Bioinformatics: the role and limitations of patents

Teresa J. Welch

The worldwide market for bioinformatics tools and services is estimated, by some, to exceed US$40 billion within the next five years. As with other biotech companies, patent protection will be key to survival in the marketplace. Although the total number of issued bioinformatics patents is still small, virtually all aspects of bioinformatics constitute patentable subject matter, and the opportunity to generate intellectual property value from investment in bioinformatics should not be neglected.

Teresa J. Welch
Michael Best & Friedrich LLP
1 South Pinckney Street
Suite 700, Madison
WI 53703
USA
tel: +1 608 257 3501
fax: +1 608 283 2275
e-mail: tjwelch@mbf-law.com

▼ The recent exponential growth in the rate of biological discovery has been nothing short of phenomenal. Biology, which was once a relatively data-poor science, has become a data-rich science. The ever-increasing surge of data has been due, in large part, to the many international genomics projects that have made available enormous amounts of genetic data. Improvements in instrumentation applications in biological systems, such as mass spectrometry, nuclear magnetic resonance spectroscopy and x-ray crystallography, have also contributed to the deluge of biological data. It is estimated that biological data are doubling every six months [1]. This massive volume of data will be the foundation of the biotechnology and pharmaceutical industries, and bioinformatics is the tool that will transform these data into therapeutic products.

Drug discovery and bioinformatics

Drug discovery is today viewed as ranging from the identification and isolation of genes and proteins, to the production of a drug candidate. The *in silico* processes and tools between these points are the focus of bioinformatics. The promise of bioinformatics is that drug discovery can be accomplished *in silico*, that is, that computational analysis of the massive volume of biological data can identify the causes of disease and enable the design of sophisticated drugs to treat them. In other words, bioinformatics will not only generate a large number of new disease targets, but also provide rational drug design based on these targets to create and test new therapeutics and redesign existing ones. Bioinformatics has the potential not only to reduce the time and money required to develop drugs, but also to reduce the rate of failure of drug candidates and provide drugs that treat disease more effectively and precisely.

The worldwide market for bioinformatics software, databases, hardware and software services, networks and computing methods used to generate and transform these biological data, is estimated to exceed US$40 billion within the next five years [2]. The entry into the bioinformatics sector of traditional information technology companies, such as IBM, Oracle and Hewlett Packard, is a good indicator of the potential for profit [2].

It is not only the genes and compounds – the end-points of drug discovery – that have inherent intellectual property value, but also the methods and structures that comprise bioinformatics. Although all possible protection mechanisms of intellectual property (i.e. patents, trademarks, copyright and trade secret) should be employed, patent protection above all will be a keystone to survival in the marketplace.

Patenting bioinformatics inventions

Identifying patentable bioinformatics inventions

The definition of what is patentable and what is non-patentable varies by country and over time. In most countries, there are certain established statutory categories of patentable subject matter [3]. In the USA, these are an apparatus [3], a process [4], an article of manufacture [3], a composition of matter [3], and a new use for a known composition or process [4]. These statutory

1359-6446/02/$ – see front matter PII: S1359-6446(02)02291-2

categories include all types of machines and instrumentation, methods of manufacture and use of machines and compositions, and consumer products.

In recent years, in the USA as well as elsewhere, the statutory categories have expanded to include inventions in the areas of biotechnology, computers and computer software, business methods, and even algorithms that have practical application [5]. The USA has led the way in patenting in these expanded areas. A US Supreme Court case set the stage more than twenty years ago by holding that 'anything under the sun made by man qualifies as patentable subject matter' [6]. More recent US cases – State Street Bank versus Signature Financial Group [5], and AT&T versus Excel Communications [7] – have cleared the way for patenting of business methods and mathematical algorithms.

The tight integration of database system architecture, peripheral software application programs and data content, is a hallmark of bioinformatics inventions. Although the data themselves are not considered patentable, the specific organization of data with a connecting physical element can be a novel and patentable data structure. Most bioinformatics inventions are computer-implemented methods or systems, and generally fall into the following basic categories: (1) database systems, (2) data analysis, and (3) prediction and design methods. Methods in these categories include generating, collecting and storing data in real-time, searching databases and comparing data, analyzing similarities and clusters, and creating models and simulations that predict and can be electronically interegated and manipulated. The entire range of bioinformatics inventions, including the algorithms that drive them, are potentially patentable.

Not all countries, however, take the US liberal view of classifying software and algorithms as patentable subject matter either in whole or in part [8]. For instance, Europe and Japan both take similar approaches to patenting of software and business methods that differs from the US viewpoint. The fundamental basis of the European patent system is that patents can be granted for inventions that have a technical character; that is, a patentable invention must have technical features. The invention must facilitate the understanding of a technical problem and the solution to that problem; in other words, the invention must make a technical contribution [9]. Although software *per se* is not considered patentable [9], a software invention that also makes a technical contribution is patentable subject matter [10].

Japan has taken a similar approach. As early as 1971, the Japanese Patent Office issued examination standards for computer-program-related inventions [11]. The standards, which were re-emphasized in the 1993 guidelines [12], state that, to be patentable, computer software must use a 'natural law' to produce technical ideas [11]. In other words, a computer-implemented invention that has technical character is not excluded from patentability. The major difference between Europe and Japan, and the USA, is that in the USA, no technical contribution is required to make an invention patentable.

Specific examples of bioinformatics inventions that constitute patentable subject matter include:

- protein modeling systems;
- sequence alignment methods;
- combinatorial library systems;
- computer-aided sequence visualization and analysis systems;
- compositions of matter on microchips; and
- methods of comparing patients' conditions.

Other examples include:

- tools for automation of laboratory experimentation and data generation;
- the design, implementation, management and integration of databases;
- the design of communications standards between databases from multiple sources; and
- data mining tools for retrieving, processing, analyzing and simulating biological data to determine a gene location within a chromosome, find similar genes or proteins from other species, and determine 3D structure and function of different proteins.

There are also display tools that alert the user to what information might be lacking and where different sources of information agree or disagree. Last, but not least, the possibilities for patentable bioinformatics inventions also include biocomputers. Biocomputers use biological materials, such as DNA, as the logic-switches for computers.

Protecting bioinformatics inventions under patent rights

The process of patent-protecting an invention involves the examination, by a patent office, of a patent application, to determine whether it meets certain legal standards (i.e. patent prosecution). A patent application essentially comprises two parts: a specification, which is a detailed description of the invention; and one or more claims, which are statements of the property-right sought to be protected [13]. Each part must meet its own legal standards. If all the requirements are met, the application is allowed and a patent is granted.

In virtually every country in the world, there are three basic legal standards that an invention, or more specifically, claims to the invention, must meet (Box 1). First, the invention must be useful [3] or have industrial or practical applicability [14]. Second, the invention must be novel [15]. Finally, the invention must not be obvious to an artisan having average knowledge in the area of technology of the invention at the time the invention was made [16], or in some countries, the invention must have an inventive step [17].

The specification of the patent application must also meet legal requirements. In the USA, for example, the invention specification must contain a written description [13,18] that is in such detail as to enable those working in the field to make and use the invention without undue experimentation [13]. In addition, the description must contain the best mode; that is, the best mode or way that the inventor contemplates to practice the invention [13]. The best mode requirement is specific to US patent law.

A crucial task when preparing the specification of either a bioinformatics patent application or any software patent application, is to identify and describe the aspects of the invention in a way that adequately covers the invention. This depends on several factors, such as the nature and scope of the invention, the 'types of claims' being sought, the directions being taken by the market and the technology, and applications of the invention. It is often helpful when dealing with bioinformatics inventions, to first describe the functional operation of the invention, without characterizing it as being software or hardware. The embodiments of the invention, such as software implementations, hardware implementations and Internet features, can then be described. If the invention is essentially a mathematical algorithm, consideration must also be given to describing its practical applications, for example, how it produces a useful, tangible result.

Typically, it is not necessary to include either the actual computer program code or flowcharts to describe the software functions [19], although diagrams that illustrate the operations performed by a computer in response to software instructions are very useful. Inventions that make functions available to a user should be described, including, typically, a user interface to enable access to those functions. Often, the user interface is a significant and commercially valuable component of the invention. In this case, the specification should include a description of the various ways and sequences whereby a user applies the functions of the invention, and the manner in which the invention displays information and enables the user to manipulate the information (e.g. screen shots of menus, dialog boxes and output screens).

As with all inventions, regardless of the business model or nature of the invention, bioinformatics inventions should be protected using a wide variety of types of claims. As bioinformatics inventions are an extremely diverse and rapidly changing body of work, 'claiming' these inventions in as many ways as practical is particularly important. The breadth of claims should be carefully considered so as to encompass potential future developments in this rapidly changing technology.

Although most bioinformatics inventions are processes and therefore amenable to method, system or computer product claims (e.g. as computer systems or computer-readable media for data), some bioinformatics inventions are amenable to composition of matter or article of manufacture claims. These include advances in chip or array technologies, libraries and populations of nucleic acids, chemical fragments, detectors and robotics. In addition, claims should be considered in view of the business model of the patent applicant, the parties who might be interested in the invention and the commercial environment applicable to the invention. Claims can be described for business methods, for activities of the parties (e.g. the user), distributors, or retailers who only distribute and sell software.

Box 1. Patent Law basics

Identification of patentable inventions

Meeting legal standards for patentability

- Novelty: new
- Nonobviousness: not an obvious change from what is already known
- Utility: useful; a specific, credible, substantial, practical utility

Protecting inventions

- Application filing
- Examination
- Issuance

Enforcing patent rights

- Infringement evaluation
- Litigation
- Licensing

Examples of bioinformatics patents

Because the USA has led the way in liberalizing patentable subject matter in the software area, it is thus also the leader in the number of issued and pending bioinformatics patents and/or applications, although that number is not large compared to other biotechnology fields. Since 1990, about 125 bioinformatics patents have been issued by the US Patent Office, and about another 400 applications are pending [20]. Thus, the history of patent prosecution in bioinformatics is brief compared with other areas of biotechnology.

Box 2 gives some examples of patents and/or applications either issued or pending in the USA, Europe and worldwide. Notably, many of the claims of bioinformatics patents read like any software or computerized-system-claims, for example, 'A computer system comprising a database…and a user interface' (US6189013) or 'A computer program product comprising a computer useable medium having computer program logic stored therein' (US6321163). Others are clearly directed to biological processes or entities, for example, 'A method of displaying the genetic locus of a biomolecular sequence,

Box 2. Examples of bioinformatics patents and applications

- US Patent Application No. 09/881,234 Apparatus and method for providing sequence database comparison.
- US Patent Application No. 09/820,662 Methods and systems for enabling efficient search and retrieval of records from a collection of biological data.
- US Patent Application No. 09/181,601 Linking gene sequence to gene function by three-dimensional protein structure determination.
- US Patent Application No. 08/792,878 Genomic analysis of tRNA gene sets.
- European Patent Application No. 01009861 Gene expression and evaluation system.
- European Patent Application No. 01002264 Method and apparatus for providing a bioinformatics database.
- European Patent Application No. 1188139 Method and apparatus for predictive cellular bioinformatics.
- European Patent Application No. 0848067 Computer-aided techniques for analyzing biological sequences.
- Japanese Patent Application No. 10-287696 Estimation of function of protein.
- International Patent Application No. PCT/US01/02316 Method, system and computer software for providing a genomic web portal.
- International Patent Application No. PCT/US00/10504 Polypeptide fingerprinting methods, metabolic profiling and bioinformatics database.
- International Patent Application No. PCT/US01/01988 System and method for modeling genetic, biochemical, biophysical and anatomical information: *in silico* cell.
- US Patent No. 6,337,181 Method of specifying vaccine components for viral quasispecies.
- US Patent No. 6,321,163 Method and apparatus for analyzing nucleic acid sequences.
- US Patent No. 6,291,182 Methods, software and apparati for identifying genomic regions harboring a gene associated with a detectable trait.
- US Patent No. 6,232,287 Systems for the analyses of gene expression data.
- US Patent No. 6,189,013 Project-based full length biomolecular sequence database.
- US Patent No. 6,085,188 Method of hierarchical LDAP searching with relational tables.
- US Patent No. 6,023,659 Database system employing protein function hierarchies for viewing biomolecular sequence data.
- US Patent No. 5,966,712 Database and system for storing, comparing and displaying genomic information.
- US Patent No. 5,966,711 Autonomous intelligent agents for the annotation of genomic databases.
- US Patent No. 5,953,727 Project-based full-length biomolecular sequence database.
- US Patent No. 5,808,918 Hierarchical biological modeling system and method.
- US Patent No. 5,795,716 Computer-aided visualization and analysis system for sequence evaluation.
- US Patent No. 5,754,524 Computerized method and system for analysis of an electrophoresis gel test.
- US Patent No. 5,733,729 Computer-aided probability base calling for arrays of nucleic acid probes on chips.
- US Patent No. 5,706,498 Gene database retrieval system where a key sequence is compared to database sequences by a dynamic programming device.

comprising…' (US5970500), or 'A method of producing a gene transcript image analysis, comprising…' (US5840484). Worth noting is a recent patent application that has a claim to a DNA sequence as a data structure [21].

Enforcing bioinformatics patent rights

Enforcing bioinformatics patent rights in an infringement suit is costly. The average cost of a patent infringement suit through trial in the USA is approximately US$1.5 million [22]. Obviously, the more that is a risk in terms of sales and/or profits, the greater the cost. For example, if US$100 million is at risk, the cost might be as much as US$10 million [22]. Evaluating the strength of a bioinformatics patent infringement case, as with all patent infringement cases, involves a number of factors. Potential infringers must first be identified after which an infringement analysis commences. The latter involves construing the meaning and scope of the claims, and determining if the infringer's activity or product falls within the interpretation of the claims. The patent owner must evaluate how much of the business is at risk, the strength of the infringement analysis, including whether a competitor can 'design around' the claims, and whether the patent can withstand a validity challenge (even though the patent is presumed valid). Other considerations include the cost of the suit versus the likely damages that might be paid, and the chances of getting an injunction that prevents the infringer from continuing the infringing activity.

Because most bioinformatics claims are method or system claims, detecting infringement is often difficult, especially when, for example, the infringer is merely offering services. Infringement investigations should include monitoring of advertising, websites, trade or journal articles, US and foreign

published patent applications, and also studying flow charts of operations, and maintenance and service manuals. Discovery of evidence in bioinformatics patent infringement suits, as well as for other software patent infringement cases, can be expensive and time-consuming. A potential litigant must be prepared for the discovery of electronic data and information storage, including archived e-mails, backup tapes and disks, and even erased documents that still might be available on PC hard drives. Nonetheless, although costly and time-consuming for the patent owner, a patent infringement suit is equally as costly and time-consuming for the competitor challenging the patent.

Alternatively, a patent owner might wish to allow others to make, use and/or sell the patented invention. Patents can be licensed to third parties, and most companies, universities and research institutions have set up licensing programs. The benefits of a licensing program include considerable revenues in the form of royalties paid by licensees, limiting of placing the patent at risk, and controlling the use of the patented invention. For some companies, royalties can be a significant proportion of the company's overall revenues, for example, those companies involved in patent pools [23].

Limitations on bioinformatics patent rights

Reasons not to use the patent route

There are reasons not to file a patent application. Some inventions might be best kept as a trade secret [24]. A trade secret is, in effect, the opposite of a patent. In patenting, inventors must provide a full account of the invention for which they are given this limited property right. A trade secret can be kept for an infinite amount of time, as long as it remains secret. It might be more lucrative to maintain an invention as a trade secret rather than to disclose it under the patenting requirements. For example, Coca-Cola originally sought patent protection for Coke, but dropped the application in light of the disclosure requirements. The formula for Coke turned out to be much more valuable as a trade secret.

In answering the crucial question of whether to maintain an invention as a trade secret or to patent it, one must consider the scope of patent claims that may be issued, the degree of difficulty anticipated in enforcing the patent, the expected commercial lifetime of the invention, the relevant business model (e.g. the sale of tools or services), and the ease of reverse engineering or independent creation or discovery. For rapidly changing technologies such as bioinformatics, a trade secret might be an important option, especially in view of the pendency period of patent applications (see later).

A crucial element of the trade secret route, is whether reasonable measures are in place to protect the invention claimed to be a trade secret, and if the trade secret finds its way to a competitor, whether a company can prove that it took the measures necessary to show that the invention was, in fact, a secret worthy of a court's protection. This requires proactive implementation of policies, which might include employing security guards and/or surveillance cameras at entrances and exits; confidentiality agreements with employees, outside consultants and potential suitors; prohibiting the use of company work on e-mail, and restricting trade secret information to only those who need to know.

The trade secret route, however, leaves little recourse against others who subsequently appropriate the invention, either by legally copying (e.g. reverse engineering) or independent creation. In the case of the latter, patenting by another who independently developed the trade secret, might prevent an owner from practicing its own invention.

Limitations on protecting database *per se*

In the USA, data are facts and are therefore not considered inventive. A collection of data, such as a database, is thus not patentable *per se*. As discussed previously, however, claims linked to computer systems and computer programs that interact with biological information stored in a computer database, are likely to be considered patentable subject matter.

Europe has taken a somewhat different approach in the recent European Union directive, which establishes a new proprietary right in databases *per se* [25]. This right is significantly different from a patent right. As mentioned previously, the European Patent Office as well as the Japanese Patent Office has been more restrictive in the patentability of computer-related inventions, although software inventions that meet the technical contribution requirement have been patented.

Timing of the patenting process

Patent prosecution can be a lengthy process. In the USA, the duration of pendency of a software-related patent application could be as long as three years or more, and publication of the application eighteen months from its filing alerts competitors to the technology. As with other computer systems and programs unrelated to biology, a lengthy duration of pendency could render a bioinformatics invention an 'old technology' by the time a patent is issued. Part of the decision to file any computer-related application should include an assessment of whether the technology is rapidly changing and whether it is possible to secure broad claims that cover future improvements.

Cost and difficulty of proving infringement

The cost and difficulty of proving infringement for many bioinformatics patents are significant problems. Associated costs could be prohibitive for a small company. Most claims for bioinformatics inventions are method or process claims. Proving infringement for any process claim is difficult, and more so for bioinformatics claims. For example, if a competitor is using a patented program to model protein function using a proprietary

database, there are few avenues to discover infringement of patented claims to that program.

Concluding remarks

In summary, virtually all aspects of bioinformatics constitute patentable subject matter. Most bioinformatics inventions can be patented with claims similar to other computer methods. Once a patent is procured, consideration must be given to whether the patent owner wishes to exclude all others from practicing the invention, or whether the owner should permit use through a licensing program. Although there are certain limitations with patents, the reality is that patents usually have a significant deterrent effect.

In a real sense, bioinformatics is the brains behind biotechnology. Bioinformatics is the only way to transform the promise of genomics and proteomics into products. The opportunity to generate intellectual property value from investment in bioinformatics should not be neglected. Companies that establish patent-protected bioinformatics platforms, and leverage these assets, can prosper in the marketplace.

References

1 Flanagan, A. (2000) Racing to conquer biodata with computers. *Genetic Engineering News* 20, p.8

2 Licking, E. *et al.* (2001) Bioinformatics. *BusinessWeek Online* (March 23, 2001; http://www.businessweek.com)

3 United States Code, Title 35, §101

4 United States Code, Title 35, §100

5 State Street Bank and Trust Co. versus Signature Finance Group Inc., 149 F.3d, 1368 (Fed. Cir. 1998)

6 Diamond versus Chakrabarty (1980) 447 US 303

7 AT&T Corp. versus Excel Communications Inc., 172 F.3d 1352 (Fed. Cir. 1999)

8 Patent Cooperation Treaty (2001) Regulations, Rule 39, World Intellectual Property Organization, Geneva

9 European Patent Convention, Article 52(2)

10 European Patent Office (1999) Technical Board of Appeal, Decision of Case T1193/97

11 Japanese Patent Office (1971) An examination standard for computer program related inventions.

12 Japanese Patent Office (1993) Examination guidelines for computer program related inventions.

13 United States Code, Title 35, §112

14 Patent Cooperation Treaty (2001) Article 33(4)

15 United States Code, Title 35, §102

16 United States Code, Title 35, §103

17 Patent Cooperation Treaty (2001) Article 33(3)

18 United States Patent and Trademark Office (1999) Revised Interim Guidelines of Examination of Patent Applications Under the 35 U.S.C. §112, para. 1, Written Description. Fed. Reg. Vol. 64, No. 244 (December 21, 1999)

19 Fonar Corp. versus General Electric Co, 41 U.S.P.Q.2d 1801 (Fed. Cir. 1997)

20 B. Quinn, Office of Public Affairs, United States Patent & Trademark Office, Communication (February 2002)

21 Fleishman, R.D. *et al.* (1996) [Human Genome Sciences, Inc., Johns Hopkins University] Nucleotide sequence of the Haemophilus Influenzae Rd genome, fragments thereof and uses thereof, International Patent Application Publication WO96/33276, (published October 24, 1996)

22 Ryan, T.P. (1999) Markman: Where are we now? An update on developments and trends on claim construction. 1999 IP Summer Conference, American Bar Association, 23–27 June 1999, San Francisco, CA, USA

23 Kulbaski, J., A primer in patenting standards and patent pools, Wisconsin Intellectual Property Law Association, Madison, WI, USA, (January 17, 2002)

24 Uniform Trade Secrets Act. Sec. (1980) 1(4)

25 European Union Database Directive (1996) Chapter III, Articles 7–11

Information Biotechnology: A Supplement to ***Drug Discovery Today***

How to cite an article in this supplement

Augen, J. (2002) Bioinformatics and information technology: reshaping the drug discovery process. *Drug Discov. Today.* 7 (Suppl.), S39–S40

How to cite the whole supplement

Lawrence, R. ed. (2002) *Information Biotechnology: A Supplement to Drug Discovery Today* (Vol. 7, No. 11), Elsevier Science