chemistry apparent from literature and interviews. More detail on how these searches were carried out is given in Chapter 2.

Chapter 2: Mapping Changes In Pharmaceutical R&D Processes

2.1 Introduction

The first key question introduced in Section 1.1, i.e. "how has bioinformatics been integrated into the drug research and development processes of the firm" poses two "hows". The first "how" relates to how new technologies sit within the existing framework of drug discovery, and is of a technical nature. The second "how" is addressed in chapter 3, and relates to organisational issues.

This chapter introduces the nature of change in pharmaceutical R&D. It identifies broadly the areas where new technologies have allowed change and explains how limiting factors in the R&D process have been affected. These are related to specific frameworks which authors have previously developed to describe the R&D process.

2.2 Traditional Routes to Drug Discovery

Discovering a compound with therapeutic effects in animal models either from random screening or from clues given from traditional remedies used to be the classical route to drug discovery. This route has been used since the "golden age" of the 1950's and 1960's. The main competencies involved were chemistry and pharmacology. Finding targets, i.e. the point at which therapeutic agents should intervene for treatment of a disease, would be the limiting step. Close integration of pharmacological and chemical competencies were needed as candidates for lead compound selection underwent validation[47] and toxicology tests before being passed on to wider trials. Often firms would accept targets derived from animal models[48] developed outside the firm, from universities or other research organisations, although other stages of the process would be continued in-house[49]. An additional limiting step

[47] See glossary for definition
[48] See glossary for definition
[49] B2 , Interview

is identified by Jean[50], who notes the average organic chemist could synthesis only 50-100 compounds per year for testing, against the limited targets available. These constraints governed that the overall process was low throughput.

All the integrated pharmaceutical firms interviewed previously subscribed to this methodology, although they noted that the process did not always begin from the same stage, and one firm stated their discovery program previously had concentrated on "busting patents" of other firms to invent new patentable versions of existing drugs. This was and still is a common strategy, with lower risk of failure. The so called "me too" strategy is used to highly profitable effect in Japan's pharmaceutical industry[51].

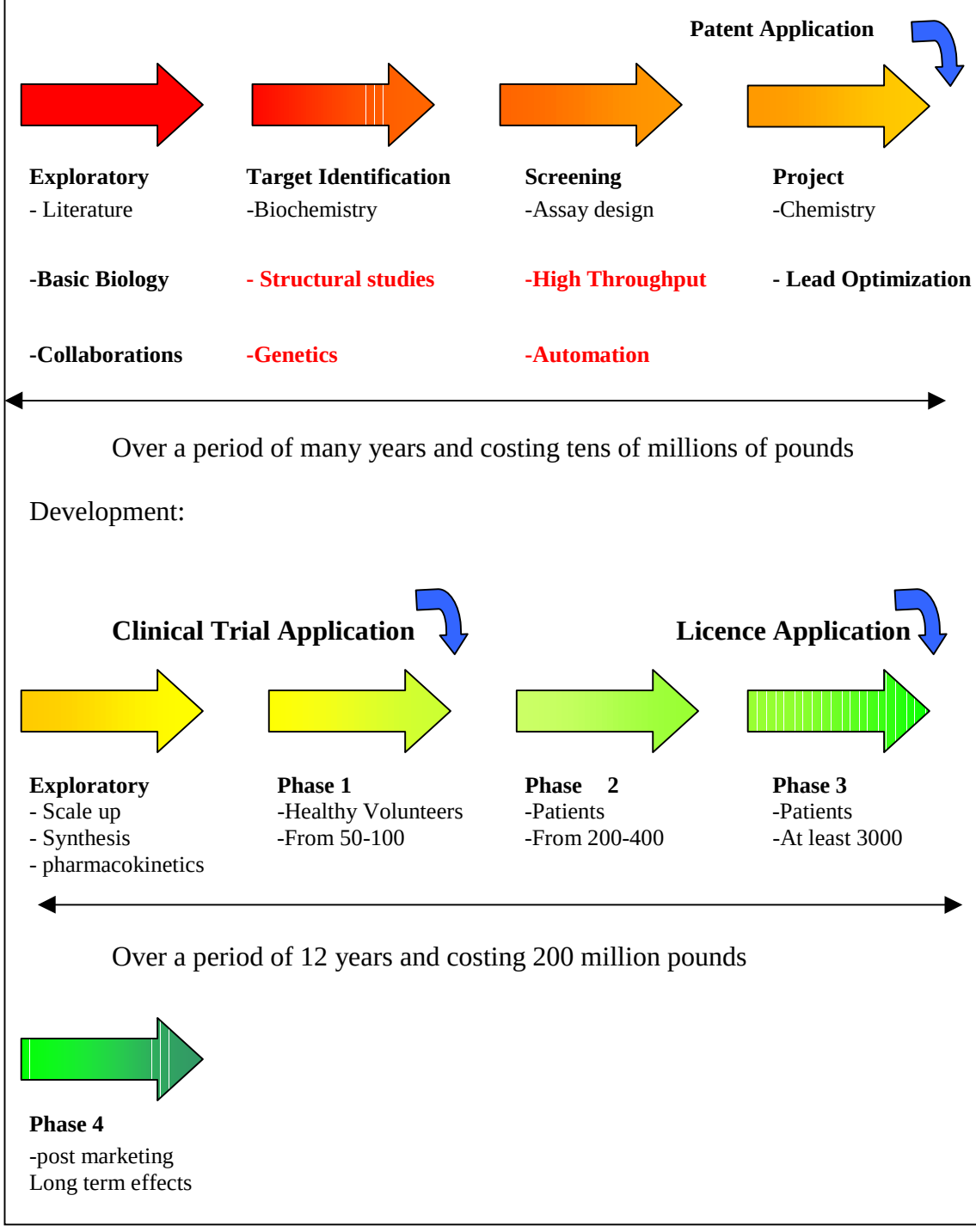
2.3 Present and Future Routes To Drug Discovery

As new molecular biology and genetic techniques have become available these have been integrated into the R&D process, along with increasing drives towards automation of target searches and screens. Figure 2.1 below gives an indication of how these new and old processes fit together in drug R&D programs. Figure 2.1 shows some new technologies, in red text, appearing to enter the process. Although this process seems linear and continuous, the odds of a newly found molecule making it through all these stages and becoming licensed for therapeutic use are slim, at 5,000 to 1, and 80% of leads entering development fail during clinical trials[52]. Drug discovery had traditionally begun with an animal model of disease, which might be used to yield a target.

[50] Jean, T, Drug Discovery CROs, European Pharmaceutical Review, September 1997, p.36-41
[51] Chisaki, O (1998) An Overview of the Japanese Biopharmaceutical Industry, European Pharmaceutical Review March, p.13-19.
[52] Anon, (1997) Testing, Testing, The Economist Vol.342, 1st February p.1847

However the advent of genomics has reversed this process. Instead a target can be found through DNA sequence analysis, and animal models are only introduced

Figure 2.1 A Typical Drug R&D Process In A Contemporary Integrated Pharmaceutical Firm



Source: p.56 Lyall, A, "Informatics – The Key to redesigning Drug Discovery" European Biopharmaceutical Review, March 1998.

later. This has moved the focus of the R&D process away from developing animal models as a primary route to establishing targets. Firms may now identify many targets first and then choose which ones they which to validate and pursue. All firms interviewed have developed competencies in molecular biology over the past decade. It is now widely acknowledged as a core competency[53]. However, the degree to which firms are accepting other changes seems to vary. Firm A and Firm B have begun to restructure their whole drug discovery programs around genomics (providing targets), combinatorial chemistry and high throughput screening (selection of leads to act on these targets). They have both invested heavily in bioinformatics professionals to support these new systems with an IT infrastructure that facilitates the highest possible throughput of targets and drug candidates. Firm C and Firm D have been more cautious in their approach. This is reflected in their smaller bioinformatics staff numbers (figure 2.2) and by the continued use of traditional techniques. However all firms interviewed do use bioinformatics substantially to mine DNA databases for targets.

[53] A1, Interview

Figure 2.2 Table Showing Selected Interview Results

Name	Firm A	Firm B	Firm C	Firm D	Firm E
Size	**Large**	**Large**	**Large**	**Medium**	**Small**
Description	**Pharmaceutical firm**	**Pharmaceutical firm**	**Pharmaceutical firm**	**Pharmaceutical firm**	**Drug Discovery**
R&D spending [54]	**$1,830 million**	**$1,270 million**	**$1,522 million**	**$45 million**	**$15.5 million**
R&D employees[55]	**10,000**	**5000**	**1800**	**-**	**60**
Dedicated bioinformaticians	**up to 60**	**up to 70**	**12**	**up to 5**	**3**
External data bases mirrored on site	**Yes**	**Yes**	**Yes**	**Yes**	**No**

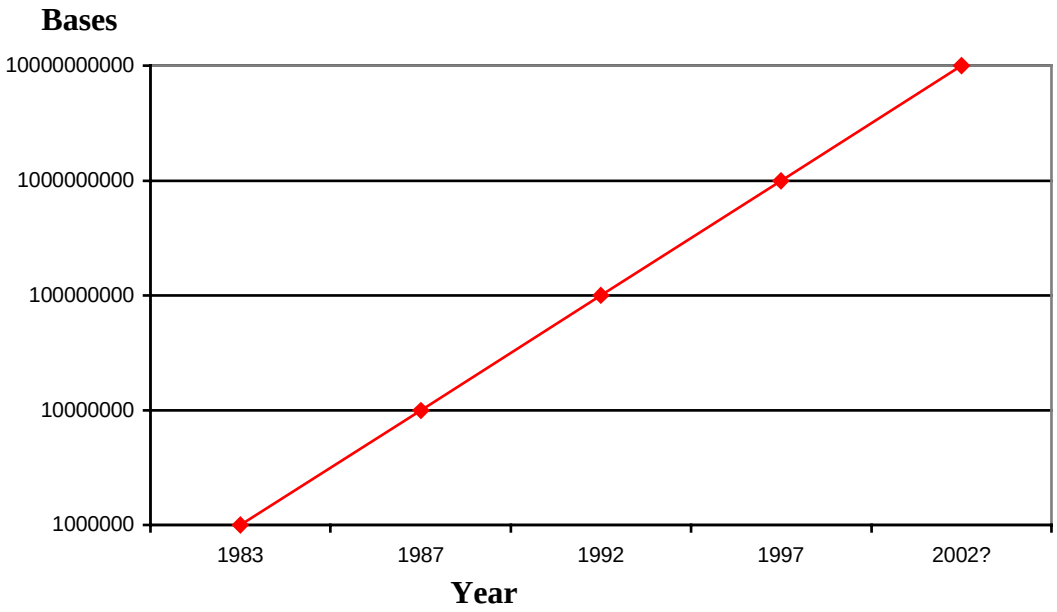
[1& 2]Based on 1997 annual reports

Figure 2.2 also shows that in the search for targets all firms interviewed, except Firm E, own copies (called mirror sites) of private and public databases which are kept in house for several reasons:

- In house database searches may be faster, without traffic problems/ busy lines
- Searches and areas of interest are hidden from the view of outsiders
- It is widely felt that those who can afford to buy material should do so to support academic research.

Such is the demand for genomics data of this kind, Incyte Pharmaceuticals Inc. (US) have sold subscriptions worth millions of dollars to 17 firms including six of the top ten pharmaceutical firms (as measured by sales in Scrip magazine, January 1997). These subscribers include two of the three large firms interviewed. However with academic databases growing, and search strategies improving Firm B has recently cancelled one similar alliance with Human Genome Sciences. Figure 3.2 shows the growth of DNA data stored on one widely used database, Genbank.

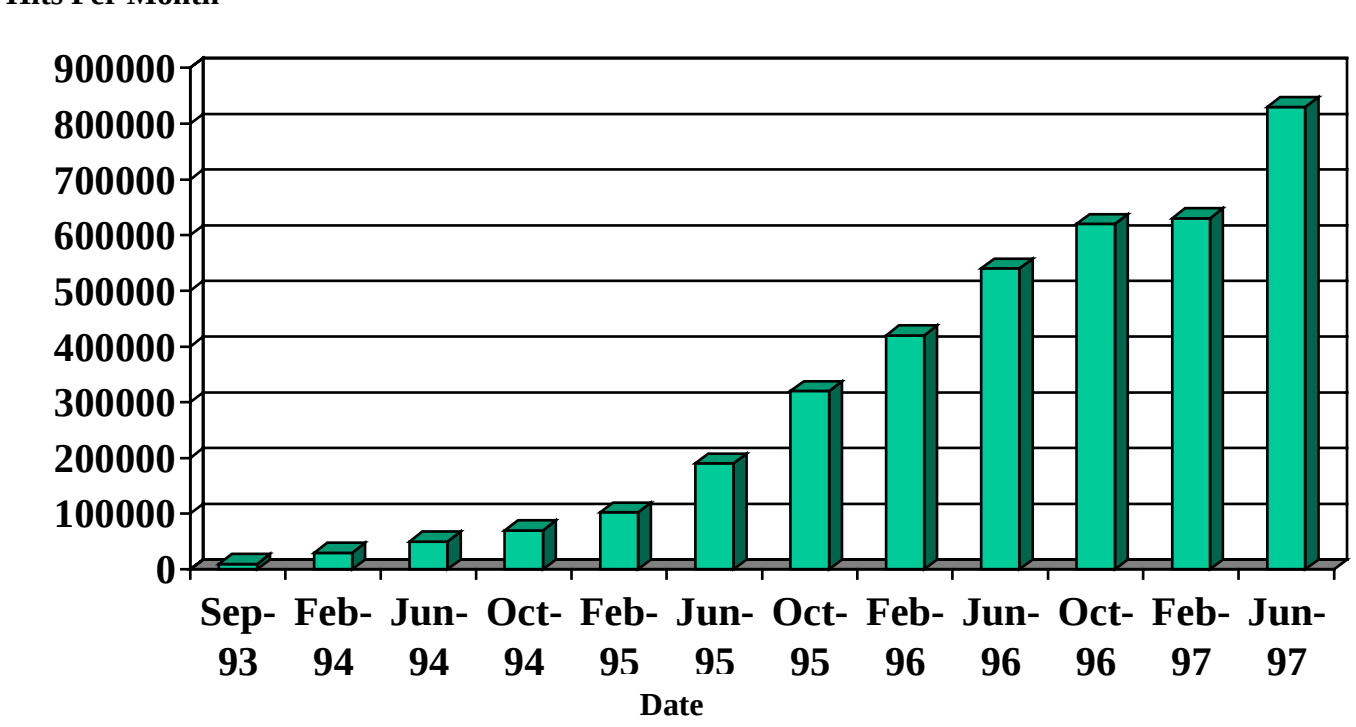
Figure 2.3 Future Projection of DNA Bases stored on Genbank Based On Previous Growth



Source: p.929, Anon , Business and regulatory news, Nature Biotechnology, Vol. 15 No.10, 1997

Figure 2.4 shows the increasing number of "hits" from individuals visiting the SwissProt on-line database, where protein sequence and structural information can be obtained. Although these hits presumably do not include searches made on mirror sites in large firms, it does show the growing importance of remote sources of data in biological research related to drug discovery.

Figure 2.4 Growth In The Number of Hits Registered On The SwissProt Database



Source: p.1255 Sansom, C, Finding A New Language For Bioinformatics, Nature Biotechnology, Vol. 15 No.12,1997.

Interviewees confirmed remote data as becoming increasingly important in the early stages of drug discovery. In particular, it was highlighted searches may negate the necessity to perform many experiments, as the experiment may have been performed elsewhere, and results might be available through on-line searches[56]. Also the importance of good informatics for literature searches should not be overlooked[57]. Increasingly journals are available in electronic form, and where they relate to sequence or structural findings, these may not even be published in hard copy at all. Electronic resources dramatically reduce the amount of time staff have to spend on basic literature searches, and allow access to a growing volume of material otherwise unreachable. Genomic and proteomic databases are interrogated using a variety of software tools, available through academics (usually for a small license fee), software suite suppliers (eg. Genetic Computing Group, GCG, and Molecular Simulations Inc, MSI), consultants, and in-house development.

Further variations in firm strategies are apparent from the ways in which firms use these software tools. Whilst all interviewed firms have these tools available to all research staff on intranets or local computer terminals, only Firm A and Firm B are attempting to integrate the tools themselves. By doing so, they hope to make them more user friendly, give them wider scope, and allow the smooth transition of data from one system to the next. This should allow the comparison of large complex data sets automatically where other firms make do with manual comparison. This should allow unique insight and understanding of the biological systems they research, which might be missed by those using non-integrated systems.

Whilst genomic data is mainly used for finding targets at present, proteomic data and structural data relating to small molecules, may be used later in the R&D process, for

[56] A1 , Interview
[57] F1, Interview

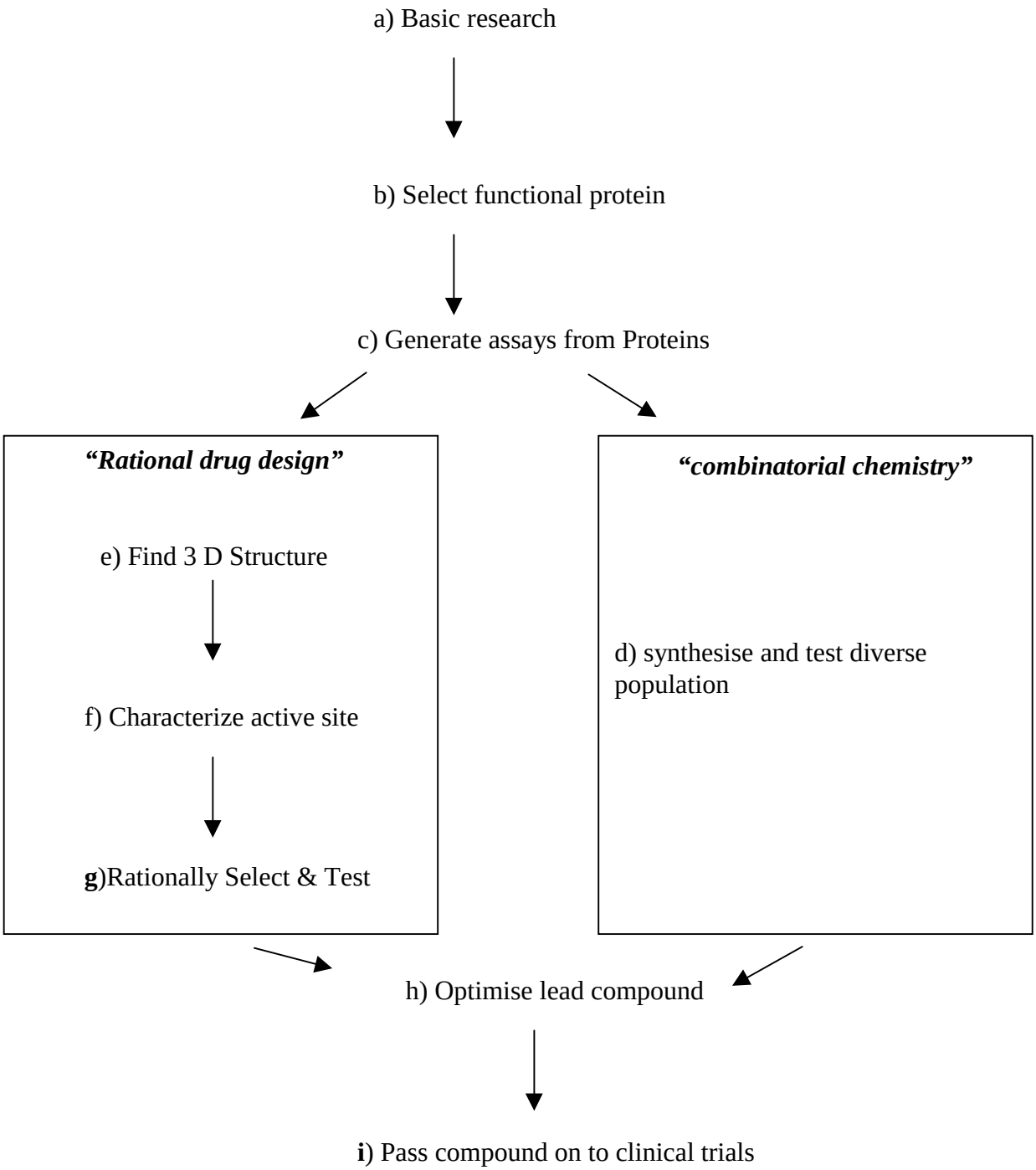
rational drug design, where models can be used for docking studies of drugs and their points of action. Once a target has been identified, screens and rational drug design are used to produce a lead candidate molecule that acts on it.

Nightingale has put forward a framework that takes into account new technologies which allow higher throughput of more effective candidates (figure 2.5) where rational drug design is applied as well as combinatorial chemistry to optimise leads.

Another author, Jean, demonstrates how the stages Nightingale identifies as d) and h) can be run by HTS and combinatorial chemistry combined with robotics and bioinformatics once a target has been identified by stages a) and b). Importantly, Jean's Model (see figure 2.6) also reflects the non-linearity of the R&D process, which Lyall's macroscopic overview does not focus on. Firm E's principle scientist also emphasized the starting point for projects may differ on a firm by firm or even a case by case basis[58]. Jean indicates the stages in which robots are used to facilitate high throughput (in blue). Robots necessitate the use of complex data handling informatics to support them. The processes set out by Jean are not the universal choice for the firms interviewed. Firm A and Firm B seem to be heading towards Jean's route, whilst others have been more cautious in their approach to combinatorial chemistry.

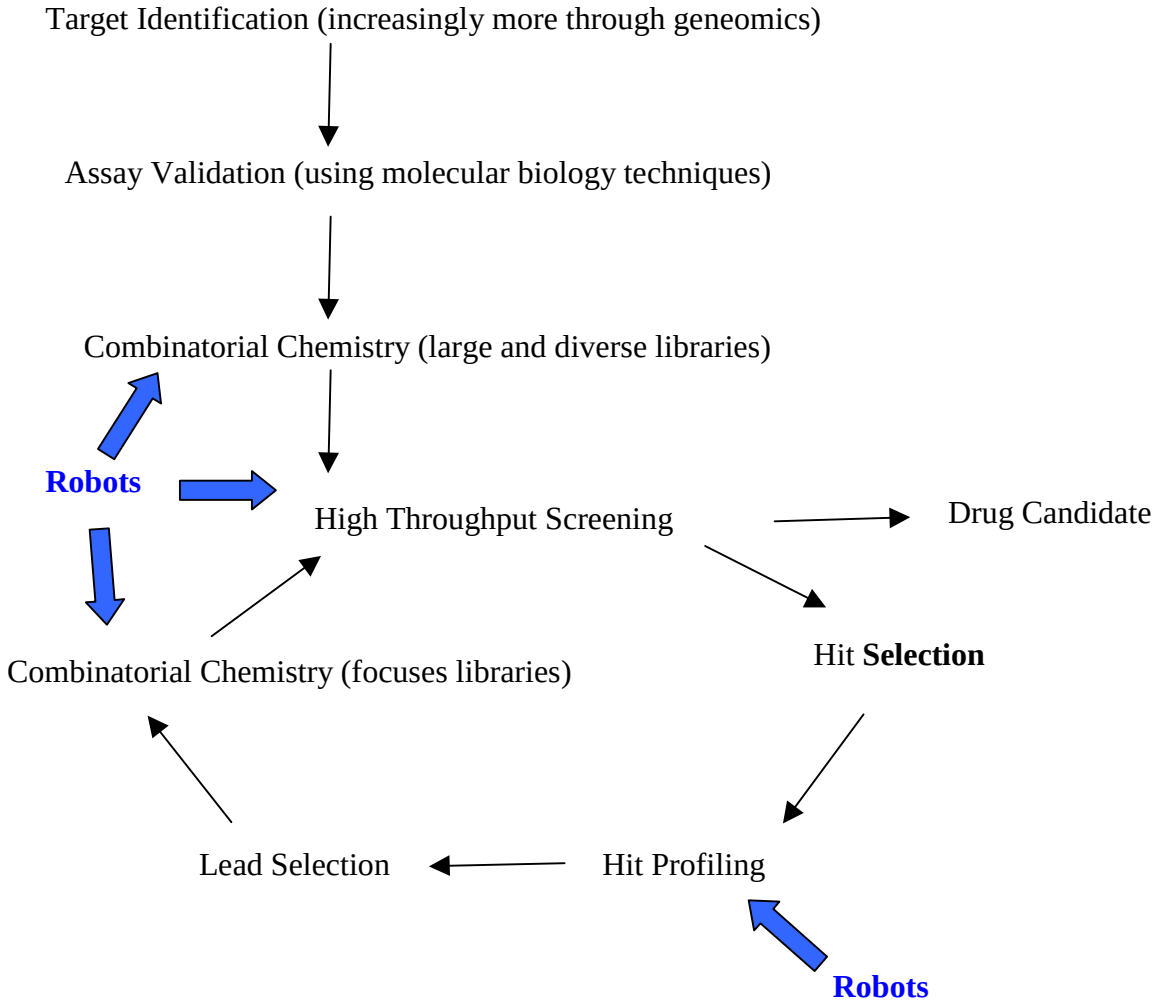
[58] E1, Interview

Figure 2.5 Nightingale's Model of Drug Discovery



Source: p.122, Nightingale, P (1997) Knowledge and Technical Change: Computer Simulations and the changing Innovation Process, University of Sussex

Figure 2.6 Jean's Model of The New Drug Discovery Process



Source: p.37 Jean, T, Drug Discovery in Contract Research Organisations, European Pharmaceutical Review, September 1997

Firm D mentioned they could always outsource this capability in the future, as there seems to be a growing number of firms specialising in this area. They went on to reveal they are continuing to use the low-throughput traditional methods[59].

[59] D1, Interview

Figure 2.7 shows the cost saving and higher rate of throughput allowed by combinatorial chemistry.

Figure 2.7 "The Power of Combinatorial Chemistry"

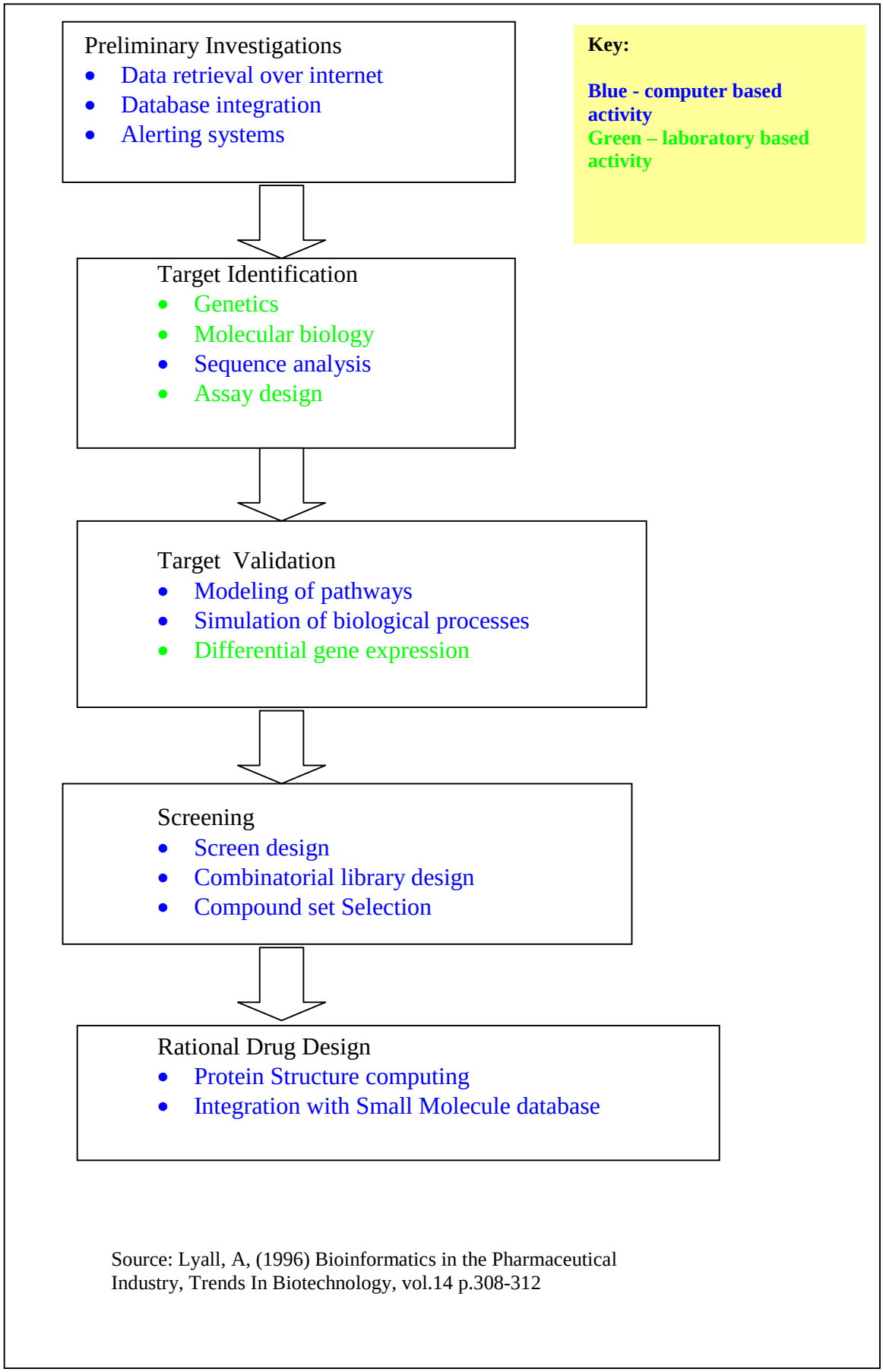
	Traditional Chemistry	**Combinatorial Chemistry**
Compounds per chemist/ month	4	3,300
Total cost	$30,000	$40,000
Cost per compound	$7500	$12

Source: p.691, Persidis, A, Combinatorial Chemistry, Nature Biotechnology, Vol.16, July 1998

Nightingale provides further evidence of the integration of combinatorial chemistry, in this case being used in Zeneca (not interviewed). Here combinatorial chemistry has allowed screening to increase from 160,000 compounds over 16 years leading up to 1992, to 500,000 compounds a month in 1995[60]. Lyall (figure 2.8) has predicted how the R&D processes may look in a typical fully integrated pharmaceutical firm of the future, both upstream and down stream of the areas Jean's model focuses on.

[60] p.120, Nightingale, P (1997) Knowledge and Technical Change: Computer Simulations and the changing Innovation Process phD Thesis

Figure 2.8 Lyall's Vision of Future Drug Discovery



Source: Lyall, A, (1996) Bioinformatics in the Pharmaceutical Industry, Trends In Biotechnology, vol.14 p.308-312

If all the technologies shown in figure 2.8 were established in house, then around two thirds of the stages listed would be carried out on bioinformatics systems. Firm A and perhaps others aim to integrate as many of these as possible. Thus it would be conceivable in the future for a researcher to identify a target gene, find its sequence and predict a likely structure for the protein it represents, and then find a family of compounds likely to show activity on the protein, without leaving their computer[61]. Others take the view that this is unlikely for many years, if it becomes possible at all[62] because the very slow X ray-crystallography technique is currently the only accurate process that can predict protein structure, and computer modeling may never be sophisticated enough to replace this technique entirely. However X ray-crystallography can be used to study homologies between structures, which can then be related to sequence data through infomatics. Thus the exact structure may not be known but a "best guess" may be available and this will be enough to bias screening, saving valuable resources[63].

2.4 Changes In The Limiting Step of The Drug R&D Process

Even though full integration has not been reached, the limiting stage of the discovery process has already been changed in many firms. The majority of interviewees in firms and academia see validation and "wet biology" as continuing to be a limiting step. In the validation stage molecular biologists must demonstrate by experiment that acting on the protein or gene selected will have the desired result, and once a lead compound has been found through screening, testing it in an animal model will also prove to be a limiting step. B1 described these stages as "bespoke biology", with no

[61] A1, Interview
[62] Woolfson, Interview
[63] Nightingale, P (1997) Knowledge and Technical Change: Computer Simulations and the changing Innovation Process, SPRU Sussex University

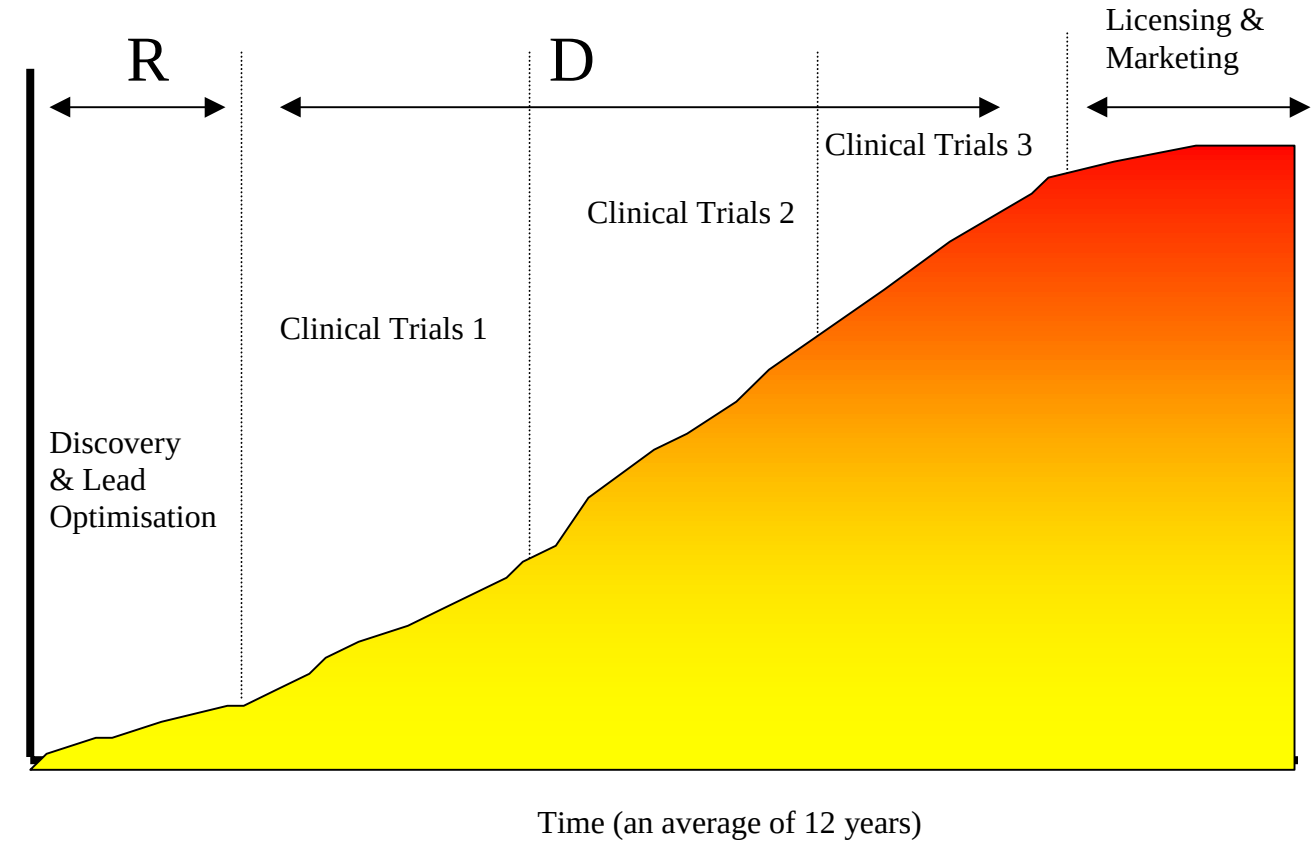
short cuts as yet, although other interviewees hinted that a degree of computer simulation might be applied in these areas in the future.

The overall effect of these process changes has been to greatly reduce the time taken for a research project to move through its initial stages, and for the volume of leads generated to increase. However, the leads may be more numerous, but they are not necessarily of as high quality (likely to succeed) as leads developed through traditional processes[64]. Another bottle neck interviewees agreed on was the boundary between research, where a candidate for trials is discovered and optimised, and development, where the candidate is prepared for scaled up production and tested in clinical trials. At this stage candidates may fail simply because the cost of mass producing them is uneconomical[65]. Furthermore trials are very expensive to run, and so only the highest quality candidates are entered into this process. Failure at a late stage in trials may cost a firm tens or even hundreds of million dollars. Figure 2.9 shows roughly the cost structure of the whole R&D process. Cost on the diagram is not indicated as a figure, due to the variability of this between projects and firms. The important point to note however, is the tiny proportion of cost spent on research, where all the process innovation and bioinformatics developments discussed here is incurred. The overall cost per drug on development is not expected to fall as a result of bioinformatics, or other new technologies used. The cost saving should result from the greater knowledge of targets and candidates, and knowledge metabolic pathways surrounding them, which will allow researchers to select candidates more likely to pass the gauntlet of trials successfully and thereby reduce the numbers of costly failures.

---

[64] B2, Interview
[65] C1, Interview

Figure 2.9: The Cost Structure of Bringing a New Drug to Market



Source: adapted from a sketch by B1 at Interview

It should also be noted that bioinformatics or other new technologies are not expected to reduce the time it takes to bring a compound through the development stage. The clinical trials structure must adhere strictly to regulatory guidelines, and so the trial periods are not flexible.

Chapter 3: Strategies For Managing The Growth of New Technological Competencies In Bioinformatics

3.1 Introduction

This Chapter addresses issues related to the second key question posed in section 1.1 which was: what strategies have pharmaceutical firms used to acquire capabilities in bioinformatics? Additionally, the organisational issues surrounding the first key question of "how bioinformatics has been integrated into the pharmaceutical R&D process" are also addressed.

The findings of interviews conducted on the cross sectional group of firms, academics and others mentioned in section 1.6, as well as data from secondary sources are examined in relation to notions in the technology strategy literature. These notions relate to the development of technological competencies and the maintaining of competitive advantage in firms.

3.2 A Review of Frameworks Used

Three theoretical frameworks from technology innovation management literature are referred to in support of the argument outlined in Section 1.2. The key points relevant to this study are outlined here.

3.2.1 Autonomous and Systemic Innovation

Chesbrough and Teece[66] examine the notion of outsourcing, and propose a framework to determine when "virtual" organisations networked to external competencies might succeed over less flexible larger organisations. They identify two type of innovation, autonomous and systemic. Autonomous innovations can be used without effecting

[66] Chesbourgh, H, & Teece, D (1996) When is Virtual Virtuous? Harvard Business Review, January – February, p.65-74

other surrounding systems (e.g. The introduction of a turbocharger for a car engine does not require the rest of the engine to be designed). On the other hand systemic innovations have complex effects on surrounding systems when they are introduced (e.g. when Polaroid introduced their new film, it needed a new camera to support it.) Chesbourgh and Teece propose that if the technology desired exists outside and is autonomous in nature, it may be suitable for outsourcing where as technologies that must be created, and are systemic should be developed in-house to ensure they are compatible with other systems. Figure 3.1 demonstrates the choices facing a firm, and the suggested action to take.

Figure 3.1 Matching Organisation to Innovation

	Innovation is autonomous	**Innovation is Systemic**
Capabilities needed exist outside	**Go virtual**	**Ally with caution**
Capabilities needed must be created	**Ally or bring in house**	**Bring in house**

Source: p.73 Chesbourgh, H, & Teece, D (1996) When is Virtual Virtuous? Harvard Business Review, January –February

3.2.2 Multi-Technology Firms and Distributed Competencies

Granstrant, Patel and Pavitt[67] describe multi-technology firms as having distributed competencies, often broader than represented by their products. Furthermore they maintain that over time it becomes necessary to develop new technological competencies to provide the driving force for corporate growth, increasing R&D investment, increasing external linkages, and allowing new technologically related business opportunities to be grasped. They argue firms developing these new technological competencies firms should be wary of four previously accepted management concepts. These are:

- firms should focus on their core competencies,
- major new innovations often bring about "competence destruction",
- firms should downsize and outsource, and that
- firms should specialize on a narrow set of core businesses.

In addressing these points, they highlight two methodologies for measuring technological competencies. One method is to measure the capacity of firms to achieve a certain level of functional performance in a generic product or system. This method is difficult to measure by quantitative means, and so the authors chose to use another methodology for their study, the measurement of patenting activity by technology class as recorded in the US patent and trademark office database.

To identify and examine holders of technological competencies in bioinformatics, I used the qualitative interview approach as well as attempting the quantitative patent measuring approach, and the findings from these are discussed in section 3.3 and section 3.4 respectively.

3.2.3 Dynamic Capabilities and Competitive Advantage

Teece and Pisano[68] describe a framework of characteristics associated with firms that possess what they call dynamic capabilities. Dynamic capabilities refer to the management strategy that allows certain firms to:

"reconfigure internal and external organisational skills resources and functional competencies towards a changing environment"

Source: p.537 Teece, D & Pisano, G "The Dynamic Capabilities of Firms" Industrial & Corporate Change, Vol.3 No.3, 1994, Oxford University Press.

The relevance of these capabilities to bioinformatics is that restructuring of core processes within pharmaceutical firms is apparently occurring, to take advantage of new technologies and overcome environmental pressures. Teece and Pisano explore the theories of previous authors, who take the view that competitive advantage stems from exploiting internal and external firm specific capabilities and developing new ones. They then examine how some firms are able to renew capabilities to adapt or even capitalize on rapidly changing markets. Although they identify factors important to product innovation through times of environmental change, these factors are equally valid when applied to process innovation in pharmaceutical R&D.

Teece and Pisano break down the relevant characteristics into processes, positions and paths.

- **Processes** focused on are organisational and managerial. They relate to the firm's ability to integrate, learn, reconfigure and transform.

---

[2]Granstrand, O, Patel, P, & Pavitt, K (1997) Multi-Technology Corporations: why they have Distributed rather than distinctive core competencies, California Management Review, Vol. 39, No. 4, p8-25.

[68] Teece, D & Pisano, G "The Dynamic Capabilities of Firms" Industrial & Corporate Change, Vol.3 No.3, 1994, Oxford University Press, p.537-556

- **Position.** The technological assets such as know-how or patents, complementary assets such as distribution networks or customer relations, financial assets, and even locational assets, i.e. geographical position.
- **Paths** refer to the choices available to a firm for future development, which are shaped by path dependencies and technological opportunities. Path dependency refers to the notion that firms previous experience helps or hinders them, thus determining their options. Not all paths will be available, due to core rigidities, and firms may prefer to do in the future what they have done best in the past. The number and scope or technological opportunities available to a firm also depends on the R&D capability it may posses, or links with external R&D. Thus firms will have different options open to them.

Teece and Pisano also suggest that the key to maintaining competitive advantage is having a capability that is difficult to imitate, and is supported by intellectual property rights. Section 3.5 relates Teece and Pisano's work to the development of bioinformatics within the pharmaceutical industry.

3.3 Assessing Competency In Bioinformatics Though Qualitative Information From Interviews

Difficulty in establishing the existence of bioinformatics competencies in pharmaceutical firms from patent data meant that more qualitative information from interviews was required to identify the extent of expertise held by these firms.

Granstrand, Patel, and Pavitt identify four stages that lie in between full integration of a technology and full-scale disintegration, where purchase of products or services relating to a technology are produced externally. By using these four stages as a

basic scale it is possible to assign a rough quantitative measure to the progress of interviewed firms in achieving a capability to create, integrate and use software tools. The stages on the scale are:

1. Full design capability – competency to design and test products comparable to those produced externally for example the ability to design bioinformatics software suites as available from specialist external suppliers.
2. Systems integration capability – competency in integrating changes and improvement in internally or externally designed or produced systems or products, For example, the ability to modify software tools and integrate them together to optimise their use within the firm.
3. Applied research capability – competency to specify and purchase externally produced products or systems, and to control changes and improvement relating to them. Thus the firm may have sufficient knowledge of software systems to allow their use within the firm, and to modify or inform external suppliers how best to modify tools to meet the firms' needs.
4. Exploratory research capability – competency in identifying and evaluating new opportunities available from a new technology, and integrate new opportunities with current activities, for example, recognizing the importance of bioinformatics and being able to introduce some software tools to enhance R&D.

At interview questions were posed to establish which of the four stages firms intended to reach and the strategies they used to achieve their goals with respect to bioinformatics.

Questions posed are shown in fig 3.2. It is clear that the firms have achieved differing levels of competency, but identifying which of the four stages above they fit with best is not straightforward. Furthermore it must be accepted that firms seek to achieve

different goals to meet their specific needs. Whilst all interviewed firms seek to discover new drugs, the therapeutic areas in which they hold expertise differ. Thus whilst a larger firm might be able to develop drugs in a number of therapeutic areas, a smaller firm may only be looking for drugs to treat neurological diseases or circulatory disorders in which they are specialised. These factors inevitably effect the strategy they might use to find and select targets, and in turn the types and characteristics of the software tool they require to meet these needs.

Firm A and Firm B require a high throughput system to meet their target of producing several successful new drugs each year. Thus both firms have invested heavily in attempting to gain the capability to integrate existing systems and tools to compare large and diverse data sets. They also possess the capability to design their own tools, however they still require software from external sources, and would be hard pressed to design many of the products designed externally. Thus they might be somewhere in between stage 1 and 2 on the scale. Pfizer has had some success with producing software tools, but also uses many externally produced tools. However they are not attempting the level of software integration planned by Firm A and Firm B, thus Firm C might be placed slightly lower than Firm A and Firm B, but still between stages 1 and 2 on the scale. Firm D and Firm E have more focused needs regarding drug candidates, and are not seeking high through put integrated systems. Both use externally produced tools. They do not produce their own software tools, although they may slightly modify existing software to fit with their needs. Neither firm has plans to integrate tools in the way Firm A or Firm B propose. Firm C and Firm D would probably sit around stage 3.

Figure 3.2: Comparing Competency Building Efforts - Selected Interview Results

Name	Firm A	Firm B	Firm C	Firm D	Firm E
Years since entry into bioinformatics	**10 years**	**7 years**	**6 years**	**3 years**	**3 years**
Training courses in house	**Yes**	**Yes**	**No**	**No**	**No**
New software tools developed in house	**Yes**	**Yes**	**Yes**	**No**	**No**
Software integration undertaken in house	**Yes**	**Yes**	**No**	**No**	**No**
Co-development alliances with software providers	**Yes**	**Yes**	**Yes**	**No**	**No**
Identified a shortage of bioinformaticians	**Yes**	**Yes**	**Yes**	**-**	**Yes**

Elements of strategy used by these firms to build up their competencies were also identified (figure 3.2). The pattern and approach to recruitment of dedicated bioinformatics staff, mainly from academic backgrounds, is a clear indicator of how firms see bioinformatics changing the R&D process internally. Firm A and Firm B have established internal bioinformatics training programs to facilitate the learning of new techniques amongst the bench scientist user community. However at the other end of the size scale, Firm E take the view that such programs are not necessary due to the more effective diffusion of such tacit knowledge through a small and informal firm, which also has close links to the near by University. Other indications of competence building are shown by efforts to enter co-development with suppliers of software solutions, such as alliance Firm A has with Oxford Molecular in which several Oxford Molecular staff work within the firm semi permanently. Firm D and Firm E do not have such alliances and are content merely to use existing tools, although they may co-operate with suppliers of software solutions in testing and feedback exercises, which might result in favourable licensing agreements later on.

Figure 3.3 summarizes the grades firms have been given on the scale and the key reasons for placing them accordingly.

Figure 3.3 Grading Interviewed Firms By Competency in Bioinformatics

Grade	Firm	Qualifying Characteristics
High Grade 1-2	Firm A Firm B	Relatively long presence High resource commitment Developing software in-house Attempting systems integration Co-development alliances
Lower Grade 1-2	Firm C	Relatively long presence Medium resource commitment Developing software in house Not attempting systems integration No co-development
Grade 3	Firm D Firm E	Relatively short presence Low resource commitment Not developing software in house Not attempting Systems integration No co-development

It appears there are few initial barriers to firms wishing to use bioinformatics. Indeed academics interviewed stressed the ease for firms and academics to access these tools and techniques with relatively small budgets, although firms would have the added expense of greater licensing fees. One barrier widely noted and revealed in figure 3.2, was the shortage of bioinformatics staff. One firm questioned was in the process of poaching staff from other firms at the time of interview.

To summarise, the evidence from interviewees suggests bioinformatics can be accessed relatively easily at a basic level. However, beyond this level, and surrounding the task of integrating tools and changing R&D processes more significantly, more complicated barriers and questions are raised. A greater level of competency may be required to derive further benefits from bioinformatics and gain a competitive advantage from it.

The numbers of staff and many years Firm A and Firm B have allocated to the task are a clear indication of this when compared to the numbers of staff Firm C use. Granstrand, Patel, and Pavitt state:

"The effective use and improvement of outside components, sub-systems and machinery requires a matching in house capability to choose, integrate and learn as well as to co-ordinate and manage systematic change."

Source: p.11 Granstrand, O, Patel, P, & Pavitt, K (1997) Multi-Technology Corporations: why they have Distributed rather than distinctive core competencies, California Management Review, Vol. 39, No. 4.

It has become clear that Firm A and Firm B have attempted to integrate bioinformatics to the extent that it is not just a set of additional tools at the fingers of bench scientists. They intend to re-design the entire R&D process around a set of technologies geared towards making the whole R&D process a high throughput drug discovery search. Whilst the changing technological aspects of the R&D process were considered in Chapter 2, the ensuing organisational implications are considered in section 3.5 and relate to Teece and Pisano's notion of dynamic capabilities.

3.4 Assessing Competence In Bioinformatics Through Patenting Activity

Granstrand, Patel and Pavitt identified distributed competencies in multi-technology firms by using patent data to show specific fields of technology in which firms were active[69]. To further support the interview findings, I carried out a patent search to identify firms actively patenting in bioinformatics.

[69]Granstrand, O, Patel, P, & Pavitt, K (1997) Multi-Technology Corporations: why they have Distributed rather than distinctive core competencies, California Management Review, Vol. 39, No. 4.

3.4.1 Search Methodology:

Searches for patents were carried out on the US PTO web site, rather than on SPRU's Megatech database, because the window of time requiring examination includes resent years not recorded on the Megatech database which contains data only up to 1994. Due to the fact bioinformatics is a relatively new field, most relevant patents are post 1994. Relevant patents where found by searching for those which fell into both chemistry and computing classifications. Patents thought to be appropriate were found by a search of several classes[70] from 1990 to the present day.

3.4.2 Limitations:

It is not clear whether the search has "hit" every bioinformatics related patent in the database, as other relevant classes or even sub classes may not have been searched. However, it appears that a sample of patents in this area are available to make some analysis of activities in the field.

3.4.3 Results:

The search yielded several hundred hits, and the abstracts were examined manually. Only those relating to DNA analysis, protein modeling, rational drug design or combinatorial chemistry databases were useful for the purpose of this study. However, many of the discarded patents show the increasing use of computers in biological and

[70] 435 chemistry: molecular biology and microbiology **OR** 436 chemistry: analytical and immunological testing **AND (**345 computer graphics processing operator interface processing, **OR** 364 Electrical computers and digital processing systems: memory **OR** 395 information processing systems organization, **OR** 702 data processing measuring calibrating and testing, **OR** 706 data processing artificial intelligence, **OR** 707 data processing database and file management, data structure and document processing**)**

chemical research, particularly in the areas of automated laboratory processes and detection/ screening systems, but not necessarily those linked to drug R&D.

The 27 patents remaining where held by firms, individuals and universities in the proportions shown below in figure 3.4.

Figure 3.4 Assignees of Bioinformatics Related US Patents From 1990 Onwards

Assignee type	Number of patents held
Large Firms	4
SMEs	13
Universities & Research Organisations	5
Individuals & Others	5

The large firms were the Japanese firm Hitachi (1) US firms IBM (2) and Perkin Elmer (1).

The SMEs were all of US origin, and included 3-D pharmaceuticals (2), Cirrus Diagnostics (1) DNAstar (1), Genex (3), Immunex (1), Ironi (1) MCC-Molecular Simulations Inc. (1) US Biochemicals Corp. (2), and Visible Genetics (1).

Additional searches for interviewed firms by name did not reveal any patents in fields relating to bioinformatics.

3.4.4 Discussion of Patent Search Results:

Integrated pharmaceutical firms, known from interview to be carrying out work in this field have not been shown to have a competency using the internet patent search technique. When questioned about patenting in the field of bioinformatics, some firms declined to answer and others mentioned the information sought was confidential. However, it must be accepted that patenting in this field would concentrate around

software protection, and this is an area of known weakness in the patenting system[71]. One firm specializing in this area, Oxford Molecular, indicated they generally use trade secrets as their primary means of intellectual property protection because of the ease with which software patents can negated[72]. Another possible reason for the lack of patent protection on work in this field is that firms may be creating software tools or links that address problems specific to their own bioinformatics programs. These might not be of use to other firms, as each sub-system created is likely to be unique and of little use to a competitor without copying the system as a whole. The so-called "embedded software"[73] created by such activity is therefore not patented because firms are not concerned about competitors imitating it, unlike software designed as a commercial product.

Firms who are patenting in the field seem to be almost entirely from the US, which is where many of the bioinformatics developments began in universities. Although large integrated pharmaceutical firms were not represented in the groups above, it is known from interview that the distribution of their staff working in bioinformatics also reflects the US is leading other regions in terms of bioinformaticians, just as it did with molecular biologists some years ago. This, coupled with the US entrepreneurial culture seems to have lead to the formation of several firms apparent in the search results, who specialize in bioinformatics software tools. Possibly because these tools are produced for use in external organisations, these firms are more likely to try to protect their intellectual property with patents.

[71] Aharonian, G, Patent News (available from srctran@world.std.com)
[72] Interview from previous project with Oxford Molecular (The Impact of Intellectual Property Rights on The Development of European Biotechnology, S M Thomas et al)
[73] Davies, A (1996) Innovation in Large Technical Systems: The Case of Telecommunications, Industrial and Corporate Change, Vol.5 No.4, Oxford University Press.

Also present in the results are Hitachi and IBM. Both these firms have patent portfolios that are broad and reflect their exploratory nature. Given their existing competencies, electronics and IT firms may be well placed to grow development capabilities in the area of bioinformatics, and these exploratory patents may indicate a future interest in working in this field as suppliers of software solutions to the pharmaceutical industry. In conclusion it seems the patent search methodology has limitations, and technological activities of firms can remain hidden unless they choose to disclose these in interview or by publication.

3.5 Assessing The Role of Dynamic Capabilities In Allowing Firms to Integrate and Derive Competitive Advantage From Bioinformatics

The findings of section 3.2 show distinct differences between the approaches of the interviewed firms towards bioinformatics. By using Teece and Pisano's framework it is possible to see which firms exhibit strategies characteristic of firms with dynamic capability and how bioinformatics creates competitive advantage.

Processes:

Integration: Of the interviewed firms, Firm A and Firm B have made considerable financial effort to facilitate the integration of bioinformatics into R&D. In Firm A this process began over ten years ago, with a plan set down for steady organic growth of bioinformatics related capabilities, in contrast to the rather rapid response Firm B made several years later. Infrastructural changes needed for the integration of these tools were established with a computer infrastructure (and one of the first intranets in the pharmaceutical industry) data management systems, and layers of information systems which would eventually support the bioinformatics systems. Firm B has also spent

several years on what Director of Bioinformatics for Firm B, interviewee B1, describes as "ground work". The other firms interviewed had not spent as much time or resources on this infrastructure, although they had all built intranets. According to researchers, without integration of tools and packages associated with bioinformatics, data becomes balkanized[74]. The result is data collected forms into islands and may be under-utilized and even need to be re-created on a different system before it can be used. Such islands can form as a result of poor planning as well as due to software incompatibilities according to a Bains[75]. To avoid these problems in the future, SB has been careful to leave "threads" to link databases, and has worked with consultants to develop software termed "wrappers" which can be used to move information between systems that cannot understand each other.

Learning: Learning processes are also very important when trying to gain new competencies, especially where software skills are involved. These are notoriously difficult to learn from a book, and tend to be learnt best "by doing", i.e. a personal process of trial and error. Indications in 1996 where that only 10% of Firm B's scientists were using bioinformatics tools available to them effectively[76]. However some firms are attempting to solve these problems and as mentioned in section 3.3, Firm A and Firm B have established in house training, whilst Firm C and Firm D train staff through external courses. But training courses may not provide an effective route in all firms, as Firm E's principle scientist, E1 noted. He prefers his staff to experiment and teach each other. Evidence from some interviewees also suggest that resistance to change amongst bench

[74] Williams, N, "How to Get Databases Talking the Same Language" Science vol.275 17 Jan 1997 p.301-302

[75] Bains, W, (1996) "Company Strategies For Using Bioinformatics" Trends in Biotechnology, Vol. 4 August, p.312-317.

scientists may be strong in some firms, and at least one large firm was indicated to be developing cultural problems between bioinformaticians and other research staff. This is partly because of problems in semantics between groups of individuals from very different scientific backgrounds, and partly because of a feeling of being "devalued" which is said to be spreading through molecular biologists, who might feel bioinformaticians are unduly influencing their research[77].

Reconfiguration and transformation in the R&D process have been discussed in Chapter 2, where it was highlighted Firm A and Firm B are attempting to substantially change their R&D processes whilst other firms are not. Bains notes some firms may be dissuaded by the costs of software, support systems, time and other resources that may be difficult to justify as future gains resulting are still uncertain[78]. Additionally for firms such as Firm C, it seems traditional methods are still yielding results[79], and a more cautious approach than Firm A and Firm B has been adopted as a result. However, being slow to adopt may also be a risky strategy. Teece and Pisano point out that timely response and rapid and flexible innovation are important in maintaining dynamic capabilities. Certainly, bioinformatics is a very fast moving field at present, and given the time taken by Firm A and Firm B on groundwork, firms which are slow to respond may find themselves left behind. Janet Thornton (University College London) noted much of the best expertise has already been hired and firms now developing capabilities may have trouble finding good staff.

---

[76] Bains, W, (1996) Using Bioinformatics In Drug Discovery, Trends in Biotechnology, vol.14, p.37-39
[77] Anon, interview
[78] Bains, W, (1996) Using Bioinformatics In Drug Discovery, Trends in Biotechnology, vol.14, p.37-39
[79] F1, Interview

Positions:

Technological assets: Many of the technological assets important for utilising bioinformatics do not have to be developed in house. However, buying in technology does not always work, as some firms may find they cannot use the technology effectively[80]. Interviewees from more than one firm indicated technical support offered by software suppliers was generally inadequate to support firms without internal know-how. As the data from Section 3.4 shows pharmaceutical firms do not seem to be developing patented technological assets in this field, but the assets they are accumulating may be non-transferable and difficult to replicate because they are tailor-made to work on internal systems.

Complementary assets: Established pharmaceutical firms have complementary assets such as the ability to run clinical trials, as well as distribution and sales capabilities which are necessary to derive benefit from strong R&D capabilities. Bioinformatics does not devalue these capabilities nor does it enhance them. Therefore firms that do not already possess these assets do not threaten the position of established firms by developing bioinformatics. Indeed it is partly because of these complementary assets that biotechnology firms previously hoping to become fully integrated pharmaceutical firms have often had to rely on established firms to help them bring their products to market.

Financial Assets: Bioinformatics and other new technologies being integrated into pharmaceutical R&D are expensive, and only firms which devote large amounts of funds can generate the tailored systems Firm A and Firm B are attempting to create. However,

[80] Bains, W, (1996) "Company Strategies For Using Bioinformatics" Trends in Biotechnology, Vol. 4 August, p.312-317.

large pharmaceutical firms are generally in a position to make this level of investment if it is required. It is smaller firms for which this may prove to be a problem.

Locational assets: Even in an age of globalised R&D, and remote databases accessible through the internet, location can still be of great importance. Interviewee E1 emphasized this point as he pointed out many of his research staff come from laboratories around Cambridge, and their proximity allows them close links to developments in the academic community there, as well as access to tacit knowledge.

Paths:

Path dependency: Although bioinformatics may help firms to discover new drugs, it does not replace the other capabilities needed in the research process. Therefore only firms with experience of fields such as pharmacology, molecular biology, organic chemistry etc. will be able to benefit from bioinformatics. Furthermore, firms that have not been successful in integrating other recent technologies are less likely to benefit from bioinformatics. Some firms may even be less inclined to attempt integration because of limited success in the past with chemi-informatics[81].

Technological Opportunities: These are available in broadest choice to firms with in house R&D in the relevant fields, and to those who have alliances or links to external groups also working in similar fields. Interview findings show that all three large firms interviewed are innovating in the field, creating new tools and thus opportunities to make better use of the available data than firms simply buying in tools. Firm A and Firm B also have extensive links with consultant firms and universities. Firm C are not using any

[81] Nightingale, P (1997) Knowledge and Technical Change: Computer Simulations and the changing Innovation Process, Sussex University

consultants at present. They have however worked closely with Incyte Pharmaceuticals, and receive preferential treatment to other licensees of Incyte's products as a result[82]. Therefore large firms appear to be in a better position to make use of the technological opportunities made available by integrating bioinformatics into R&D processes. However as yet the advantages of using bioinformatics are unproven, and no drugs yet licensed are the product of genomics or combinatorial chemistry, two key technologies made accessible through bioinformatics.

Competitive Advantage and Apropriability Regimes:

Of the firms interviewed, only Firm A and Firm B made a specific point of stating bioinformatics will give them a competitive advantage. The basis of this competitive advantage will be the integration of bioinformatics tools with each other to allow data to flow between systems, and thus complex analysis will be possible more easily, and will produce more valuable research as a result[83]. Teece and Pisano suggest this ability becomes a distinctive competency of competitive advantage if it is hard to replicate and if the intellectual property upon which it is built is easily protected. From the limited information available from interviews, it does seem that such a bioinformatics capability, in its most developed form would indeed be very hard to replicate, and thus could make bioinformatics a core competency of strategic importance, and providing competitive advantage. However, even the most advanced systems available to large pharmaceutical firms are far from completion at present.

[82] Incyte is a key producer of genomic data, and has issued licenses for use of its databases to several large pharmaceutical firms.
[83] B1 and A1 interviews

Chapter 4: Conclusions

4.1 The Changing Nature of Pharmaceutical R&D

In this study an insight has been gained into the changing nature of pharmaceutical R&D processes, as effected by bioinformatics and a host of other new technologies and environmental factors. Developments in a number of fields including molecular biology, cellular chemistry, combinatorial chemistry, high throughput screening, structural biology and computer based information handling systems offer new opportunities to streamline in-house drug R&D. Bioinformatics is key to unleashing and optimising research results from waves of complex data generated by these new fields of research and stored in large databases, available internally, and externally from private and public sources. Interview findings show bioinformatics is being used almost universally in pharmaceutical firms.

At the same time internal and external pressures are forcing change in the way R&D is carried out within the pharmaceutical industry. The overall cost of R&D has been steadily increasing as regulatory guidelines tighten. New developments in the field of biotechnology centered around the activities of DBFs and universities have forced pharmaceutical firms to form complex webs of alliances in order to access new technological fields.

4.2 Review of How Theoretical Frameworks Relate To Key Findings

This study has addressed the questions: what strategies have pharmaceutical firms used to acquire capabilities in bioinformatics, and how has bioinformatics been integrated into the drug research and development process of the firm?

Interview data collected from pharmaceutical firms, DBFs, bioinformatics consultants, and university departments as well as secondary data sources, supports theoretical frameworks relating to innovation by firms in new technological fields:

- Large firms attempting to benefit from tailor-made integrated systems have found in-house capabilities must be created to support bioinformatics because of the systemic rather than autonomous nature of the complex software infrastructure required, as theorised by Chesbourgh and Teece (1996). Whereas firms using software tools in a piecemeal fashion do not require as substantial capabilities in house, as the tools fit more with the autonomous model Chesbourgh and Teece put forward.
- Firms seeking innovative design of their R&D processes are forced to develop capabilities (i.e. software development) beyond those represented in their products, as suggested by Granstrand, Patel, and Pavitt (1997). Without this capability firms are less able to direct innovation of software, in-house or from specialist suppliers, and therefore less able to exploit other related platform technologies such as genomics.
- Teece and Pisano (1994) demonstrate the importance of paths, positions and processes in allowing a firm to derive benefit from technological change. The characteristics they identify as significant have been examined in firms interviewed, demonstrating areas of strength and weakness. Most crucially, Teece and Pisano maintain competitive advantage is best kept by demonstrating difficult to replicate competencies. The ability to create integrated bioinformatics systems is demonstrated by interview data to be a highly complex task, requiring much investment of time and resources. Furthermore the resulting system is tailored to user

requirements and so even direct imitation may not allow a competitor similar advantages.

Additionally the frameworks set out by Granstrand, Patel, and Pavitt (1997) have allowed a scale to be applied to qualitative data to grade firms' comparative competencies, in the absence of sufficient quantitative data such as patent statistics.

The scale identifies a continuum of competency levels in interviewed firms, from which two distinct groups emerge. Whilst some firms are using bioinformatics as additional tools in a piecemeal manner and at relatively low cost, other firms are attempting to integrate bioinformatics tools into tailored systems allowing more complex analysis of separate data sets automatically, that firms without integrated systems can only manage manually.

The development of the two groups is partly due to the bioinformatics requirements of interviewed firms being different, however factors such as cost and risk also play an important role in dissuading some firms from following the route to full systems integration. Some pharmaceutical firms such as Firm D[84] and Roche[85] (not interviewed) are content to limit the extent of in-house expertise in favour of outsourcing to Contract Research Organisations, even in traditionally key areas such as drug discovery and screening. Where such moves are considered as an alternative to maintaining in-house capabilities, they should not be confused with alliance formation, often used to boost internal capabilities in the long term. Instead this outsourcing should be seen as a cost cutting measure, chosen simply because external organisations offer a cheaper alternative. These pharmaceutical firms may choose to focus on their core competencies

[84] Interview finding
[85] as stated in an advert for contract managers, p.73 New scientist No..2148, 22nd August 1998

for reasons of cost. However, such firms risk reliance on firms in the DBFs sector to supply targets or leads for research. For example, a pharmaceutical firm might continue to run clinical trials and outsource research instead of broadening their competencies in new platform technologies such as combinatorial chemistry or bioinformatics. However, technological changes have not resulted in competence destruction, as predicted by early DBFs and hopeful venture capitalists. Instead firms with strong competencies in pharmaceutical R&D will now have the opportunity to develop new competencies and enhance their R&D. A number of these firms are attempting to integrate new techniques to maintain as much of the R&D process as possible in-house[86].

Interview evidence has shown the initial barriers to entry for bioinformatics are relatively low in the experience of firms and academics using the piecemeal approach to derive localised benefits to researchers. However staying up to date and developing robust systems to quickly interpret fast changing sources of information, and to act as a platform supporting the backbone of a research program in a large firm is proving more a costly and complex problem to address. This form of integration is much more complicated to manage, and requires a higher level of time commitment and resources. The resulting tailored systems take many years to develop, will be hard to replicate externally, and thus may become a source of competitive advantage as suggested by Teece and Pisano's framework. Firms attempting to follow this path, such as Firm A and Firm B view bioinformatics as a new core competency necessary for maintaining pharmaceutical product pipelines at affordable levels as well as a source of competitive advantage.

[86] **A1 and B1, Interview**

4.3 The Impact of Bioinformatics: Benefits and Limitations

Bioinformatics will enable R&D staff to use a wealth of new information to make more informed choices at a number of stages in the research process. This should result in stronger candidates entering the expensive human trials stage, with less likelihood of costly failure resulting, according to interviewees[87] and as argued in Section 1.2 of this study. Additionally, when integrated these new tools will allow the faster progression of initial phases of research, thus the volume of candidates developed will increase. Optimizing the use of information that bioinformatics provides will derive future competitive advantages for firms with more advanced systems, whilst firms without such systems may be less able to produce high quality candidates at high rates, and will still face increasing financial pressures as discussed in section 1.5 .

As a result of bioinformatics, the focus of initial research has moved away from traditional methods. The computer becomes a vital part of the research process right from the outset, providing details of external research quickly via the internet, and allowing targets to be found from data mining in sequence databases. The growth in targets and the speed of screening has lead to a change in the limiting factors in drug R&D processes. These are expected to remain in the areas of "wet biology", where real experiments are needed to validate findings from computer based research. Optimistic researchers at firms such as Pharsight, and MGA (both US SMEs) see even these bottlenecks reducing in the future through advanced simulation techniques[88]. However, it is still debatable as to whether further software developments will allow the simulation of findings from x-ray crystallography and drug metabolism studies. Although research projects may move

[87] A1 and B1, Interview
[88] Anon, (1997) Testing, Testing,The Economist Vol.342, 1st February p.1847

more quickly through their early stages as a result of bioinformatics and other new technologies, development processes will remain largely unaffected by these changes. Rather, hopes for reducing development times lie with the FDA, who after pressure by US Congress are allowing new rules for rapid progress of certain drug through trials. One example is Saquinavir, an AIDS protease inhibitor[89], which was passed in under three years of trials as opposed to the average 7-9 years, and licensed in less than 100 days (less than a third of average licensing time). This was facilitated by the introduction of "surrogate" end points in trials, predicting improvements in the long term through milestone changes over short time periods.

Although much work is being undertaken in-house by large pharmaceutical firms, the academic community has been the major contributor both of software and expertise in bioinformatics. The availability of some software for free or at little expense from the public sector has facilitated firm's entry into this area at low cost. More costly commercial software houses and consultants also contribute to the supply of tools available to firms. These are often provided with more robust, user-friendly interfaces and integration mechanisms. Due to the wealth of expertise within the academic community, firms close to this community stand to gain from enhanced knowledge transfer in this field. Thus study of interactions between academics, suppliers of software solutions and pharmaceutical firms is important in understanding how firms are building competencies in bioinformatics.

[89] Anon, (1997) Testing, Testing,The Economist Vol.342, 1st February p.1847

It has been stated previously a shortage of trained bioinformatics professionals exists[90], and interview data has confirmed firms still see recruitment of staff as one of the most major barriers to entry for bioinformatics. Researchers from the Biological and Biotechnological Science Research Council, were even able to estimate there is demand for 50% more employees in large firms operating bioinformatics departments[91].

The exact role of bioinformaticians in these departments may vary in large firms seeking integration, but generally they take part in four activities: developing new tools, acting as internal consultants to help bench scientists using tools, assessing external offerings, and working on the infrastructure to link tools. In firms following a piecemeal approach to using tools their activities may be more limited in terms of development and focus on using the software tools for research and assisting other to do so, as well as advising on the purchase of new tools.

Whatever the exact requirements they face, the number of bioinformaticians available to firms at present places a strain on late entrants to replicate the developments of established teams of early entrants. Furthermore, firms who do not build in-house competencies may lose the ability to derive maximum benefit from externally produced technology, as noted by Bains[92] and Granstrand, Patel, Pavitt[93] and as a result may miss new technological opportunities as noted by Teece and Pisano[94].

It may be necessary to carry out more research into differences between firms in new technological capabilities that relate to pharmaceutical R&D process change to discover

[90] Gershon, D, (1997) Bioinformatics In A Post Genomics Age, Nature Vol 389, 25 Sept. p.417-422
[91] MacLean and Harding, Interview
[92] Bains, W, (1996) "Company Strategies For Using Bioinformatics" Trends in Biotechnology, Vol. 14 p.312-317.
[93] Granstrand O, Patel P, Pavitt K (1997) Multi-Technology Corporations: Why They Have Distributed Rather Than Distinctive Core Competencies, California Management Review Vol.39 No.4 p.8-25

the extent to which new competitive advantages are being developed, and how that may affect the structure of the industry as a whole. Furthermore it must be highlighted the benefits of bioinformatics are yet to be fully realised, as fully integrated systems do not currently exist, and their effectiveness in producing licensed drugs are unproven as no drugs have been developed solely as a result of them. Others suggest the full impact of bioinformatics will be apparent only after the establishment of pharmacogenomics[95&96]. This could prove to be an entirely new dimension in healthcare, that is presently impractical but might be possible in the medium term to long term.

It is must also be stated other mechanisms of drug discovery based around imitation will still support firms not seeking to develop entirely novel treatments. Complex bioinformatics may be hard to imitate, but it does not provide a means of preventing competitors imitating final products.

[94] Teece D, Pisano, G (1994) "The Dynamic Capabilities of Firms", Industrial & Corporate Change, Vol.3 No. 3, p537-556.
[95] Poste, G, (1998) Molecular Medicine and information- based targeted healthcare, Nature Biotechnolgy Suppliment, Vol..16, May, p.19-21
[96] See Glossary for definition

Glossary of Terms

Animal Model

Development of a disease condition in animal subjects, then used to prove the effectiveness of therapeutic agents prior to testing on humans.

Bioinformatics / Computational biology

The storage, retrieval and analysis of data for scientific research aimed at providing answers to biological questions

Candidate (drug candidate)

Term used to describe a promising new chemical entity suitable to be passed into clinical trials**.**

Combinatorial chemistry

The systemic and repetitive covalent connection of a set of different building block of varying structure to each other yielding a large array of diverse molecular entities. This technique may used to generate many new chemical entities to rapidly screen against targets.

CRO (Contract Research Organisation)

A firm undertaking research, typically screening or validation, instead of (or to fund) their own research activities

DBF **(Dedicated Biotechnology Firm)**

A firm producing products or services based on biotechnological techniques rather than those of traditional synthetic chemistry.

DNA sequencing

A process, generally automated, where by DNA can be read to produce a series of bases equivalent to a binary code. The coding bases are read in units of three bases called "codons". Each DNA Codon represents one amino acid which in turn are the basic

building blocks of proteins. Therefore from DNA sequencing it is possible to determine the type of protein encoded, and often other characteristics too.

Drug Discovery Firm

Firms dedicated to research generating targets or leads, with the intention of selling research results to other organisations, rather than attempting to produce licensed drugs.

FDA (Food & Drug Administration)

The American government organisation responsible for the licensing of drugs. Because America offers the largest single market for most commercial drugs, and has one of the harshest regulating authorities, firms generally aim for FDA approval standards.

Genomics

The study of genes and genomes (the sum of an organism's genes) through data collected by DNA sequencing. Genomics allows not only the number of copies and their positions in the genome to be known, but increasingly function and mutation (and thus pre-disposition to disease) can be identified.

Lead compound

Term used to describe a promising new chemical entity in the early stages of the discovery process. Leads successfully completing screening and toxicology stages may then become drug candidates.

Pharmacogenomics

The use of genetic information from the patient to determine which of a group of treatments will have the most beneficial effect on their disease. The cause of disease may stem from one of several genetic pre-dispositions according to the specific illness, thus finding out which is the source of a patient's illness may in the future lead to the prescribing of more effective drugs to a patient population sub-divided categories according to genetic profile.

Proteomics

The study of proteins, their structure and function. Increasingly Protein structures are available, and may be allowing proteins to be classified by function and relationships to other known proteins.

Screening (High Through Put and Combinatorial)

A process of testing new chemical entities against large numbers of targets to see if they show any desired result. Increasingly screening is becoming automated, allowing the routine processing of large libraries of compounds.

SME (Small to Medium sized Enterprise)

A firm of not more than 500 employees.

Software tool

Software capable of performing a multiple or single tasks (eg searching for a DNA sequence, or recognising patterns in data). Software suites or systems may be made up of numerous software tools that may interact.

Target

The point a therapeutic agent intervenes at a molecular level to interrupt the disease process. Targets may be genes, proteins, or other molecules.

Validation

Confirmation experimentally that a target or lead is responsible for an observed effect in the metabolic pathway of interest.

Appendix 1: Interviewees

Representing Pharmaceutical Firms

Large Firm A, Interviewee A1, Unit head, Advanced Technologies and Informatics

Large Firm B, Interviewee B1, Director of Bioinformatics UK
Interviewee B2, Alliance and Technology Group

Large Firm C, Interviewee C1, Head of Genetic Technologies

Medium Firm D, Interviewee D1, Molecular Biologist
Interviewee D2, Molecular Modeller

SME Firm E, Interviewee E1, Principle Scientist

Representing Universities:

Thornton Janet, Prof. Biomolecular Structure and Modelling Group
Dept. Biochemistry and Molecular biology, University College London

Woolfson, Dek, Biochemistry, University of Sussex

Representing Others:

Harding, Debbie (Researcher) Biological and Biotechnological Science Research Council

Maclean, Marlie (Researcher) Biological and Biotechnological Science Research Council

Informatics Consultants firm F, Interviewee F1 (Bioinformatics consultant)

Appendix 2: Interview Questionnaire to Pharmaceutical Firms

Section 1: General Background

Company name:

Person Interviewed and Position held:

Section 2: What Strategies are firms using to access the technologies and tools of bioinformatics?

a) How Long has bioinformatics been used in your firm and why did they move into the area?

b) Before Bioinformatics what did the drug discovery and development pathway in your firm look like? Please indicate limiting steps, sourcing and barriers (eg. cost high failure rate).

c) How is the pathway in your firm set to look (short to medium term)? Again please indicate systems you use, sourcing, limiting steps, and barriers to progress and reasons why.

d) Is integration of software sytems occurring yet (as in banking sector), or is technology still in separate "boxes"?

e) Are you developing internal capacities or out sourcing? Is software developed in house or bought off the shelf (is it developed by IT consultants, Public sector researchers, or internal programmers?)

f) Is there a complex relationship with suppliers of hardware and software involving after sales care and feed back for R&D, alliances and licences?

g) Does innovation in this field make hardware obsolete quickly and is this a problem for buying equipment for your programs.

Section 3: How are these technologies being integrated into development process in studied firms?

a) How far off is the "holy grail" of structural models gained purely from sequence data?

b) Who are the end users of bioinformatics and what skill training do they have?

c) Will the end user change in profile (e.g. specialist today, but bench scientists in future)?

d) Will different systems in separate organisations be able to “talk” to each other?

e) What do you see as the major barriers to integration of bioinformatics systems in your firm? What reasons make them barriers? Is there in innovation paradox (biologists don’t know what systems can do therefore can’t ask engineers to make it)

Section 4: How will bioinformatics change the focus of activities in those firms?

a) Will the limiting step of the drug discovery process change from being target discovery to a later stage?

b) Will they out-source more or downsize R&D departments if the work is less intensive in terms of man-hours?

c) Will it become easier for small less integrated firms to develop targets on their own? Could this affect the structure of the industry?

d) What do you see as being the major impact of bioinformatics in the medium to long term?

Appendix 3: Interview Questionnaire to Non-Firms

Section 1: General Background

Organisation:

Person Interviewed:

Background: (how long has bioinformatics been used in your dept, and what skill training do staff have.)

Section 2: What Strategies are firms using to access the technologies and tools of bioinformatics?

h) Are pharmaceutical firms all developing internal informatics capabilities or out-sourcing portions to external sources?

i) Is the software that firms are developing developed in house or bought off the shelf (is it developed by IT consultants, Public sector researchers, or internal programmers?)

j) Who is designing the hardware and software systems used (bioinformatics specialists or old familiar IT firms, new IT firms or a mixture)?

k) Is there a complex relationship with suppliers of hardware and software involving after sales care and feed back for R&D, alliances and licences?

l) Does innovation in this field make hardware obsolete quickly and is this a problem for those buying equipment to run programs in house?

m) When academics claim bioinformatics is easy and relatively cheap to access, why are large pharma spending so much - do large firms have to replicate external developments? Why?

Section 3: How are these technologies being integrated into development process in studied firms?

f) Are software systems being integrated (as in banking sector), or is technology still in separate "boxes"?

g) How far off is the "holy grail" of structural models gained purely from sequence data?

h) Will there be a “windows” type standard, and who will make it?

d) Who are the end users and what skill training do they have?

e) Will the end user change in profile (e.g. specialist today, but bench scientists in future)?

f) Will different systems in separate organisations be able to “talk” to each other?

g) What do you see as the major barriers to integration of bioinformatics systems in firms? Is there still a design paradox?

h) Is there any variation in how the firms are using Bioinformatics?

Section 4: How will bioinformatics change the focus of activities in pharmaceutical firms?

a) **Will the limiting step of the drug discovery process change from being target discovery to a later stage?**

b) **Will they out-source more or downsize R&D departments if the work is less intensive in terms of man-hours?**

c) **Will it become easier for small less integrated firms to develop targets on their own?**

d) **Will there come a time when the industry does it's own in house development and does not need software firms?**

e) **What do you see as being the major impact of bioinformatics in the medium to long term?**

f) **Does it seem likely some of these will fail to fulfil future needs, thus causing competitive differences between firms in the future.**