BIOINFORMATICS: PATENTING THE BRIDGE BETWEEN INFORMATION TECHNOLOGY AND THE LIFE SCIENCES

CHARLES VORNDRAN, Ph.D.* **AND**

ROBERT L. FLORENCE**

INTRODUCTION

Over the past two decades, advances in the fields of genomics and proteomics have enabled life scientists to amass an enormous amount of complex biological information. The ability to gather such information, however, has surpassed the ability and speed at which it can be interpreted. The growing field of bioinformatics holds great promise for making sense of this biological information through the development of new and innovative computational tools for the management of biological data. As with any developing field, the continued advancement and commercial viability of such innovations often depends on successful patent protection. This article, therefore, discusses the patentability of computer and software related inventions in this unique and evolving field.

Part I begins by defining bioinformatics and bioinformatic-tools. A brief discussion of the main segments of the bioinformatics market is provided with examples from each segment. An explanation of how our bodies store and regulate the flow of genetic information is also provided in order to

* Associate in the Atlanta office of Troutman Sanders LLP; J.D. 1997, Southern Methodist University; Ph.D. Biological Sciences 1995 concentrating on Cellular/Developmental/Molecular Biology, University of Texas at Austin; B.Sc. Zoology, 1988, University of Texas at Austin. The views and opinions expressed herein are solely those of the authors and not necessarily those of Troutman Sanders LLP.

** J.D. Candidate 2002, Franklin Pierce Law Center; B.Sc. Biology, 1998, University of Utah.

demonstrate the usefulness of bioinformatics related inventions. Following this explanation the scientific principles of genomics and proteomics are discussed. Particular attention is given to the vast amounts of data being generated by new techniques within these fields and the problems caused by the rapid accumulation of such data. Bioinformatics is offered as a solution to many of these problems with the recommendation of protecting bioinformatics related inventions by patent.

Part II explains the requirements of patent protection on such inventions as they relate to this unique field. A large portion of this section is dedicated to providing a history of the patentability of computer related inventions along with an examination of related case law. Individual discussion of each of the statutory patentability requirements follows thereafter. Although other forms of intellectual property protection can and should be sought, they are beyond the scope of this article and are, therefore, not discussed.

Part III explores the future of bioinformatics with emphasis on showing why bioinformatics is not just another fad. In addition, Part III provides some predictions and trends. This part begins with a brief look at the unusual problems caused by the complex nature of bioinformatics inventions. Specifically, these problems are due to the multidisciplinary nature of the field. Following next is a discussion regarding how bioinformatics has permanently changed the way life scientists conduct research along with academia's reaction to the change. Finally, current business trends in the bioinformatics industry are examined along with government involvement and some promising future innovations.

I. WHAT IS BIOINFORMATICS?

A. Defining "Bioinformatics" and "Bioinformatic-Tools"

Just what is bioinformatics? In the broadest sense, it is the science of using information to understand biology.[1] Essentially, bioinformatics originated from computational biology, which is "the application of quantitative analytical techniques in modeling biological systems."[2] Bioinformatics draws researchers from the disciplines of biology, computer science, statistical

---

1 Cynthia Gibas & Per Jambeck, *Developing Bioinformatics Computer Skills*, 3 (O'Reilly & Assoc. 2001).

2 *Id.*

mathematics, and linguistics.[3] Now that bioinformatics has achieved "buzzword status," however, its definition has varied depending on who is using it.[4] For the purpose of this article, therefore, bioinformatics will be defined more precisely as using computers and computer related tools for the management of biological information.[5]

Using this definition, bioinformatic-tools include any products that store, organize, evaluate, integrate, analyze, and/or distribute biological data.[6] The bioinformatics market can be divided into four main segments: databases, software, hardware, and custom consulting.[7] Databases are the "heart" of bioinformatics because the first step in the bioinformatics process is usually the accumulation of biological data for storage.[8] Bioinformatics software can be used to convert the raw biological data by sorting, integrating, analyzing, and distributing database contents.[9] The software segment is currently dominated by data mining software due to its capability of searching databases for relationships, patterns, and functions between biological components.[10]

Bioinformatics hardware is similar to traditional computer hardware such as servers, desktop computers, and other storage devices except that it is typically designed with faster processing ability and greater storage capacity.[11] Finally, the smaller but growing segment of custom consulting deals with designing customized systems that integrate all aspects of bioinformatic capabilities.[12] These four segments combined exceeded an estimated $700 million in the year 2000 alone, making bioinformatics one of the fastest growing areas of all life science related markets.[13]

---

3 *Id.*

4 *Id.*

5 Strategic Directions International Inc., *Bioinformatics: The Rosetta Stone of Life Science Research*, Instrument Business Outlook no. 21, vol. 9, p.1 (February 15, 2001).

6 *Id.*

7 *Id.*

8 *Id.* at 2.

9 *Id.*

10 *See* Strategic Directions International Inc., *supra* n. 5, at no. 21, vol. 9, p. 2. Visualization software is also extremely popular with its ability to model molecular structures for analysis. *Id*.

11 *Id.*

12 *Id.*

13 *Id.* at 1.

B. The Flow of Genetic Information

To better understand and appreciate the usefulness of bioinformatics, it is helpful to review how our bodies store and regulate the flow of genetic information. Other reasons for such a review include the complex nature of the biological information that the bioinformatics industry seeks to manage and the multidisciplinary nature of the field itself with bioinformaticians hailing from many different backgrounds. For simplicity, the following discussion will focus on eukaryotes. Residing inside every organism is a genetic blueprint for building and maintaining that organism's physical structures throughout its lifetime.[14] This complete set of instructions is known as an organism's genome.[15] Our genetic information is stored in the famous[16] double helix molecule, deoxyribonucleic acid, or DNA, residing in the nucleus of most eukaryotic cells.[17] DNA is a polymer made up of the four different nucleotide bases: adenine (A), guanine (G), cytosine (C), and thymine (T).[18] It is the sequence of A's, G's, C's, and T's that allows DNA to be the information carrying material known as genes.[19] Genes are "generally arranged in linear arrays along the chromosomes of a cell."[20] Understanding exactly how our cells utilize genetic information requires further discussion of the structure and organization of our DNA.

The double helix molecule of DNA is formed from two very long, helical polynucleotide chains running in opposite directions while coiled around a common axis.[21] Each chain, or strand, is constructed of a sugar-phosphate backbone forming the outside of the helix, whereas the aforementioned nucleotide bases, the A's, G's, C's, and T's, face the inside.[22] The two chains are held together by hydrogen bonds between complimentary base

---

14 Sir John Kendrew et al., *The Encyclopedia of Molecular Biology* 421 (Blackwell Science 1995).

15 *Id.*

16 *Id.* at 268. Oswald Avery first identified DNA as the cell's heredity material in 1944. *Id.* It was James Watson and Francis Crick, however, who later elucidated DNA's three-dimensional structure in 1953 using x-ray diffraction techniques. *Id.*

17 *Id.* at 267.

18 *Id*. at 267-268.

19 Kendrew et al., *supra* n. 14, at 267-268.

20 *Id*.

21 Lubert Stryer, *Biochemistry* 80 (4th ed., W.H. Freeman & Co. 1995).

22 *Id*. at 81.

pairs.[23] Adenine (A) always pairs with thymine (T), while guanine (G) is always paired with cytosine (C).[24] Thus, one strand of the double helix is the exact compliment of the other.[25]

The total number of these complimentary base pairs usually determines the size of an organism's genome.[26] For instance, there are roughly 3 billion base pairs in the human genome.[27] Scientists also estimate the number of genes in the human genome to be between 35,000 and 45,000.[28] A gene is a sequence of DNA that codes for a particular protein and is expressed when the cell uses the DNA sequence to make the protein.[29] In turn, the resultant proteins serve many vital functions in cellular and non-cellular organisms.[30] Significant protein functions include everything from making up cellular structure, catalyzing biochemical reactions, acting as receptors for hormones and other signaling molecules, to regulating gene expression itself.[31] Therefore, the flow of genetic information from DNA to protein is highly regulated and often tightly coordinated.[32]

The genetic information stored in a gene is regulated through two sequential steps: transcription, where a linear portion of DNA is copied, and translation, where the cell uses the copied DNA to make the corresponding protein.[33] Transcription is the principle point of controlling gene expression by using a class of information-carrying intermediates called messenger RNA's or mRNA's.[34] Although the exact details of transcription are quite complex,

---

23 *Id*.

24 *Id*.

25 *Id*.

26 Kendrew et al., *supra* n. 14, at 124. This number is also referred to as the C-value. *Id.*

27 *Id.*

28 Elizabeth Pennisi, *The Human Genome: News*, 291 Sci. 1177, 1178 (2001). This number is considerably less than earlier estimates of 100,000. *Id.*

29 Kendrew et al., *supra* n. 14, at 400.

30 *Id*. at 879.

31 *Id*.

32 *Id*. at 400.

33 *Id*. at 1078 and 921.

34 Stryer, *supra* n. 21, at 95. RNA is similar to DNA except that it is single stranded and the thymine nucleoside (T) is transcribed into uracil (U) in RNA. Sir John Kendrew et al., *supra* n. 14, at 975. It should be noted that there are other forms of RNA involved in the flow of genetic information, mainly transfer RNA (tRNA) and ribosomal RNA (rRNA). *Id.*

the cell essentially uses the chromosomal DNA as a template for transcribing molecules of mRNA.[35] The resulting mRNA molecules are complementary to the corresponding DNA, thus the coded sequence is transferred to the mRNA.[36] Because genomic DNA contains sections of DNA that do not code for any part of a protein, mRNA is processed to remove the complementary non-coding regions called "introns."[37]

Transcribed mRNA is then transported from the nucleus to the cytoplasm of the cell for translation into proteins.[38] In the cytoplasm, the cell uses its machinery to translate the already transcribed genetic code into corresponding amino acids, which are then assembled into proteins.[39] The cell determines the appropriate amino acids by translating the code in triplets (three consecutive nucleotides of code) called codons.[40] Each codon specifies an individual amino acid and the sequence of these codons determines the amino acid sequence required for manufacturing the desired protein.[41]

Because there are four kinds of nucleotide bases in DNA, there are theoretically sixty-four possible codon triplets capable of coding for separate amino acids.[42] In reality, however, proteins are built from a basic set of only twenty different amino acids.[43] This is due to the fact that for most amino acids, there is more than one codon.[44] Thus, the genetic code is highly degenerate with many amino acids being coded by more than one triplet.[45]

The resulting protein molecules are typically long polypeptide chains of these amino acids containing several thousand atoms.[46] These chains are

---

35 Stryer, *supra* n. 21, at 101.

36 *Id.*

37 *Id.* at 113. While the complementary sections retained in the mRNA are called "exons." *Id.*

38 Kendrew et al., *supra* n. 14, at 921.

39 *Id.*

40 *Id*.

41 *Id.*

42 Stryer, *supra* n. 21, at *Biochemistry* 103. This number is based on the mathematical combination of $4^3 = 64$. *Id.*

43 *Id*.

44 *Id*. at 109. For example, the codon sequences GCU, GCC, GCA and GCG all code for the amino acid alanine. *Id.* at Table 5-4.

45 *Id.*

46 Kendrew et al., *supra* n. 14, at 911.

not necessarily the final product, however, as they may undergo several modifications before taking on final functional forms.[47] One such modification is acquiring complex three-dimensional folding patterns.[48] Such folding is the point where the flow of genetic information is complete leaving behind the "one-dimensional" storage medium of the genetic code and entering the functional "three-dimensional world" of which we are all familiar.[49] Other modifications include attaching various substances to the proteins such as sugars or fats.[50]

C. Genomics

Over the years, scientists have developed several techniques that facilitate manipulation of our genetic information. One important innovative scientific method, called recombinant DNA technology, began emerging in the 1970s.[51] Based on nucleic acid enzymology, recombinant DNA technology uses a combination of different enzymes with the ability to cut, join and replicate DNA, as well as reverse-transcribe RNA.[52] Another foundation of recombinant DNA technology is the complimentary base pairing of nucleic acids.[53] New DNA combinations can be formed due to this base pairing as well as locating and replicating desired sequences.[54] Without question, this new technology has revolutionized the life sciences by enabling scientists to disrupt the flow of genetic information, and actually manipulate it by moving DNA sequences from one molecule to another.[55]

It should be noted that recombinant DNA technology is presently a blanket term for many different technologies, of which only a few will be discussed herein. The first such technology involves sequencing techniques. One of the most common methods for sequencing DNA is the Sanger

---

47 *Id.* at 923-24.

48 *Id.* at 911. Although there are numerous folding patterns, some common underlying features include α-helices and β-sheet conformations and dense packing of any hydrophobic residues in the protein's interior. *Id.*

49 *Id.*

50 *Id.* at 870.

51 Stryer, *supra* n. 21, at 119.

52 *Id.*

53 *Id.*

54 *Id.*

55 Stryer, *supra* n. 21, at 119.

method.[56] The Sanger method uses enzymes to make new DNA chains from a target molecule while using "dideoxy" reagents to randomly stop the chains while growing.[57] Thus, chains of different lengths are produced which can then be analyzed using electrophoresis.[58] This technique is of specific importance because it was the first method used to determine the entire sequence of a DNA genome.[59]

Another important technology involves DNA cloning. A number of methods have been developed to purify and isolate individual fragments of DNA for cloning by a bacterial host.[60] Due to the parasitic nature of viruses, both plasmids[61] and λ phage[62] are often used as a carrier DNA[63] into which scientists splice an isolated fragment of DNA for cloning.[64] These carriers can be used to transform the genetic information in bacterial cells by infecting them with the carrier.[65] Once infected, the carrier replicates, making many copies of the inserted fragment of DNA.[66] The result is a collection of cloned DNA from a single source called a library.[67] Scientists have used this and similar techniques to genetically engineer various plants, animals, and fungi, leading to a number of practical applications in agriculture, industry, and medicine.[68]

Finally, both DNA sequencing and cloning have become greatly simplified by a technique called polymerase chain reaction, or ("PCR"). PCR

---

56 *Id.* at 123. This technique was developed by Frederick Sanger and his co-workers and is also called the dideoxy method. *Id.*

57 *Id.*

58 *Id.*

59 *Id.* at 124. Sanger used the method to determine the complete sequence of ϕX174 (a bacteriophage) DNA in 1977. *Id.* Several years later, the complete sequence of human mitochondrial DNA was also determined using the technique. *Id.*

60 Kendrew et al., *supra* n. 14, at 272. Yeast artificial chromosomes and eukaryotic viruses have also been used as hosts. *Id.*

61 Plasmids are small, naturally occurring, extrachromosomal pieces of DNA that can autonomously propagate in cells. *Id.* at 856.

62 Lambda (λ) is a bacterial virus first isolated in 1951. *Id.* at 573.

63 Such carriers are called vectors. *Id.* at 272.

64 *Id*.

65 *Id.* at 272.

66 *Id.*

67 *Id.*

68 *Id.* at 417-418.

is a method[69] for amplifying specific DNA sequences *in vitro* utilizing thermostable enzymes to catalyze rapid synthesis.[70] The method is carried out in three main steps: strand separation of the target DNA, hybridization of primers, and DNA synthesis at a particular temperature, all of which can be carried out simply by repeating the temperature cycle of the mixture.[71] A key feature of PCR is that each newly synthesized strand acts as a template for successive amplification.[72] Thus, amplification of DNA sequences is exponential with the number of synthesized strands doubling every cycle.[73]

These revolutionary techniques, along with many others, are the means for discovery in the field of genomics. Genomics focuses on the study of whole sets of genes and the interactions between them.[74] Genomic scientists have made numerous important contributions and discoveries. For example, over a thousand disease-causing genes have been identified since 1981.[75] Studying the causes of such inherited single-gene disorders has become easier by mapping[76] chromosomes to identify the genes responsible.[77] One of the largest undertakings of genomics thus far is the mapping of the human genome with the goal of increasing our understanding of biological processes and our ability to treat diseases.[78]

In 1988, the National Institutes of Health ("NIH") and the Department of Energy ("DOE") agreed to collaborate on an effort to sequence the entire human genome that became known as the Human Genome Project ("HGP").[79] The HGP's main focus from the start was on

---

69 Kary Mullis invented PCR in 1983 while working for Cetus Corporation. *Id.* at 864.

70 *Id*. at 864-865.

71 *Id.* at 864.

72 *Id.*

73 *Id.*

74 Leslie Roberts, *A History of the Human Genome Project*, 291 Sci. 1195, 1200 (2001).

75 Leena Peltonen and Victor A. McKusick, *Genomics and Medicine: Dissecting Human Disease in the Postgenomic Era*, 291 Sci. 1224, 1225 (2001).

76 To map a genome means to determine the positions and spacing of genes or defined DNA sequences along each chromosome. Kendrew et al., *supra* n. 14, at 494. Physical mapping usually measures the actual spacing between landmarks in terms of base pairs. *Id.* While another type of mapping called genetic mapping gives the relative positions of genes, or other markers that follow inheritance patterns, on a chromosome. *Id.* at 499.

77 *Id.*

78 *Id.*

79 Leslie Roberts, *Controversial from the Start*, 291 Sci. 1182, 1185 (2001).

mapping the human chromosomes instead of sequencing.[80] This focus changed in May of 1998, however, when Craig Venter[81] announced the formation of a new company, later named Celera Genomics, with the goal of producing the complete human genome sequence within the next three years using a new technology[82] recently developed.[83] Faced with sudden competition, the NIH and DOE responded by announcing a new goal of having a working draft of ninety percent of the human genome by 2001.[84] Thus, the race was on to be the first to provide the world with the definitive "book of life."[85]

The race ultimately ended in a tie with each group agreeing to publish separate drafts of the sequence in February of 2001 after failing to put aside their differences long enough to pool any data.[86] Lack of cooperation aside, the reaching of this sequencing landmark is being heralded as a phenomenal achievement considering that only fifteen years ago most researchers did not believe it was even possible.[87] Added to this is the fact that the human genome is more complex and close to twenty-five times larger than any previously sequenced genome.[88] Indeed, the sequence making up the human genome is large enough to fill two hundred New York City phone books.[89]

---

80 *Id.* at 1184.

81 Venter originally worked for NIH on the HGP until he left in 1991 when he was offered $70 million to try out a new gene identification strategy called expressed sequence tags at the Institute for Genomic Research. *Id.* at 1185-1186.

82 The new technology was whole-genome shotgun sequencing which Venter and his colleagues had already used to sequence the genome of *Haemophilus influenzae* in 1995. *Id.* at 1186. Venter further tested the shotgun strategy by sequencing the genome of the fruit fly, *Drosophila melanogaster*, proving the method could work on a large, complex genome. *Id*. at 1188.

83 *Id.* at 1187.

84 *Id*. at 1188. It should be stressed that the mapping and sequencing methods utilized by the HGP differed from those utilized by Celera Genomics and an informative article detailing HGP's methods can be found at The International Human Genome Mapping Consortium, *A Physical Map of the Human Genome*, 409 Nat. 934-941 (2001).

85 *Id.* at 1187-1188.

86 Pennisi, *supra* n. 28, at 1178. Celera's results were published in the February 16th issue of *Science* while the HGP's results were published in the February 15th issue of *Nature. Id.*

87 *Id.*

88 *Id.*

89 *Id.*

In addition to the initial sequencing of the human genome, the complete genomes of over sixty other species are also now available in databases.[90] Thus, in a short period of time, scientists have come from knowing almost none of life's genetic details to possessing an immense amount of information regarding the structure[91] of individual genes.[92] It is hoped that by studying and comparing the genomes of various species, scientists will learn how variations in our genetic instructions ultimately cause disease.[93]

D. Proteomics

The data scientists have gathered from mapping various genomes is only a small layer in an ever-increasing mountain of information. Researchers continue to generate gigantic databases containing information on how our genes are turned on and off, the proteins they encode, how those proteins interact, and the different roles played by such interactions.[94] Although complete sequences provide a genetic blueprint, it is the proteins that do the actual work, so the "future" really belongs to proteomics.[95] Proteomics is "the analysis of complete complements of proteins," which includes "the identification and quantification of proteins" through "the determination of their localization, modifications, interactions, activities and . . . functions."[96]

Overall, proteomics may be much more challenging than the study of DNA. A cell often makes embellishing changes to proteins after their initial manufacture.[97] Depending on a protein's ultimate destination, it may become sulphated, ubiquitinated, glycosylated, phosphorylated, acetylated, linked to various types of anchors, or changed in many other ways.[98] Added to this

---

90 Peltonen and McKusick, *supra* n. 75, at 1224.

91 For instance, less than a year and a half prior to the publishing of the human genome sequence, scientists knew the true position of only ten percent of the sequences, whereas, ninety percent were represented at the time of publishing. Pennisi, *supra* n. 28, at 1178.

92 Peltonen and McKusick, *supra* n. 75, at 1224.

93 *Id.* at 1224-1225.

94 *See* Ken Howard, *The Bioinformatics Gold Rush*, Scientific Am. 58, 58 (July 2000).

95 Stanley Fields, *Proteomics: Proteomics in Genomeland*, 291 Sci. 1221, 1221 (2001).

96 *Id*.

97 *Id.*

98 *Id.*

complexity is the fact that single genes can, and often do, encode more than one type of protein.[99]

Most proteins are only functional when interacting with "coordinated networks" of other proteins.[100] A major problem with studying proteins involved in metabolic pathways is that only a fraction of these networks have been identified and characterized.[101] Another difficulty is that the proteome (the total protein content of a cell at a single moment in time) is a moving target that is different in every cell.[102] As a cell's environment changes, so do the number and functions of proteins.[103] Finally, some proteins may control more than one process, while multiple proteins regulate similar processes.[104] Despite these difficulties, scientists are developing systematic methods and approaches for identifying new proteins leading to the promise of additional discovery.[105]

One of the greatest areas of promise for proteomics is that of drug discovery. Because "[m]ost drugs are small molecules that modify the function of a specific protein", proteomics may speed the drug development process, thus, saving both time and money.[106] The process of getting a drug to market can take years.[107] Identifying an initial compound with therapeutic possibility is a time and money intensive process.[108] Once a promising compound has been identified, the process of pre-clinical testing and FDA approval eliminates most drugs from ever making it to the market at all.[109] As

---

99 *Id.* Possibilities for this can be alternative splicing of the transcribed mRNA, frameshifting of the mRNA during translation, and varying of stop and start sites. *Id.*

100 Peltonen and McKusick, *supra* n. 75, at 1226.

101 *Id*.

102 *See* BioVenture Publishing, *Getting to the Business End of Biology*, BioVenture View n. 4, vol. 15, p. 5 (April 1, 2000).

103 Fields, *supra* n. 95, at 1221.

104 *Id.*

105 *Id.* at 1224.

106 John Thackray, *Bioinformatics Grows Legs: The Exploding Need for Tools to Analyze and Mine Life-Science Data Promises a Bonanza for some IT Firms*, Electronic Business 76, 80 (July 1, 2001).

107 *See* Arti K. Rai, *The Information Revolution Reaches Pharmaceuticals: Balancing Innovation Incentives, Cost, and Access in the Post-Genomics Era*, U. Ill. L. Rev. 173, 181 (2001).

108 *Id.*

109 *Id*.

a result, the process of discovering and developing just one marketable drug can take over twelve years at a cost of hundreds of millions of dollars.[110]

Bioinformatics may speed discovery by finding better drug targets earlier in the development process.[111] If a pharmaceutical company gets a drug to market even one year sooner, it could mean as much as $500 million alone in incremental sales.[112] Bioinformatics may also aid in trimming the number of targets being tested by drug companies, thus, saving additional time and money.[113] The resulting savings may increase the incentive for additional drug development, as companies are able to lengthen the time a drug is on the market before the expiration of any patent rights.[114] Thus, the pharmaceutical industry's need for more efficient drug development has become a driving force behind the massing of information on a variety of data, including protein sequences, structures, and biomolecular pathways.[115]

E. Finding the Right Tool for the Right Job

Developments within the fields of genomics and proteomics have reached the point where scientists are now amassing vast amounts of data faster than they can interpret it.[116] To put the amount of information into perspective, consider an analogy to a symphony of music. Scientists can clearly hear the music being played and can even determine some of the instruments but they cannot see the notes. In fact, the sequencing of the human genome was recently compared to collecting and reading all the music ever published.[117] While documenting the proteome, the more difficult task, has been compared to capturing and understanding every musical performance ever made.[118] The key to capturing and understanding this symphony of information is bioinformatics.

Bioinformatics provides algorithms, databases, user interfaces, and statistical methods, all of which are important tools that will enable scientists

---

110 *Id*. *See also*, Thackray, *supra* n. 106, at 80.

111 Howard, *supra* n. 94, at 58.

112 *Id.*

113 *Id.*

114 *Id.*

115 *Id.*

116 *See* David S. Roos, *Bioinformatics – Trying to Swim in a Sea of Data*, 291 Sci. 1260, 1260 (2001).

117 *See* BioVenture Publishing, *supra* n. 102, at p. 5.

118 *Id*.

to sift through and make sense of all this information. Such tools are essential to scientists interested in comparative genomics, or the comparison of different organisms' genomes.[119] Without the ability to effectively mine these genomes for information, the sequences themselves are less useful because they don't reveal what genes do.[120] Another complication is the fact that less than three percent of the human genome sequence is thought to contain coding regions for genes.[121] The other ninety-seven percent contains unidentified sequences thought to be important in gene regulation such as promoters, which turn genes on and off, and transcription regulating sequences.[122] Therefore, new computational methods and tools are needed to explore this area of the sequence as well.

Although bioinformatics is a developing discipline, it is already becoming both "the mother and daughter of invention."[123] The need for new and innovative research methods is driving the development of sophisticated bioinformatic tools designed for speeding discovery.[124] A prime example is the decade of computer innovation dedicated to the mapping of various organisms' genomes.[125] Celera used one of the largest supercomputer systems ever created to unravel the DNA sequences during their sequencing of the human genome and is currently building gigantic databases from the resulting information.[126] The rapid sequencing technique utilized by Celera required more than 600 Alpha processors from Compaq each capable of a trillion operations a second.[127] The final computations also required 64 gigabytes because of the algorithms involved and the enormous amount of data.[128]

The pharmaceutical industry has similar computational needs, as it now requires petabytes (approximately 1,000 terabytes) of storage for the

---

119 Elizabeth Pennisi, *What's Next for Genome Centers?*, 291 Sci. 1204, 1207 (2001).

120 *Id.*

121 *See* Gretchen Vogel, *Objection #2: Why Sequence the Junk?*, 291 Sci. 1184, 1184 (2001); *see also* Ewan Birney et. al., *Mining the Draft Human Genome*, 409 Nat. 827, 827 (2001).

122 Birney et. al., *supra* n. 121, at 827.

123 Thackray, *supra* n. 106, at 76.

124 *Id*.

125 *Id*.

126 Worldwide Videotex, *Compaq Technology Enables Completion of Human Genome*, 13 Computer Workstations n. 8, vol. 13 (August 2000).

127 *Id.*

128 *Id*.

complex biomedical data.[129] This level of technological advance rivals that of government agencies, who are traditionally supercomputing's biggest customer.[130] Handling such data requires new and expanded methods beyond those currently available.[131] As a result, information technology firms are struggling to provide the pharmaceutical industry the tools it needs to "de-bottleneck" the vast amounts of data.[132]

Successful protection of innovations in bioinformatics is crucial to the continued advancement of the field and the commercial viability of the businesses sprouting up to fill the computational needs of the most basic life science researcher. Recent court decisions in the Court of Appeals for the Federal Circuit ("Federal Circuit") have dramatically changed the landscape of the patent system particularly for software and business method patents.

II. BIOINFORMATICS PATENTS

A. A Software Patent by any Other Name

1. Patentable Subject Matter Under *State Street*

As previously discussed, patents on bioinformatics are directed to processes combining the use of a computer in combination with biological information to identify, for example, the expression of specific genes or the function of a protein or protein subunit.[133] Thus, patents on bioinformatics are typically process or method patents relating in some aspect to computers or software and are typically associated with the catch phrases "data mining" or "predictive modeling."[134] Using a computer to "mine data" or in predictive modeling is not new, and at first blush, begs the question as to how bioinformatics can be the subject matter of a patent at all.

---

129 *See* Hank Simon, *Solving the Mystery*, Intelligent Enter. n. 18, vol. 3, p. 56 (Dec. 5, 2000).

130 Thackray, *supra* n. 106, at 79-80.

131 *Id*.; *see also* Simon, *supra* n. 129, at 56.

132 *See* Chemical Week Associates, *Debottlenecking R & D*, Chemical Week 30 (Jan. 31, 2001).

133 See U.S. Pat. No. 6,185,561 (issued Feb. 6, 2001) and US Pat. Application 20010016314.

134 *The Bioinformatics Gold Rush—Molecular Mining in the 21st Century* Business Wire June 30, 2000.

All patentable inventions, including inventions directed to the field of bioinformatics, must fall within statutory subject matter, be novel or unanticipated, non-obvious, and enabled.[135] Historically, patents to software alone, and methods of doing business were believed unpatentable because they were not considered proper statutory subject matter.[136] Software has been held as patentable subject matter, however, if claimed as a machine, claimed using means plus function language, or claimed in a tangible medium.[137] Recently, the Federal Circuit specifically held that software alone is patentable, and that methods of doing business can be the subject matter of a patent.[138] The U.S. Supreme Court has yet to grant certiorari challenging such patentability. The result of the Supreme Court's decisions to let *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* and *AT&T v. Excel Commu.s, Inc*. stand is that pure software patents are here to stay whether they are directed to methods of doing business or predicting the three dimensional structure of a protein via a software application.[139] Because of the significance of this decision, the *State Street* case is briefly discussed in greater detail below.

In *State Street,* the Federal Circuit considered whether the subject matter of U.S. Patent No. 5,193,056 (the '056 patent) fell within the statutory subject matter of 35 U.S.C. § 101.[140] The Court described the '056 patent as

---

135 35 U.S.C. §§ 101 – 103, 112 (1994), respectively.

136 *Gottschalk v. Benson*, 409 U.S. 63, 71-73, 175 U.S.P.Q. 673, 676-677 (1972). Business method patents were believed unpatentable because of the judicially created "business method exception" overturned in *State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 1375, 47 U.S.P.Q.2d 1596,1602-1603 (Fed. Cir. 1998)(holding that business method patents should be "subject to the same legal requirements for patentability as applied to any other process or method"), *cert. denied*, 525 U.S. 1093 (1999).

137 *Diamond v. Diehr*, 450 U.S. 175, 187, 209 U.S.P.Q. 1, 8 (1981); *In re Iwahashi*, 888 F.2d 1370, 1375, 12 U.S.P.Q.2d 1908, 1911-1912 (Fed. Cir. 1989); *see also* Joseph Robert Brown, Jr., *Software Patent Dynamics: Software as Patentable Subject Matter After State Street Bank & Trust Co.*, 25 Okla. City U. L. Rev. 639 (2000).

138 *AT & T Corp. v. Excel Communs., Inc.*, 172 F.3d 1352, 1360-1361, 50 U.S.P.Q.2d 1447, 1454 (Fed. Cir. 1999), *cert. denied*, 528 U.S. 946 (1999); *State Street*, 149 F.3d. at 1375, 47 U.S.P.Q.2d 1596, 1602-1603.

139 Julie E. Cohen and Mark A. Lemley, *Patent Scope and Innovation in the Software Industry,* 89 Cal. L. Rev. 1 (2001).

140 149 F.3d at 1370, 47 U.S.P.Q.2d at 1598. 35 U.S.C. § 101 provides, "Whoever invents or discovers any new and useful process, machine, manufacture, or composition of

being directed to a data processing system for mutual funds.[141] In particular, the data processing system monitored and recorded the information flow and data, while making all calculations necessary for maintaining a partnership portfolio and partner fund financial services configuration.[142]

The Court began its analysis of the '056 patent by construing claim 1 of the patent to be directed to a machine rather than a process.[143] The Court could have stopped its analysis there because a machine is clearly statutory subject matter. The Court went on to hold, however, that whether the claim was directed to a machine or a process was irrelevant so long as the patent

---

matter, or any new and useful improvement thereof, may obtain a patent therefore, subject to the conditions and requirements of this title."

141 *Id*. at 1370, 47 U.S.P.Q.2d at 1598.

142 *Id*. at 1371, 47 U.S.P.Q.2d at 1598; *see also* U.S. Pat. No. 5,193,056 (issued Mar. 9, 1993). The first independent claim, "claim 1", of the patent is as follows:

> 1. A data processing system for managing a financial services configuration of a portfolio established as a partnership, each partner being one of a plurality of funds, comprising:
>
> (a) computer processor means for processing data;
>
> (b) storage means for storing data on a storage medium;
>
> (c) first means for initializing the storage medium;
>
> (d) second means for processing data regarding assets in the portfolio and each of the funds from a previous day and data regarding increases or decreases in each of the funds, assets and for allocating the percentage share that each fund holds in the portfolio;
>
> (e) third means for processing data regarding daily incremental income, expenses, and net realized gain or loss for the portfolio and for allocating such data among each fund;
>
> (f) fourth means for processing data regarding daily net unrealized gain or loss for the portfolio and for allocating such data among each fund; and
>
> (g) fifth means for processing data regarding aggregate year-end income, expenses, and capital gain or loss for the portfolio and each of the funds.

143 *State Street*, 149 F.3d at 1371, 47 U.S.P.Q.2d at 1599.

claim was directed to one of the categories of statutory subject matter.[144] Because the District Court held that the claimed invention fell within one of the judicially created exceptions to statutory subject matter, the Federal Circuit extended its analysis to address the applicability of these exceptions.[145]

The judicially created exceptions to statutory subject matter include "laws of nature, natural phenomena, and abstract ideas."[146] In *State Street*, the Appeals Court emphasized that the transformation of data in the '056 patent is a patentable application of an algorithm.[147] This holding was consistent with an earlier 1948 Supreme Court holding in the case of *Funk Bros. Seed Co. v. Kalo Inoculant Co.*.[148] In *Funk Bros*., the Supreme Court held that "patents cannot issue for the discovery of the phenomena of nature . . . . If there is to be invention from such a discovery, it must come from the application of the law of nature to a new and useful end."[149]

The patent at issue in *Funk Bros*. was directed to a novel combination of Rhizobia bacteria used to inoculate leguminous plants for fixing nitrogen.[150] Although the combination of bacteria was novel, the Court held the patent invalid because there was no transformation of a natural phenomenon into something new and useful.[151]

Subsequently, in *Diamond v. Chakrabarty* the Supreme Court held that bacteria genetically engineered to degrade hydrocarbons was patentable subject matter because the bacteria was not "nature's handiwork" but that of the inventor because the bacteria as claimed did not exist in nature.[152] It appears that whether a particular aspect of biology is patentable subject

---

144 *Id*. at 1372, 47 U.S.P.Q.2d at 1600.

145 *Id*. These exceptions are the "mathematical algorithm" and "business method" exceptions. *Id*.

146 *Id*. at 1373, 47 U.S.P.Q.2d at 1600, quoting *Diamond v. Diehr*, 450 U.S. 175, 182, 209 U.S.P.Q. 1, 20 (1981).

147 149 F.3d at 1373, 47 U.S.P.Q.2d at 1601. "[T]he transformation of data, representing discrete dollar amounts, by a machine through a series of mathematical calculations into a final share price, constitutes a practical application of a mathematical algorithm, formula, or calculation, because it produces a 'useful, concrete and tangible result'—a final share price momentarily fixed for recording and reporting purposes and even accepted and relied upon by regulatory authorities and in subsequent trades." *Id*.

148 333 U.S. 127, 130, 76 U.S.P.Q. 280, 281 (1948).

149 *Id.*

150 *Id.*

151 *Id.*

152 447 U.S. 303, 310, 206 U.S.P.Q. 193, 197 (1980).

matter rests on a combination of human intervention and the transformation of the natural phenomenon into a new, useful, concrete, and tangible result. Mixing bacteria is not enough to merit a patent, but genetically transforming bacteria clearly is enough to merit a temporary monopoly. Thus, it should not be surprising that the application of an algorithm to transform data would be found to be patentable subject matter.

2. The Demise of the Business Method Exception

The *State Street* decision by the Federal Circuit is also significant because the case ended the belief that a business method could not be patented. The Federal Circuit stated:

> We take this opportunity to lay this ill-conceived exception [the business method exception to statutory subject matter] to rest. Since its inception, the business method exception has merely represented the application of some general, but no longer applicable legal principle, . . . Since the 1952 Patent Act, business methods have been, and should have been, subject to the same legal requirements for patentability as applied to any other process or method.[153]

Thus, the Federal Circuit made it dear that business methods are patentable subject matter. The effect of the resulting business method and software patents on their respective industries is a hotly debated subject.[154] The effect such patents may have on the overall U.S. patent system, however, is beyond the scope of this article.

3. *AT&T v. Excel Communs., Inc.*

In *AT&T Co. v. Excel Communs., Inc.*[155] the Federal Circuit once again addressed the scope of 35 U.S.C. § 101 in the context of the application of an algorithm.[156] AT&T owned U.S. Patent No. 5,333,184 (the '184 patent), which was directed toward a process for the addition of a data

---

153 *State Street*, 149 F.3d at 1375, 47 U.S.P.Q.2d at 1602.

154 Cohen and Lemley, *supra* n. 139, at (2001); Chad King, *Abort, Retry, Fail: Protection for Software-Related Inventions in the Wake of State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 85 Cornell L. Rev. 1118 (2000); Leo J. Raskind, *Symposium: The State Street Bank Decision: The Bad Business of Unlimited Patent Protection for Methods of Doing Business*,10 Fordham Intell. Prop., Media & Ent. L.J. 61 (1999).

155 *AT&T v. Excel Communs., Inc.*, 172 F.3d 1352, 50 U.S.P.Q.2d 1447 (Fed. Cir. 1999), *cert. denied*, 528 U.S. 946 (1999).

156 *Id.* at 1356, 50 U.S.P.Q.2d at 1450.

field into a standard message record to indicate whether a call involves a particular primary long-distance service (interexchange) carrier.[157] The lower court invalidated the '184 patent because it found the patent's claims were not directed to statutory subject matter.[158] The Federal Circuit reversed and remanded holding that "[b]ecause the claimed process applies the Boolean principle to produce a useful, concrete, tangible result without pre-empting other uses of the mathematical principle, on its face the claimed process comfortably falls within the scope of § 101."[159]

Excel argued, unconvincingly, that the '184 patent claims fell outside the statutory subject matter because there was no physical transformation.[160] The Federal Circuit noted that the physical transformation concept can be misunderstood and proceeded to focus their inquiry on whether the application of an algorithm produced a useful, concrete, tangible result rather than a mathematical abstraction.[161] The Court acknowledged the use of the transformation language in *Diehr*,[162] and that the same language was echoed

---

157 Claim 1 of the '184 patent provides:

> 1. A method for use in a telecommunications system in which interexchange calls initiated by each subscriber are automatically routed over the facilities of a particular one of a plurality of interexchange carriers associated with that subscriber, said method comprising the steps of:
>
> generating a message record for an interexchange call between an originating subscriber and a terminating subscriber, and
>
> including, in said message record, a primary interexchange carrier ("PIC") indicator having a value which is a function of whether or not the interexchange carrier associated with said terminating subscriber is a predetermined one of said interexchange carriers.

Id. at 1354, 50 U.S.P.Q.2d at 1449.

158 *Id.* at 1353, 50 U.S.P.Q.2d at 1448.

159 *Id.* at 1358, 50 U.S.P.Q.2d at 1452.

160 *Id.*

161 *See id.* at 1358-1360, 50 U.S.P.Q.2d at 1452-1453.

162 *Id.* at 1359, 50 U.S.P.Q.2d at 1452. *See also* Cathy E.Cretsinger, I. Intellectual Property; B. Patent; 4. Patentability; a) Computer Software: *AT&T Corp. v. Excel Communications, Inc.*, Berkeley Tech. L.J. 165 (2000).

in *In re Schrader*.[163] The Court failed to acknowledge that data transformation was important in the *State Street* holding and reasoned that data transformation was "confirmation" of patentable subject matter, but not a requirement for patentable subject matter in the context of an algorithm.[164] Thus, the Court limited the judicially created exceptions of statutory subject matter to mere laws of nature, natural phenomena, and abstract ideas, while confirming the use of the concrete, tangible result test.[165]

4. Application of *State Street* and *AT&T* to the Field of Bioinformatics

In light of *State Street* and *AT&T*, it is clear that computer or software-related inventions are patentable subject matter—either under the process or the machine category, provided the application of an algorithm performs a useful, concrete, and tangible result. *AT&T* also confirms that the transformation of data is sufficient, but not required, for claims applying algorithms to fall within statutory subject matter. Therefore, as a practical matter, whether drafting a software patent application addressed to bioinformatics or Internet banking, a prudent patent applicant would include at least one independent claim capturing a feature for the transformation of data. The inclusion of such a feature in a claim should ward off challenges based on lack of statutory subject matter. Other claims without the transformation feature should also be included. The truly conservative applicant would go one step further and include at least one independent claim directed to a system or machine, as well as a process. Including a computer or processor in such a claim should be sufficient to direct the claim to a machine, rather than a process or method. Bioinformatics-related patent applications, therefore, should include claims directed to a process that transforms data to produce a useful, concrete, tangible result, and a system to increase the likelihood of obtaining claims directed to statutory subject matter.

---

163 22 F.3d 290, 30 U.S.P.Q.2d 1455 (Fed Cir. 1994). In *Schrader*, the Court held that the "claims, except for incidental changes to a 'record,' do not reflect any transformation or conversion of subject matter representative of or constituting *physical activity or objects*." *Id.* at 294, 30 U.S.P.Q.2d at 1458 (emphasis in original).

164 *AT&T*, 172 F.3d at 1359, 50 U.S.P.Q.2d at 1452. The Court referenced *Arrhythmia Research Tech., Inc. v. Corazonix Corp*., 958 F.2d 1053, 22 U.S.P.Q.2d 1033 (Fed. Cir. 1992) for the proposition that data transformation is sufficient, but not required for statutory subject matter. *Id.*

165 172 F.3d at 1360, 50 U.S.P.Q.2d at 1452.

Conversely, claims directed to the "information" contained in a naturally occurring protein would fall outside statutory subject matter because such information already exists in nature, and is essentially nonfunctional.[166] This sequence information would be abstract and more akin to a natural law. For example, a naturally existing form of the protein actin can be easily isolated and sequenced. That protein's sequence is essentially information obtained from nature, which has not been transformed or used to generate a concrete, tangible result. Thus, information in the sequence itself would not likely fall within statutory subject matter. Because genomic DNA is interspersed with non-coding regions of DNA, this argument is not as strong for amino acid sequences of proteins. Although amino acid sequences of proteins may be truncated or otherwise modified during post-translation modifications, they are not shuffled as with RNA splicing.

The sequence information should not be confused with a physically-isolated oligonucleotide or protein having a particular sequence. Patents on isolated compositions of matter can, and often are, used to prevent others from using the compositions unless they pay a license fee or the like. Likewise, methods of sequencing DNA are patentable. Thus, though an applicant may obtain a patent on DNA compositions of matter or methods of sequencing DNA, an applicant likely would not be successful in obtaining a patent on the actual sequence information of amino acids to prevent others from using that information.

But what about a scenario involving engineered proteins and/or organisms? Recombinant proteins that ordinarily do not exist in nature contain combinations of naturally occurring sequence information if different functional regions of unlike proteins are combined. Although the composition of matter is clearly patentable subject matter, the sequence information itself is probably not. Here, as in *Funk*, naturally occurring information is combined to perform its natural functions. For example, 'subunit a' is combined with 'subunit b' to produce a protein that has the functional characteristics of both proteins. The sequence information was not modified; rather, a scientist merely used the lexicon of DNA and proteins to produce a product having a desired result. At the end of the day, the information alone, genetic or proteomic, is not a composition of matter, an article of manufacture, a machine, or a process.[167]

The United States Patent and Trademark Office ("USPTO") has announced the following regarding raw genetic information:

---

166 66 Fed. Reg. 1092, 1093 (2001).

167 The USPTO has indicated that patents do not confer ownership of genetic information. 66 Fed. Reg. at 1093. However, a novel and useful composition of matter or a novel and useful method of producing the composition of matter can be patented. *Id.*

> The genetic sequence data represented by strings of the letters A, T, C, and G alone is raw, fundamental sequence data, i.e., nonfunctional descriptive information. While descriptive sequence information alone is not patentable subject matter, a new and useful, purified and isolated DNA compound described by the sequence is eligible for patenting, subject to satisfying the other criteria for patentability.[168]

If the patent system does not protect information, should the information be protected by copyright or potential database protection? Although this question is outside the scope of this article, it is an important question that should be evaluated.

If a new protein or gene is discovered and the sequence information is obtained as the target molecule is being characterized, then is there a way to regulate the use of the sequence information using the patent system? After claiming the composition of matter having a specific sequence information, it might be useful to also claim methods of using the sequence information. The patent specification should fully describe embodiments detailing how such information can be used. For instance, when claiming the composition of matter of a novel isolated protein, it is now routine practice to describe an antibody to that protein in the same application. This aids in addressing obviousness rejections on claims to the antibody, should the protein data be disclosed to the public.

Similarly, embodiments and claims for using biological information will also become routine in biotechnology applications. Even if the claims to the methods of using the information never make it to patent, the published application detailing suggested uses may serve as a bar preventing others from patenting those uses.

5. Utility

A patentable invention must be useful or have utility.[169] The utility requirement stems from the United States Constitution, which provides that the patent system is intended to promote the useful arts.[170] Additionally, 35 U.S.C. § 101 describes patentable subject matter as new and useful. In the

---

168 *Id*.

169 35 U.S.C. § 101 (1994). For a discussion of utility as it relates to biotechnology inventions *see* Phanesh Koneru, *To Promote the Progress of Useful Articles?: An Analysis of the Current Utility Standards of Pharmaceutical Products and Biotechnological Research Tools*, 38 IDEA 625 (1998).

170 U.S. Const. art. I, § 8, cl. 8.

seminal case of *Brenner v. Manson*,[171] the Supreme Court held that specific and substantial utility is a requirement for patentability.[172]

In *Brenner*, the patent application at issue was directed to a new process for making a known steroid compound.[173] The applicant, Manson, sought to bring an interference with an earlier granted patent on the same process.[174] Manson stated in an affidavit that the compound produced by his claimed process was known in the art and that the compound had obvious utility, at least to him.[175] The USPTO, however, declined to declare the interference because no utility was specified.[176] The Court of Customs and Patent Appeals ("CCPA") reversed, holding that a process producing a desired product has utility.[177]

On appeal to the Supreme Court,[178] Manson argued that homologues of the steroid compound were known in the art to have tumor inhibiting effects in mice.[179] The Court rejected this argument on the premise that minor changes in chemical structure can have significant effects.[180] Manson also argued that the claimed process had utility if it actually produced the desired compound or, alternatively, if the product of the claimed process belonged to a class of compounds subject to serious scientific research.[181] The Court rejected these arguments as well, holding that the process was only patentable if the compound it produced had substantial utility or a specific benefit.[182] If the only utility of a compound is for scientific research, then there is no substantial utility.[183] Recently, the USPTO has promulgated Utility Examination Guidelines ("Guidelines").[184] In the Guidelines, the

---

171 *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. 689 (1966).

172 *Id*. at 534-535, 148 U.S.P.Q. at 695-696.

173 *Id.* at 521, 148 U.S.P.Q. at 690.

174 *Id.*

175 *Id.*

176 *Brenner*, 383 U.S. at 521-522, 148 U.S.P.Q. at 690.

177 *Id.*

178 The Supreme Court granted certiorari in order to "resolve [the] running dispute over what constitutes 'utility' in chemical process claims. *Id*. at 522, 148 U.S.P.Q. at 691.

179 *Id.* at 531, 148 U.S.P.Q. at 694.

180 *Id.* at 532, 148 U.S.P.Q. at 694.

181 *Id.*

182 *Id*. at 533-535, 148 U.S.P.Q. at 695-696.

183 *Id.* at 535, 148 U.S.P.Q. at 695-696.

184 66 Fed. Reg. 1092. For a general discussion on the Guidelines and DNA technology *see*

USPTO enunciated three requirements for utility: specific, substantial, and credible.[185] Evidence of practical utility is considered equivalent to specific and substantial utility.[186] Thus, all three requirements for utility will be met when the application contains a practical utility that appears credible to a person of ordinary skill in the art.[187]

B. 35 USC § 102: Novelty

1. Section 102(a)

In addition to the statutory subject matter and utility requirements, a patentable invention must be novel.[188] To be novel, the invention must not be the subject of a prior art reference. Prior art is defined in 35 U.S.C. § 102.[189] Section 102 is broken into subsections (a)-(g), and brief descriptions of each follow. Section 102(a) provides:

> A person shall be entitled to a patent unless—
>
> the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for patent [ ].[190]

This section establishes that a patent cannot issue on what was already publicly known in the United States of America.[191] Prior knowledge or use by others must be accessible to the public before it can invalidate a patent under § 102(a).[192] Prior art typically refers to printed publications and patents that predate the effective filing date of an issued patent. "If one prior art reference completely embodies the same process or product as any claim

---

Timothy A. Worrall, *The 2001 PTO Utility Examination Guidelines and DNA Patents*, 16 Berkeley Tech. L.J. 123 (2001).

185 66 Fed. Reg. at 1098.

186 *Id.*

187 *Id.*

188 35 U.S.C. § 102 (1994).

189 *Id.*

190 *Id.* at § 102(a) (1994).

191 *Woodland Trust v. Flowertree Nursery, Inc.*, 148 F.3d 1368, 1370, 47 U.S.P.Q.2d 1363 (Fed. Cir. 1998).

192 *Id.* A printed publication or patent can be from outside the U.S..

of the patent in suit, the process or product recited by that claim is said to be 'anticipated' by the prior art and the claim is therefore invalid under § 102 for want of novelty."[193] If the reference discloses each and every element of the claimed invention, there is an identity between the two.[194]

It is important to note that all elements need not be explicitly present in the prior art. The judicially created doctrine of inherency applies when an element or aspect of the invention is deemed to be part of the reference because that aspect is necessarily associated with the thing described in the reference.[195] Inherency is typically used in situations in which common knowledge is not expressly recited in the prior art reference.[196] It should be noted, however, that the possibility that a specific result or characteristic may occur, or be present, in the prior art is not sufficient to establish the inherency of that result or characteristic.[197]

The novelty requirements for inventions relating to bioinformatics have more significance now that the issue of subject matter has been resolved. Because bioinformatics is still in its infancy, there are few instances of anticipatory prior art in patents and printed publications – but they are growing. One difficulty regarding such prior art for bioinformatics, however, is that bioinformatics inventions involve more than one technology. Thus, prior art may include both computer related references as well as those involving the life sciences.

2. Section 102(b)

Section 102(b) provides that "the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of the application for patent in the United States [ ]."[198]

This section is also referred to as the statutory bar section. Publications, patents, public uses, and sales that occur more than one year prior to the date of applying for a patent in the United States will absolutely bar the issuance of a patent for an invention disclosed in the reference or

---

193 *Shatterproof Glass Corp. v. Libbey-Owens Ford Co.*, 758 F.2d 613, 619, 225 U.S.P.Q.2d 634, 637 (Fed. Cir. 1985).

194 *Id*.

195 *See Continental Can Co. v. Monsanto Co*., 948 F.2d 1264, 1269, 20 U.S.P.Q.2d 1746, 1749 (Fed. Cir. 1991).

196 *Id.*

197 *In re Rijckaert*, 9 F.3d 1531, 1534, 28 U.S.P.Q.2d 1955, 1957 (Fed. Cir. 1993).

198 35 U.S.C. § 102(b) (1994).

subject to a public use, sale, or offer for sale. Unlike § 102(a), an inventor's own work can bar the issuance of a patent if the work is published more than one year before the filing date of the patent application.[199] Furthermore, an applicant for a patent cannot submit an affidavit to establish an earlier conception date to rebut a statutory bar rejection.[200]

The public use or on-sale aspect of § 102(b) has been explained by the courts. For example, public use can be established even if only one person uses the invention.[201] Courts have interpreted the public use and on-sale bars as means to prevent the inventor from commercially exploiting the exclusivity of his invention substantially beyond the statutorily authorized period.[202]

The Supreme Court has also recently addressed the issue of the on-sale bar in the case of *Pfaff v. Wells Electronics Incorporated*.[203] In *Pfaff*, the Supreme Court articulated a two-part test for determining whether an invention was "on-sale" within the meaning of 35 U.S.C. 102(b).[204] The two-part test is met, and the on-sale bar triggered, when (1) the product is the subject of a commercial offer for sale, and (2) the invention is ready for patenting.[205] The plaintiff, Pfaff, accepted a purchase order for the invention more than one year before the filing date of his patent application.[206] The invention, therefore, was subject to a commercial offer for sale.[207] Pfaff also fully disclosed his invention more than one year prior to filing by sending drawings to the manufacturer. This act, consequently, satisfied the test's second condition.[208] The offer for sale must meet the level of an offer for sale in the contract sense.[209]

---

199 Manual of Patent Examining Procedure § 2133.02 at 2100-06 (2000) (citing *De Graffenried v. U.S.*, 16 U.S.P.Q.2d 1321, 1330 n.7 (Cl. Ct. 1990)).

200 37 C.F.R. § 1.131 (2001).

201 *Egbert v. Lippmann*, 104 U.S. 333, 336 (1881).

202 *RCA Corp. v. Data Gen. Corp.*, 887 F.2d 1056, 1062, 12 U.S.P.Q.2d 1449, 1454 (Fed. Cir. 1989).

203 *Pfaff v. Wells Elect., Inc.*, 525 U.S. 55, 48 U.S.P.Q.2d 1641 (1998).

204 *Id*. at 67-68, 48 U.S.P.Q.2d at 1646-1647.

205 *Id.*

206 *Id.*

207 *Id.*

208 *Id.*

209 *Group One, Ltd. v. Hallmark Cards, Inc.*, 254 F.3d 1041, 1047, 59 U.S.P.Q.2d 1121, 1126 (Fed. Cir. 2001).

Just as with any other invention, one must be mindful of any § 102 statutory bars. These may be particularly relevant in the area of bioinformatics, where the biological information is being provided on databases and/or other tools over the Internet. Of equal importance for bioinformatics is the "on-sale" test provided in *Pfaff*. As the bioinformatics industry continues to grow, the demand for new and innovative bioinformatics-tools will undoubtedly strain the resources of companies attempting to meet the high demand. Therefore, it is critical for bioinformatics companies, inventors, and patent practitioners to understand and recognize possible events triggering the "on-sale" bar in order to timely file patent applications.

3. Section 102(c)

Section 102(c) provides, in pertinent part, that a person may obtain a patent unless "he has abandoned the invention [ ]."[210] The abandonment must be intentional, but intent can be implied.[211] Section 102(c) is self- explanatory; the intentional abandonment of any invention, including one involving bioinformatics, will bar the subsequent issuance of a patent for that invention.

4. Section 102(d)

Section 102(d) provides that a person shall be entitled to a patent unless:

> the invention was first patented or caused to be patented, or was the subject of an inventor's certificate, by the applicant or his legal representatives or assigns in a foreign country prior to the date of the application for patent in this country on an application for patent or inventor's certificate filed more than twelve months before the filing of the application in the United States [ ].[212]

A section 102(d) rejection is proper if the applicant applied for a patent in a country other than the United States more than one year before filing the U.S. application. Importantly, the foreign patent must ultimately be granted to trigger 102(d). Section 102(d) is designed to encourage prompt filing of United States patent applications. Thus, it is best to file U.S. applications as soon as possible after foreign filing, if not concurrently, to avoid a 102(d) rejection.

---

210 35 U.S.C. § 102(c) (1994).

211 *See In re Gibbs*, 437 F.2d 486, 489, 168 U.S.P.Q.2d 578, 581 (CCPA 1971).

212 35 U.S.C. § 102(d) (1994).

5. Section 102(e)

Section 102(e) provides that a person is entitled to a patent unless:

> The invention was described in--
>
> (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effect under this subsection of a national application published under section 122(b) only if the international application designating the United States was published under Article 21(2)(a) of such treaty in the English language; or
>
> (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that a patent shall not be deemed filed in the United States for the purposes of this subsection based on the filing of an international application filed under the treaty defined in section 351(a).[213]

This new amended section 102(e) applies to applications filed on or after November 29, 2000, or any application that has been voluntarily published.[214] Formerly restricted to U.S. Patents, 102(e) rejections can now be made using a published U.S. patent application.[215] Thus, if (1) a U.S. patent issues after the filing date of the application in question and has an earlier effective filing date or (2) a prior published U.S. patent application can both serve as basis for a 102(e) rejection and anticipate the application in question, provided every element of the invention (as defined by the claim in the application in question) is disclosed, then a section 102(e) rejection is proper.

Although section 102(e) is relevant for all inventions, it is specifically mentioned here because bioinformatics is an emerging field. Thus, the bulk of bioinformatics patents will be filed under the amended 102(e) requirements.

6. Sections 102(f) and 102(g)

These sections provide:

> (f) he did not himself invent the subject matter sought to be patented, or

---

213 35 U.S.C. § 102(e) (1994).

214 Pub. L. No. 106-113, § 1000(a)(9), 113 Stat. 1501A, 1565 (1999).

215 *Id.*

> (g) (1) during the course of an interference conducted under section 135 or section 291, another inventor involved therein establishes, to the extent permitted in section 104, that before such person's invention thereof the invention was made by such other inventor and not abandoned, suppressed, or concealed, or (2) before such person's invention thereof, the invention was made in this country by another inventor who had not abandoned, suppressed, or concealed it. In determining priority of invention under this subsection, there shall be considered not only the respective dates of conception and reduction to practice of the invention, but also the reasonable diligence of one who was first to conceive and last to reduce to practice, from a time prior to conception by the other.[216]

Section 102(f) is generally used when an applicant has derived the invention from another,[217] while section 102(g) is the basis for interference proceedings.[218] Interference proceedings determine who was first to invent by establishing right of priority.

Sections 102(f) and 102(g) will likely play a significant role as the bioinformatics industry expands and evolves. Because patent protection is often necessary for commercial viability, fierce competition between emerging players in the bioinformatics field is sure to intensify. With considerable sums of money already spent on research and development and even greater sums on the line, developers of bioinformatic-tools will compete for patent protection and licensing contracts. The demand for talented bioinformaticians will also cause the movement of scientists from one company to another. Thus, problems involving who invented what, where, and when are all highly probable.

A larger problem for patents relating to bioinformatics may be posed by the non-obvious requirement. Because computer science has grown so rapidly, many publications and patents already exist describing the specific software used in areas other than bioinformatics. The question then arises: given the computation demands in the life sciences, would it have been obvious to apply existing (or similar) computer algorithms already used in other computational areas to the life sciences?

C. 35 USC § 103: Obviousness

A new and useful product or process is not patentable if the differences between the subject matter sought to be patented and the prior art are such that the subject matter, as a whole, would have been obvious at the time the invention was made to a person having ordinary skill in the art.[219]

---

216 35 U.S.C. § 102(f)-(g) (1994).

217 *Ex parte Kusko,* 215 U.S.P.Q. 972, 974 (Bd. App. 1981)

218 *See* Manual of Patent Examining Procedure § 2138 at 2100-81 – 2100-89 (2000).

219 35 U.S.C. § 103.

The non-obviousness requirement serves to limit the grant of a patent monopoly to only those inventions that are beyond the grasp of those of ordinary skill in the art.[220] The Supreme Court interpreted and applied section 103 in *Graham v. John Deere Company*.[221]

In *Graham*, the Supreme Court enunciated what have come to be known as the *Graham* Factual Inquiries for assessing obviousness. Parsing out the *Graham* opinion presents four factors that are to be considered: (1) determination of scope and content of prior art, (2) ascertaining the differences between the prior art and the claims, (3) resolving the level of ordinary skill in the pertinent art, and (4) considering certain secondary considerations such as long felt but unresolved need, commercial success, and failure of others.[222]

As a preliminary inquiry, "it must be known whether a patent or publication is in the prior art under 35 U.S.C. § 102."[223] Prior printed publications and patents related to the invention fall within the scope of the prior art. Thus, only prior art that relates to the invention or is analogous to it may serve as the basis for a section 103 rejection. Therefore, the scope of the prior art includes that which is the same as the inventor's art, and those arts that logically relate to the inventor's concern.[224]

To determine the differences between the prior art and the invention as claimed, the invention must be considered as a whole - not just the differences.[225] Inherent properties or aspects of the invention are also considered.[226] Finally, portions of a reference cannot be selectively used as the basis of a section 103 rejection; rather, as with the determination of the invention, the reference as a whole must be used even if parts of the reference teach away from the claimed invention.[227]

The person of ordinary skill in the art is a fiction or ghost, "not unlike the reasonable man and other ghosts in the law."[228] Courts have enunciated

---

220 Donald S. Chisum, *Chisum on Patents* vol. 2, § 5.01, 5-11 (Matthew Bender 2001).

221 383 U.S. 1, 148 U.S.P.Q. 459 (1966).

222 *Id*. at 17-18, 148 U.S.P.Q. at 467.

223 *Panduit Co. v. Dennison Mfg. Co.*, 810 F.2d 1561**,** 1568, 1 U.S.P.Q.2d 1593, 1597 (Fed. Cir. 1987), *cert. denied*, 481 U.S. 1052 (1987).

224 *Shatterproof Glass Co.*, 758 F.2d at 620, 225 U.S.P.Q. at 637-638.

225 *Stratoflex, Inc. v. Aeroquip Corp.*, 713 F.2d 1530, 1537, 218 U.S.P.Q. 871, 877 (Fed. Cir. 1983).

226 *In re Antonie*, 559 F.2d 618, 620, 195 U.S.P.Q. 6 (CCPA 1977).

227 *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 1551, 220 U.S.P.Q. 303, 311 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984).

228 *Panduit Co.*, 810 F.2d at 1566, 1 U.S.P.Q.2d at 1595.

six factors to help determine the level of the person of ordinary skill in the art.[229] These factors include: (1) the educational level of the inventor; (2) type of problems encountered in the art; (3) prior art solutions to those problems; (4) rapidity with which innovations are made; (5) sophistication of the technology; and (6) the educational level of active workers in the field.[230]

Obviousness inquires may also arise during the prosecution of a patent, which may serve as a basis to reject a claim. In establishing a *prima facie* case for obviousness, three elements must be met: (1) there must be some motivation or suggestion, in the references cited or in the general knowledge of ordinary skill in the art, to modify the reference or to combine the references to produce the claimed invention; (2) there must be a reasonable expectation of success that the modification or combination of the references will make the invention as claimed; and (3) the reference or combination of references must teach or suggest all the claim limitations.[231] Once a *prima facie* case of obviousness has been made, the applicant can rebut the rejection with arguments and evidence of the secondary considerations referenced above.

As mentioned previously, obviousness inquires are likely to take on significant importance as references relating to computational methods in other fields are applied to inventions in bioinformatics. In particular, references suggesting the use of a particular algorithm with imaging, predictive modeling, and the like, in combination with biological data, will be extremely problematic. Such references may be considered prior art if they relate to or are analogous to the bioinformatics invention in question. Because the use of biological information with sophisticated hardware and software will generate innovations in bioinformatics with unique applications, another worry is that initial broad patents will cripple the industry.

D. 35 USC § 112 ¶ 1: Written Description, Enablement, and Best Mode

The last hurdle of patentability relates to the written specification.[232] The first paragraph of section 112 has three separate requirements for

---

229 *Envtl.l Designs, Ltd. v. Union Oil Co. of Cal.*, 713 F.2d 693, 696, 218 U.S.P.Q. 865, 868 (Fed. Cir. 1983), *cert. denied*, 464 U.S. 1043 (1984).

230 *Id*.

231 *In re Vaeck*, 947 F.2d 488, 493, 20 U.S.P.Q.2d 1438 (Fed. Cir. 1991).

232 For a general discussion of the recent PTO guidelines applied to biotechnology applications *see* Lisa A. Karczewski, Comments: *Biotechnological Gene Patent Applications: The Implications of the USPTO Written Description Requirement Guidelines*

patentability: (1) a written description of the invention; (2) a description of how to make and practice the claimed invention; and (3) the best mode contemplated by the inventor for practicing the invention.[233]

The first requirement, also known as the written description requirement, requires the applicant to clearly convey to those skilled in the art that he is in possession of the invention as claimed.[234] The written description requirement limits the applicant to claims directed to subject matter originally contained in the application, when filed.

The second requirement is known as the enablement requirement. To satisfy this requirement, the specification must describe the invention as claimed in a manner that enables a person skilled in the art to make and use the invention.[235] For bioinformatics inventions, which are essentially software or process patents, sufficient detail must be described to enable a programmer to write an algorithm and practice the claimed invention after reading the specification. The actual source code, however, need not, and probably should not, be included in the specification. Generally, logic flow diagrams illustrating the steps of the process are sufficient to enable the disclosure.

The rationale of the enablement requirement is to ensure that the public receives all necessary information to practice the invention in exchange for the limited exclusionary rights granted by the patent.[236] The test for enablement is whether any undue experimentation is necessary to practice the invention as claimed.[237] Factors to consider when assessing enablement include: "(1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims."[238] These factors have been referred to as the Wand's factors.

The last requirement concerns the best mode. Section 112 requires the applicant to disclose the best mode of practicing the invention known to

---

on the Biotechnology Industry, 31 McGeorge L. Rev. 1043 (2000).

233 *See* 35 U.S.C. § 112, ¶ 1.

234 *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563-64, 19 U.S.P.Q.2d 1111, 1117 (Fed. Cir. 1991).

235 *See* 35 U.S.C. § 112, ¶ 1.

236 Manual of Patent Examining Procedure § 2164 at 2100-129 (2000).

237 *Minerals Separation, Ltd. v. Hyde*, 242 U.S. 261, 270 (1916); *see also In re Wands*, 858 F.2d 731, 737, 8 U.S.P.Q.2d 1400, 1404 (Fed. Cir. 1988).

238 *In re Wands*, 858 F.2d 731 at 737, 8 U.S.P.Q.2d at 1404.

the inventor, at the time the application was filed.[239] The applicant is required to explicitly identify the best mode known to him, however, if after filing the inventor discovers a better mode, then he is not required to update the patent application.[240] Indeed, the examiner will presume that the best mode is contained in the application, unless there is evidence to the contrary.[241] Generally, the best mode requirement is to ensure that inventors are completely disclosing their invention to the public in exchange for the grant of a patent, and not withholding critical information from the public.

For bioinformatics inventions involving software, adequately stating the functions of the applicable software in the patent specification should satisfy the best mode. The courts have generally held that where the best mode involves software, it is well within the skill of the art to write code for software as long as it has been functionally disclosed in the specification.[242]

III. PROBLEMS, PREDICTIONS, AND TRENDS

A. Ordinary Skill in What Art?

As discussed earlier, the person of ordinary skill in the art is a legal fiction, and courts have enunciated factors for helping determine the level of skill for this hypothetical individual. Nevertheless, the requirements of sections 103 and 112 must both be met with this person in mind. The field of bioinformatics, however, presents some unique challenges to determining exactly who this hypothetical person is. This difficulty can be attributed to two main factors. The first is that bioinformatics is a multidisciplinary field, while the second is that the field, itself, is fairly new, yet developing rapidly.

Many scientists with classical training in the life sciences lack the quantitative and computational skills required for understanding the intricacies of bioinformatic-tools and their related inventions.[243] The opposite is true of computer scientists. Most computer scientists and statisticians simply don't

---

239 *See Transco Prod. Inc. v. Performance Contracting, Inc.*, 38 F.3d 551, 32 U.S.P.Q.2d 1077 (Fed. Cir. 1994).

240 *Spectra-Physics, Inc. v. Coherent, Inc.*, 827 F.2d 1524, 1535, 3 U.S.P.Q.2d 1737, 1745 (Fed. Cir. 1987).

241 Manual of Patent Examining Procedure § 2165.03 at 2100-146 (2000).

242 *Fonar Corp. v. General Electric Co.*, 107 F.3d 1543, 1549, 41 U.S.P.Q.2d 1801, 1805 (Fed. Cir. 1997).

243 Brad Stone, *Wanted: Hot Industry Seeks Supergeeks; To Build Better Drugs, the Exploding Field of Bioinformatics is Looking for Highly Trained Workers Comfortable with Supercomputing and Biology*, Newsweek 54, 55 (Apr. 30, 2001).

have enough training in biology to identify and truly understand the important problems facing the life sciences.[244] Yet, due to the complex interrelationships between biological data, it is critical for bioinformaticians to understand both the biological problems and possible computational solutions.[245]

Unfortunately for the field of bioinformatics, scientists with such a collection of skills are hard to find.[246] Compounding this problem is the growing need for scientists specializing in bioinformatics because of the rapid growth of the field.[247] One study has already estimated that the bioinformatics field will need 20,000 of these new specialists by 2005.[248] Many colleges and universities are responding to this need by creating bioinformatics centers and/or offering advanced degrees in bioinformatics.[249] There is a real worry, however, that many of the academics qualified to teach the required interdisciplinary courses will leave teaching for the very industries driving the need in the first place.[250]

This shortage of "those of ordinary skill in the art" of bioinformatics may also translate into a shortage of patent practitioners possessing the requisite skills for understanding the new technologies emerging in the field. Arguably, the most important function of any patent attorney is being able to understand the invention whether preparing a patent application, conducting due diligence, advising clients, or litigating an infringement suit. Indeed, this is the reason the USPTO requires attorneys practicing before it to have a scientific or engineering background. Patent practitioners, therefore, will need to adapt in order to meet the evolving needs of their clients. Keeping current no longer involves just reading the latest journals and trades; keeping current must include a willingness to learn and understand disciplines beyond one's initial training or background. With the growth and increasing demands

---

244 *Id.*

245 Cynthia Gibas & Per Jambeck, *Developing Bioinformatics Computer Skills*, 3 (O'Reilly & Assoc. 2001).

246 Stone, *supra* n. 243, at 54.

247 *Id.* at 54-55.

248 *Id.* at 55.

249 *Id.* For instance, the University of California, Davis, recently developed a new $95 million dollar bioinformatics program. *Id.* at 54. Other schools, such as UCLA, UC Berkeley, and Cornell already have bioinformatics centers or are planning to start them. *Id.* at 55.

250 *Drowning in Data: Like so Many Others, Biologists are Confronted by a Tidal Wave of Information. Unfortunately, Few of Them Know How to Swim*, The Economist (June 26, 1999).

of bioinformatics, patent practitioners need to take notice because bioinformatics is here to stay.

B. "In Silico" or Bust

It must be stressed that bioinformatics is not a mere researching fad. Developments in the field are being driven by decades of genomics research culminating in enormous amounts of information. Related industries, in turn, are demanding new and advanced tools to make sense of all the information. As a result, ever increasing amounts of scientific research are being carried out through the use of computers, an act that has come to be known as "in silico" biology.[251] Thus, those scientists already possessing the requisite computational ability are at a significant advantage. They are being offered their pick of jobs by the various industries demanding their talents.[252] Conversely, other scientists are finding their grant proposals turned down for lacking a bioinformatics component.[253] This has led some to predict that traditional biologists will be forced to adapt.[254] As one researcher put it, "[a]s [they] see their colleagues outpace them because of technology, they will change."[255]

Undoubtedly, these scientists will rise to the occasion, but such change will take time. After all, molecular biology has already weathered several technology revolutions over the past two decades.[256] The revolutionary techniques of DNA sequencing, cloning, and PCR are prime examples.[257] The majority of change, however, must take place within the ranks of academia. Although many universities are developing bioinformatics programs, there is much to be done. Some believe that a multidisciplinary approach, cutting across departments and even major research universities, is what is truly needed to create the environment required for bioinformatics to

---

[251] *See* David S. Roos, *Bioinformatics – Trying to Swim in a Sea of Data*, 291, Sci. 1260, 1260 (2001); *see also*, Ken Howard, *The Bioinformatics Gold Rush*, Scientific Am. 58, 61 (July 2000).

[252] Stone, *supra* n. 243, at 54.

[253] Declan Butler, *Are You Ready for the Revolution?*, 409 Nat. 758, 758 (2001).

[254] *Id.* at 759.

[255] *Id.*

[256] *Id.*

[257] *Id.* A brief discussion of each of these techniques has already been provided herein. *Supra*, part I C.

fully flourish.[258]

C. Competition, Collaboration, and Innovation

Bioinformatics is already growing at an estimated rate of twenty-five to fifty percent a year.[259] This growth is mainly due to the increased need for data mining tools and visualization software, which are currently the fastest growing segments of the market.[260] Companies most involved in bioinformatics currently include all of the big name computer companies and another fifty to one hundred stand alone companies, some of which are backed by venture capitalists.[261] These companies are doing everything from building better bioinformatic-tools, facilitating gene discovery, conducting gene analysis, and specializing in data content, to outsourcing the needs of the pharmaceutical industry.[262]

Many companies are already aggressively seeking patent protection for their bioinformatics inventions, with some attempting to corner the market. The computer giant Compaq played a key role in helping to speed up human genome mapping.[263] As a result of this seed investment, Compaq now expects to build a life-sciences portfolio of $100 million in just a few years.[264] IBM is following suit by investing a comparable sum in "Big Gene," a bioinformatics system dedicated to modeling the folding patterns of human proteins.[265] Although Big Gene is expected to take five years to build, it will be 5000 times faster than the typical PC by implementing innovative water-cooled silicon chips that contain fewer instructions.[266] Of even greater usefulness, Big Gene is expected to self-diagnose its own hardware and software for glitches, breakdowns, and other malfunctions, and take appropriate action.[267]

---

258 Butler, *supra* n. 253, at 758.

259 John Thackray, *Bioinformatics Grows Legs: The Exploding Need for Tools to Analyze and Mine Life Science Data Promises a Bonanza for some IT Firms*, Electronic Business 76, 78 (July 2001).

260 *Id.*

261 *Id.* at 79.

262 *Id.*

263 *See supra* part I E.

264 Thackray, *supra* n. 259, at 79.

265 *Id.*

266 *Id.*

267 *Id.*

As competition among various bioinformatics companies increases, a future trend in the field is expected to be collaborations, alliances, and acquisitions.[268] Indeed, the need to integrate technologies has already resulted in several partnerships among many of bioinformatics most notable players.[269] A prime example is the recent settlement of patent infringement lawsuits between Affymetrix and Hyseq, dating back to 1997.[270] The patents involved were for various DNA analysis techniques.[271] Rather than waste money on costly litigation, the two companies decided to collaborate on the formation of a new subsidiary.[272] The new company will focus on developing a new high-speed DNA sequencing chip that could be capable of sequencing up to ten thousand bases at a time.[273]

Entire governments are also making bioinformatics investments. One of the most notable examples was the funding of the Human Genome Project by the NIH and the DOE.[274] With most of the mapping complete, the DOE has now set its sights on a program titled "Genomes to Life," which is slated to begin in 2002.[275] This program will be built on a computational infrastructure combining hardware and software development, which will be dedicated to exploring regulatory networks, determining how protein complexes operate, and assessing the function of microbial communities.[276] Other countries, such as Japan, China, Germany and the United Kingdom, are also making major investments in bioinformatics.[277]

One of the most intriguing possibilities for bioinformatics is the prediction that computers will utilize DNA as a storage medium some day.[278]

---

268 Strategic Directions International Inc., *Bioinformatics: The Rosetta Stone of Life Science Research*, Instrument Business Outlook no. 21, vol. 9, 1, 3 (Feb. 15, 2001).

269 *Id.* These partnerships are between Viaken Systems and Hewlett Packard, Incyte Pharmaceuticals and SGI, and Gene Logic and Affymetrix to name just a few. *Id.*

270 Karen Young, *Hyseq Forms New Subsidiary, Settles Affymetrix Litigation*, 12 Bioworld (Oct. 26, 2001).

271 *Id.* At issue were Hyseq's sequencing-by-hybridization technology and Affymetrix's GeneChip technology. *Id.*

272 *Id.*

273 *Id.*

274 *See* discussion *supra* part I C and related footnotes.

275 Elizabeth Pennisi, *So Many Choices, So Little Money*, 294 Sci. 82, 85 (2001).

276 *Id.*

277 Strategic Directions International Inc., *supra* n. 5, at no. 21, vol. 9, p.3.

278 Fiona Harvey, *The Processors at the Heart of Healthcare*, London Financial Times, Inside Track section at p. 20 (Nov. 29, 2000).

As impossible as the idea may seem, scientists at the University of Southern California have already used a test-tube of DNA to solve a mathematical problem.[279] In addition, other scientists are working on designing bacterial cells with logic gates.[280] DNA, after all, is an incredibly efficient storage molecule capable of holding all the information required for building an entire organism. Yet, the method of storage is incredibly similar to the digital one used by computers.[281] The only real differences being a four-base code instead of the two-base binary system, and groupings of code as codons rather than eight-bit bytes.[282] This has led some to believe that just as computer technology has enabled huge strides in biological research, DNA computers will some day return the favor.[283]

IV. CONCLUSION

The once fledgling field of bioinformatics is beginning to flourish. New and innovative tools are emerging for the management of biological data. The ability to organize, integrate, analyze, evaluate, and distribute this data holds great promise for the future. As technology has advanced over the years, so has scientists' understanding of genetic information. Such advancements, however, have accelerated to the point where genetic information is being gathered at a rate far exceeding the current ability to interpret it. Prime examples are the technological advances of the last decade that culminated in the mapping of various organisms' genomes.

Scientists are hoping that by studying these genomes through the use of comparative genomics, important disease causing secrets will be revealed. The field of proteomics may prove to be even more promising. But proteomics, however, may also prove to be more challenging than the study of DNA, due in large part to the post-translation modifications of proteins and the complex interactions between them. Nevertheless, proteomics may greatly improve drug discovery by efficiently identifying potential targets earlier in the development process, saving considerable amounts of time and money.

---

279 *Id.*

280 *Id.*

281 *Drowning in Data: Like so Many Others, Biologists are Confronted by a Tidal Wave of Information. Unfortunately, Few of Them Know How to Swim*, The Economist, (Sat., June 26, 1999).

282 *Id.*

283 Harvey, *supra* n. 278.

Before such promises can be realized, however, new innovations and tools must be developed. Successful patent protection of these innovations will provide the required incentive for advancement in the field and commercial viability for companies struggling to fill the computational needs of today's life scientists. Bioinformatics inventions are unique, however, because they combine the use of a computer and/or software with biological information. It is crucial, therefore, for patent practitioners to fully understand the case law relating to computers and software as well as biotechnology. Certain aspects of the statutory requirements for patentability are also affected by blending these disciplines. Determining the identity of the person of ordinary skill in the art can be particularly difficult due to the multidisciplinary nature of the field and its rapid rate of change. As bioinformatics evolves, patent practitioners must be willing to adapt if they want to meet the needs of a bioinformatics client. Such willingness must include stretching beyond one's initial training or background to be effective.

Although bioinformatics is currently evolving, it is here to stay. It has already become firmly established as one of the fastest growing areas of all life science-related markets. Academia, the biotechnology industry, pharmaceutical companies, and even entire governments are all demanding the latest and most innovative bioinformatic-tools. As the bioinformatics market develops, competition will increase, alliances will be made, and new players will emerge. Much of this, however, will not happen without the value-added step of patent protection.