Bio-informatics Companies in India 2002: The Search for Business Models

Abstract

Bio-informatics is the application of information technology to the life sciences. Given the established strength in IT, several Indian firms are looking at entering this market. The case profiles the strategies of established IT firms as well as start-ups. Highlights the challenges of finding a business model in an emerging industry, extending capabilities into new domains, and building high technology firms in the developing world.

Subjects Covered:

R&D outsourcing, Business models, Information Technology, Biotechnology.
Bio-informatics Companies in India 2002: The Search for a Business Model

In the nineties, Indian software firms had achieved considerable success by adopting a business model that linked the low-cost, high-quality technical manpower in India to profitable opportunities in the developed world. R&D outsourcing was considered by many to be the next major opportunity for India’s technical workforce. In 2002, there were a handful of firms in India that provided R&D services in areas like hardware development, chip design, and consumer and automotive electronics.¹ Many argued that the life sciences provided opportunities for the success of a similar business model, as the country had an inexpensive talent pool of highly qualified scientists.

Bioinformatics (BI) is the application of information technology to research and development in the life sciences. Many viewed it as an important opportunity for Indian firms, as it involved the confluence of the life sciences and information technology (IT). The Confederation for Indian Industry, an apex industry association, had voiced the opinion that BI probably represented “the biggest opportunity for the IT industry since Y2K” (the Year 2000 problem)². This case profiles the business models of established IT firms as well as start-ups in India that had entered the bioinformatics sector by the end of 2002.

The increasing importance of biotechnology in drug discovery

Biotechnology is the science of using living organisms – bacteria, fungi, plant and animal cells – to produce substances of clinical and commercial value. In the eighties and nineties, pharmaceutical biotechnology firms were focused on techniques like recombinant DNA (rDNA) and hybridoma to produce biologically active substances with therapeutic properties, like antibodies and enzymes. rDNA is a technique to produce a desirable protein in large volumes by “tricking” an organism into behaving as if this production task were part of its genetic instructions. Hybridoma involves fusing antibodies for particular diseases with cancerous cells, so that when the cells multiply, they also produce the antibody. The U.S. government–funded Human Genome Project was completed in 2001, and even during its progress had changed the landscape of biotechnology research. From the production of desirable biological substances through the modification of genetic material, the focus shifted to understanding the genetic basis for disease.

Genomics is the study of genes and their function. A genome is the entire set of hereditary instructions for building, running, and maintaining an organism, and passing life on to the next generation. The genome is made of a chemical called DNA (DeoxyriboNucleic Acid). The genome contains genes, which are packaged in chromosomes and affect specific characteristics of the organism. One could think of these relationships as analogous to a written page of text. The page would represent

¹ Product majors shifting more R&D work to India; The Hindu, Business Line, June 19, 2002
² CII website, http://www.ciionline.org/busserv/biotechnology/inv_genomics.html
the whole genome, which would in turn consist of multiple paragraphs, the chromosomes. Words make up the paragraph in the manner that genes make up the chromosome, and individual letters, analogous to DNA, are the ingredients for the basic building blocks of life. The problem is that the page is written in a language that scientists do not fully understand; identifying the individual words (genes) in the page (genome) was not easy. The task of “annotating” DNA sequences to identify individual genes was notoriously difficult, and could be likened to “reading this sentence without any punctuation.” After annotation, genetic data from healthy and diseased individuals are compared to identify which genes were missing or mis-located on the genome across the two. Proteomics is the study of a full set of proteins encoded by a particular genome. The study of the 3-D structure of proteins provides a much better understanding of how a gene “expresses” itself -- produces particular biological substances -- and its consequent effect on the physiology of the cell.

Genomics and proteomics were stimulating the discovery of breakthrough healthcare products by giving scientists innovative ways to design new drugs, vaccines and diagnostics. Genomics-based therapeutics included "traditional" chemical drugs, protein drugs, and gene therapy -- an experimental procedure aimed at replacing, manipulating, or supplementing non-functional or malfunctioning genes with healthy genes. Research in genomics and proteomics had created a host of new companies around the world that specialized in genetic testing, diagnostics, and the creation of drugs targeted toward particular disease-related genes. The massive information processing requirements created by genomics and proteomics had given rise to the latest branch of biotechnology, bioinformatics (BI). It involved the creation and analysis of vast, complex databases of genetic information that could be annotated and compared, and visualisation techniques that allowed the structure of proteins to be deciphered. IT was playing an increasingly important role in biotechnology as the volume and complexity of data generated increased incessantly. Many saw the future of biotechnology research approaching a stage where design and simulation exercises could be run on computers to speed up experiments and confirm theories.

The sequence of steps in the discovery, development and commercialization of drugs generally followed a similar pattern in the major pharmaceutical markets worldwide (Exhibit 1).

DRUG DISCOVERY: The process of finding a marketable drug starts with understanding how a disease develops. Once the mechanism for the development of a disease has been discovered, the location in the body where a drug or active compound can act is identified. These locations (targets) are usually proteins in the body that play a key role in the biological process. Once a protein is identified, one can attempt to influence its activity with certain active compounds in order to eliminate the cause of the disease. After identifying the target, it is validated. In this process the essential or causal involvement of a gene or protein in the disease is confirmed. In addition, it is determined whether a protein is suited for the development of the medicine. At this stage, a chemical compound that fits the

---

3 Experts estimated that there were between 4-6 terabytes of publicly available genetic data in 2002.
identified and validated target has to be found. This compound, known as a **lead**, is then further optimized until it binds exclusively to the selected target and exhibits the desired effect. The active compound has to influence the gene or protein that triggers the disease in such a manner that it no longer has a disturbing effect on the metabolism. This process is known as **lead generation** and **lead optimization**.

**CLINICAL DEVELOPMENT:** After optimizing the active compound, in the pre-clinical phase lab and animal testing is used to verify whether the compound is suited for use in the human body. To determine this, toxicologists examine how the body absorbs the compound, how it is excreted, and how it affects the organs. In addition, they examine whether the compound has any toxic effects. On the basis of the research findings, the potential effect on human beings and the potential risk of the first use on human beings are assessed. If the lead compound discovered proves successful and also fulfills the strict legal requirements of pre-clinical trials, larger quantities are produced for use in clinical trials. This involves testing it directly on human beings in three clinical phases. In Phase I, volunteers undergo tests to see if the drug causes any harm to human beings. In Phase II trials, controlled experiments are carried out to test the efficacy and optimal doses for the drug. Phase III trials are conducted on a large sample of subjects, and are intended to gather the additional information about effectiveness and safety that is needed to evaluate the overall benefit-risk relationship of the drug. Legal authorities typically monitor each phase of the trials, and will not allow the next phase to begin unless satisfied with prior phase results. The drug cannot be marketed until final approval has been obtained.

**PRODUCT AND THERAPEUTIC FRANCHISE:** After obtaining all regulatory approvals, the product is manufactured on a large scale and distributed and marketed through the usual channels.

The nineties were marked by the ever-increasing importance of biotechnology in drug discovery. Biotechnology was a $50 billion global industry in 2001, with a collective market capitalization estimated at over $200 billion; nearly 5,000 firms worldwide engaged in biotechnology research and product development. The majority of these were concentrated in the U.S. and Europe. In the late 1990s, biotechnology companies were entering into alliances with pharmaceutical partners in unprecedented numbers. Between 1990 and 1998, the top 20 pharmaceutical companies had invested approximately $21 billion in collaborations with biotechnology companies, and between 1996 and 2000 the number of alliances had increased to an average of 616 a year. These collaborations offered young biotechnology firms the access to resources that they often lacked, such as regulatory expertise and manufacturing and marketing capabilities. In return, pharmaceutical companies gained access to emerging technologies, proprietary products, and the bright minds behind them. This collaborative approach proved successful, resulting in biotechnology product sales of $39 billion and a growing pipeline of new products.

---

4 Beyond borders: The global biotechnology report 2002, Ernst & Young
5 Value Drivers in Licensing Deals; Katie Arnold, Anthony Coia, Scott Saywell, Ty Smith, Scott Minick, and Alicia Loffler; *Nature Biotechnology* Nov 2002.
In 2002, the pipeline of prescription drugs derived from biotechnology was growing faster than the traditional pharmaceutical pipeline. Pharmaceutical firms outsourced more and more of their R&D budgets; some estimated that as much as 25 percent of the discovery-stage R&D budgets of large pharmaceutical firms found its way into small biotechnology firms through various research partnership agreements. It was also estimated that in 2002, 14 of the 55 blockbuster drugs marketed by the 10 largest pharmaceutical companies had been in-licensed from biotechnology companies, and their reliance on revenues from in-licensed molecules as a proportion of their total revenues had increased from 24 percent in 1992 to about 40 percent in 2002. 6

On average, it took 10 to 15 years to develop a new drug (Exhibit 2). Most drugs did not survive the rigorous development process—one estimate suggested that only 20 in 5,000 compounds that were screened finally entered pre-clinical testing, and only one out of five drugs that entered clinical trials was approved for use. Patent protection lasted a finite number of years from the date of filing, and the opportunity cost of each day’s delay in drug development was estimated to be about $1.3 million.

The bioinformatics (BI) industry worldwide

BI played an increasingly significant role in managing and harvesting valuable knowledge from the volumes of genetic and clinical data generated in each part of the drug discovery and development process. Target identification, validation and lead generation were all managed more efficiently if existing information on DNA sequences was available in easily searchable databases, and through the creation of graphic modeling and visualization tools that helped in the prediction of protein structure. As one industry observer put it, “through all these stages of the discovery process, the sine qua non of BI is knowledge management, integration and dissemination.” 7 Exhibit 3 shows the various life sciences and IT-enabled technologies that were involved in drug discovery and development.

BI software applications in the drug discovery phase needed deep domain knowledge -- knowledge of genomics and proteomics -- in addition to traditional IT skills. BI aided clinical development with specialized tools for data management such as drug and patent information databases, and software platforms that enabled the simultaneous analysis of clinical trials data with genomic data (also known as pharmacogenomics). Experts believe that except for some pharmacogenomic applications, BI tools and techniques in the clinical development stage were generally less demanding of domain knowledge than applications in the drug discovery phase.

Generally speaking, BI tools and techniques fell into one or more of the following categories:

**Data integration** tools created user interfaces that allowed scientists to view multiple types of data from multiple sources within a single environment. For instance, an

---

6 Pharmaceutical R&D Outsourcing Strategies: An Analysis of Market Drivers and Resistors to 2010 – Reuters Business Insights
application could help a scientist pull together published papers, patents, sequence and structure data all related to a particular problem, rather than search multiple databases and pool the information. Other data integration applications included software for managing data generated in laboratory activities from multiple instruments. High throughput screening (HTS), which allowed large numbers of drug candidates to be screened simultaneously, depended critically on such data integration applications.

**Analysis** tools were essentially algorithms that performed particular kinds of searches and manipulations on sequence and structure data. These algorithms called for mathematical and programming expertise in addition to knowledge of genomics.

**Visualization** tools were graphics packages that rendered 2D and 3D images of proteins, and provided graphic representations of sequence data.

In addition to the tools themselves, their application could yield “content” -- databases of genomic information arranged in a particular way. For instance, a content product could be a database of sequence repeats in the human genome. Firms would be willing to purchase such information as they could then explore its commercial possibilities through lab research.

Estimates of the global BI market ranged between $1.5 billion and $5 billion in 2002.\(^8\) In mid-2000, there were an estimated 50 private and public companies that offered BI products and services. Industry experts believed the number may have more than tripled by 2002. BI was widely used in the drug discovery divisions of major pharmaceutical and biotechnology companies and well-funded research institutions. These organizations had internal teams of bioinformaticians who assisted scientists in analysing and organizing their data. Software written by academics and scientists while solving their own research problems was often available as freeware over the Internet, though the programs tended to be highly specialized and poorly documented.

Stand-alone public and private BI companies mushroomed in the latter half of the nineties in anticipation of a huge market for genomic software and databases. Unlike pure-play IT companies, BI companies spent large sums of money on acquiring domain expertise to understand the science behind the data and applications. They hired highly skilled molecular biologists, mathematicians, statistical geneticists, scientific application developers and database designers for product development, project implementation and customer support. In the late nineties, some standalone BI firms like **Informax** ([www.informaxinc.com](http://www.informaxinc.com)) and **Accelrys** ([www.accelrys.com](http://www.accelrys.com)) (formerly Oxford Molecular Group) had developed business models around selling shrink-wrapped BI software. However, by 2002, both had begun selling enterprise-wide data integration platforms in addition to analytical software for PCs. **LION BioSciences**, ([www.lionbiosciences.com](http://www.lionbiosciences.com)) for instance, had a business model similar to that of many enterprise resource planning (ERP) software vendors. They sold an enterprise-wide data integration platform and generated additional revenues through

---

\(^8\) International Data Corporation's (IDC) White Paper, "The Convergence of Biology and IT: Market Impacts of the Bio-Sciences Era"
customisation. The informatics division developed and implemented integrated solutions to aid the drug discovery process, and its professional services division helped in customizing LION’s software to client-specific needs. LION focused on selling enterprise-wide solutions to large pharmaceutical firms and academic institutes, and emphasized the scale and customization of their product and services. Initially LION had worked with a fee-for-services model, but later shifted to the ERP model. LION was founded in March 1997 by a group of leading scientists from the European Molecular Biology Laboratory in Heidelberg, Germany, and had about 400 employees. In 2002 Lion Biosciences incurred losses of $54 million on revenues of $40 million.

Other standalone BI firms specialized in providing “content” derived from their own research. Incyte Genomics (www.incyte.com) and Celera Genomics (www.celera.com) were pioneers in the study of human genes and how they caused or prevented disease. Both companies had developed gene-mapping technologies that revolutionized the way the pharmaceutical industry thought about developing new drugs. Both compiled databases of genomic information and sold the information to drug companies. One could visualize them as scientific prospectors mining the human genome, and then shipping the ore through sluice gates to pharmaceutical companies for sifting and evaluation. Celera’s revenues came principally from two sources—subscription revenues from users of its online information, and service revenues from activities such as collaborative gene discovery and contract sequencing. Commercial and academic institutions subscribed to Celera’s data through the Celera Drug Discovery System and engaged in discovery collaborations and contracts. Celera focused on expanding the content and functionality of its integrated information and discovery system, which it believed included the most comprehensive and integrated databases of genomic and related biological and medical information currently available. For the quarter ended June 30, 2002, Celera Genomics reported a net loss of $211 million on revenues of $121 million. Like Celera, Incyte sold online research solutions and licensed out gene mapping technologies. Incyte provided such genomic and proteomic products and services to pharmaceutical, biotechnology and academic institutions worldwide, and had developed a comprehensive patent portfolio and leading-edge technology databases. Incyte Genomics reported a loss of $136 million on revenues of $101 million in 2002. Both Celera and Incyte were frequently in the news and were viewed favourably for their scientific expertise.

DoubleTwist was a classic example of the difficulty of finding an appropriate business and revenue model to make money in BI. Founded by two Stanford graduate students, Joel Bellenson and Dexter Smith in 1991 and based in Oakland, California, DoubleTwist began life offering enterprise software solutions to pharmaceutical companies and academic groups. In 1997, John Couch, a former executive with Apple Computer, joined the company as CEO, hoping to make BI tools as accessible to the life science community as Apple had helped make the desktop computer accessible to the population at large. The company moved to an application service provider (ASP) business model in 1999, in which software and data were resident on DoubleTwist’s

---

9 Celera vs. Incyte: A Biotech Battle Royal, BusinessWeek, September 20, 2000
own computers, and customers accessed both through the Web. The company secured generous funding from a bevy of blue-chip backers, including the Mayfield Fund, Kleiner Perkins Caufield & Byers, and the Burrill Agbio Capital Fund. In all, DoubleTwist successfully raised more than $75 million since its inception. In March 2001, amidst a precipitous downturn in biotech stock valuations, DoubleTwist withdrew its initial public offering six months after it had been filed. In January 2002, former COO Robert Williamson replaced Couch as CEO, and two months later, DoubleTwist closed down. "No one was surprised by this," Williamson told the San Francisco Chronicle, "but everyone was disappointed. We had a great product and a great team — we just didn't have the revenues."

**Overview of the Indian BI industry**

By late 2002, large established players in the IT services industries from India, such as Infosys, Satyam Computers, Tata Consulting Services and Wipro, had all established basic operations in BI, though investments by these firms in their BI divisions could still be considered insignificant compared to their other “vertical” product groups like finance, insurance, telecom and enterprise software. In addition, several smaller players, including start-ups, were also turning their attention to what they saw as emerging opportunities in BI. Most of these BI companies were located in southern India and were clustered in and around Bangalore, Hyderabad and Chennai, thanks to the presence of large IT players, software technology parks, government-sponsored biotechnology research institutes and local government support (Exhibit 4). Few of these operations had revenues in 2002, and even fewer released details of those revenues. By late 2002, it was estimated that about 30 to 40 firms in India had BI operations.

In spite of the relatively small scale and early stage of development of the BI industry in India, it had already attracted considerable interest and excitement. Many experts saw BI as the new cash cow for the Indian software industry. The argument was that in order to restrict spiralling costs of drug discovery and development, large pharmaceutical firms would have to outsource (initially) generic IT services, followed eventually by BI software and services, from countries like India. Indian IT firms could hire programmers and biotechnologists at a fraction of what it would cost in the United States (Exhibit 5). The Confederation for Indian Industry viewed database development as the area where Indian companies would be sought to offer complete solutions to major pharmaceutical and genomics-based biotech companies of the West.

Several questions remained unresolved in 2002 about the future of this nascent industry. Which stage of the biopharmaceutical value chain should Indian BI firms target? Did it make sense to compete in providing BI services to the drug discovery, clinical trials or development stages? How would they balance the need for domain and IT knowledge required to compete in these stages? How easy would it be to leverage their existing capabilities in IT in the new domain of BI? Should they train

---

10 An Indian adventure - Frontline, September 2000. Also see CII website, http://www.ciionline.org/busserv/biotechnology/inv_genomics.html
their IT experts in biology, or would the converse be a better solution? Should they focus on a product-based revenue model or on the more traditional services model that plays to the labor cost advantages in India? Should the product be tools or content? How could Indian firms avoid replicating the generally poor performance of stand-alone BI companies in the West? As pharmaceutical companies had not embraced the offshore outsourcing model that had worked in many other industries, it would be no mean feat to break into an essentially new market.

**IT services majors and their involvement in BI**

The Indian software services industry enjoyed its defining period in the early 1990s when the demand for IT services in the developed world outstripped the available supply of skilled labor. About 150,000 English-speaking engineers graduated in India each year. Companies in the developed world began to look toward India to leverage its low-cost, English-speaking IT manpower. Given the low levels of initial investment required to enter the software services business and the minimal regulatory interference from the government, there were few, if any, entry barriers or constraints in the early years. As the industry witnessed very rapid growth, there was a general consensus that the initial competitive advantage and growth of Indian companies was largely based on their access to cheap but well qualified manpower emerging from the government-funded engineering college system, supported by some improvements in IT-related infrastructure (e.g., high-bandwidth communication lines) set up by private and state-owned companies.

In 2002, the Indian software industry comprised about 3,000 firms and posted revenues of $10.2 billion of which exports accounted for $7.7 billion. Most Indian software exporters concentrated on custom application development and maintenance, and application outsourcing, regarded as fairly low-end niches. Industry experts were optimistic that Indian firms would have significant presence in higher end markets such as packaged software support and installation, systems integration and network infrastructure management by 2008. There was some basis for this optimism. The leading Indian firms had shown a capacity to compete on the terms of companies from the West. More than 20 Indian software firms had been rated at Level 5 of the Capability Maturity Model, a global standard for quality and excellence in software development, by 2002; in fact, five of the first 10 firms to attain this quality certification were Indian. Among the approximately 590 software and programming firms traded in the U.S. capital markets in 2002, Indian firms were among the top in rankings based on return on equity (e.g., Infosys Technologies ranked 6, and Wipro Ltd. ranked 12). In 2002, 22 percent of the export revenues for Indian software firms came from the banking and financial services verticals, with 16 percent and 14 percent coming from manufacturing and telecom equipment makers, respectively. Experts believed that health care offered many unexplored opportunities. Some of the larger players located their BI ventures within the health care vertical, and saw BI as part of a portfolio of IT services.

The remaining part of this section profiles four major players in IT services from India: TCS, Infosys, Satyam and Wipro, which had all achieved the highest international industry benchmarks for quality, such as Level 5 CMM certification, and
were the largest revenue earners. All were also involved in BI to varying degrees by the end of 2002.¹¹

Wipro Technologies was ranked 21st worldwide among software services companies, and 86th in terms of best performing technology companies by BusinessWeek in 2002 and 2001. The company’s IT business began in 1980, and by 2002 had about 12,000 employees operating out of 27 offices worldwide. With $736 million in revenues in 2001-2002 and a CAGR of 45 percent over the last five years, Wipro was one of the major success stories of the Indian IT industry (http://www.wipro.com). In April 2002, Wipro CEO Azim Premji created Wipro Health Care and Life Sciences, a wholly-owned subsidiary located in Bangalore. This business addressed the requirements of the “bio-IT market,” which Wipro defined to include traditional IT services to hospitals, health insurance companies, and medical and analytical devices companies as well as to players in the drug discovery value chain. Wipro estimated this market to be about $25 billion, growing at over 20 percent annually. Wipro’s application of IT to drug discovery (or “pure BI”) was still a relatively small part of its business, and most of their efforts were in the traditional IT service areas of outsourcing operations, IT consulting and enterprise package implementation in health care sector companies.

Wipro Healthcare and Life Sciences had about 250 employees. Over the course of 2002, the company’s plans to attack the pure BI market underwent some changes. The original plan to go after large pharmaceutical companies to offer customized software had evolved into a strategy based on partnering with other firms to integrate domain and IT knowledge. The CEO of the health care and life sciences business had said, “Though we have experience in this area, gained from our association with GE Medical, we would like to focus more on the domain knowledge side, where partners would bring in the necessary strength.”¹² The company was interested in concentrating on IT support for clinical trials, data management and statistical analysis. The partners would help in providing the necessary support services to accelerate the drug discovery process and generate data on chemical compounds. In the drug discovery stage, for example, Wipro wanted to work as a technology partner for large consortia of life science equipment vendors who supplied the equipment that aids the drug discovery process. Wipro would help in synchronizing the data that these equipment churned out so that customers would not need to custom design them using their in-house IT resources. In the drug approval phase, Wipro intended to capitalize on the FDA's plans to allow electronic filing of Investigational New Drug (IND) reports by pharmaceutical companies. Wipro would use its knowledge of Web solutions, data security and imaging technology to assist customers in the electronic filing process.

¹¹ The section on the IT services majors relies extensively on secondary data.
¹² The Financial Express, Hyderabad, October 22, 2002
Tata Consultancy Services (TCS) was India’s first and Asia’s largest independent software company (www.tcs.com). In 2002, TCS had about $1,200 million in annual revenue, which represented a five-year CAGR of 42 percent. It had 20,000 employees located in 106 offices in 31 countries. A part of the Tata industrial group, by the late nineties it was estimated that TCS alone would represent half the value of the portfolio of companies owned by the group. Its CEO, S. Ramadorai, has been named one of the “top 25 most influential consultants” in the world. The BI effort in TCS was housed within the Advanced Technologies Center located in Hyderabad. The BI initiative within TCS was led by Dr. M. Vidyasagar, and involved about 40 engineers. Vidyasagar took pains to point out on many occasions that any software engineer would not be able to work on BI projects; TCS itself put its engineers through a nine-month course in the life sciences before they began working in BI. By 2002, TCS had already invested around $1.5 million in a project in collaboration with the Council for Scientific and Industrial Research (the apex body for government-funded research labs in India) to develop an end-to-end BI product, “Bio-Suite,” for use by Indian academic institutions. The company was expected to launch the alpha version of Bio-Suite in May 2003. The software suite could be used for analysis and visualization of genomic and proteomic data.

Satyam Computer Services Limited is another important IT services company from India. Incorporated in 1987, by 2002 it had grown to about 10,000 employees, with revenues of $363 million, representing an 87 percent CAGR since 1997. Satyam is headquartered in Hyderabad (http://www.satyam.com). In 2002, Satyam had established a team of about 30 individuals in a small BI group. The team included domain experts with Ph.D.’s in the life sciences, and was led by Dr. Sudhakar Varanasi, a molecular biologist by training. Varanasi’s group had begun some collaborative work with the Hyderabad-based Center for Cellular and Molecular Biology (CCMB). The CCMB had one of the best genomics and proteomics research groups in the country, as well as a small BI group, and already collaborated with Satyam’s scientists to identify a drug target for malaria through comparative analysis of gene sequences and of metabolic pathways; patenting for this “content” product was in progress. Among other collaborative projects with CCMB, they were now developing an automated system for statistical analysis of micro-array data, and an automated human genome annotation methodology for identifying novel genes specific to neurological diseases. Both these applications required considerable depth of domain knowledge. By the end of 2002, Satyam’s BI group had begun to line up clients.

Infosys Technologies had grown to over 10,000 employees in 2002 since its incorporation in 1981. In 2002, it had revenues of $545 million, and was the “Best Employer” in India (for the second year running) in a survey conducted by Hewitt Associates. Its charismatic founder-CEO, Narayana Murthy, was the public face of the company, and after stepping down from the position of CEO, was the chief mentor in addition to being chairman of the board. Infosys was headquartered in Bangalore, and had offices worldwide (http://www.infosys.com). In 2002, industry experts believed that Infosys was seriously considering the BI market, and that it had assembled a small group of people within a new venture unit to examine the market potential for BI. In August 2002, Narayana Murthy, referring to their biotechnology
initiative, said at the annual analysts’ meeting for Infosys, "We have put in place necessary things and we will talk about it after a few wins... after we reach a critical mass."

Other BI players

The IT services majors were not the only firms interested in BI. Several smaller firms also had start-up operations in BI in 2002. This section profiles three of the smaller players.

Ingenovis was the bio-sciences division of the technology incubator, iLabs (http://www.ingenovis.com). iLabs was founded by Srinivasa Raju, one of the co-founders of Satyam Computers. Headquartered in Hyderabad, Ingenovis saw itself as an “IT company focused purely on life sciences.” It had about 40 employees, of whom eight were domain specialists in biotechnology. Satish Kumar, who headed Ingenovis, used to run an IT start-up company of his own before coming to iLabs in January 2001. “When I came in,” he said, “what we knew was that the life sciences were increasingly becoming an information-driven business. Fortunately or unfortunately, that’s more or less all we knew!” Most of his prior experience was in product development, and Kumar had limited knowledge of the life sciences area. As he began to understand the business opportunity, Kumar realized three things: “First, the market is not as big as we thought initially. The market for IT products and services in pharmaceuticals and biotechnology industries is about $500 million in 2002, though it is expected to grow to $2 billion by 2005. Now, this includes hardware and equipment, which means the component for pure software is much smaller. Second, there exists a well-established system of Western firms outsourcing their IT operations to Indian firms, but this practice is rare in the pharmaceutical sector. They haven’t had the experience of outsourcing—one of the issues may be the credibility of India as a provider of such services for specialized requirements in the life sciences. Another may be our poor intellectual property regime -- unlike credit card companies outsourcing their back operations to India, BI projects can often involve extremely valuable intellectual property. Third, we realized that to have any sort of credibility as a BI player, we had to gain significant domain knowledge.” The point about India’s intellectual property regime was particularly relevant, as it was considered weak and inadequate by Western standards, though significant improvements were expected after 2005 (Exhibit 6).

Kumar believed that fortunately for Ingenovis, Hyderabad was also home to one of India’s premier biotechnology research institutes, the Centre for Clinical and Molecular Biology (CCMB). Around the time Kumar and his colleagues were trying to understand the opportunity space in life sciences, they were approached by scientists at CCMB who wanted help isolating the occurrence of a particular sequence of nucleic acids, GATA, in the human genome. The hypothesis was that the location and frequency of occurrence of this pattern in the genome was associated with a particular disease. To find all such occurrences was quite a memory- and

---

13 The sections on Ocimum and Ingenovis rely on primary data, while that on Strand Genomics relies on secondary data.
computation-intensive task. In their first meetings, the software engineers and scientists did not have an easy time communicating. “In fact, I heard it as ‘ta ta’ -- ‘goodbye’!” recalled Kumar. “Of course, being IT professionals, whether it is GATA or TATA doesn’t really matter in the end; what we needed was an algorithm that can locate any pattern in a very, very large string of data in a memory-efficient manner.” Kumar’s team did manage to develop such an algorithm, and the partnership with CCMB had been a strong one ever since. The two had collaborated to produce a database product that contains information about simple sequence repeats in the human genome, SSRD (Simple Sequence Repeats Database). More importantly, the experience served to build confidence in their basic assumption -- that a team of skilled IT professionals could in fact develop applications for biotechnologists. It also provided a valuable learning opportunity for Kumar’s team -- to understand some of the vocabulary biotechnologists used and the kind of computational and data management problems they were likely to encounter. “This can be quite critical -- when we initially approached private companies and offered our services, we often heard them say that they didn’t doubt our IT skills at solving the problem, but it was not worth their time to explain the problem to us in a format that we could then solve it in. Convincing the customer is never easy -- they want to know about our backgrounds, our qualifications. What we sometimes did was to offer to do a pro bono project as a sort of test of skills. We would say, give us a small project, one that may not be critical to your operations and does not involve any significant IP. Let us show you what we can do,” said Kumar. In late 2002, Ingenovis seemed set to acquire a handful of new clients, including a few biotechnology companies in the UK and the U.S.

Ingenovis had product offerings in all tool categories -- integration, analysis and visualization. In addition to tools, Ingenovis also offered a content product, the Simple Sequence Repeats Database it developed in collaboration with CCMB. “The best kinds of content products would involve proprietary data converted into proprietary content, like what Celera and Incyte do -- so far we have been able to develop proprietary content starting with publicly available data, as with the SSRD database,” said Kumar. He believed it was easier to take an IT professional and train her or him in the necessary biology rather than the other way around. “We are ultimately an IT company – I am very clear that we need good biotechnology domain expertise as well, but our revenue comes from our IT skills.” In terms of the future, Kumar believed that ultimately, the best returns were in being directly involved in drug discovery itself, like Celera and Incyte. That would of course represent a great deal of deepening of domain knowledge. For the present, Ingenovis had enough on its hands in trying to work its way up from a pure services model to one that mixed products and services.

Strand Genomics, based in Bangalore, specialized in designing algorithms for analysing and visualizing genomic data (www.strandgenomics.com). It was founded by a group of four computer scientists -- Chandru, Hariharan, Manohar and Vinay, from the Indian Institute of Science (IISc) in Bangalore, and was the country’s first venture-funded company spun out of a research university. The “Fab Four,” as they were soon dubbed by the media, made news as they were allowed to keep their jobs as professors at the Indian Institute of Science, something unheard of in India. The founders were quick to realize the need for professional management, and brought in
Dr. Srinivasan Seshadri, who was formerly CTO of Lucent’s Network Management Business in the U.S. Strand had obtained first-round funding of about $5 million from WestBridge partners, an Indo-American venture capital fund, in September 2002.

The company enjoyed a reputation for cutting-edge scientific capabilities, enhanced by the continued academic status of its founders. The team was also behind the widely hailed invention called the Simputer, a hand-held computer that cost less than $200 and was aimed at bringing the Internet to rural India. The *New York Times* called it “computing as it would have looked if Gandhi had invented it, then used Steve Jobs for his ad campaign.” In 2002, Strand seemed clearly headed in the direction of product development based on creating intellectual property, though its founders stated that they would eventually like to move into services based on the products and high-end consulting. Strand’s suite of products, “Oyster,” included tools for analysis, integration and visualization. Biotechnology firms from the West as well as within India had begun working with Strand’s products. Seshadri expected to have about 40 employees by early 2003.

Ocimum Bio-solutions was a part of the $50 million Saraca Group (sales figure from fiscal year 2002). Ocimum offered products and contract services covering data integration, analysis and visualization (http://www.ocimumbio.com). The company had about 45 employees by the close of 2002, and was run by two young entrepreneurs, Subhash Lingareddy and Anuradha Acharya. The two had backgrounds in mathematics and physics. Like many other entrants to this industry, Ocimum was experimenting with multiple business models. Among these was a 20-week certification course in BI that it offered in collaboration with Michigan Technical University. Ocimum planned to hire the best of its students for projects, thus developing technical manpower at a profit. In addition to training, Ocimum had also developed a suite of proprietary products in the areas of data integration and analysis. Lingareddy and Acharya believed that their products represented a price difference of as much as 75 percent compared to equivalent products on the market in the West. This was essentially a function of their lower development costs as well as some algorithm innovations. The problem was getting users to switch, as BI software applications tended to generate considerable switching costs for the user. Scientists were used to working with BI freeware that was not very user-friendly or flexible, but the question the founders of Ocimum asked themselves was whether the benefits their software offered were adequate to make scientists willing to pay a price and learn how to use new software.

Ocimum had long-term plans to offer contract R&D services, by investing in “wet-labs” and production facilities. Short term, what worried Lingareddy and Acharya was what would worry any start-up: cash flow. They had even considered the need to possibly partner with a pharmaceutical firm to be a serious player in the discovery value chain. Like Kumar at Ingenovis, they believed that ultimately the best returns were in being involved directly in some way in drug discovery. In the meantime, the search for a viable business model continued.

---

14 *New York Times*, Dec 9, 2001
Exhibit 1: Sequence of activities in drug discovery, development and commercialization
Exhibit 2: The evolution of drug development timelines in the U.S.

Source: Reuters Business Insight
Exhibit 3: Life sciences and IT in drug discovery and development

Drug Discovery > Clinical Development

- Pure Genomics & Proteomics
- Functional Genomics & Proteomics
- Functional Assays
- Virtual Screening
- Combinatorial Chemistry
- Molecular Modelling
- Structural Biology

Pharmacogenomics

HIGH THROUGHPUT SCREENING

BIOINFORMATICS

IT based technologies
Life sciences
Exhibit 4: Map of India showing location of software technology parks

Source: http://www.mapsofindia.com/general
Exhibit 5: Indian labor cost advantages in IT and life sciences

Source: Case writer’s estimates
Exhibit 6: Indian patent law history

<table>
<thead>
<tr>
<th>THE CURRENT LAW</th>
<th>WTO REQUIREMENTS</th>
<th>TRANSITION STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Patents Act, 1970 Allows only process patents</td>
<td>Both process and product patents must be available in all fields of technology</td>
<td>Product patent applications go into a mailbox to be opened by 2005 at the latest</td>
</tr>
<tr>
<td>Duration of process patents is 5 years from the date of approval or 7 years from the date of filing, whichever is less</td>
<td>Duration of patents must be 20 years</td>
<td>Exclusive Marketing Rights (EMR) for mailbox patents can be granted for 5 years</td>
</tr>
</tbody>
</table>

Source: Dr. Reddy’s Laboratories (LBS case).