

BIOINFORMATICS AND INTELLECTUAL PROPERTY PROTECTION

By M. Scott McBride[†]

ABSTRACT

This article describes the nature of bioinformatics and how the various components of bioinformatics relate to intellectual property law. The article begins by “decomposing” bioinformatics into three categories: (A) biological sequences such as DNA, RNA, and protein sequences; (B) databases in which these sequences are organized; and (C) software and hardware designed to access, organize, and analyze information contained within these sequences and databases. Next, the article analyzes how each of these components relates to patent law, copyright law, and trade secret law. In particular, the article analyzes whether the various components qualify as protectable subject matter under these areas of law. Where protection may be available, the article discusses whether such protection is practical. The article concludes with a policy discussion of whether intellectual property protection should be available for bioinformatics, where bioinformatic inventions may promote advances in human health care.

I. INTRODUCTION

Advances in biotechnological techniques, such as DNA, RNA, and protein sequencing,¹ and more widespread application of these techniques,² have led to a huge accumulation of information in the past two decades. The DNA of the human genome has now been sequenced,³ and

© 2002 M. Scott McBride

[†] Associate, Foley & Lardner, Milwaukee, WI. J.D. (*summa cum laude*), Marquette University; Ph.D. (Cell and Molecular Biology), University of Wisconsin-Madison; B.S. (Biochemistry), Colorado State University. The author would like to thank Professor Irene Calboli, Marquette University, for reading the text and providing helpful comments.

1. 1 GENOME ANALYSIS: A LABORATORY MANUAL 1-36 (Bruce Birren et al. eds., 1997) (describing numerous techniques for isolating and sequencing DNA and RNA) [hereinafter GENOME ANALYSIS].

2. See CYNTHIA GIBAS & PER JAMBECK, DEVELOPING BIOINFORMATICS COMPUTER SKILLS ix-x (O'Reilly & Assoc. 2001) (describing the increase in accessibility to computers during the past two decades and how this increase in accessibility has given rise to bioinformatics).

3. See Leslie Roberts, *A History of the Human Genome Project*, SCIENCE, Feb. 16, 2001, at 1195 (describing the history of the Human Genome Project and containing a map of the human genome).

the entire human genome will likely be assembled and determined in the near future.⁴ Much of this information is in “raw form” and must be analyzed, organized, and stored.⁵ Bioinformatics is the “[r]esearch, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data.”⁶ It is an amalgamation of biology and information technology.

Bioinformatics is estimated to generate more than a billion dollars of revenue per year worldwide.⁷ Several publicized deals demonstrate that companies value bioinformatics very highly,⁸ perhaps because of the promise it holds for human medicine. Moreover, “genomics information [is nearly] a commodity these days.”⁹

Because companies are willing to invest large sums to reap the benefits bioinformatics holds, it is important to understand the nature of bioinformatics,¹⁰ whether bioinformatics may be subject to intellectual property protection, and what the scope of that protection may be when available. In cases where bioinformatic components are protected by multiple areas of intellectual property law, it is important to determine which form of

4. See Elizabeth Pennisi, *The Human Genome*, SCIENCE, Feb. 16, 2001, at 1177-80. (noting that drafts of the human genome “have yet to be finished, with all the i’s dotted and the t’s crossed.”).

5. See *id.*

6. National Institutes of Health, Office of Extramural Research, Bioinformatics at the NIH, available at <http://grants1.nih.gov/grants/bistic/bistic.cfm> (visited May 5, 2002).

7. See John Thackray, *BIOINFORMATICS Grows LEGS*, ELEC. BUS., July 2001 (citing a report by Strategic Direction International (“SDI”) stating “Bioinformatics generated worldwide revenue [in 2000] of more than \$700 million . . . and total bioinformatics volume could exceed \$2 billion [in 2001]”).

8. See GARY ZWEIGER, *TRANSDUCING THE GENOME* 161 (2001) (describing the growth in biotech companies attempting to capitalize on bioinformatics); *BIO Session: What Does It Mean to Be a Genomics Company in a ‘Post-Genomics’ Era?*, BioSpace.com, at http://www.biospace.com/articles/bio_genomics.cfm (visited Oct. 10, 2002) (noting that in 1993, SmithKline Beecham entered into a \$125 million deal for access to Human Genome Sciences’ biological information, and in 1999, Bayer AG entered into a \$465 million agreement for identification and validation of drug targets with Millennium Pharmaceuticals); *Exelixis in Deal for Genomica*, THE N.Y. TIMES, Nov. 20, 2001, at C8 (noting a \$110 million deal); Press Release: Compaq Announces \$100 Million Investment Program for Life Sciences Start-Up Companies: Targeted for Genomics, Bioinformatics, and Related Areas, Sept. 26, 2000, available at <http://www.compaq.com/newsroom/pr/2000/pr2000092604.html>.

9. David Shook, *Celera: A Biotech That Needs a Boost*, BUS. WK. ONLINE, Mar. 1, 2002, at http://www.businessweek.com/bwdaily/dnflash/mar2002/nf2002031_8351.html.

10. See Sana Siwolop, *INVESTING: A Hunt for the Gems in Genomics*, THE N.Y. TIMES, Oct. 29, 2000, sec.3 C8 (describing how many investors lack the basic knowledge of what genomics companies do).

protection is most practical. These issues are critical in performing a cost-benefit analysis of an investment in bioinformatics; if protection is available and practical, then high investment costs may be justified.¹¹ However, the resolution of these issues is unclear, as exemplified by the recent dispute over the patentability of human genes,¹² and the recent dispute between Celera Genomics, Corp. and the International Human Genome Sequencing Consortium (“IHGSC”) over Celera’s attempt to commercialize a database of the human genome.¹³ This article explores some of these issues by providing a survey of the intellectual property protection currently available for bioinformatic components, and the practicality of the available protection. First, Part II defines the term “bioinformatics.” Part III separates bioinformatics into three components: (A) the information contained within biological sequences, (B) biological databases comprised of this information, and (C) software and hardware designed to access, organize, and analyze this information. Next, Parts IV, V, and VI discuss whether these components are proper subject matter for protection under patent law, copyright law, and trade secret law, respectively.¹⁴ Part VII

11. Intellectual property protection generally includes a right to exclude others from “practicing” the protected subject matter. *See, e.g.*, 35 U.S.C. § 271 (stating patent protection’s basic right to exclude). As such, the owner of a protected “product” may extract a higher price for the “product” to recoup any investment costs, because the owner is the only source.

12. For arguments against gene patents, see generally Public Comments of the United States Patent and Trademark Office “Revised Utility Examination Guidelines; Request for Comments,” 64 Fed. Reg. 71,440 (Dec. 21, 1999), corrected 65 Fed. Reg. 3425, (Jan. 21, 2000), *available at* <http://www.uspto.gov/web/offices/com/sol/comments/utilguide/index.html>. Based on these comments, opposition to the patentability of human genes arises in part because of the mistaken impression that a DNA *sequence* is patentable in lieu of a DNA *composition*. *See id.* (arguing against patents for genomic *sequences*). These opponents may not realize that the sequence itself is probably unprotectable “information,” whereas only the isolated “composition” would be protectable. *See infra* Part III.A.

13. *See* Jasper A. Bovenberg, *Should Genomics Companies Set Up Database in Europe? The E.U. Database Protection Directive Revisited*, E.I.P.R. 23(8) 361[2001] (discussing Celera’s claims that its database is protected by copyright law); Justin Gillis, *Celera to Share Human Genetic Map: Scientists Will Be Able to Download Some Information From Web*, WASH. POST, Feb. 8, 2001, at E18 (noting the IHGSC’s concern over Celera’s *limited* agreement to allow academic scientists access to its database); *Row Over ‘Book of Life,’* BBC NEWS, Feb. 12, 2001, *available at* <http://news.bbc.co.uk/1/hi/sci/tech/1164014.stm>. (noting the IHGSC’s accusation that Celera is “holding back science by imposing commercial restrictions on its data”).

14. Traditionally, intellectual property includes patents, trademarks, copyrights, and trade secrets. *See generally* MARK A. LEMLEY et al., *SOFTWARE AND INTERNET LAW* 50 (Richard Epstein et al. eds, 2000) [hereinafter LEMLEY]. This article excludes trademark protection because a trademark is generally a “source indicator,” and, as such, trademark

concludes with a discussion of public policy issues in regard to intellectual property protection for bioinformatics.

II. BIOINFORMATIC COMPONENTS

Before one can understand intellectual property protection for bioinformatics, it is necessary to understand the nature of the various components that comprise the field of bioinformatics. Bioinformatics involves the acquisition, organization, storage, analysis, and visualization of information contained within biological molecules.¹⁵ For the purposes of this article, bioinformatics is analyzed according to the following categories: (A) biological sequences such as DNA, RNA, and protein sequences, (B) databases in which these sequences are organized, and (C) software and hardware designed to create, access, organize, and analyze information contained within these sequences and databases.

A. DNA, RNA, and Protein Sequences

Scientists classify biological molecules into four general classes that include nucleic acids (which comprise DNA and RNA), proteins, lipids, and carbohydrates.¹⁶ Bioinformatics is currently focused on the biology of DNA, RNA, and protein.

DNA is the material whereby genetic traits are transmitted from one generation to the next. Genes are comprised of DNA.¹⁷ Before DNA is “expressed,” *i.e.*, effects a genetic trait, DNA serves as a “template” to create an RNA molecule.¹⁸ The information within this RNA molecule is then interpreted by cellular machinery to create a protein.¹⁹ As such, RNA is an intermediary molecule within the process of genetic expression.²⁰ The protein created from the RNA molecule is typically the final effector of the genetic trait.²¹ Based on the information within the DNA molecule, a protein folds into a three-dimensional structure, which ultimately deter-

protection might not raise any unique issues for bioinformatics. For a discussion of the goals of trademark law; see generally JANE C. GINSBURG et al., TRADEMARK AND UNFAIR COMPETITION LAW: CASES AND MATERIALS 44-47 (2d ed. 1996).

15. NIH, *supra* note 6.

16. See BENJAMIN LEWIN, GENES VI (6th ed. 1997) (noting that the study of lipids and carbohydrates are largely reserved to biochemists) [hereinafter LEWIN].

17. *Id.* at 71-79. However, Lewin indicates that some viruses use RNA as their genetic material. *Id.*

18. *Id.* at 153-55.

19. *Id.* at 179-81.

20. *Id.* at 153-55.

21. LEWIN at 61-63.

mines its function.²² For example, most enzymes are composed of protein, and many diseases, *e.g.*, lactose intolerance, are the result of defective enzymes created from a mutated DNA. In conclusion, the central dogma of molecular biology is described by the expression:



Each of these three molecules are described using a fairly simple code: DNA by A,C,G,T; RNA by A,C,G,U; and protein by twenty different amino acids.²⁴

DNA, or deoxyribonucleic acid, is a large molecule comprised of four different repeating units called nucleotides.²⁵ DNA nucleotides contain one of four nitrogenous bases (adenine (“A”), guanine (“G”), cytosine (“C”), or thymine (“T”)),²⁶ and the sequence of a particular DNA is typically described by using the single-letter designation of the nucleotides within the DNA sequence, *e.g.*, ATTGGCATGGA.²⁷

RNA, like DNA, is comprised of a chain of nucleotide molecules.²⁸ However, RNA differs from DNA because it contains RNA nucleotides,²⁹ rather than DNA nucleotides. RNA nucleotides, like DNA nucleotides, may contain adenine, guanine, or cytosine, but unlike DNA nucleotides, RNA nucleotides use uracil (“U”) instead of thymine.³⁰ In a simplistic way, an RNA molecule is a copy of the DNA where “T” is replaced with “U.” Therefore, a DNA molecule with the sequence “ATTGGCATGGA,” would have a corresponding RNA molecule with the sequence “AUUGG-

22. *Id.* at 13-19.

23. *Id.* at 154.

24. *Id.* at 76-79 (describing the DNA and RNA codes); *id.* at 10-11 (describing the amino acid code).

25. LEWIN at 76-77.

26. *Id.* at 76-77.

27. *See id.* at 81.

28. *Id.* at 76-77.

29. *Id.*

30. LEWIN at 76-79. In addition to using uracil instead of thymine, the nucleotides of an RNA molecule use ribose instead of deoxyribose as a sugar moiety. *Id.* at 76-77. This difference, while conceptually simple, actually has drastic implication for the stability of RNA as compared to DNA. While DNA is relatively stable and resistant to enzymes that degrade nucleic acid called nucleases, RNA is inherently unstable and sensitive to nucleases. *See id.* at 173-77. The cell can utilize RNA’s instability as a mechanism for regulating the expression of a corresponding gene. *Id.* For example, after an RNA has been synthesized and a gene has been expressed, the cell can rapidly and easily degrade the RNA to prevent further expression until the cell synthesizes new RNA. *Id.*

One additional difference between RNA and DNA is that RNA typically exists as a single-stranded molecule while DNA is typically double-stranded. *See id.* at 81.

CAUGGA.” This RNA molecule is used as a template to synthesize the encoded protein.

Proteins are comprised of twenty different amino acids described by the single letter designations A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y,³¹ and a protein molecule contains a sequence of any combination of these twenty amino acids, *e.g.*, P-A-T-E-N-T-L-A-W-I-S-G-R-E-A-T. Each of these twenty amino acids is specified by three nucleotides of RNA, *e.g.*, AUG corresponds to methionine or “M”.³² Such triplets comprise codons. Because there are sixty-four different combinations of nucleotide triplets, *i.e.*, $4^3 = 64$, and there are only twenty amino acids, there are more codons than necessary to code for the twenty amino acids.³³ As such, more than one codon can code for a particular amino acid, thereby leading to redundancy in the genetic code.³⁴ Because of this redundancy, it is not always possible to determine the correct codon sequence for a given amino acid, while it is always possible to determine the correct amino acid for a given codon sequence.³⁵

Gene expression, or the route from gene to protein, is regulated within cells. Thus, two genetically identical cells, such as a skin cell and a nerve cell, may express a different complement of proteins³⁶ and hence exhibit different traits. One aspect of bioinformatics is the study of gene expression through functional genomics (*e.g.*, studying the expression of genes at the mRNA level), and functional proteomics (*e.g.*, studying the expression of genes at the protein level).³⁷

In summary, DNA, RNA, and protein are large molecules comprised of repeating units of DNA nucleotides, RNA nucleotides, and amino acids, respectively. DNA, RNA, and protein can be described by the sequence of these repeating units, and the sequence of these repeating units ultimately determines the function of the DNA, RNA, or protein. Therefore, the sequence of the DNA, RNA, or protein contains functional information.

31. LEWIN at 8. The twenty amino acids are alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine, tryptophan, and tyrosine, respectively. *Id.* These twenty amino acids can also be designated by three-letter designations (Ala = “A”, Cys = “C”, Asp = “D”, Glu = “D”, Phe = “F”, Gly = “G”, His = “H”, Ile = “I”, Lys = “K”, Leu = “L”, Met = “M”, Asn = “N”, Pro = “P”, Gln = “Q”, Arg = “R”, Ser = “S”, Thr = “T”, Val = “V”, Trp = “W”, and Tyr = “Y”). *Id.*

32. *Id.* at 213-15.

33. *Id.*

34. *Id.*

35. LEWIN at 8.

36. *See id.* at 811-13.

37. *See GIBAS & JAMBECK, supra* note 2, at 310-21.

B. Biological Databases

As more DNA, RNA, and protein sequences are reported, scientists are developing biological databases to catalog and store the sequence information.³⁸ These databases are valuable if the stored information can be readily searched, accessed, and analyzed. For instance, scientists can use these databases to compare and assign biological functions to particular or characteristic sequences (*i.e.*, “motifs”).³⁹ Then, when a scientist obtains a sequence from an unknown DNA, RNA, or protein molecule, the scientist can use these databases to identify the unknown molecule and determine its function.⁴⁰ Scientists are encouraged to contribute to these databases.⁴¹ For instance, most scientific journals expect the scientist to submit the sequence of a novel biological molecule to a public database prior to publication.⁴² Failure to submit a sequence may result in the scientist being denied the opportunity to publish the article.⁴³

Although several databases are available to the general public,⁴⁴ private companies are not required to make their databases freely available. For example, one company working on sequencing the human genome, Celera, generally charges for access to its database,⁴⁵ although it provides

38. For example, the National Center for Biotechnology Information (“NCBI”) offers several databases that are available to the general public. *See* NCBI, Submit to GenBank, available at <http://www.ncbi.nlm.nih.gov/Genbank/index.html> (visited May 5, 2002).

39. *See* NCBI, BLAST: Basic Overview, available at http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul_1.html (visited May 5, 2002).

40. Search programs, such as BLAST®, can be used to search databases for similar proteins. *See id.*

41. *See* NCBI, Submit to GenBank, available at <http://www.ncbi.nlm.nih.gov/Genbank/index.html> (visited May 5, 2002). (“The most important source of new data for GenBank® is direct submissions from scientists.”) There is a “20-year old convention within genomics research of placing data in GenBank[®] or similar large publicly run databases as a condition of academic publication.” Pete Moore, *Publication with a Pinch of Privatisation*, THE SCIENTIST, Apr. 4, 2002, available at <http://www.biomedcentral.com/news/20020404/04>.

42. *See* Moore, *supra* note 41.

43. *Id.* However, SCIENCE recently broke with tradition and published two articles even though “the genomic data underpinning the publications” are kept in private databases. *Id.* SCIENCE’s break with tradition caused “20 eminent scientist to write a letter of protest . . . saying that the action poses ‘a serious threat to genomics research.’” *Id.* (reprinting the letter of protest in its entirety).

44. The NCBI offers several databases besides GenBank®, including “RefSeq,” “PDB,” and “Entrez Genomes,” for nucleotide sequences, and “SwissProt,” “PIR,” “PRF,” and “PDB” for amino acid sequences. *See* NCBI, DATABASES, available at <http://www.ncbi.nlm.nih.gov/Databases/index.html> (visited May 5, 2002).

45. Bovenberg, *supra* note 13.

free access to “qualified academic users.”⁴⁶ Celera claims that its database is subject to patent and copyright protection,⁴⁷ an issue disputed by Celera’s noncommercial competitor, the IHGSC.⁴⁸ Celera’s case exemplifies the necessity of analyzing whether databases such as Celera’s should be subject to IP protection.⁴⁹

C. Bioinformatic Software and Hardware

To utilize information contained in these databases, software developers have developed bioinformatic programs to organize, access, analyze, and view sequence information.⁵⁰ One such program, BLAST[®] (“Basic Local Alignment Search Tool”),⁵¹ compares sequences for similarity by first aligning the two sequences at areas of local identity or similarity and then calculating a “similarity score.”⁵² Such algorithms can be designed to incorporate scientific principles based on the molecular biology of DNA, RNA, and protein. For example, an algorithm may be created to compare two nucleotides or amino acids that are not identical but function similarly based on their molecular biology.⁵³ Such programs are useful in predicting

46. *Id.*

47. *See id.* *See also* Celera Free Public Access Click-On Agreement, Heading 4.a., at <http://www.celera.com/genomics/academic/pubsite/terms.cfm> (“The Celera Data, both the *primary sequence assembly* and the representation thereof, is a copyrighted work . . .”) (emphasis added).

48. Bovenberg, *supra* note 13. The IHGSC further argues that Celera is not entitled to intellectual property protection for its database of the human genome because its database contains sequences that are within the public domain. Philip Cohen, *Rivals Dismiss Celera’s Human Genome Draft*, NEW SCIENTIST, Mar. 5, 2002, available at <http://www.newscientist.com/news/news.jsp?id=ns99991999> (visited May 5, 2002).

49. Incyte Genomics also offers subscriptions to its databases. *See* <http://www.incyte.com>. Other gene database companies include Human Genome Sciences and Millennium Pharmaceuticals. *See* Matthew Herper, *Stock Focus: Genomics Companies*, FORBES.COM, Apr. 4, 2001, at <http://www.forbes.com/2001/04/11/0411sf.html>.

50. *See, e.g.*, NCBI, Tools for Data Mining, available at <http://www.ncbi.nlm.gov/Tools/index.html> (visited May 5, 2002) (listing several bioinformatics programs including BLAST[®], MapViewer, LocusLink, UniGene, ORF Finder, Electronic PCR, VAST Search, and VecScreen).

51. *See* NCBI, BLAST: Basic Overview, available at http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul_1.html (visited May 5, 2002) (describing the algorithm used by BLAST[®] to compare biological sequences).

52. *Id.*

53. For nucleotide sequences, because of redundancy in the genetic code, two genes may use different nucleotides and still encode the same amino acid. *See supra* notes 32-35 and accompanying text. For proteins, certain amino acids may be interchangeable. *See id.* (describing how certain amino acids may be grouped as “hydrophobic” or “hydrophilic,” or alternatively described as “acidic,” “basic,” or “neutral”).

the function of an unknown gene or protein, or to draw evolutionary relationships.⁵⁴

Engineers have also developed computer hardware and machines that facilitate the acquisition and storage of biological information. For example, machines called “thermocyclers” amplify small amounts of DNA or RNA to provide a scientist with a workable amount for sequencing.⁵⁵ Other machines rapidly determine the sequence of DNA, RNA, or protein molecules.⁵⁶ One of the most promising recent inventions is the “gene chip.” A gene chip contains many different DNA sequences organized in a grid or microarray on the chip.⁵⁷ By exposing the chip to a test sample of DNA, a scientist determines whether the test sample corresponds to any of the sequences on the chip through a process called “hybridization.”⁵⁸ The gene chip is advantageous because it is a “high throughput device,” meaning that a scientist can obtain a large amount of information from a single input or experiment, and furthermore, the gene chip is suitable for automation.⁵⁹

The next three sections analyze whether these defined components of bioinformatics, *i.e.*, (A) DNA, RNA, and protein sequences, (B) biological databases, and (C) bioinformatic software and hardware, are proper subject matter for patent, copyright, or trade secret protection.

54. See, *e.g.*, L. Feng et al., *Aminotransferase Activity and Bioinformatic Analysis of L-Aminocyclopropane-1-Carboxylate Synthase*, *BIOCHEMISTRY*, Dec. 12, 2000, at 15242-29 (describing the use of BLAST® to draw an evolutionary connection between the aminocyclopropane carboxylate synthases and the aminotransferases).

55. Brinkmann Company sells popular thermocyclers, described on its website: <http://www.brinkmann.com/product.asp?ref=86&tb=Description> (visited May 5, 2002).

56. These machines are aptly named “sequencers.” Applied Biosystems sells rapid DNA sequencers, described on its website: http://www.appliedbiosystems.com/products/productdetail.cfm?prod_id=41 (visited May 5, 2002).

57. GIBAS & JAMBECK, *supra* note 2, at 311-17. For a description of the technology underlying “gene chips,” see also <http://www.gene-chips.com> (visited May 5, 2002) [hereinafter *Gene Chips*].

58. See *Gene Chips*, *supra* note 57. See also *GENOME ANALYSIS*, *supra* note 1 (describing numerous techniques for DNA analysis including “hybridization”). “Hybridization” refers to the process of identifying a particular DNA or RNA sequence by using a probe that is complementary to the identified sequence. For example, DNA and RNA form double-stranded molecules like a “zipper” by binding to a complementary molecule. Complementarity relies on the fact that A binds to T (or U in RNA’s case) and G binds to C. To detect the DNA target sequence AGCTTCGA, one would use the probe TCGAAGCT labeled with radioactivity or photo-emitting moieties. Gene chips are useful because a scientist can adhere many nucleotide sequences to a single gene chip, and use the chip to obtain a large amount of information from a single “hybridization.” See *Gene Chips*, *supra* note 57.

59. See *Gene Chips* *supra* note 57.

III. PATENT PROTECTION: ELIGIBLE SUBJECT MATTER MUST BE A PROCESS, MACHINE, APPARATUS, OR COMPOSITION OF MATTER

One of the most critical questions regarding whether bioinformatic components are patentable is whether they qualify as statutory subject matter under § 101 of the Patent Act.⁶⁰ Under § 101, “Whoever invents or discovers any new and useful process, machine, manufacture [apparatus], or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor.”⁶¹ Thus, to determine whether bioinformatic components qualify as statutory subject matter, one must determine whether bioinformatic components are a “new and useful process, machine, manufacture [apparatus], or composition of matter, or any new and useful improvement thereof.”⁶²

A. Patent Protection for DNA, RNA, and Protein Sequences

Section 101 permits the patentability of “composition[s] of matter.”⁶³ Courts have held this as including “all compositions of two or more substances and . . . all composite articles, whether they be the results of chemical union, or of mechanical mixture, or whether they be gases, fluids, powders or solids.”⁶⁴ The USPTO has specifically interpreted this to include DNA, RNA, and protein *compositions*⁶⁵ because they are composed of two or more substances—DNA and RNA are composed of nucleotides while proteins are made up of amino acids. Indeed, many DNA, RNA, and protein molecules have been patented as *compositions*.⁶⁶

60. 35 U.S.C. § 101 (1998). To be patentable, subject matter must also possess “utility” under § 101, and it is well-known that the subject matter of an invention must also meet the statutory requirements under 35 U.S.C. §§ 102 (novelty) and 103 (non-obviousness). However, a thorough discussion of the “novelty” or “nonobviousness” of bioinformatic components is beyond the scope of this article. For such a discussion, see Charles Vorndran & Robert L. Florence, *Bioinformatics: Patenting the Bridge Between Information Technology and the Life Sciences*, 42 J.L. & TECH. 93 (2002).

61. 35 U.S.C. § 101. For a discussion of the scope of patentable subject matter see CHISUM ON PATENTS §§ 1.01-1.06[5] (2002).

62. 35 U.S.C. § 101.

63. *Id.*

64. *Diamond v. Chakrabarty*, 447 U.S. 303, 308 (1980) (citing *Shell Dev. Co. v. Watson*, 149 F. Supp. 279, 280 (D.D.C. 1957) (citing WALKER ON PATENTS § 14, at 55 (1st ed. 1937))).

65. *See* 66 Fed. Reg. 1092-97 (Jan. 5, 2001) (noting the USPTO response to comments regarding the patentability of genes).

66. *See, e.g.*, U.S. Patent No. 6,348,348 (issued Feb. 19, 2002) (claiming the nucleotide and deduced amino acid sequences of the Human Hairless gene); U.S. Patent No. 6,284,492 (issued Sept. 4, 2001) (claiming viral nucleic acid).

However, it was not always clear that biological molecules were patentable subject matter. Only after the Supreme Court's decision in *Diamond v. Chakrabarty*⁶⁷ did patents on biological molecules become widespread. Writing for the majority, Chief Justice Burger concluded that § 101 permitted the patenting of genetically modified bacteria,⁶⁸ stating, "Congress intended statutory subject matter to 'include anything under the sun that is made by man.'"⁶⁹ Since then, the USPTO has permitted the patenting of biological molecules under the premise that a biological molecule is a "composition made by man," where the biological molecule has been isolated and purified from its natural setting.⁷⁰

While biological molecules are themselves patentable as *compositions*, the information within the composition, *i.e.*, the abstract biological sequence itself, arguably is not patentable subject matter.⁷¹ Based on the Supreme Court's holding in *Diamond v. Diehr*,⁷² to qualify as patentable subject matter the biological sequence would have to be categorized as a process, machine, apparatus, or composition, and do more than describe a "natural phenomenon."⁷³ The *Diehr* Court also excluded "laws of nature . . . and abstract ideas" from patent protection.⁷⁴ "An idea of itself is not patentable,"⁷⁵ and neither is "[a] principle, in the abstract[,] a fundamental truth[,] an original cause[, or] a motive."⁷⁶ As "Einstein could not patent his celebrated law that ' $E = mc^2$ ' [and] Newton [could not] have patented the law of gravity,"⁷⁷ it is unlikely that one could patent a biological sequence since it may be characterized as a natural phenomenon. Therefore, patent protection for DNA, RNA, or protein extends only to the physi-

67. 447 U.S. 303.

68. *Id.* at 308-09.

69. *Id.* at 309 (citing S. REP. NO. 82-1979, at 5 (1952); H.R. REP. NO. 82-1923, at 6 (1952)).

70. *See* 66 Fed. Reg. 1092-99 (Jan. 5, 2001).

71. For a discussion of the distinction between DNA as a molecule versus DNA as information, *see* Rebecca S. Eisenberg, *Re-examining the Role of Patents in Appropriating the Value of DNA Sequences*, 49 EMORY L.J. 783, 786-89 (2000).

72. 450 U.S. 175, 191-93 (1981).

73. *Id.* In *Diehr*, the Court held that a process for curing rubber was patentable, even though the process relied on an unpatentable mathematical formula to calculate the amount of time that the rubber needed to "cure." *Id.*

74. *Id.* at 185 (citing *Parker v. Flook*, 437 U.S. 584 (1978); *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972); *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948)).

75. *Id.* (citing *Rubber-Tip Pencil Co. v. Howard*, 20 Wall. 498, 507 (1874)).

76. *Id.* (citing *LeRoy v. Tatham*, 14 How. 156, 175 (1853)).

77. *Id.* (citing *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980) (quoting *Funk Seed Bros. Co.*, 333 U.S. 127 (1948))).

cal/biological composition, and not to the abstract biological sequence information that describes the composition. Thus, a patentee could only prevent another from using the composition itself and not the information within the molecule.⁷⁸

B. Patent Protection for Biological Databases

For the same reason that patent protection is unavailable to biological sequences, patent protection may also be unavailable for biological databases. Biological databases are compilations of biological sequences. If biological sequences are unpatentable information, then a biological database is a compilation of unpatentable information.⁷⁹ Thus, for a database to be patentable, the process of compiling and organizing the biological sequences into a database must convert the unpatentable information into statutory subject matter for a patent, *i.e.*, a “tangible product.”⁸⁰ Whether a database is a “tangible product” might be debatable,⁸¹ but the USPTO has said that if the database is merely a “data structure” or “nonfunctional descriptive material,” it is not patentable.⁸²

Even if the database itself does not constitute patentable subject matter, the *manner of creating the database* may constitute a patentable process. For instance, in *State Street Bank and Trust Co. v. Signature Financial Group, Inc.*,⁸³ the Federal Circuit held that a data processing system for a financial fund was patentable subject matter where the system produced a “useful, concrete, and tangible result”⁸⁴ even though the data processing system produced only information. Of course, to reconcile the Federal Circuit’s holding and semantics in *State Street Bank* with prior

78. Eisenberg, *supra* note 71, at 788 (“Patent claims on DNA sequences as ‘compositions of matter’ give patent owners exclusionary rights over tangible DNA molecules and constructs, but do not prevent anyone from perceiving, using, and analyzing information about what the DNA sequence is.”).

79. *See id.* at 787. (“The traditional statutory categories of patent-eligible subject matter . . . seem to be limited to *tangible products* and processes, as distinguished from information as such.”) (emphasis added).

80. *See id.*

81. *See id.* (“Although many cases have used the word ‘tangible’ in defining the boundaries of patentable subject matter, neither the language of the statute nor judicial decisions elaborating its meaning have *explicitly* excluded ‘information’ from patent protection.” However, “such a limitation is *implicit* in prior judicial decisions.”) (emphasis added).

82. *See* U.S. Department of Commerce, Manual of Patent Examining Procedure 2106-11 to -35 (8th ed. 2001) (citing *In re Warmerdam*, 33 F.3d 1354, 1360-61 (Fed. Cir. 1994)).

83. 149 F.3d 1368 (Fed. Cir. 1998).

84. *Id.* at 1375 (citing *In re Alappat*, 33 F.3d 1526, 1544 (Fed. Cir. 1994)).

courts' holdings,⁸⁵ we may have to assume that a "tangible result" is not necessarily to be equated with a "tangible product."⁸⁶ Nonetheless, under *State Street Bank*, even if information *per se* is not patentable as a "tangible product," a process of producing information may be patentable if it produces a "tangible result."⁸⁷ Applying this principle to bioinformatic databases, we can conclude that if the *process* of creating a bioinformatic database produces a "useful, concrete, and tangible result," *i.e.*, a database that has numerous applications, then the process of creating the database may be patentable.

However, such a process patent would be limited in at least two ways. First, the process must satisfy the other requirements of the Patent Act. In particular, the process must be novel under § 102,⁸⁸ and nonobvious under § 103.⁸⁹ In conforming to §§ 102 and 103, the scope of the patent's claim undoubtedly would be narrowed. Because scientists have been producing and cataloguing biological information for many years, a patentee would have to draft process claims narrowly to avoid the prior art; and even if the patentee could draft process claims narrowly so as to be novel, the patent claims may yet be obvious in light of the prior art.

Second, patent protection would extend only to the process for creating the database and not to the database itself. This would limit the value of the patent because a competitor wanting to create an identical database could avoid infringing the patent simply by creating the database by a non-infringing process,⁹⁰ *i.e.*, creating the database by performing different steps than those recited within the claimed method.⁹¹ Even if the competi-

85. See Eisenberg, *supra* note 71, at 787 (noting that prior courts have implicitly held that only "tangible products" are patentable and information is not a "tangible product").

86. See *id.*

87. See *State St. Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 1375 (Fed. Cir. 1998).

88. 35 U.S.C. § 102 (1998).

89. *Id.* § 103.

90. In contrast, machine claims, apparatus claims, and composition claims implicitly include method of making claims, because under 35 U.S.C. § 271, an infringer is one who "without authority makes, uses . . . any patented invention." *Id.* § 271(a).

91. Because the patent claims define the invention, an infringer must perform the equivalent of each step of the claimed process to infringe the process under the "all elements rule." *ATD Corp. v. Lydall Inc.*, 159 F.3d 534, 552 (Fed. Cir. 1998) (Clevenger, J., concurring in part and dissenting in part) ("A claim of infringement by equivalents cannot succeed unless each limitation of a claim is met by an equivalent.") (citing *Warner-Jenkinson Co. v. Hilton Davis Chem. Co.*, 520 U.S. 17, 41 (1997) (adopting *sub silentio* the "all elements" rule of *Pennwalt Corp. v. Durand-Wayland, Inc.*, 833 F.2d 931 (Fed. Cir. 1987) (en banc)).

tor infringes the patented process, it may be more difficult to prove infringement of a process claim than a machine, apparatus, or composition claim, because the plaintiff would have to prove that the database was *created* by the patented process, not just used or sold. If the database itself is more valuable than the patented process, the patent would offer only token protection. Therefore, while patent protection for creating databases may be available under a *State Street Bank* theory, the protection may be narrow, easily evaded, and of questionable value.

C. Patent Protection for Bioinformatic Software and Hardware

In contrast to biological sequences and databases, computer software does constitute patentable subject matter if the software produces a “useful, concrete, and tangible result.”⁹² The Supreme Court and Federal Circuit have indicated that so long as a software program is more than a mere algorithm, the program may be eligible for patent protection.⁹³ Bioinformatic software should be no exception.⁹⁴ The results produced by bioinformatic software have a biological application and are therefore most definitely “useful, concrete, and tangible.” Because bioinformatic software can be used to make medical diagnoses, design drugs, or draw evolutionary conclusions, it would be difficult to hold bioinformatic software as unpatentable under a *State Street Bank* regime.

Likewise, patent protection is available for bioinformatic hardware, where the hardware qualifies as patentable subject matter under § 101 as a “machine” or an “apparatus.”⁹⁵ Bioinformatic hardware may be used to acquire bioinformatic information (*e.g.*, as a sequencer or a gene chip), and/or store, access, or organize bioinformatic information (*e.g.*, as a computer system). However, because a patent would only protect the patentee from an infringer who uses a machine or apparatus that contains *all the elements* of the claimed invention,⁹⁶ the patentee could not protect a biological sequence or a database *that is only a component* of a protected machine or apparatus.

92. *State St. Bank*, 149 F.3d at 1375 (citing *In re Alappat*, 33 F.3d 1526, 1544 (Fed. Cir. 1994)).

93. *See* *Diamond v. Diehr*, 450 U.S. 175, 185 (1981); *In re Alappat*, 33 F.3d at 1544; *State St. Bank*, 149 F.3d at 1373.

94. However, see Part VI for a discussion of the “Open Informatics” petition, which would require that all publicly-funded, bioinformatic software be made freely available to the public.

95. 35 U.S.C. § 101 (1998) (“Whoever invents . . . any new and useful . . . machine, manufacture [apparatus] . . . may obtain a patent therefor . . .”).

96. *See supra* note 91.

IV. COPYRIGHT PROTECTION: ELIGIBLE SUBJECT MATTER MUST BE AN ORIGINAL EXPRESSION

The Copyright Act defines the requirements for copyrightable subject matter.⁹⁷ Under § 102, “[c]opyright protection subsists . . . [i]n original works of authorship fixed in any tangible medium of expression.”⁹⁸ Copyright protection is available for “works of authorships,” such as “literary works,” but copyright protection does not extend to “any idea, procedure, process, system, method of operation, concept, principle, or discovery. . . .”⁹⁹ This latter limitation severely restricts the scope of copyright protection available for bioinformatic components.

A. Copyright Protection for DNA, RNA, and Protein Sequences

Arguably, the originator(s) of the DNA code nomenclature (who used A, G, C, and T to describe a DNA’s sequence), the RNA code nomenclature (who used A, G, C, and U to describe an RNA’s sequence), and the protein code nomenclature (who used A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y to describe a protein’s sequence) may have had a legitimate claim to copyright protection for their original expression.¹⁰⁰ However, as the law now stands “the Copyright Office has unofficially stated that it will not grant copyright registration to gene sequences or DNA molecules because they are not copyrightable subject matter.”¹⁰¹ Furthermore, a contemporary scientist discovering a biological molecule probably would not be entitled to copyright protection for the sequence of the newly discovered molecule or information contained therein for several reasons.¹⁰²

First, the scientist is not the original author of the biological code nomenclatures. Although the scientist is the first to report the sequence of the

97. 17 U.S.C. §§ 101-1332 (1998).

98. *Id.* § 102(a).

99. *Id.* § 102(a)(1), (b).

100. Using any one of these codes to describe the respective biological molecule might be considered an “original work of authorship” under § 102. *See id.* § 102. However, these codes have been in use at least since the 1930s, and any “work of authorship” that was published before 1923 and was never registered has fallen into the public domain. *See id.* § 301 (describing the duration of copyrights that had not fallen into the public domain prior to January 1, 1978, the effective date of the act. Prior to the Copyright Act of 1976, the term of a copyright was 56 years, and copyrights initiated before 1923 would have expired before the January 1, 1978 effective date of the 1976 revisions.).

101. *See* James G. Silva, *Copyright Protection of Biotechnology Works: Into the Dustbin of History?*, B.C. INTELL. PROP. & TECH. F. (2000) (citing MICHAEL A. EPSTEIN, MODERN INTELLECTUAL PROPERTY, Ch. 11, II, C 458-59 (2d ed. 1992)).

102. *See id.*

novel molecule and the reported sequence may therefore comprise “original expression” under copyright law, the originality of his expression is minimal because the biological codes have been used for decades to report sequences.¹⁰³ Second, the sequence or information that the scientist seeks to protect is a “discovery” or “idea,”¹⁰⁴ neither of which is entitled to copyright protection. Third, because of the limited ways to express a DNA, RNA, or protein sequence, these biological codes have become standard techniques for describing molecules and are therefore not “creative expression.” Under the doctrine of scenes à faire, “when similar features . . . are ‘as a practical matter indispensable, or at least standard in the treatment of a given [idea], they are treated like ideas and are therefore not protected by copyright.’”¹⁰⁵ Where there is simply no other way to describe a natural phenomenon, there is no room for “creative expression.”

Even if the scientist were to obtain copyright protection for the sequence of a discovered biological molecule, an accused infringer might assert the defense of “fair use” under § 107.¹⁰⁶ In determining “fair use,” courts use four balancing factors including (1) “the purpose and character of the use,” *e.g.*, commercial versus not-for-profit, (2) “the nature of the copyrighted work,” *e.g.*, fiction versus nonfiction compilation, (3) “amount and substantiality of the portion used,” *e.g.*, an entire work versus a small portion of a large work, and (4) “effect of the use upon the potential market.”¹⁰⁷ For example, “fair use” would arguably exist where the accused infringer shows that he used the sequence of a single gene from a large copyrighted compilation (assuming that the compilation is copyrightable¹⁰⁸) where his purpose was “criticism, comment, news reporting, teaching, scholarship, or research”¹⁰⁹ in a not-for-profit, academic setting. In this regard, many critics of IP protection for bioinformatics have been

103. The chemical composition of DNA was found in 1909, and DNA was made artificially in 1956. Damian Carrington, *The History of Genetics*, BBC NEWS, May 30, 2000, available at http://news.bbc.co.uk/1/hi/english/in_depth/sci_tech/2000/hman_genome/newsid_749000/749026.stm (visited May 5, 2002).

104. See 17 U.S.C. § 102(b).

105. *Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1444 (9th Cir. 1994) (citing *Frybarger v. IBM Corp.*, 812 F.2d 525, 530 (9th Cir. 1987) (quoting *Atari, Inc. v. N. Am. Philips Consumer Elec. Corp.*, 672 F.2d 607, 616 (7th Cir. 1982), *cert. denied*, 459 U.S. 880 (1982)).

106. 17 U.S.C. § 107 (1998).

107. *Id.* See also *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994) (discussing the four “fair use” factors). “All [four factors] are to be explored, and the results weighed together . . .” *Id.* at 578.

108. See *infra* Part IV.B.

109. 17 U.S.C. § 107.

academic researchers,¹¹⁰ for whom the “fair use” is more likely to apply.¹¹¹ In summary, copyright protection for biological sequences is probably unavailable,¹¹² and were it to become available, it might be evaded by some of its strongest critics under the “fair use” exception.

B. Copyright Protection for Biological Databases

The Copyright Act of 1976 specifically describes compilations as copyrightable subject matter; therefore if a database is described as a compilation, it may qualify for copyright protection. The Supreme Court explored the boundaries of copyright protection for compilations in *Feist Publications, Inc. v. Rural Telephone Service Co.*¹¹³ In *Feist* the work at issue was a telephone book, for which the creator sought copyright protection.¹¹⁴ Justice O’Connor, writing for the majority, described the issue in the case: “[F]acts are not copyrightable [but] compilations of facts generally are.”¹¹⁵ However, the compilation must be sufficiently original, *e.g.*, in *selection* or *arrangement* of the compiled facts.¹¹⁶ Where a compilation is copyrighted, copyright protection does not extend to every element of the work.¹¹⁷ “Originality is the *sine qua non* of copyright [and] copyright protection may extend only to those components of a work that are original to the author.”¹¹⁸

110. *See generally* Public Comments on the United States Patent and Trademark Office “Revised Interim Utility Examination Guidelines,” 64 Fed. Reg. 71440 (Dec. 21, 1999). Many of those responding to the USPTO’s request for comments regarding its new “utility” guidelines for patentability were academic researchers who echoed Dr. Steven E. Sherer’s comments: “I believe that at least [the] human genomic sequence goes to the core of what it means to be human and no individual or corporation should control or have ownership of something so basic.”

111. *See* 17 U.S.C. § 107.

112. The DNA Copyright Institute (“DNACI”) Inc. might disagree. *See* <http://www.dnacopyright.com> (visited May 5, 2002). For a fee, the DNACI will collect a sample of your DNA, determine your unique “DNA profile,” and report your profile to you so that you can establish copyright protection. *Id.* However, nowhere on the DNACI website does the DNACI persuasively establish that copyright protection is available for one’s “DNA profile” under the Copyright Act. *Id.* Furthermore, one can argue that we are not the “authors” of our DNA profiles. Our parents or maybe even a “higher authority” may be the true authors.

113. 499 U.S. 340 (1991).

114. *Id.* at 342-43.

115. *Id.* at 344.

116. *Id.* at 346, 348.

117. *Id.* at 348.

118. *Id.* *See also* N.Y. Times Co. v. Tansini, 533 U.S. 483, 494 (2001) (discussing the elements of an electronic database compilation which are subject to copyright). Copyright in a compilation “is limited to the compiler’s original ‘selection, coordination, and

Applying *Feist's* principles, biological databases are copyrightable, provided they contain the requisite originality. For example, a scientist might obtain copyright protection if he chooses an original set of genes or proteins for a database or arranges the database in an original way. However, the copyright protection would not extend to all the genes or proteins in the database. Rather, copyright protection would extend only to his original selection or arrangement. Thus, a competitor who creates his own database using individual elements of the scientist's copyrighted database would not infringe the scientist's copyright so long as the competitor does not use the same selection or arrangement as the scientist's copyrighted database. Therefore, copyright protection for databases is limited.

Certain databases might have qualified for *sui generis* protection¹¹⁹ under bills that were debated in the U.S. House of Representatives in 1998 and 1999.¹²⁰ These bills contemplated *sui generis* protection for databases and borrowed elements from the Patent Act, *e.g.*, a short defined term,¹²¹ and elements from the Copyright Act, *e.g.*, a research exception comparable to "fair use."¹²² To date, this legislation has not been enacted. However, because some members of the European Union have enacted *sui*

arrangement." *Id.* (quoting *Feist*, 499 U.S. at 358). *See also* *Torah Soft Ltd. v. Drosnin*, 136 F. Supp. 2d 276, 286 (S.D.N.Y. 2001) (analyzing which elements of a compilation of the Hebrew Bible are subject to copyright). "A work comprised of material which in itself is not protected may become protectable as a compilation if the copyright holder has utilized sufficient creativity in selecting and arranging the material." *Id.* (quoting *Feist*, 499 U.S. at 358). The *Torah Soft* court found that the Hebrew Bible compilation was not protectable because the compilation possessed only *de minimis* creativity and incorporated only functional changes that were merely *scenes à faire*. *Id.* at 287-88.

119. "Sui generis protection" refers to protection "of its own kind or class." BLACK'S LAW DICTIONARY 1434 (2d ed. 1990).

120. These bills include H.R. 2652, 105th Cong. (1998), which later became H.R. 2281, 105th Cong. (1998) as part of the Digital Millennium Copyright Act ("DMCA"), and H.R. 354, 106th Cong. (1999). *See* J.H. Reichmann and Paul F. Uhlir, *Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology*, 14 BERKELEY TECH. L.J. 793, 802 (1999). H.R. 2281 was dropped prior to enactment of the DMCA and was reintroduced as H.R. 354, 106th Cong. (1999). *See id.*

121. Under H.R. 2652 and H.R. 354, the term for protection would have been 15 years. H.R. 2652 § 1207(C); H.R. 354 § 1408(c). Although some have argued that the owner could extend the term indefinitely by "invest[ing] in maintenance or updates of a dynamic database." *Id.* at 809-10.

122. "[N]o person shall be restricted from making available or extracting information for nonprofit educational, scientific, or research purposes in a manner that does not materially harm the primary market for the product or service referred to . . ." H.R. 354 § 1403(b). Similar language is included in H.R. 2652 § 1202(D) and H.R. 2281 § 1303(D). H.R. 354 also lists five factors similar to the four factors in 17 U.S.C. § 107. *See* H.R. 354 § 1403(a)(1)-(5).

generis protection for databases under an E.C. Directive,¹²³ Congress may feel pressure to harmonize U.S. law and enact some form of database protection in the future.¹²⁴

C. Copyright Protection for Bioinformatic Software and Hardware

Courts have construed the term “literary works”¹²⁵ liberally to encompass computer software.¹²⁶ Thus, copyright protection is available for computer software and by extension to bioinformatic software where either the object code¹²⁷ or the source code¹²⁸ represents an original form of expression.¹²⁹ However, copyright protection for computer software is not as robust as patent protection. For instance, copyright protection extends only to the “original expression” contained within the software, and not to the functional elements or methods.¹³⁰ Typically, “original expression” is found in the literal code of the software,¹³¹ and to avoid infringement, a competitor need only use different object or source code to achieve the same result. Therefore, copyright might not protect functional elements of the software, such as a hierarchical structure of the bioinformatic pro-

123. See Xuqiong (Joanna) Wu, *Foreign and International Law: E.C. Database Directive*, 17 BERKELEY TECH. L.J. 571 (2002).

124. *Id.* at 572 (“The database industry has been lobbying Congress to strengthen database protection in the United States.”).

125. 17 U.S.C. § 102.

126. See *Torah Soft, Ltd. v. Drosnin*, 136 F. Supp. 2d 276, 284 (S.D.N.Y. 2001) (“It is well-established that computer programs are protected by copyright law as *literary works*.”) (emphasis added) (citing *Computer Assoc. Int’l, Inc. v. Altai, Inc.*, 982 F.2d 693, 702 (2d Cir. 1992); *Whelan Assoc., Inc. v. Jaslow Dental Lab., Inc.*, 797 F.2d 1222, 1233 (3d Cir. 1986) (citing *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852, 855 n.3 (2d Cir. 1982) (extending copyright protection to source code); *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1246-47 (3d Cir. 1983) (extending copyright protection to source and object code), *cert. dismissed*, 464 U.S. 1033 (1984); *Williams Elecs., Inc. v. Artic Int’l, Inc.* 685 F.2d 870, 876-77 (3d Cir. 1982) (extending copyright protection to object code)).

127. Computers respond to instructions embodied in “object code,” which is a binary language consisting of “0’s” and “1’s.” LEMLEY, *supra* note 14, at 85. However, because it is difficult for a programmer to write a program in object code, programmers rely on intermediate languages called “source code,” which is more akin to a written language with word instructions. *Id.* The computer “compiles” the source code into object code to obtain its binary instructions. *Id.*

128. See *id.*

129. See cases cited *supra* note 126.

130. *Lotus Dev. Corp. v. Borland Int’l*, 49 F.3d 807, 815 (1st Cir. 1995) (holding that Lotus’ computer menu command hierarchy consisted of a “method of operation,” and as such, it was not subject to copyright protection).

131. See cases cited *supra* note 126.

gram.¹³² For programs like BLAST[®], which searches a database of biological molecules to find those similar to a particular molecule,¹³³ copyright protection might not extend to the functional elements, such as BLAST's search and comparison method.

Copyright protection may be available for a bioinformatic machine or apparatus,¹³⁴ but the protection would extend only to the aesthetic, non-functional elements. For example, in *Carol Barnhart Inc. v. Economy Cover Corp.*,¹³⁵ the court held that certain elements of mannequin display forms could be copyrightable, but because the forms were functional, the functional and nonfunctional elements first needed to be "conceptually separable."¹³⁶ After conceptual separation, only the nonfunctional elements could be copyrightable.¹³⁷ For bioinformatic machines or apparatuses, most of their commercial value lies within their functional elements and not their aesthetic qualities. Thus, copyright protection may be inapplicable to bioinformatic machines or apparatuses.

One particular bioinformatic apparatus, the "gene chip," may qualify for *sui generis* protection under the Semiconductor Chip Protection Act ("SCPA").¹³⁸ The SCPA borrows concepts from the Patent Act¹³⁹ and the Copyright Act¹⁴⁰ and protects a semiconductor chip where it contains an original "mask work."¹⁴¹ "Mask work" refers to the layers of a chip that are built up by deposition and etching to create the functional chip,¹⁴² so in

132. See *Lotus*, 49 F.3d 807, 815 (arranging the code in a particular manner, *i.e.*, hierarchical structure, might be described as a patentable "method.").

133. See *supra* note 51 (describing BLAST[®], its principles, and its algorithm).

134. See *Mazer v. Stein*, 347 U.S. 201, 214-15 (1954) (holding that a sculptural lamp base could be copyrighted).

135. 773 F.2d 411, 415 (2d. Cir. 1985). Copyrightable elements might reside in the aesthetic features but not in the functional features of the mannequins.

136. See *id.* at 414.

137. See *id.* at 418. Even after determining that the elements are copyrightable subject matter, the elements would also have to be an original form of expression. 17 U.S.C. § 102.

138. 17 U.S.C. §§ 901-914 (2002).

139. Like patent protection, protection under the SCPA is for a short, finite term, *i.e.*, 10 years. 17 U.S.C. § 904(b).

140. Compare 17 U.S.C. § 102(b), with 17 U.S.C. § 902(c) (similarly limiting the scope of protection). See also text accompanying *infra* note 152.

141. 17 U.S.C. § 902(b) ("Protection under this chapter shall not be available for a mask work that . . . is not original.").

142. 17 U.S.C. § 901(a)(2) defines a "mask work" as "a series of related images, however fixed or encoded—(A) having or representing the predetermined, three-dimensional pattern of metallic, insulating, or semiconductor material present or removed from the layers of a

some ways a “mask work” may be considered a “creative work.” While the traditional idea of a gene chip is a microarray of DNA molecules imbedded or immobilized on a solid substrate, and not necessarily a semiconductor chip, recently developed gene chips do incorporate DNA onto a semiconducting chip.¹⁴³ If such a gene chip contains an original “mask work,” the chip may be eligible for protection under the SCPA.¹⁴⁴ However, like copyright, the protection afforded to any mask work does not “extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which in which it is described, explained, illustrated, or embodied in such work.”¹⁴⁵ This provision significantly limits the scope of protection under the SCPA and may preclude practical applicability of the SCPA to gene chips, where the value of such a chip likely resides in its “method of operation” and not its “mask work”—if it contains a “mask work” at all.¹⁴⁶ Therefore the SCPA may not afford significant protection to gene chips.

V. TRADE SECRET PROTECTION: ELIGIBLE SUBJECT MATTER MUST BE SOMETHING OF VALUE KEPT CONFIDENTIAL

Where federal patent or copyright protection is unavailable, state trade secret law may provide protection for bioinformatic components. Trade secret protection derives from the common law of tort,¹⁴⁷ but most states have enacted the Uniform Trade Secrets Act (“UTSA”) in some form.¹⁴⁸ The UTSA defines a “trade secret” as:

semiconductor chip product; and (B) in which series the relation of the images to one another is that each image has the pattern of the surface of one form of the semiconductor chip product”

Id.

143. See, e.g., J.P. Cloarec et al., *Immobilization of Homooligonucleotide Probe Layers onto Si/SiO₂ Substrates: Characterization by Electrochemical Impedance Measurements and Radiolabelling*, 17(5) *BIOSENSORS & BIOELECTRONICS* 405-12 (May 2002); <http://linkage.rockefeller.edu/wli/microarray/> (discussing microarray); Gene Chip, *supra* note 57.

144. 17 U.S.C. § 902.

145. See *supra* note 140.

146. See Gene Chip, *supra* note 57, Cloarac, *supra* note 143.

147. LEMLEY, *supra* note 14, at 50.

148. A recent survey notes that 44 of the 50 states had enacted some form of trade secrets law. See Andrew Beckerman-Rodau, *Trade Secrets—The New Risks to Trade Secrets Posed by Computerization*, 28 *RUTGERS COMPUTER & TECH. L.J.* 227, 230-33; Uniform Law Commissioners, Uniform Trade Secrets Act (“UTSA”), at http://www.nccusl.org/nccusl/uniformact_why/uniformacts-why-utsa.asp (visited Nov. 3,

information, including . . . a compilation, program, device, method . . . that:

- (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and
- (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.¹⁴⁹

As such, if a bioinformatic component, such as a sequence, database, software or hardware can derive “independent economic value . . . from not being generally known,” it can qualify for trade secret protection.¹⁵⁰ Trade secret protection offers a distinct advantage over patent or copyright protection because the protection is for a potentially infinite term. However, trade secret protection exists only as long as the subject matter remains “secret.”¹⁵¹ A confidentiality agreement can be used to prevent the contracting party from disclosing the trade secret, but if breached, it cannot be used to “regain” a trade secret that is released into the public domain; *i.e.*, a plaintiff could recover damages for breach of contract,¹⁵² but the trade secret, once exposed to the public, is lost forever.¹⁵³ Similarly, trade secret protection does not prevent independent creation¹⁵⁴ or (perhaps more importantly) “reverse engineering,”¹⁵⁵ and like confidentiality agreements, contracts that prohibit licensees from reverse engineering may be futile because of the inability to “regain secrecy” in the event of breach.

2002). Even where a state has not enacted statutory protection, common law protection may be available.

149. The National Conference of Commissioners on Uniform State Laws approved its final draft of the Uniform Trade Secrets Act in 1985, *available at* <http://www.law.upenn.edu/bll/ulc/fnact99/1980s/utsa85.html>.

150. *Id.*

151. *See* UTSA *supra* note 149; MILGRIM ON TRADE SECRETS § 1.01. (To remain a trade secret, the subject matter must not become “generally known.”).

152. “A suit to redress theft of trade secret is grounded in tort damages for breach of a contract” *Kewanee Oil, Co. v. Bicron Corp.*, 416 U.S. 470, 498 (1974) (Douglas, J., dissenting).

153. *Bonito Boats v. Thunder Craft Boats*, 489 U.S. 141, 155 (1989) (“[T]he policy that matter once in the public domain must remain in the public domain is not incompatible with the existence of trade secret protection.”) (citing *Kewanee Oil*, 416 U.S. at 484).

154. *See Kewanee Oil*, 416 U.S. at 490; *Bonito Boats*, 489 U.S. at 160 (citing *Kewanee Oil*, 416 U.S. at 476).

155. *See Kewanee Oil*, 416 U.S. at 490. For example, if a scientist commercializes a product containing an embodiment of the trade secret, the scientist cannot prevent one from purchasing the product, discovering the trade secret therein by reverse engineering, and subsequently releasing the trade secret into the public domain.

Therefore, the feasibility of trade secret protection for bioinformatic components may depend on the ease with which they can be reverse engineered.

A. Trade Secret Protection for DNA, RNA, and Protein Sequences

Trade secret protection may be the only viable form of IP protection available for biological sequences.¹⁵⁶ However, trade secret protection may also be impractical. First, it is relatively easy to determine the sequence of a biological composition, so others could independently obtain the sequence information if the biological composition is made readily available. Likewise, trade secret protection of a discovered function encoded in the biological sequence information might be equally futile if the owner intends to market and capitalize on that very function he is trying to protect. For example, assume that a biotechnology company discovers that a particular biological sequence is predictive for a particular disease, and the company develops a corresponding diagnostic test. To establish the validity of its testing services, the company would probably have to submit its test to some form of “peer review.” However, by doing so, it might lose its trade secret protection because validation usually entails general dissemination and widespread acceptance,¹⁵⁷ and as such, the company might have to reveal the basis of its test. While the company might find a small group of experts willing to validate its test under a confidentiality agreement, it may be difficult to market a test for which the validity is not *generally and widely accepted*. Therefore, trade secret protection may be impractical to protect biological sequences or their encoded functions if the scientist seeks commercialization thereof.

B. Trade Secret Protection for Biological Databases

Trade secret protection is also available for databases if the database can be shown to derive “independent economic value . . . from not being generally known.”¹⁵⁸ If the owner of a database wishes to commercialize it by selling access or even the database itself, the creator runs the risk that the information within the database will be disclosed and released to the public domain. To avoid such a risk, the owner might engineer or acquire security devices that allow access to the database without revealing the

156. See *supra* Parts III.A and IV.A (discussing unlikely protection of biological sequences under patent and copyright law, respectively).

157. Some “disclosure” is permitted under trade secret law, but the subject matter of the trade secret must not become “generally known.” See MILGRIM ON TRADE SECRETS § 1.01.

158. See LEMLEY, *supra* note 14, at 52.

entire contents.¹⁵⁹ Nevertheless, these devices can be circumvented and the database content released into the public domain, thereby forever destroying the trade secret status of the information. Notwithstanding such risks, some databases owners have attempted to “exploit [their] databases commercially by controlling access to them, in effect using contracts and trade secrecy to protect their intellectual property.”¹⁶⁰ Even where database owners have controlled access and secrecy through contracts, *third party release* of independently acquired information into the public domain has hampered efforts to commercialize these databases.¹⁶¹ For instance, as part of its policy, the Wellcome Trust, called the “world’s largest medical charity,”¹⁶² releases the DNA sequence information that it gathers from the human genome into the public domain.¹⁶³ Likewise, pharmaceutical giant Merck sponsored human DNA sequencing research by Washington University for “instantaneous dedication [of the results] to the public domain.”¹⁶⁴ This policy increases the amount of information that is *freely available* and, therefore, may diminish the value of fee-based databases.¹⁶⁵ Merck nonetheless believes that release of such information into the public domain will benefit its own development efforts in the long run.¹⁶⁶ Data released by the Wellcome Trust or companies like Merck may be incorporated into the free databases offered by the NIH.¹⁶⁷ Therefore, if the owner of a database wishes to maintain the database as a trade secret, the owner must protect against not only unlicensed access, but also erosion of the database’s value through third party disclosures and the growth in the number of free databases.

159. For example, companies that offer on line databases typically require that subscribers use passwords, and subscribers may have limited access based on the subscription agreement.

160. Rebecca S. Eisenberg, *Intellectual Property at the Public-Private Divide: The Case of Large-Scale cDNA Sequencing*, 3 U. CHI. L. SCH. ROUNDTABLE 557, 563 (1996).

161. *Id.* at 570 (describing Merck’s collaboration with Washington University to release data to the public domain). Some observers suggest cynically that Merck’s goal is “to undermine the value of investments already made in existing sequence databases by its commercial competitors.” *Id.* See Alexander K. Haas, *The Wellcome Trust’s Disclosures of Gene Sequence Data into the Public Domain & the Potential for Proprietary Rights in the Human Genome*, 16 BERKELEY TECH. L.J. 145, 152 (2001) (describing the Wellcome Trust’s release of biological information into the public domain).

162. Haas, *supra* note 161, at 151.

163. *Id.* at 152.

164. Eisenberg, *supra* note 160, at 559.

165. *Id.* at 564.

166. *Id.* at 570. Because Merck does not have the resources to investigate every biological sequence that it discovers, it has chosen to release the sequence into the public domain, hoping to “capture an adequate share of [the] resulting products.” *Id.*

167. See *supra* note 44 (listing some of the databases offered by the NIH).

C. Trade Secret Protection for Bioinformatic Software and Hardware

Trade secret protection is available for computer software, and bioinformatic software is no exception.¹⁶⁸ Many software developers maintain the source code of their programs as a trade secret, releasing only the object code for sale or license.¹⁶⁹ However, the software developer still runs the risk of disclosure by reverse engineering if the object code is decompiled into source code.¹⁷⁰ Again, the use of security devices or contracts to prevent reverse engineering is insufficient if the devices are circumvented or the contracts breached. Despite these risks, developers have utilized trade secret protection effectively, where decompiling is difficult and produces errors.¹⁷¹

Trade secret protection is also available for a bioinformatics machine or apparatus. However, if the owner intends to sell the machine or apparatus, the risk of disclosure is very high because machines and apparatuses, once freely distributed and in “plain view,” can be reverse engineered by disassembling them and determining how they function.¹⁷²

VI. ARGUMENTS AGAINST IP PROTECTION FOR BIOINFORMATIC COMPONENTS

Even where IP protection is available and practical for bioinformatic components, some argue that bioinformatic components should be excluded from IP protection for policy reasons. For instance, some argue against IP protection for bioinformatics because they believe that the human genome belongs to everyone and should not be kept as a property

168. *But see infra* Part VI for a discussion of the “Open Informatics” petition which argues for free licenses for bioinformatic software.

169. LEMLEY, *supra* note 14, at 61-62 (“[S]oftware developers generally distribute their programs only in object code form and keep the source code . . . as [a] trade secret[], licensing [it] only rarely and only under agreements of confidentiality.”). For an example of a Microsoft licensing agreement see *Microsoft Corp. v. Commissioner*, 115 T.C. 228, 235-38 (2000). *See also* <http://www.compaq.com/Cas-Catalog/das055hm.html> (exemplifying a Compaq license agreement for its “trade secret diagnostic software.” “Source code . . . is not sold, licensed, nor otherwise distributed without prior approval . . . Object code, in binary form, . . . is available for sale or license.”).

170. “Decompiling” involves “translat[ing] the 1s and 0s into some form of assembly language and then into readable source code.” LEMLEY, *supra* note 14, at 85.

171. *Id.* (citing Andrew Johnson-Laird, *Reverse Engineering of Software: Separating Legal Mythology from Actual Technology*, 5 SOFTWARE L.J. 331, 342-43 (1992)).

172. Patent protection is probably more suitable for bioinformatics machines and apparatuses. *See supra* Part III.C.

right.¹⁷³ As previously noted, many of these arguments incorrectly equate the human genome *sequence* with the human genome *composition*. Others argue against IP protection for bioinformatics because it relates to human medicine. For example, some commentators argue against patenting medical procedures,¹⁷⁴ *i.e.*, patents on medical procedures hinder medical research where the patentee excludes others from practicing the patented procedure. By analogy, some may argue that patents on bioinformatic components hinder medical research where bioinformatic components released into the public domain would advance human medicine more rapidly. However, these arguments do not address the need to create incentives for biological sequence research and development.¹⁷⁵ The USPTO¹⁷⁶ and many patent scholars¹⁷⁷ argue that *patents spur invention* and that it is wrong “to single out any area of subject matter and deny rewards for creativity in that area.”¹⁷⁸ Furthermore, patent protection encourages the *full disclosure* of bioinformatics components,¹⁷⁹ which *promotes* progress in medical research.

173. *See supra* note 12.

174. *See, e.g.*, Linda Rabin Judge, *Comment: Issues Surrounding the Patenting of Medical Procedures*, 13 *COMPUTER & HIGH TECH. L.J.* 181, 194 (1997). Compare Wendy W. Yang, *Note: Patent Policy and Medical Procedure Patents: The Case for Statutory Exclusion from Patentability*, 1 *B.U. J. SCI. & TECH. L.* 5 (1995), with Joel Garriss, *Note and Comment: The Case for Patenting Medical Procedures*, 22 *AM. J.L. & MED.* 85 (1996). These articles were written in response to H.R. 1127, 104th Cong. (1995) and S. REP. NO. 1334 104th Cong. (1995).

175. *See* DONALD S. CHISUM ET AL., *PRINCIPLES OF PATENT LAW* 62-67 (1998) (describing four policy goals of our patent system including (1) to create an “incentive to invent,” (2) to create an “incentive for full disclosure,” (3) to create an “incentive to commercialize,” and (4) to create an “incentive to design around”). *See also* 66 Fed. Reg. 1092, 1094 (Jan. 5, 2001) (stating that “[t]he incentive to make discoveries and inventions is generally *spurred* . . . by patents.”) (emphasis added).

176. *See* 66 Fed. Reg. 1092, 1092-97 (Jan. 5, 2001) (USTPO, addressing arguments raised by opponents of DNA patents).

177. For example, testifying against the banning of patent protection for medical procedures, Donald R. Dunner, former Chairman of the Intellectual Property Law Section of the American Bar Association, stated that the goal of the patent system is to provide incentives for innovation for “any and all subject matter.” Linda Rabin Judge, *supra* note 174.

178. *Id.* (quoting Donald R. Dunner, former Chairman of the Intellectual Property Law Section of the American Bar Association).

179. In order to obtain a patent, the applicant must fulfill a disclosure requirement. *See* 35 U.S.C. § 112 (discussing the disclosure requirements for obtaining a patent). *See also* M. Scott McBride, *Note: Patentability of Human Genes: Our Patent System Can Address the Issues Without Modification*, 85 *MARQ. L. REV.* 511, 527-28 (2001) (describing how our patent system encourages full disclosure of DNA sequences in regard to patent applications for genes). Even though the patentee has the right to exclude others from

Copyright protection would hinder medical research only to a limited degree because it would not extend to a component's *function*.¹⁸⁰ However, because copyright protection is available for the literal code of bioinformatic software, opponents of copyright protection for publicly-funded bioinformatic software have recently argued for "open source licenses" as further discussed below.¹⁸¹

With regard to trade secret protection, any arguments against protection for bioinformatic components are essentially arguments for forced disclosure, which for privately-funded entities would be unworkable and unconstitutional.¹⁸² For publicly-funded entities, disclosure is encouraged, and the Bayh-Dole Act¹⁸³ and current NIH guidelines preclude an NIH-funded scientist from keeping bioinformatics information as a trade secret as further discussed below.¹⁸⁴

Others argue against IP protection of bioinformatics components because their discovery is publicly funded and thus belong to the public.¹⁸⁵ For patents, this persuasive argument has been largely muted by enactment of the Bayh-Dole Act.¹⁸⁶ The Bayh-Dole Act amended the Patent

practicing his claimed invention, see 35 U.S.C. § 271, the "right to exclude" is for a limited period of time. See 35 U.S.C. § 154 (describing the finite length of a patent term). In exchange, society receives full disclosure of the claimed invention. See 35 U.S.C. § 112. See also Rebecca S. Eisenberg, *Proprietary Rights and the Norms of Science in Biotechnology Research*, 97 YALE L.J. 177, 181-84, 197-205, 207-17 (1987) (discussing disclosure in regard to biotechnology).

180. See *supra* notes 134-137 and accompanying text. The *functional elements*, not the aesthetic or literal elements, typically drive medical advances. For example, most would agree that the functional elements of a magnetic resonance imaging machine ("MRI") are more important than its aesthetic qualities.

181. See Jason E. Stewart & Harry Mangalam, *The Open Informatics Petition*, O'REILLY NETWORK (Jan. 14, 2002), available at <http://www.oreillynet.com/pub/a/network/2002/01/11/openinfo.html> (visited Nov. 3, 2002).

182. For example, the inventor, creator, or discoverer could simply keep his invention, creation, or discovery secret, and any attempt to force disclosure would violate the U.S. Constitution, *e.g.*, the First Amendment's right to free speech. U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech . . ."). See *Bartnicki v. Vopper*, 532 U.S. 514, 533 n.20 (2000) (citations omitted). A forced disclosure might also violate the Fifth Amendment's prohibition against "takings." U.S. CONST. amend. V ("[N]or shall private property be taken for public use, without just compensation."). Where a trade secret is "property," forced disclosure might constitute a "taking." *Id.*

183. 35 U.S.C. §§ 200-212 (1998).

184. See *infra* notes 210-220 and accompanying text.

185. For example, grants from the National Institutes of Health ("NIH") or the National Science Foundation ("NSF") often fund research that may lead to the development of bioinformatics components.

186. 35 U.S.C. §§ 200-212 (1998).

Act to indicate that inventors¹⁸⁷ are entitled to their inventions, even if the invention was funded by public sources from a federal agency.¹⁸⁸ However, the Bayh-Dole Act also provides that the federal government has “march-in” rights in limited circumstances,¹⁸⁹ although the government rarely exercises these rights.¹⁹⁰ In order to secure rights in their inventions, the inventors must “disclose each subject invention to the [funding] Federal agency within a reasonable time,”¹⁹¹ although the inventors may ultimately choose not to pursue patent protection. This “disclosure requirement” supplements the disclosure requirements of 35 U.S.C. § 112, which an applicant must satisfy before receiving a patent. Therefore, in enacting the Bayh-Dole Act, Congress effectively addressed the argument against patent protection for publicly-funded research by permitting protection in exchange for further disclosure.¹⁹²

However, the Bayh-Dole Act applies to “inventions,” and therefore does not apply to copyright law,¹⁹³ which (as noted) is applicable to some bioinformatic components such as literal elements of software.¹⁹⁴ Recently, software developers circulated a petition that required the free licensing of bioinformatic software developed with public funds.¹⁹⁵ These developers and others signing the petition “believe that publicly funded

187. “Inventors” includes “small business firm[s] [and] nonprofit organizations.” 35 U.S.C. § 202(a) (2002). However, the GAO asserts that “inventors” was extended to include large businesses under Exec. Order No. 12,591, 52 Fed. Reg. 13414 (Apr. 10, 1987). See Peter S. Arno & Michael H. Davis, *Why Don't We Enforce Existing Drug Price Controls? The Unrecognized and Unenforced Reasonable Pricing Requirements Imposed upon Patents Deriving in Whole or in Part from Federally Funded Research*, 75 TUL. L. REV. 631, 642 n. 60 (2001).

188. 35 U.S.C. § 201(b) (1998) (defining “funding agreement” as an agreement between any Federal Agency and any contractor).

189. 35 U.S.C. § 203

190. “‘The Government does not use its march-in rights one in a million times . . . I think that is a paper tiger. I think we can forget [march-in rights] as a realistic protection for the public.’” Arno & Davis, *supra* note 187, at 658 (quoting Representative Jack Brooks, “perhaps the harshest critic of the proposed legislation,” during congressional hearings). “Brooks’s statement proved to be prophetic—the NIH has never exercised its march-in rights.” *Id.*

191. 35 U.S.C. § 202(c)(1) (1998).

192. *Id.*

193. 35 U.S.C. § 201(d) (“The term ‘invention’ means any invention or discovery which is or may be patentable [under the Patent Act]”).

194. See *supra* Part III.C.

195. See *The Open Source Definition*, at <http://www.opensource.org/docs/definition.html> (visited May 5, 2002). The circulation of this petition was initiated by developers Jason Stewart, Harry Mangalam, and Jiaye Zhou. *Id.*

research should be made available to all.”¹⁹⁶ The “Open Informatics” petition, as it is called, “would require that publicly funded researchers publish any source code under an open source or free software license.”¹⁹⁷ The purported advantages to such a policy include:

- Promoting cooperation between academic and commercial organizations;
- Promoting standardization;
- Promoting peer-review of software to allow “bugs . . . to be found and corrected”; and
- Promoting a mechanism for more rapid improvement and development of code.¹⁹⁸

Andrew Dalke, co-founder of the Biopython Project, “an international association of developers of *freely available* . . . tools for computational molecular biology,”¹⁹⁹ opposes the “Open Informatics” petition.²⁰⁰ First, Dalke states that an “open source” policy would not promote cooperation between academic and commercial organizations.²⁰¹ Indeed, an academic scientist under an “open source” obligation may not be able to work with commercial organizations that require license agreements that are apposite to the “open source” policy.²⁰² Further, the petition may be “dead on arri-

196. David Malakoff, *Petition Seeks Public Sharing of Code*, 294 SCIENCE 27 (Oct. 5, 2001).

197. Andrew Dalke, *Why I'm not Supporting the Open Informatics Petition*, O'REILLY NETWORK, Jan. 14, 2002, at <http://www.oreillynet.com/pub/a/network/2002/01/12/dalke.html> (visited Oct. 13, 2002).

198. Jason E. Stewart & Harry Mangalam, *The Open Informatics Petition*, O'REILLY NETWORK, Jan. 14, 2002, at <http://www.oreillynet.com/pub/a/network/2002/01/11/openinfo.html> (visited Oct. 13, 2002).

199. <http://www.biopython.org> (visited Oct. 13, 2002).

200. Dalke, *supra* note 197.

201. *Id.*

202. *Id.* See also Bernadette Toner, *Legal Pitfalls of Free Bioinformatics Software May Loom Large*, GENOME WEB NEWS, Aug. 17, 2001, at <http://www.genomeweb.com/articles/view-articles.asp?Article=200181784719> (discussing the case of Steve Brenner, assistant professor of computational genomics research at the University of California, Berkeley). Because Dr. Brenner's work on open source was incompatible with the university's default software license, Dr. Brenner had to request a formal variance to continue his work.

val”²⁰³ because funded entities have proprietary rights to their inventions under the Bayh-Dole Act and cannot be made to “release software under [free] licenses that contradict federal law.”²⁰⁴ Second, while supporters argue that the petition will promote standardization, Dalke disagrees that standardization is desirable in all instances.²⁰⁵ For example, in some instances, verifying one scientist’s results by using different bioinformatic software might bolster the scientist’s results. Third, Dalke disagrees with equating “open source” to “peer review.”²⁰⁶ Although a policy of “open source” might facilitate detection and correction of “bugs,” it does not necessarily follow that the author of the code should be denied IP protection. For instance, Dalke argues that we allow copyright protection and peer review of scientific papers, and by analogy that he should not be “allowed to take a peer-reviewed [scientific] paper, modify a paragraph, and republish it.”²⁰⁷ This would violate the author’s copyright,²⁰⁸ and yet this is exactly what the proponents of “open source” would permit in regard to bioinformatics software. Finally, Dalke disagrees that “open source” would promote improvement and development, and he cites article I, section 8, clause 8 of the U.S. Constitution for the argument that IP protection “promote[s] the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”²⁰⁹ Proponents of “open source” should note that while “open source” may promote improvement and development through widespread analysis and constructive criticism, “open source” would remove the incentive to create where the author of the code could not obtain a proprietary right in his work.

Even in the absence of an “open source” requirement, publicly-funded entities are encouraged to disclose their work not only under Bayh-Dole²¹⁰ but also under NIH policy. It is the NIH’s stated goal to “promote free dissemination of research tools without legal agreements whenever possi-

203. Justin Hibbard, *The Open-Source Debate Enters the Genomics Arena*, RED-HERRING, Feb. 25, 2002, available at <http://www.redherring.com/insider/2002/0225/1805.html>.

204. *Id.*

205. Dalke, *supra* note 195.

206. *Id.*

207. *Id.*

208. *Id.*

209. U.S. CONST. art. I, § 8, cl. 8.

210. In fact, if the publicly-funded entity seeks patent protection, the entity must disclose its finding to the Federal agency which funded its work within a reasonable amount of time. 35 U.S.C. § 202(c)(1).

ble.”²¹¹ A “research tool . . . includes[s] DNA sequences [and] databases,”²¹² and scientists are expected to make “intellectual property, such as computer programs,” accessible as well.²¹³ Further, the NIH is developing a new policy on data sharing and has requested comments on its Draft Statement of Sharing Research Data.²¹⁴ Under the proposed policy, researchers who “submit . . . an NIH [grant] application will be required to include a plan for data sharing or to state why data sharing is not possible.”²¹⁵ The recipient of a grant will be subject to NIH policy as a condition of the grant.²¹⁶ Therefore, under current and proposed guidelines a scientist would violate NIH policy by attempting to keep as a trade secret any sequence or database developed with an NIH grant.²¹⁷ However, while the current NIH guidelines may prohibit a researcher from maintaining “research tools” as a trade secret, the guidelines do not prevent the researcher from obtaining patent protection²¹⁸ or copyright protection²¹⁹

211. <http://www.nih.gov/news/researchtools/index.html>. See also 64 Fed. Reg. 72,090 (Dec. 23, 1999) (Department of Health and Human Services, Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice), available at http://ott.od.nih.gov/NewPages/Rtguide_final.html (visited May 5, 2002) (responding to comments submitted in regard to the NIH’s then proposed policy on Sharing Biomedical Research Resources)[hereinafter NIH Dissemination Policy].

212. NIH Dissemination Policy, *supra* note 211, at 72092 n.1.

213. NIH Grants Policy Statement (Mar. 2001) at 121. http://grants.nih.gov/grants/policy/nihgps_2001/nihgps_2001.pdf [hereinafter NIHGPS].

214. National Institutes of Health, Office of Extramural Research, NIH Draft Statement of Sharing Research Data, available at http://grants1.nih.gov/grants/policy/data_sharing/index.html (visited May 5, 2002)[hereinafter NIH Draft Statement].

215. *Id.* Unfortunately, because researchers are often judged by their publication record—*i.e.*, “publish or perish”—some are loathe to share their research data for fear that a competitor may “scoop” them. The NIH Draft Statement does not describe acceptable circumstances where data sharing might not be possible, but presumable, the risk of “being scooped” will not be a permissible circumstance.

216. National Institutes of Health, Office of Extramural Research, Award Conditions and Information for NIH Grants, at <http://grants1.nih.gov/grants/policy/awardconditions.html> (visited May 5, 2002).

217. NIHGPS, *supra* note 213, at 121, “Investigators are expected to submit unique biological information to the appropriate data banks.” Further, the NIHGS has enforcement provisions, and noncompliance with the terms and conditions of an award may result in “[s]uspension, [t]ermination, or [w]ithholding of [s]upport.” *Id.* at 144-45.

218. See Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice, 64 Fed. Reg. 72090 (Dec. 23, 1999), available at http://ott.od.nih.gov/NewPages/RTguide_final.html “[W]here patent protection is necessary for development of a research tool as a potential product for sale and distribution to the research community, Recipients are not discouraged from seeking such protection, but should license the intel-

where available. While the NIH promotes accessibility, the NIH leaves it up to “*recipients* to determine the appropriate means of effecting prompt and effective access to research tools,”²²⁰ and as such, the NIH is not necessarily advocating “free licenses” as proposed by the “Open Informatics” petition.

In summary, it appears that our laws seek to strike a balance, *i.e.*, promote full disclosure and cooperation to encourage development, but permit IP protection to create incentives for development as well. It remains to be seen whether opponents of IP protection for bioinformatics will tip the scale in their direction, *i.e.*, to require full disclosure with no proprietary rights.

VII. CONCLUSION

Bioinformatics comprises a wide array of components, and it follows that a wide array of protection might be available, depending on the particular nature of the bioinformatic component and its intended use. Because of the tremendous growth and investment in the field of bioinformatics, it is important to consider whether IP protection is available to offset the cost of development.

With regard to biological sequences, trade secret protection may be the only practical protection. This holds best where the owner effectively maintains confidentiality agreements or does not intend to commercialize the corresponding biological composition, because sequences can be easily determined or “reverse engineered” where compositions are available.

Likewise, trade secret protection may provide the best protection for biological databases, but only if adequate security measures can reliably limit access and the owner effectively maintains confidentiality agreements. Copyright protection for databases is minimal and is unlikely to extend to the information contained within the database.

With regard to bioinformatic software, the inventor can obtain patent protection on the method within the program, provided the method produces tangible results; and the author can obtain copyright protection, but only for the literal elements of the bioinformatic software code. Although

lectual property in a manner that maximizes the potential for broad distribution of the research tool.”).

219. NIHGPS, *supra* note 213, at 119 (“Except as otherwise provided in the terms and conditions of the award, the grantee is free to copyright without NIH approval when publications, data, or other copyrightable works are developed under, or in the course of, work under an NIH grant.”).

220. *Id.* at 121.

trade secret protection is available for bioinformatic software, again, like many bioinformatic components, the owner runs the risk that the code will be reverse engineered and the trade secret will be lost to the public domain.