

The Use of Computational Methods to Describe and Establish Utility of a DNA Sequence for Purposes of Patenting

Tamara Fraizer

- I. INTRODUCTION
- II. THE LEGAL AND POLITICAL CONTEXT
 - A. Patent Law and Gene Patents
 - 1. Some Basic Tenets of Patent Law
 - 2. The Patentability of Genes
 - B. The Written Description Issue: Inferring Structure from Function
 - 1. The Court's Interpretation of Gene Descriptions
 - 2. The Public Debate about Treating DNA as a Chemical Compound
 - 3. The PTO's Guidelines for Examination of the Written Description
 - C. The Politics of EST Patents
 - D. The Utility Issue: Inferring Function from Structure
 - 1. The Courts' Interpretation of Utility in the Chemical Arts
 - 2. The PTO's Guidelines for Examination of Utility
 - 3. The Public Debate about Using Genomics to Establish Utility
- III. THE USE OF COMPUTATIONAL TECHNIQUES IN GENE PATENTS
 - A. Finding EST Patents
 - 1. Methods for Searching the Databases
 - 2. Quantitative Search Results
 - C. Satisfying the Written Description Requirement:
 - 1. Synopsis of Legal Criteria
 - 2. Observed Uses of Computational Methods
 - a. Use of the Genetic Code and Combinatorics
 - b. Use of Percent Sequence Identity
 - c. Use of Structural Variants Having Similar Function
 - D. Satisfying the Utility Requirement:
 - 1. Synopsis of Legal Criteria
 - 2. Observed Uses of Computational Methods
 - a. To Identify the Polypeptide Encoded by a Sequence
 - b. To Show that the Polypeptide is Unknown
 - E. Discussion and Critique
 - 1. Satisfying the Written Description Requirement
 - 2. Satisfying the Utility Requirement
- V. CONCLUSION

I. INTRODUCTION

Computers and the internet have, in the matter of a few decades, changed the nature of personal communication, business, and scientific research. The creation of large gene and protein databases and the development of sophisticated methods for analyzing sequence data via the web have, for example, transformed certain aspects of molecular biology and genetics into the information sciences now known as genomics and bioinformatics. Indeed, the typical research biologist now combines work at the bench with work online, and knows both chemistry and computational methods. Meanwhile, companies such as Incyte and Celera are specializing in the production and analysis of genetic information, leaving other companies to pursue the development of particular pharmaceutical products.¹

The recent changes in methods of biological research and business create significant challenges for the definition and defense of intellectual property rights relating to genetic research.² Our legal system, an institution of resilience rather than reform, is adapting to the new world. Together, the Court of Appeals for the Federal Circuit³ (CAFC) and the United States Patent and Trademark Office (USPTO) are establishing the precedents and procedures needed to assess whether and how particular genetic discoveries can be patented. The process is slow and imperfect, though, and the pace of scientific advancement has made many of the CAFC's rulings appear inadequate if not obsolete. Nonetheless, the USPTO has responded in timely and pertinent ways, interpreting the CAFC's rulings in guidelines that are used by examiners in evaluating patent applications.

The CAFC and the USPTO are struggling most notably to adapt precedents and procedures to a fundamentally new type of invention: myriad isolated cDNA sequences whose functions are inferred from computational analysis of existing annotated databases of genetic sequences. Many people have argued that such inventions are merely "information about the natural world" and therefore should

¹ Randall Scott, President and Chief Scientific Officer of Incyte Genomics, described the new pharmaceutical industry as being vertically rather than horizontally integrated. Thus, instead of one company conducting all aspects of research and development, some companies provide the data needed for early stages of R&D while other companies direct the commercial development of particular products. Randall Scott, President and Chief Scientific Officer, Incyte Genomics. Prepared Statement at Hearing Before the House of Representatives Subcommittee on Courts and Intellectual Property of the Committee on the Judiciary, on Gene Patents And Other Genomic Inventions. 106th Congress, 2nd Session, July 13, 2000 [hereinafter Congressional Hearing on Genomic Inventions]. Interestingly, the agricultural industry appears to remain horizontally integrated, with most aspects of the industry dominated by companies such as DuPont and Pioneer.

² The National Academy of Science has acknowledged the significance of these challenges and is conducting a two-phased project study on "Intellectual Property Rights in the Knowledge-Based Economy". See <http://www4.nas.edu/cp.nsf>.

³ The CAFC was created in 1982 as a speciality court that would hear appeals from all the federal district courts involving patent issues. Many of the CAFC judges have technical backgrounds and all are more familiar with patent issues than the typical court of appeals judge. Thus, the creation of the CAFC has helped to create a systematic and sensible body of patent law.

not be patentable.⁴ Participants at the 1996 International Strategy Meeting on Human Genome Sequencing endorsed the idea that “all genomic DNA sequence information should be “freely available and in the public domain in order to encourage research and development and to maximize its benefit to society”.⁵

The ease with which researchers can now obtain cDNA sequences of unknown function, and compare them to sequences of known functions, stands in contrast to the state of the art in the 1980s, when researchers worked diligently to determine the actual nucleotide sequence for proteins of known function. These contrasting states of the science have raised two legal issues: (1) whether the invention, i.e. nucleotide sequence, is possessed by the inventor and adequately described, and (2) whether the invention, i.e. cDNA fragment, has a real world utility.

The issue associated with the earlier situation—i.e. patents claiming an unknown (but knowable) sequence of experimentally known function—has been addressed by the CAFC. In the early 1990s, the CAFC chose to assess DNA as it would any chemical compound. To claim a chemical compound as a composition of matter, the inventor must describe the compound’s structure. Therefore, the court found that describing a protein’s function and a method for isolating its DNA was not enough to claim the gene. Rather, the inventor had to describe the DNA, which was most obviously done by giving its nucleotide sequence. Recently, in January 2001, the USPTO published guidelines for assessing the adequacy of the description of inventions, consistent with the CAFC’s decisions, and applied them to contemporary scientific scenarios in associated but not yet revised training materials. XXX MORE?

The issue associated with the latter situation—known sequences with function inferred from the computational analysis of annotated databases—has not been addressed specifically by the CAFC. However, the USPTO announced in 1997 that it would allow claims on cDNA fragments or expressed sequence tags (ESTs) based on their utility as probes.⁶ In January 2001, after responding to considerable public debate about the matter, the USPTO published guidelines requiring a specific, substantial, and credible real world utility for every claimed invention. Associated but interim training materials provide examples of contemporary scenarios, including the use of computational analyses of annotated sequences to establish the utility of a claimed EST or cDNA fragment—so-called “genomic patents”. However, the guidelines emphasize that utility is evaluated on a case-by-case basis, according to scientific principles, and many remain skeptical of the validity of genomics patents.⁷

⁴ Antonio Regalado, *The Great Gene Grab*, 103 THE TECHNOLOGY REVIEW 48 (2000) (quoting Professor Rebecca Eisenberg).

⁵ David R. Bentley, *Genomic sequence information should be released immediately and freely in the public domain*, 274 SCIENCE 5287 (1996).

⁶ John Murray, *Owning Genes: Disputes Involving Dna Sequence Patents*, 75 CHI.-KENT. L. REV. 231, 239 (1999).

⁷ Arti K. Rai, *The Information Revolution Reaches Pharmaceuticals: Balancing Innovation Incentives, Cost, and Access In The Post-Genomics Era*, 2001 U. ILL. L. REV. 173, 194 fn100 (2001).

Computational methods are undoubtedly an essential and accepted tool in molecular biology. The patent office, moreover, has been evaluating patent applications that rely on computational methods to describe the claimed sequence and define its utility since at least 1998, and probably for as long as scientists have been using them. Many of these patents have now issued. Nonetheless, whether and how computational methods may be used to establish the patentability of a genetic sequence has not been addressed by the courts, and is not apparent in the legal or scientific literature.

In this paper, I review the law, politics, and administrative procedure relating to “genomic patents”; i.e. patents claiming gene sequences whose utility is based upon similarity to sequences of known function. I then review recently issued patents to assess whether or how computational methods are currently used to (1) describe the claimed gene or nucleotide sequence, and (2) establish the utility of an EST or cDNA fragment. I critique these current practices and respond to criticisms. I find that, in general, the patents are legally and scientifically sound; they may, however, be undesirable for social and political reasons.

II. THE LEGAL AND POLITICAL CONTEXT

To appreciate the use of computational methods in describing and defining the utility of EST patents, some background is necessary. In this section, I provide a simplistic account of the relevant features of patent law and explain why and how genes are patentable. I then consider the written description issue, reviewing the CAFC’s assessment of the written description as it applies to gene patents, considering the public’s reaction to the ruling that genes are chemical compounds and must be described (preferably by sequence), and summarizing the USPTO’s efforts to summarize, update, and implement the law in its *Written Description Guidelines*. To provide context for the debate about patenting ESTs, I next discuss some politics and history. Finally, I consider the utility issue, reexamining a single but important court case, summarizing the USPTO’s new *Utility Guidelines*, and noting the public’s reaction and predictions about the use of computational methods to define the utility of ESTs.

A. Patent Law and Gene Patents

Patents are issued by the USPTO in accordance with the Patent Statute of 1952 and the courts’ interpretations of that statute. An isolated gene sequence is suitable subject matter for a patent, and may be claimed as a “composition of matter.” I review the basics of patent law and the logic for patenting gene sequences here.

1. Some Basic Tenets of Patent Law

A patent confers intellectual property rights on an inventor, giving the inventor the right to exclude others from making, using, or selling the claimed

invention for a period of twenty years. Because a patent prevents others from capitalizing on the inventor's ingenuity and investment, it provides the inventor with an incentive to make and develop the invention. However, in order to obtain the temporary monopoly created by a patent, the inventor must disclose the invention. Thus, the patent also assures that new inventions are made available to the public.

The patent application and issued patent comprise a specification and claims. The specification provides background for the invention describes the invention in general and specific terms, and provides examples. It is the technical part of the patent and it tends to be very detailed and comprehensive. The claims are the legal part of the patent. They define the scope of the property claim, much as a surveyor's assessment defines the bounds of a land claim. They are carefully crafted in light of legal precedents and with reference to the invention as disclosed in the specification.

To obtain a patent, the inventor files an application (i.e. a specification and claims) with the United States Patent and Trademark Office (USPTO) and pays certain fees. The application is assessed by an examiner with technical knowledge of the field of the invention. The application must meet criteria established by Congress and clarified by the courts. If the examiner finds that the patent application meets all the applicable requirements, the patent will issue—typically about twenty-four months after the application was filed.

The criteria were established by Congress, acting under the explicit authority of the United States Constitution,⁸ in the Patent Act of 1952. Under this statute a patent may be obtained for any (1) process, (2) machine, (3) manufacture, or (4) composition of matter,⁹ so long as it satisfies the requirements of (a) utility, (b) novelty, (c) nonobviousness, and (d) description. That is, the invention must have real world utility;¹⁰ it must be novel or new¹¹ and nonobvious in light of the prior art;¹² and there must be a written description of the invention that shows the inventor's possession of the claimed invention and is sufficient to enable others to practice it.¹³

The patent system is neutral with respect to technology; that is, the same norms apply to all types of inventions. Nonetheless, the USPTO and the CAFC may determine how the general rules will apply to particular areas such as biotechnology and gene patents. It is not uncommon for the USPTO and the CAFC to differ in their interpretations of the statute. At least one scholar has

⁸ United States Constitution, Art. I, Sect. 8[8]. [The Congress shall have power] To promote 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.”

⁹ 35 U.S.C. §101 (1998). “Whoever invents any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.” *Id.*

¹⁰ *Id.*

¹¹ 35 U.S.C. §102.

¹² 35 U.S.C. §103.

¹³ 35 U.S.C. §112¶1. “The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use he same. . . .”

advocated that the CAFC defer to the informed and technically competent opinion of the USPTO,¹⁴ but the CAFC officially has the final word.¹⁵

The dynamic typically begins at the administrative level. The USPTO develops policies and sometimes publishes guidelines to be used by patent examiners in assessing patent applications. Both are based on the statute and the CAFC's previous decisions. An inventor may appeal a decision of the examiner to the Board of Patent Appeals and, thereafter, to the CAFC. Disputes arising over patent rights are also taken to a federal district court and, thereafter, to the CAFC. If the CAFC disagrees with the USPTO's decision, the USPTO must revise its policies so that they are in line with the views of the CAFC.

2. The Patentability of Genes

Many people object to the idea of gene patents, arguing that genes are natural and therefore should not be "owned" by anyone.¹⁶ Others object to the consequences of gene patents, arguing that restrictions on access to genetic tools will impede the progress of research.¹⁷ Some seek to limit gene patents to "process" rather than "composition of matter" claims. Irrespective of these public sentiments, policy concerns, and suggestions, isolated genes are simply not per se unpatentable, in any way. However, the information content of genes is probably unpatentable.

In 1980, in the seminal case of *Diamond v. Chakrabarty*, the United States Supreme Court found that genetically engineered bacteria were patentable.¹⁸ The Court cited the Congressional Report accompanying the 1952 Patent Act when it said that the subject matter of patents was meant to include "anything under the sun that is made by man".¹⁹ Thus, the key to the patentability of naturally occurring products of nature is human intervention. Genetic engineering had created organisms whose genomes were manipulated "by man"; therefore, those organisms were patentable.

Eleven years later, in the important case of *Amgen v. Chugai*, the CAFC established that it would treat DNA as a chemical compound: "A gene is a chemical compound, albeit a complex one."²⁰ Chemicals may be claimed as a

¹⁴ Arti K. Rai, *Intellectual Property Rights in Biotechnology: Addressing New Technology*, 34 WAKE FOREST L. REV. 827 (1999).

¹⁵ Rarely, intellectual property cases may be appealed to the United State Supreme Court, whose opinion trumps the opinion of the CAFC. Moreover, the Supreme Court ruled on several important issues in patent law prior to the creation of the CAFC in 1982. In some cases, the Supreme Court and the CAFC have held distinctly different opinions and have ignored the previous decisions of the other court.

¹⁶ See, e.g., Mark Christopher Farrell, *Designer DNA for Humans: Biotech Patent Law Made Interesting for the Average Lawyer*, 35 GONZ. L. REV. 515, 529 (1999/2000) (asserting the common view that "[l]egal protection for the mere discovery of a genetic code sequence already existing in nature seems incorrect."); Murray, *supra* note 6 (providing a general review of gene patenting controversies).

¹⁷ GET CITATION.

¹⁸ *Diamond v. Chakrabarty*, 447 U.S. 303 (1980).

¹⁹ *Id.* at 309.

²⁰ *Amgen, Inc. v. Chugai Pharmaceutical Co.*, 927 F.2d 1200, 1206 (Fed. Cir. 1991).

composition of matter if they are “made by man”—i.e. created in the lab or isolated from nature. In general, matter in its naturally occurring state cannot be patented, but isolated and purified “products of nature” are eligible for patent protection. Thus, it is now clear that “a DNA sequence itself is not patentable. . . [but a] purified DNA molecule isolated from its natural environment . . . is a chemical compound and is patentable if all the statutory requirements are met.”²¹

Some people advocate that patent claims involving DNA should be limited to applications or methods of using the DNA; i.e. that patents on the DNA as a composition of matter should not be allowed.²² However, there is no basis in law for such a limitation on gene patents. As the USPTO recently noted, “Patentable subject matter includes both “process[es]” and “composition[s] of matter.” . . . [and p]atent law provides no basis for treating DNA differently from other chemical compounds that are compositions of matter.”²³

For strategic reasons, patents that claim isolated genes as compositions of matter are preferred to patents that claim a particular process for making or using a DNA sequence. A process patent gives the patentee the right to prevent others from using that particular process, but it cannot be used to prevent others from making the resulting product in other ways. However, “a patent on a product per se will be infringed by a competitor making the same product—no matter what process is used to make that product,” as was found in the recent case of *Amgen v. Hoescht*.²⁴ Moreover, a composition patent can be used to prevent others from using the product in any way whatsoever, “. . . even if the inventor disclosed only a single use for the composition.”²⁵

In short, genes that have been isolated may be patented as a composition of matter, and such patents are extremely powerful weapons in the business world. It is probably not possible, though, to patent pure genetic information.²⁶ For example, patents on sequences as information stored on a computer readable medium would prevent storage and retrieval of the information. Such patents are unlikely to ever issue, in part because electronic compilations of data are not patentable.²⁷

The policy of patents grants the inventor a monopoly in exchange for public disclosure of the invention. Prof. Eisenberg, a noted authority on biotech law, concludes that “[p]atent 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.” Thus, once a patent issues on an

²¹ U.S. Patent and Trademark Office, Utility Examination Guidelines, 66 Fed. Reg. 1092, 1094 (January 5, 2001), available at <http://wais.access.gpo.gov> [hereinafter Utility Guidelines].

²² Utility Guidelines, *supra* note 21, at 1094-95.

²³ *Id.*

²⁴ *Amgen, Inc. v. Hoechst*, 126 F. Supp. 2d 69 (D. Mass, 2001); *see also* Jennifer Van Brunt, *The Next Move in the Patent Game*, Signals Magazine (April 4, 2001), <http://www.signalsmag.com/signalsmag.nsf> (discussing the import role of composition of matter gene patents in the business world).

²⁵ Utility Guidelines, *supra* note 21, at 1095.

²⁶ Rebecca S. Eisenberg, *Re-Examining the Role of Patents in Appropriating the Value of DNA Sequences*, 49 EMORY L.J. 783 (2000).

²⁷ *Id.* at 790.

isolated sequence, the information content of that sequence is freely available, “subject only to the inventor's right to exclude others from making, using, and selling the claimed materials.”²⁸

B. The Written Description Issue: Inferring Structure from Function

Section 112 of the Patent Act sets forth the requirements for the specification, and says that it “shall contain a written description of the invention.” This seemingly simple requirement has been interpreted by the courts to require a description that is sufficient to indicate that the inventor had “possession” of the invention.²⁹ That is, the inventor must fully set forth the claimed invention, providing “sufficient detail that one skilled in the art can clearly conclude that the inventor invented the claimed invention.”³⁰

An inventor who has reduced his or her invention to practice is clearly in possession of it and will easily satisfy the written description requirement by describing what was done. If an inventor has merely conceived an invention, the inventor must clearly demonstrate conception in order to show possession and satisfy the description requirement. When an invention is not obvious in light of what is described, the requirement is not satisfied. Thus, the written description requirement is often intertwined with the issue of obviousness.

I review here the CAFC’s early rulings on the description of genes and their obviousness in light of knowledge of an amino acid sequence, and I consider some criticisms of the court’s approach and findings. I then summarize the recent *Guidelines* developed by the USPTO and show how they sensibly address most of the concerns.

1. The Court’s Interpretation of Gene Descriptions

In the early 1990s, the CAFC determined that a “biomolecule sequence described only by a functional characteristic, without any known or disclosed correlation between that function and the structure of the sequence, normally is not a sufficient identifying characteristic for written description purposes, even when accompanied by a method of obtaining the claimed sequence.”³¹ That is, a claim to a nucleotide sequence could not be supported by merely naming the protein for which it codes and a method for isolating it.

The court first addressed the issue in 1991 in the case of *Amgen v. Chugai*.³² It considered the validity of Amgen’s patent claim to a “purified and isolated DNA sequence consisting essentially of a DNA sequence encoding human

²⁸ *Id.* at 787.

²⁹ *Vas-Cath Inc. v. Mahurkar*, 935 F.2d 1555, 1563-64 (Fed. Cir. 1991) (to satisfy the written description requirement, the specification must “reasonably convey to the artisan that the inventor had possession at that time of the ... claimed subject matter.”).

³⁰ *Lockwood v. American Airlines*, 107 F.3d 1565, 1572 (Fed. Cir. 1997)

³¹ U.S. Patent and Trademark Office, *Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, para. 1, “Written Description” Requirement*, 66 Fed. Reg. 1099, 1108 fn14 [hereinafter *Written Description Guidelines*].

³² *Amgen, Inc. v. Chugai Pharmaceutical Co.*, 927 F.2d 1200 (Fed. Cir. 1991).

erythropoietin.”³³ Amgen had not isolated and sequenced the gene and the polypeptide sequence of human erythropoietin was unknown. The court decided that knowing a method to isolate and sequence the gene was not enough—Amgen needed to know and describe the sequence; that is, it needed to actually reduce the invention to practice.³⁴

The court based this decision on its assessment of the DNA as a chemical compound. It noted that “conception of a chemical compound requires that the inventor be able to define it so as to distinguish it from other materials.”³⁵ It then concluded that “[i]t is not sufficient to define [the erythropoietin gene] solely by its principal biological property, e.g., encoding human erythropoietin, because an alleged conception having no more specificity than that is simply a wish to know the identity of any material with that biological property.”³⁶ Rather, the inventor must have “a mental picture of the structure of the chemical, or [be] able to define it by its method of preparation, its physical or chemical properties, or whatever characteristics sufficiently distinguish it.”³⁷

The court addressed the issue again in 1993 in the case of *Fiers v. Revel*.³⁸ In this interference action between parties seeking similar but as yet unissued patents, the court addressed the validity of a potential claim to a “DNA which consists essentially of a DNA which codes for a human fibroblast interferon-beta polypeptide.”³⁹ The court cited *Amgen* in holding that “conception of any chemical substance, requires a definition of that substance other than by its functional utility” and then elaborated that “[c]onception of a substance claimed per se without reference to a process requires conception of its *structure, name, formula, or definitive chemical or physical properties*” (emphasis added).⁴⁰ In short, the court found that “[a]n adequate written description of a DNA requires . . . a description of the DNA itself.”⁴¹

At about the same time that it was addressing the written description requirement as applied to gene patents, the CAFC addressed the issue of the obviousness of a DNA sequence when the amino acid sequence of the polypeptide for which it codes is already known.⁴² To the surprise of many biologists, the CAFC determined that knowing the amino acid sequence of a polypeptide and a

³³ *Id.* at 1204.

³⁴ *Id.* at 1206.

³⁵ *Id.*

³⁶ *Id.*

³⁷ *Id.*

³⁸ *Fiers v. Revel*, 984 F.2d 1164 (Fed. Cir. 1993).

³⁹ *Id.* at 1166.

⁴⁰ *Id.* at 1169. The court elaborate on the connection between conception and description, noting that “[i]f a conception of a DNA requires a precise definition, such as by structure, formula, chemical name, or physical properties, as we have held, then a description also requires that degree of specificity. To paraphrase the Board, one cannot describe what one has not conceived.” *Id.* at 1171.

⁴¹ *Id.* at 1171.

⁴² An invention must be nonobvious to qualify for a patent. See note 12, *supra*, and accompanying text; Jeffrey S. Dillen, *DNA Patentability - Anything but Obvious*, 1997 WIS. L. REV. 1023 (1997) (reviewing case law related to the issue of the obviousness of a DNA sequence if the amino acid sequence for which it codes is known).

general method of cloning does not make the naturally occurring nucleotide sequence obvious. The logic is, however, consistent with the court's assessment of the written description requirement as it applies to claims to DNA.

The court first addressed the issue of obviousness in 1993 in a case called *In re Bell*.⁴³ Bell sought to claim the sequences which code for *human* insulin-like growth factors (IGF) I and II; the amino acid sequence of these proteins was already known. The CAFC again focused on the DNA molecules as chemical compounds rather than assessing the methods used to isolate the DNA. The court acknowledged that, "knowing the structure of the protein, one can use the genetic code to hypothesize possible structures for the corresponding gene" but it also acknowledged the vast number of sequences that could code for a protein.⁴⁴ Because it was not known which of the possible sequences would be found *in humans*, the court found that the human sequence was not obvious.⁴⁵

The court addressed the issue again in 1995 in a case called *In re Deuel*.⁴⁶ Deuel claimed: "A purified and isolated DNA sequence consisting of a sequence encoding human heparin binding growth factor of 168 amino acids having the following amino acid sequence: Met Gln Ala ... [remainder of 168 amino acid sequence]." ⁴⁷ The court saw that the claim was "tantamount to the general idea of all genes encoding the protein, all solutions to the problem." And it wisely acknowledged that this set of sequences "might have been obvious from the complete amino acid sequence of the protein, coupled with knowledge of the genetic code" explaining that "this information may have enabled a person of ordinary skill in the art to envision the idea of, and, perhaps with the aid of a computer, even identify all members of the claimed genus."⁴⁸ However, because the amino acid sequence was previously unknown, the court found that the claim was not invalid for obviousness.

These rulings of the CAFC may be summarized as follows: A claim to a DNA must describe the DNA; it cannot be inferred by naming the protein for which it codes and a method for isolating the DNA. Even if the amino acid sequence of the protein is known, the actual sequence that codes for the protein in a particular organism is not. Therefore, the DNA sequence must be established to claim a gene specific to a particular organism. However, if the amino acid sequence of the protein is newly discovered, then the entire class of DNAs that could code for the protein is also newly discovered. In this case, a set or "genus" of DNA sequences may be claimed by acknowledging the genetic code and describing the polypeptide sequence.

In 1997, the court expanded these precedents to address the description of a set, or genus, of DNAs (rather than a single molecule, or species) in a case known commonly at *U.C. v. Eli Lilly*.⁴⁹ The University of California sought to claim

⁴³ *In re Bell*, 991 F.2d 781 (Fed. Cir. 1993).

⁴⁴ *Id.* at 784 (Fed. Cir. 1993). The court acknowledged the possibility that a known amino acid sequence is specified exclusively by unique codons, in which case the gene would be obvious. *Id.*

⁴⁵ *Id.*

⁴⁶ *In re Deuel*, 51 F.3d 1552 (Fed. Cir. 1995).

⁴⁷ *Id.* at 1555.

⁴⁸ *Id.* at 1560.

⁴⁹ *Regents of University of California v. Eli Lilly*, 119 F.3d 1559 (Fed. Cir. 1997).

mammalian and vertebrate insulin cDNA based upon a description of human insulin cDNA. The court found the description inadequate. It said that a “written description of an invention involving a chemical genus, like a description of a chemical species, “requires a precise definition, such as by structure, formula, [or] chemical name,” of the claimed subject matter sufficient to distinguish it from other materials.”⁵⁰ It concluded that “a generic statement such as “vertebrate insulin cDNA” or “mammalian insulin cDNA,” without more, is not an adequate written description of the genus because . . . it does not define any structural features commonly possessed by members of the genus that distinguish them from others.”⁵¹

The *Lilly* court asserted that DNA claims would require “a kind of specificity usually achieved by means of the recitation of the sequence of nucleotides that make up the DNA” and, by analogy, that claims to a genus of cDNAs would require reciting a “representative number of cDNAs, defined by nucleotide sequence.” The court refused, however, to “speculate in what other ways a broad genus of genetic material may be properly described”⁵²

2. The Public Debate about Treating DNA as a Chemical Compound

The courts’ treatment of genes as chemical compositions has been debated extensively, both as it relates to the issue of obviousness and the written description.⁵³ By treating DNA as a chemical, the CAFC has simultaneously lowered the bar for non-obviousness (by finding that knowledge of an amino acid sequence and a general method for identifying genes with the use of nucleotide probes does not make the DNA sequence obvious) and raised the bar for the written description (by requiring that genes are actually isolated and sequenced before being patented).⁵⁴

Rai, for example, argues that the CAFC’s treatment of DNA as a subset of chemical technology is “fundamentally misconceived” and reflects the court’s failure to recognize DNA-based technologies “as involving information first and foremost.”⁵⁵ She says that, as a result, “the courts have thereby made patent protection too strong in some respects and too weak in others.”⁵⁶ Eisenberg also emphasizes the importance and value of DNA sequences as information.⁵⁷ She finds that “the chemical analogy is of little value as a strategic guide to exploiting this information as intellectual property.”⁵⁸

⁵⁰ *Id.* at 1568.

⁵¹ *Id.*

⁵² *Id.* at 1569. START HERE

⁵³ See, e.g., Todd R. Miller, *Motivation and Set-Size: In Re Bell Provides a Link Between Chemical and Biochemical Patent Claims*, 2 U. BALT. INTELL. PROP. J. 89 (1993) (drawing upon and citing previous participants in the debate).

⁵⁴ See Part II.B.1; see also Rai, *supra* note 14.

⁵⁵ Rai, *supra* note 14, at 836 (“Although DNA is, obviously enough, a chemical compound, it is more fundamentally a carrier of information.”).

⁵⁶ *Id.*

⁵⁷ Eisenberg, *supra* note 26.

⁵⁸ *Id.* at 785.

There are defenders of the court's approach. Margaret Sampson suggests that the heightened description approach helps prevent overly broad patents.⁵⁹ Sampson argues that the heightened description requirement prevents an inventor from restricting the use of "homologs, alleles, polymorphisms, and isoforms found in the same gene family, all of which have a high degree of sequence identity with the gene, but not 100% identity," and limits the ability of inventors to assert rights to sequences of which they have no knowledge, in organisms with which they have never worked.⁶⁰ As discussed in Part III.B.2, this does not appear to be the case.

Perhaps more importantly, the court's approach may be good policy if it encourages inventors to establish nucleotide sequences for known proteins and prevents them from asserting rights to genes without ever revealing their sequences. Indeed, by treating DNA as a chemical compound and requiring inventors to describe its structural attributes, the court has effectively required inventors to (1) determine the critical information attribute of a DNA (i.e. the nucleotide sequence) and (2) reveal it to the public. These rulings may therefore promote the discovery of genetic information—by providing an incentive to discover gene sequences, as well as the dissemination of genetic information—by requiring that the information is revealed to the public in the patent.

3. The PTO's Guidelines for Examination of the Written Description

The USPTO published its *Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, para. 1, "Written Description" Requirement* (Written Description Guidelines) on January 5, 2001. This document reflects the USPTO's understanding of the law on the statutory requirement of a written description, and was created to provide guidance to the examiners who must evaluate patent applications in light of the law. An interim version of the document was previously made available to the public for comments; in the final version, the USPTO summarizes and responds to those comments, but does not change the guidelines substantially. The document provides a comprehensive, accurate, and accessible summary of the law, and indicates how the USPTO has applied and will apply the law—at least until the CAFC contradicts its interpretation.

The Written Description Guidelines provides a sensible restatement of the law, noting that "[a]n adequate written description of the invention may be shown by any description of sufficient, relevant, identifying characteristics so long as a person skilled in the art would recognize that the inventor had possession of the claimed invention."⁶¹ It also acknowledged the finding of the *Amgen* court, i.e. when "an invention is described solely in terms of a method of its making coupled with its function and there is no described or art-recognized correlation or

⁵⁹ Margaret Sampson, *The Evolution of the Enablement and Written Description Requirements Under 35 U.S.C. 112 in the Area of Biotechnology*, 15 BERKELEY TECH. L.J. 1233, 1261 (2000).

⁶⁰ *Id.*

⁶¹ Written Description Guidelines, *supra* note 31, at 1105.

relationship between the structure of the invention and its function,” the description is inadequate.⁶²

According to the Written Description Guidelines, an invention may be sufficiently described by disclosure of “complete or partial structure, other physical and/or chemical properties, *functional characteristics when coupled with a known or disclosed correlation between function and structure*, or some combination of such characteristics (emphasis added).⁶³ For at least some biomolecules, such characteristics include “a sequence, structure, binding affinity, binding specificity, molecular weight, and length” but “other identifying characteristics or combinations of characteristics may demonstrate the requisite possession”.⁶⁴

Thus, the Written Description Guidelines acknowledge that molecules may be described not only by sequence, but also by functional attributes when such attributes are clearly associated with structural attributes. Indeed, the *Guidelines* instruct examiners to consider “the level of skill and knowledge in the art, partial structure, physical and/or chemical properties, [as well as] *functional characteristics alone or coupled with a known or disclosed correlation between structure and function.*”⁶⁵

The Written Description Guidelines also address the court’s interpretation of the written description as it applies to a claimed genus, noting that a claim to a genus is satisfied “through sufficient description of a representative number of species by actual reduction to practice. . . , reduction to drawings. . . , or by disclosure of relevant, identifying characteristics.” The Written Description Guidelines again indicate that such characteristics include “structure or other physical and/or chemical properties, . . . *functional characteristics coupled with a known or disclosed correlation between function and structure*, . . . [and a combination of such identifying characteristics, . . .”

The most novel and interesting direction in the Written Description Guidelines pertains to the adequacy of the description of a genus of DNAs by reference to an amino acid sequence. The USPTO notes two comments asserting that, “if the amino acid sequence for a polypeptide whose utility has been identified is described, then the question of possession of a class of nucleotides encoding that polypeptide can be addressed as a relatively routine matter using the understanding of the genetic code.” The suggestion was incorporated into the Written Description Guidelines as follows: “if an applicant disclose[s] an amino acid sequence, it [is] unnecessary to provide an explicit disclosure of nucleic acid sequences that encode[] the amino acid sequence. Since the genetic code is widely known, a disclosure of an amino acid sequence . . . provide[s] sufficient information such that one would accept that an applicant was in possession of the full genus of nucleic acids encoding a given amino acid sequence, but not necessarily any particular species.”⁶⁶

⁶² *Id.*

⁶³ Written Description Guidelines, *supra* note 31, at 1106.

⁶⁴ Written Description Guidelines, *supra* note 31, at 1110 fn 42.

⁶⁵ Written Description Guidelines, *supra* note 31, at 1106.

⁶⁶ Written Description Guidelines, *supra* note 31, at 1111 fn 57.

The Written Description Guidelines note, though, that “this does not mean that applicant was in possession of any particular species of the broad genus.”⁶⁷ Such claims may therefore be allowed, but may fail to preclude subsequent claims to sequences that are, e.g., specific to a particular organism.

C. The Politics of EST Patents

In the early 1990s, when the courts were assessing the legal implications of claiming DNA whose sequence was not yet known, scientists were beginning to produce large numbers of cDNA fragments known as expressed sequence tags, or ESTs. These short nucleic acid sequences were relatively easily discovered, but their function was usually unknown—in sharp contrast to the situation of the previous decade, when sequences of known function were sought and obtained after substantial focused effort.

The community was divided about the merits of patenting ESTs.⁶⁸ The National Institute of Health and then Craig Venter sought to patent them, but the Human Genome Organization (HUGO) vehemently opposed any and all such efforts.⁶⁹ HUGO believed that ESTs were research tools, and thought they and all sequences should be viewed as part of pre-competitive information.⁷⁰ Nonetheless, by 1996, the USPTO was deluged with over half a million applications for patents on ESTs. At that point, the office stopped tracking them.⁷¹

Fortunately for the USPTO, the flood abated, with the number of EST patent applications dropping dramatically around 1998.⁷² Various PTO officials have characterized three cycles or generations of EST patents: The first generation comprises applications that do not disclose the gene associated with the EST. The second generation comprises applications where the function of the protein being expressed by the gene is determined by homology searches. In the third-generation patents, “[the inventors] have actually found the function by doing the science,” piecing together the complete open reading frame (ORF) for the gene. In April 2001, it was estimated that the PTO had received as many as 25,000 third generation applications.⁷³ XXXCHECK

The arguments about patenting ESTs have focused on utility. As Professor Eisenberg noted in 1992, “the argument against allowing NIH to patent the sequences is not really that these sequences are useless, but rather that NIH does not yet know what they are good for and should not be able to claim patent rights ahead of subsequent researchers who figure it out. It is the as yet undiscovered

⁶⁷ Written Description Guidelines, *supra* note 31, at 1102.

⁶⁸ Gary Zweiger provides a cogent and timely review of the history of genomics, including an assessment of the companies and individuals who sought to patent ESTs and those who opposed such business tactics. Gary Zweiger, *TRANSDUCING THE GENOME: INFORMATION, ANARCHY, AND REVOLUTION IN THE BIOMEDICAL SCIENCES* (2001). *See also* Murray, *supra* note 6.

⁶⁹ Human Genome Organization (HUGO), *Statement on Patenting of DNA sequences - In Particular Response to the European Biotechnology Directive* (April 2000).

⁷⁰ *Id.*

⁷¹ Van Brunt, *supra* note 24.

⁷² *Id.*

⁷³ Todd Dickinson. Comments at Congressional Hearing on Genomic Inventions, *supra* note 1; Van Brunt, *supra* note 24.

utility of the sequences, rather than the uses that are disclosed in the patent application, that makes NIH's patent claims worth fighting about.⁷⁴ The general thinking is that ESTs should be patentable if the full gene sequence and its function are known.⁷⁵ If so, the first generation EST patent applications will not satisfy the utility requirement, but the third generation applications will.

Patent applications for ESTs in the second generation, where utility is inferred from the computational analysis of genomic databases, are most difficult to assess. The Director of the USPTO explained to members of Congress in July 2000: "The question comes down to ... how much utility can be inferred from the computer modeling that is used now to determine the utility associated with a particular EST. The question is what percentage of that analogous information—it's called percent homology in the term of the art—is sufficient, in order to justify the utility."⁷⁶ In short, the question is whether a finding of homology of an EST with a known gene is sufficient to establish utility, and hence patentability, of the EST.

The second generation EST patents are politically contentious because they provide patent rights to early stage research tools. Such patents could affect both the pace of genetics research and the structure of industry. If the patenting of ESTs restricts researchers' access to them, such patents could impede complete characterization of genes and delay or restrict exploration of genetic materials for the public good.⁷⁷ Whether or not this is true may depend upon the business methods adopted in the relevant industries. For example, the use of non-exclusive licenses and the creation of patent pools could facilitate the widespread use of patented ESTs.⁷⁸ On the other hand, such patents may provide incentives for research and development of gene fragments, and could foster the development of companies that specialize in genomics research.

Randall Scott of Incyte, a company that focuses on the accumulation and analysis of early stage research information, argues for EST patents—even when the precise biological activity of the gene is unknown. Scott rightly emphasizes that "a patent should be rewarded for commercial utility, not for biological function, and there's an important distinction."⁷⁹ He argues that ESTs are useful "as tools, as diagnostics, as markers for disease and drug therapy," and such uses do not require knowledge of their biological function. Thus, he says, "the real world utility of genes is not just buried in their biological function and what they do naturally in the body."⁸⁰

⁷⁴ Rebecca S. Eisenberg, *Genes, Patents, and Product Development*, 257 *SCIENCE* 5072 (1992).

⁷⁵ See, e.g., Murray, *supra* note 6 (1999).

⁷⁶ Todd Dickinson. Comments at Congressional Hearing on Genomic Inventions, *supra* note 1.

⁷⁷ Murray, *supra* note 6 at 254.

⁷⁸ *But see* Rai, *supra* note 17 (critiquing the idea that the market can compensate for the blocking effect of patents on early stage research tools).

⁷⁹ Dr. Randal W. Scott, President And Chief Scientific Officer, Incyte Genomics. Statement at Congressional Hearing on Genomic Inventions, *supra* note 1 (noting as an example that the common indicator of prostate cancer is the observation of a certain protein in the blood; the function of the protein is unknown, but tests for the protein clearly have significant commercial utility).

⁸⁰ *Id.*

Scott's view contrasts sharply with the view of officials at Genentech, a company that is involved in the development of pharmaceutical products. Genentech officials believe that "the utility of a particular gene or protein cannot be known unless one has determined its [biological] function."⁸¹ And such determination requires laboratory research, not genomic analysis. Dennis Henner, of Genentech, told members of Congress that "computer modeling is not sufficiently accurate to predict protein function based solely on gene comparisons."⁸² Therefore, he said, "the utility of a particular gene or polypeptide rarely can be demonstrated until there has been a sufficient characterization of the function of a gene or its expression product . . . through relevant biological assays."⁸³

D. The Utility Issue: Inferring Function from Structure

Section 101 of the Patent Act establishes that "[w]hoever invents any . . . *useful* . . . composition of matter . . . may obtain a patent therefor . . ."⁸⁴ This so-called utility requirement historically was and, in many cases still is, trivial. In 1817 it was interpreted to mean only that an invention could not be mischievous or immoral. Today, the utility requirement reflects more general policy concerns. Utility became an issue in the chemical arts in 1966, when the court ruled that a chemical compound with no known practical use could not be patented. It is now a major issue in the patenting of ESTs.

The issue of the utility of ESTs implicates the validity of structure-function relationships in biochemistry, and the consequences of such patents for further discoveries relating to the associated gene. As the Director of the USPTO acknowledged in July 2000, "legitimate questions have been raised about just what genomic discoveries, if any, should be patentable and whether genomic patents will inhibit researchers' access to the data, materials, and methods needed to develop new tools for the diagnosis and treatment of disease"⁸⁵

In the section, I review the courts' general rulings on utility, and the USPTO's guidelines for applying the utility requirement to biotech inventions. I consider

⁸¹ Dennis J. Henner, Ph.D., Senior Vice President, Research, Genentech, Inc. Statement at Congressional Hearing on Genomic Inventions, *supra* note 1.

⁸² *Id.*

⁸³ *Id.* He elaborated as follows:

The degree of homology can be an important indicator that the sequence being analyzed is similar to, or within a class of known proteins based on the degree of identity it shares with the known sequence. . . . Homology analysis, however, is a limited tool for predicting results. In our experience, homology analysis, standing alone, is not a sufficiently reliable indicator to base scientific or business decisions upon. . . .

Accordingly, where a particular biological activity is the only basis for the utility of a particular gene or expression product, a homology-based prediction should not be capable of satisfying the requirements of our law in a majority of situations.

Id.

⁸⁴ 35 U.S.C. §101 (1998); *see also* note 9, *infra*, and accompanying text.

⁸⁵ Todd Dickinson, Under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office. Prepared Statement at Congressional Hearing on Genomic Inventions, *supra* note 1.

public commentary and attempt to determine (1) whether or when, under the guidelines, an inventor must know the biological function of the protein coded by the gene associate with the claimed EST, and (2) whether the inventor can establish that function by analyzing sequence similarity to genes of known function.

1. The Courts' Interpretation of Utility in the Chemical Arts

No court has yet addressed the application of the utility requirement to partial nucleotide sequences. Thus, it is possible that the opinions of academics and the policies of the USPTO will be found irrelevant and inapplicable, respectively. The USPTO has purportedly arranged for two interested parties to present the issue of the utility of ESTs to the court in a "test case."⁸⁶ In July 2000, this case was purportedly set to go to the Board of Appeals; if so, it could appear before the CAFC as early as January 2002.⁸⁷

The United State Supreme Court did, however, address the issue of utility as it applies to the chemical arts in the 1966 case of *Brenner v. Manson*.⁸⁸ Manson had devised a method for making a certain steroid compound. The Court found that Manson had failed to assert any utility for the process, other than its use in research by chemists. Because the invention did not have practical benefits for the public, and because a patent on the process could "confer power to block off whole areas of scientific development, without compensating benefit to the public," the court found that it failed to meet the utility requirement.⁸⁹ In summary, the Court declared that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion."⁹⁰

The *Brenner* court explicitly rejected Manson's argument for utility based upon the observation that a compound similar to the one produced by his process (an "adjacent homologue") had been shown to inhibit the growth of tumors in mice. The USPTO had found that Manson had not disclosed "a sufficient likelihood that the steroid yielded by his process would have similar tumor-inhibiting characteristics," and the Court accepted its finding.⁹¹ In short, because Manson had failed to provide a convincing argument for the function of the steroid based upon its structural similarities to compounds with known functions, he had failed to assert a practical utility.

The Court's reliance on the USPTO's determination that Manson could not reasonably infer the function of his steroid from its structure is important. It suggests that assertions for the utility of ESTs based upon their structural similarity to genes coding for proteins of known function depends upon the USPTO's determination of the scientific validity of such an inference.

2. The PTO's Guidelines for Examination of Utility

⁸⁶ *Id.*

⁸⁷ *Id.*

⁸⁸ 383 U.S. 519 (1966).

⁸⁹ *Id.* at 534.

⁹⁰ *Id.* at 536.

⁹¹ *Id.* at 532.

The USPTO published its *Utility Examination Guidelines* (Utility Guidelines)⁹² on January 5, 2001. This document reflects the USPTO's understanding of the law on the statutory requirement of utility and was created to provide guidance to the examiners who must apply it.⁹³ As for the Written Description Guidelines, the USPTO summarizes and responds to comments on a previously published version.⁹⁴

The Utility Guidelines have been more contentious than the Written Description Guidelines, because the utility of ESTs is the key factor in assessing their patentability.⁹⁵ Excellent synopses of the document, with critical commentary, are already available.⁹⁶

The Utility Guidelines require the inventor to identify a *specific, substantial, and credible* utility for the claimed invention, unless such a utility is already well established.⁹⁷ This three-part test raises the bar for showing utility because previous guidelines required only a credible utility. However, a “specific” and “substantial” utility has been required by the courts. Thus, the new guidelines are more in line with case law than previous guidelines.

An asserted utility is *credible* unless (1) the logic underlying the assertion is seriously flawed, or (2) the facts upon which the assertion is based are inconsistent with the logic underlying the assertion. The credibility of an asserted utility is assessed from the standpoint of a person of ordinary skill in the art, but the presumption favors the inventor. For example, since at least some nucleic acids can be used as probes, chromosome markers, or diagnostic markers, the assertion that any particular DNA can be used in this is accepted.

An asserted utility is *substantial* if it defines a “real world” use. If further research is required to confirm or identify the use, the use is not substantial. Thus, claims that a nucleic acid is useful for studying the properties of the gene itself are not substantial.

⁹² Utility Guidelines, *supra* note 21.

⁹³ The USPTO emphasizes this point in the Utility Guidelines, clarifying that it is not free to develop its own rules about the patentability of DNA. Utility Guidelines, *supra* note 21, at 1095 (“The USPTO must administer the laws as Congress has enacted them and as the Federal courts have interpreted them. Current law provides that when the statutory patentability requirements are met, there is no basis to deny patent applications claiming DNA compositions, or to limit a patent's scope in order to allow free access to the use of the invention during the patent term.”).

⁹⁴ U.S. Patent and Trademark Office, Revised Interim Utility Guidelines, 64 Fed. Reg. 71440 (Dec. 21, 1999); correction at 65 Fed. Reg. 3425 (Jan. 21, 2000).

⁹⁵ Expressed Sequence Tags are “patentable to the same extent that any other invention is patentable, so long as they meet the test of patentability. And the question that it basically comes down to . . . is the question of utility and the ability to demonstrate sufficient utility to meet the section 101 standard.” Todd Dickinson. Comments at Congressional Hearing on Genomic Inventions, *supra* note 1.

⁹⁶ *E.g.* Timothy A. Worrall, *The 2001 PTO Utility Examination Guidelines and DNA Patents*, 16 BERKELEY TECH. L.J. 123 (2001); *The Fate of Gene Patents Under the New Utility Guidelines*, 2001 Duke L. & Tech. Rev. 0008 (2001).

⁹⁷ Utility Guidelines, *supra* note 21; *see also* U.S. Patent and Trademark Office, Revised Interim Utility Guidelines Training Materials, *available at* <http://wais.access.gpo.gov> [hereinafter Utility Training Materials]; Todd Dickinson. Comments at Congressional Hearing on Genomic Inventions, *supra* note 1.

An asserted utility is *specific* when it is particular to the subject matter claimed. For example, asserting that an EST is useful as a “gene probe” or “chromosome marker” is not sufficiently specific; the inventor must disclose a particular gene for the probe, or chromosome target for the marker. By the same logic, asserting that an EST has diagnostic utility is typically insufficient; the inventor must identify the condition that is diagnosed.

The Utility Guidelines are widely viewed as having raised the bar on utility as it applies to the patenting of ESTs. However, they appear to clearly indicate that ESTs are patentable, even if the function of the encoded gene product is unknown. They state unequivocally that “[t]he utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have a specific and substantial utility because, e.g., it hybridizes near a disease-associated gene or it has a gene-regulating activity.”⁹⁸ And they clearly suggest that computational methods such as sequence comparisons may be used to identify the relevant gene and thereby provide the required specific utility.

3. The Public Debate about Using Genomics to Establish Utility

In July 2000, Todd Dickinson told members of Congress that officials at the USPTO believed the new “heightened standard of utility w[ould] allow appropriate patents on genomic inventions, while also resulting in the rejection of hundreds of genomic patent applications, particularly those that only disclose *theoretical utilities*”⁹⁹ (emphasis added). As one reporter described it, researchers “take a gene, or even just a piece of a gene, plug it into a computer, and instantly turn up vast amounts of intriguing but *theoretical* information about it”; they then file for patents “without doing a single experiment or ‘getting [a] pipette wet’”.¹⁰⁰

These comments reflect a not uncommon sentiment that knowledge acquired by experimentation in the lab is superior to knowledge acquired through the analysis of databases. John Golden recently argued that “the science of “bio-informatics” [is] still in its infancy, [and] current computer-based methods for studying genetic sequences have failure rates as high as 95%.”¹⁰¹ He objected to the USPTO’s idea that “computer-based analogy to a known useful sequence is presumptively sufficient for patentability” and concluded that “installing a presumption in favor of the reliability of computer-based studies could . . . ultimately give away most of what a meaningful utility requirement is meant to protect.”¹⁰²

Clearly, assessing the results of database analyses can be difficult and the need to interpret findings that are typically associated with probabilities may be unfamiliar and non-intuitive to scientists who are accustomed to interpreting the typically binary feedback of laboratory results. Nonetheless, even some officials recognize that searching sequence databases for similar genes is common practice

⁹⁸ Utility Guidelines, *supra* note 21, at 1095.

⁹⁹ Todd Dickinson. Statement at Congressional Hearing on Genomic Inventions, *supra* note 1.

¹⁰⁰ Merril Goozner, *Patenting Life*, THE AMERICAN PROSPECT (December 18, 2000).

¹⁰¹ John M. Golden, *Biotechnology, Technology Policy, and Patentability: Natural Products and Invention in the American System*, 50 EMORY L.J. 101, 188 (2001).

¹⁰² *Id.*

and is “very well established and very well accepted in the academic community.”¹⁰³

Patent experts believe the USPTO new Utility Guidelines are unlikely to be overturned by the court, perhaps because the court has traditionally failed to enforce the utility requirement very strictly.¹⁰⁴ Perhaps the more interesting question is whether utilities asserted by database analyses can be justified scientifically.

III. THE USE OF COMPUTATIONAL TECHNIQUES IN GENE PATENTS

There is currently no published study of patent office decisions examining claims to ESTs or the implementation of the new Written Description and Utility Guidelines.¹⁰⁵ Thus, it is not known how or to what extent the Guidelines have affected the type or style of patents issuing on ESTs.

In this section, I rely on various searches of recently issued patents, a close reading of more than twenty patents issuing on ESTs, and the examples provided in the USPTO’s Training Materials to determine how scientists and their patent attorneys are using computational methods to satisfy the written description and utility requirements. I then critique these uses from both scientific and legal perspectives.

A. Finding EST Patents

It is not known how many patents have issued on genes in general or ESTs in particular, but from all accounts and consistent with all estimates, there are likely tens of thousands of gene patents and hundreds if not thousands of EST patents. Furthermore, there is no easy way to identify a patent as an EST patent, short of reading and considering it in its entirety. I describe here my search approach and some suggestive data on trends in the issuance of patents using computational methods.

1. Methods for Searching the Databases

I used various combinations of key word searches of the Lexis¹⁰⁶ patent database, with various field and date restrictions, to identify a manageable number of recent patents on ESTs that I could examine closely. My search methods were exploratory, and the sample of patents that I chose to examine closely may not be representative of EST patents in general.

¹⁰³ Martin Enserink, *Patent Office May Raise The Bar on Gene Claims*, 287 SCIENCE 5456 (2000) (citing Doll). The reported statement of an Incyte representative that “Everybody uses these techniques and they are virtually 100% correct” overstates the case and fails to acknowledge the importance of interpreting probabilities. *Id.*

¹⁰⁴ *Id.*

¹⁰⁵ Goozner, *supra* note 100.

¹⁰⁶ This database is available by subscription only; however, all patents examined here are available in their entirety at the USPTO website, <http://www.uspto.gov>.

There is a classification system for patents, and all patents list one and usually several class/subclass categories. I examined the classification of several patents that I had determined by various means to be EST patents, and found that most (although not all) of them listed Class 536, Subclass 23.1 Class 536 is “organic compounds” and within it, subclass 23.1 is “DNA or RNA fragments or modified forms thereof”, and subclasses 23.2 to 23.7 are DNA or RNA fragments that encode a particular type of protein. Thereafter, I restricted my searches to this Class and Subclass.

Given the large number of EST patents, I focused on patents issued most recently in the summer of 2001, between June 1 and August 15.

After examining a number of gene patents, I found that claims to sequences as compositions of matter invariably referred to a sequence given in the specification as “SEQ ID” followed by a identification number. These claims usually also specified that the claimed compound was a “polynucleotide.” I therefore restricted my searches to patents with these terms in the claims. I also restricted several searches to claims that included the term “percent identity” for reasons that will be obvious later. The most useful combination of keywords for identifying EST patents, given the previously noted restrictions, were the terms “EST” or “cDNA” in conjunction with “fragment” or “partial”.

With these search criteria, I obtained 17 patents, each assigned to one of four companies. Because 11 of the 17 patents were assigned to DuPont and only 2 were assigned to Incyte,¹⁰⁷ I searched for patents issued to various companies prior to June 1, 2001. I include 4 additional patents issued to Incyte, because they have been a vocal participant in the debate about genome patents.¹⁰⁸ I also

¹⁰⁷ The patents assigned to Dupont included: Patent Number 6,255,090: Plant aminoacyl-tRNA synthetase (July 3, 2001); Patent Number 6,271,441: Plant aminoacyl-tRNA synthetase (August 7, 2001); Patent Number 6,255,114: Starch biosynthetic enzymes (July 3, 2001); Patent Number 6,252,137: Soybean homolog of seed-specific transcription activator from *Phaseolus vulgaris* (June 26, 2001); Patent Number 6,242,256: Ornithine biosynthesis enzymes (June 5, 2001); Patent Number 6,262,345: Plant protein kinases (July 17, 2001); Patent Number 6,274,379: Plant sorbitol biosynthetic enzymes (August 14, 2001); Patent Number 6,248,584: Transcription coactivators (June 19, 2001); Patent Number 6,251,668: Transcription coactivators (June 26, 2001); Patent Number 20010005749: Aromatic amino acid catabolism enzymes (June 28, 2001); Patent Number 20010010909: Chromatin Associated Proteins (August 2, 2001).

The patents assigned to Incyte were: Patent Number 6,277,568: Nucleic acids encoding human ubiquitin-conjugating enzyme homologs (August 21, 2001); Patent Number 20010010913: Extracellular adhesive proteins (August 2, 2001).

The remaining patents included one patent assigned to Dendreon Corporation: Patent Number 6,194,152: Prostate tumor polynucleotide compositions and methods of detection thereof (February 27, 2001); two assigned to Bayer Corporation: Patent Number 6,262,333: Human genes and gene expression products (July 17, 2001); Patent Number 6,262,334: Human genes and expression products: II (July 17, 2001); and one assigned to a foreign corporation, Zeneca: Patent Number 6,265,560: Human Ste20-like stress activated serine/threonine kinase (July 24, 2001).

¹⁰⁸ The patents are: Patent Number 5,912,130: Human Homolog of the rat G protein gamma-5 subunit (Jun. 15, 1999); Patent Number 5,783,418: Human homolog of the rat G protein gamma-5 subunit (Jul. 21, 1998); Patent Number 5,932,442: Human regulatory molecules (Aug. 3, 1999); Patent Number 5,840,544: DNA encoding rantes homolog from prostate (Nov. 24, 1998).

examined the patent that Incyte claims to be the first issued EST patent,¹⁰⁹ and a patent thought to be an EST patent that issued earlier.¹¹⁰

2. Quantitative Search Results

I conducted some systematic searches of patents issued over the past five years to assess temporal trends in the number of EST patents and the use of various computational methods in those patents.

I looked at the temporal variability in patents in patents listing Class 536, Subclass 23.1 and 23.2-.7 (Table 1) to assess trends in the number of EST patents over the last five years. cursory inspection of patents in subclass 23.1 showed that not all but many of the patents listing this subclass were EST patents. The number of patents in these classes increased about three-fold from 1996 to 1998, and then remained fairly constant, with an average of 175 to 200 patents in subclass 23.1 issuing per month.¹¹¹

Table 1. The number of patents in Class 536 by subclass (23.1) or set of subclasses (23.1 to 23.7) for various two month intervals. Tallies for the two periods early in 2001 and 2000 are shown in parentheses.

YEAR	PERIOD	23.1	23.1 - .7
2001	6/1-8/1	425	233
2001	1/1-3/1	(333)	(197)
2000	6/1-8/1	351	186
2000	1/1-3/1	(396)	(188)
1999	6/1-8/1	329	154
1998	6/1-8/1	374	166
1997	6/1-8/1	235	91
1996	6/1-8/1	135	60

The USPTO declared in 1997 that it would issue patents on ESTs, and Incyte claims to have received the first EST patent in 1998 (Pat. No. 5,817,47)—although at least one patent that was issued in 1996 claims ESTs in addition to a full-length gene (Pat. No. 5,552,281). If patents in these subclasses prior to 1997 were not EST patents, then it is likely that a third of the patents in these classes after 1997 are not EST patents. If so, these numbers suggest that tens and perhaps a hundred EST patents issue every month.

I estimated references to various computational methods in EST patents as follows. I restricted my searches to patents in Class 536, Subclass 23.1 that claimed a polynucleotide sequence and included the terms “est” or “partial”

¹⁰⁹ Patent Number 5,817,479: Human kinase homologs (Oct. 6, 1998).

¹¹⁰ Patent Number 6,194,152: Prostate tumor polynucleotide compositions and methods of detection thereof (February 27, 2001).

¹¹¹ The number of patents issuing could be limited by the number of examiners or the general availability of resources for examination of patents at the USPTO.

within 2 words of the terms (sequence or cDNA). I then searched for each of the following terms by monthly intervals: BLAST, Clustal (to indicate reference to the Clustal W method), Waterman (to indicate reference to the Smith-Waterman method), Markov (to indicate reference to a Hidden Markov Model), and GCG (to indicate the use of GCG software). I present the total number of patents in each category by year except for 2001; in 2001, I estimated the tally for the year by doubling the number of patents in each category for the period from January 1 to July 1.

Table 2. The number of likely EST patents per year that mentioned each of several methods of computational analysis. *Twice the number observed for the period January 1 to July 1.

YEAR	M E T H O D				
	BLAST	Waterman	Clustal	Markov	GCG
2001*	48	28	22	8	32
2000	53	23	20	15	43
1999	96	50	17	12	88
1998	41	19	0	0	35
1997	3	2	0	0	1
1996	0	1	1	0	1

The data clearly show that references to BLAST and Smith-Waterman began to be incorporated into patents issuing in 1998. The following year, patents began to issue that provided reference to Clustal W analysis and Markov Models. The number of references was similar for all methods in all remaining years, except that BLAST and Smith-Waterman methods were referenced about twice as many times in patents that issued in 1999 as in other years after 1997.

As shown in Table 1, the number of patents in Class 536, Subclass 23.1 did not change significantly from 1998 to 2001. The tallies for the number of patents in the restricted set used to examine the computational methods was not made, but is likely similar. Thus, the tallies shown here may estimate the frequency of mention of the various methods in patents in this restricted set of patents. However, the data suggest that the USPTO did begin issuing EST patents after it announced in 1998 that it would do so. Because this announcement came mid-year, the tallies for 1998 may underestimate the rate of mention of the methods in this year.

These preliminary data indicate that the USPTO began, in 1998, to issue a significant number of patents in Class 536, Subclass 23.1 that claimed nucleotide sequences, likely mentioned partial cDNA or EST, and referenced a method of sequence alignment. Furthermore, the USPTO has continued to issue such patents, at a seemingly similar rate, since 1998.

C. Satisfying the Written Description Requirement

The quantitative data suggest that the USPTO is issuing EST patents that rely on computational methods. To assess whether, and if so, how these or other computational methods are being used to address the written description requirement, I examined twenty-three patents and the USPTO Training Materials for evaluation of the written description.¹¹² I review the legal criteria and then assess the patents in light of the law.

1. Synopsis of Legal Criteria

An invention must be adequately described to qualify for a patent. The written description requirement is set forth in Section 112 of the Patent Act, its application to gene patents was addressed by the CAFC in several cases in the early 1990s, and the USPTO published guidelines in January 2001 explaining how to apply the requirement to various biotech claims, including claims to ESTs.¹¹³ I briefly review the requirement here.

In general, the statute requires that the inventor describe the invention well enough to show “possession” of it. That is, the inventor must describe the invention in sufficient detail that a person “skilled in the art” would conclude the inventor actually invented the claimed invention.¹¹⁴

The CAFC determined in the early 1990s that, in order to describe a gene, an inventor must describe the DNA, purportedly in “structural” terms. For example, it is not enough to name the protein that the gene encodes and a method for isolating and sequencing the gene (even if it would be scientifically obvious how to isolate and sequence the gene). The inventor must give a “precise definition [of the DNA], such as by structure, formula, chemical name, or physical properties.”¹¹⁵ The rule was often (and inaccurately) simplified as requiring a description of the nucleotide sequence.

The CAFC acknowledged that a set of nucleotide sequences encoding a particular amino acid sequence could be deduced using the genetic code, but it emphasized the difference between deducing a set of possible sequences and knowing a naturally occurring sequence: If the amino acid sequence was newly discovered but the nucleotide sequence unknown, the inventor could claim only the set of all possible nucleotide sequences encoding it. But regardless whether the amino acid sequence for a protein was known or unknown, an inventor could discover and claim the nucleotide sequence that actually occurs in a particular organism.

The USPTO’s Written Description Guidelines acknowledge these basic points. They emphasize, though, that “there is no basis for a per se rule requiring disclosure of complete DNA sequences or limiting DNA claims to only the

¹¹² U.S. Patent and Trademark Office, Synopsis of Application of Written Description Guidelines, available at <http://wais.access.gpo.gov> [hereinafter Written Description Training Materials].

¹¹³ See Part II.B.1 further discussion of the CAFC’s rulings and Part II.B.3 for further discussion of the USPTO’s guidelines.

¹¹⁴ See, e.g., *Lockwood v. American Airlines*, 107 F.3d 1565, 1572 (Fed. Cir. 1997)

¹¹⁵ *Univ. of California v. Eli Lilly & Co.*, 119 F.3d 1559, 1556 (Fed. Cir. 1997).

sequence disclosed.”¹¹⁶ They therefore instruct examiners to consider “the level of skill and knowledge in the art, partial structure, physical and/or chemical properties, [and] *functional characteristics alone or coupled with a known or disclosed correlation between structure and function*” (emphasis added) in assessing the adequacy of a written description.¹¹⁷

2. Observed Uses of Computational Methods

An isolated DNA sequence that has utility may be claimed directly and is adequately described by giving its nucleotide sequence. Such a patented claim could easily be avoided, though, by changing a nucleotide so that the encoded amino acid sequence remains the same, or by changing an amino acid so that the function of the protein remains the same. Most inventors would like to state a claim that encompasses all these variants, and computational methods make that possible.

Computation methods cannot be used, though, to describe a set of nucleic acids that could vary in unpredictable ways. For example, a nucleotide sequence of a cDNA fragment or EST can often be shown by various sequence alignment methods to be homologous to a known DNA molecule that encodes a known protein of known function. However, if “gene” is defined to include naturally occurring regulatory elements and untranslated regions necessary and sufficient to mediate the expression of a cDNA, then the description of the cDNA fragment does not adequately describe the homologous gene. The USPTO Training Materials explain that the description is inadequate because “there is no known or disclosed correlation between th[e protein’s] function and the structure of the non-described regulatory elements and untranslated regions of the gene.”¹¹⁸

In short, computational methods can be used to describe a claimed set of nucleic acids when all the members of the set are expected to have the same function because of structural similarities. I found three methods for expanding the scope of a claim to a DNA sequence: by using the genetic code to define all the nucleic acids encoding the same polypeptide, by using percent identity to describe structurally similar sequences, and by identifying functional variants of particular amino acids. I discuss each in turn.

a. Use of the Genetic Code and Combinatorics

The most obvious way to define a set of nucleic acids that vary structurally but not functionally takes advantage of the degeneracy of the genetic code. Because there is more than one codon for many of the amino acids, there may be

¹¹⁶ Written Description Guidelines, *supra* note 31, at 1101 (“Describing the complete chemical structure, i.e., the DNA sequence, of a claimed DNA is one method of satisfying the written description requirement, but it is not the only method.”).

¹¹⁷ Written Description Guidelines, *supra* note 31, at 1106

¹¹⁸ Written Description Training Materials, *supra* note 112. Even if “gene” is not so defined, the description of a single cDNA is probably inadequate to claim all nucleic acids comprising it because it is not necessarily representative of that class; a “representative number” of such fragmnts are needed. *Id.* at 31-32.

a large number of nucleotide sequences that code for the same amino acid sequence. Defining that set of nucleotide sequences is a straightforward matter of mapping and combinatorics—even though there may be a very large number of nucleic acid sequences coding for a particular amino acid sequence (especially if the amino acid sequence comprises more than a few amino acids).

The USPTO Training Materials¹¹⁹ acknowledge the reliability of this association between nucleotide structure and polypeptide structure. They explain that a claim to “[a]n isolated DNA that encodes protein X (SEQ ID NO: 2). . . .” adequately describes a genus of molecules because “a person of skill in the art could readily envision all the DNAs degenerate to SEQ ID NO:1 by using a genetic code table” and “[o]ne of skill in the art would conclude that [the] applicant was in possession of the genus based on the specification and the general knowledge in the art concerning a genetic coding table.”¹²⁰ Thus, the genetic code and combinatorial methods can be used to describe and claim the set of DNAs that encode a particular polypeptide.

The code is thus used to infer a set of nucleic acids encoding an experimentally determined amino acid sequence. For example, Incyte determined the amino acid sequence of a human ubiquitin-conjugating enzyme (“SEQ ID NO:2”) and then patented the set of nucleic acids encoding that enzyme by claiming “[a]n isolated and purified polynucleotide encoding a polypeptide comprising an amino acid sequence of SEQ ID NO:2.”¹²¹ The code can also be used to infer the amino acid sequence from an experimentally determined nucleic acid sequence. For example, Incyte inferred the amino acid sequence of a protein it called “prostate expressed chemokine” from the cDNAs sequences it identified in a prostate cDNA library, and then claimed all the nucleic acids encoding that enzyme.¹²²

All of the recently issuing patents that were assigned to DuPont or Incyte used this technique to claim a genus of DNAs encoding a given amino acid sequence. (XXX Add excerpts from patents explaining this.)

b. Use of Percent Sequence Identity

Perhaps the simplest way to define a set of similar amino acid sequences—or a set of nucleic acids encoding a set of similar amino acid sequences—relies on the similarity of their sequences to a described sequence. Such similarity is usually defined by the percentage of nucleic acids or amino acids that are identical (“percent identity”) when a sequence in the set is aligned in some way with the described sequence.¹²³ The definition of a set of sequences by percent

¹¹⁹ Written Description Training Materials, *supra* note 112, at 41-42.

¹²⁰ *Id.*

¹²¹ Patent Number 6,277,568.

¹²² Patent Number 5,840,544 (claiming “A purified polynucleotide encoding a polypeptide with an amino acid sequence shown in SEQ ID NO:2.”).

¹²³ All of the EST patents assigned to DuPont noted simply that “[s]ubstantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art.”

identity presumes the use of some method of sequence alignment, and the percent identity depends upon how the sequences are aligned.¹²⁴ If gaps are introduced to align the sequences, the corresponding amino acids or nucleic acids are typically ignored in calculating the percent identity.¹²⁵

The USPTO Training Materials provide an example of the valid use of measures of percent identity to describe a set of proteins.¹²⁶ In the example, the inventor claims all variants of a protein having amino acid sequence X “that are at least 95% identical to X *and* catalyze the reaction of A B” (emphasis added).¹²⁷ This example thus alludes to a potential problem: Proteins that are at least 95% identical to X in structure might *not* be functionally similar. The example given addresses this problem by constraining the set of structurally similar proteins to those that are also functionally similar; it does not discuss any particular method alignment.

Alignment methods are used most simply to describe a set of nucleic acids that are similar to one or more specified nucleic acid sequences. For example, two patents issued recently to Bayer claim “[a]n isolated nucleic acid molecule consisting of a nucleotide sequence at least 85% identical to a sequence selected from the group consisting of SEQ ID Nos. [1, 2, . . . X]”.¹²⁸ Alignment methods are also used to describe a set of polypeptides that are similar to one or more specified amino acid sequences. Those amino acid sequences may be deduced from an isolated nucleic acid sequence using the genetic code. For example, one of several EST patents issued to DuPont claims “[a]n isolated polynucleotide comprising . . . a nucleotide sequence encoding an isoleucyl-tRNA synthase, wherein the amino acid sequence of the synthase and the amino acid sequence of [sequence 2, 4, 6, or 8] have at least 80% identity based on the Clustal alignment method . . .”¹²⁹

The method used to align the sequences is often but not always specified in the claims.¹³⁰ However, the minimum degree of similarity between the given

¹²⁴ The specification will usually describe at least one such method of sequence alignment. One patent noted several, including FASTA, BLAST, or ENTREZ (as part of the GCG package), Needleman and Wunsch, and Smith-Waterman methods. Patent Number 6,262,333.

¹²⁵ Percent Identity is defined in one patent as “the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the native sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity,” Patent Number 6,194,152. As explained in another patent, “[t]he percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity.” Patent Number 6,277,568.

¹²⁶ Written Description Training Materials, *supra* note 112, at 54.

¹²⁷ *Id.*

¹²⁸ Patent Number 6,262,333 and Patent Number 6,262,234.

¹²⁹ Patent Number 6,271,441. Very similar claims are made in Patent Number 6,251,668 and Patent Number 6,255,090.

¹³⁰ Almost all of the DuPont patents specify the use of a Clustal alignment in the claims and do not describe any methods in the specification; others describe several in the specification but mention none in the claims. In contrast, a patent issued recently to Dendrion is probably unnecessarily

sequence and the sequences in the claimed set must be specified in the claims. This cutoff is clearly arbitrary. Requiring a higher degree of sequence identity means that the claimed sequences are less likely to differ functionally, all equal. Thus, it is common to see a series of claims that differ only in the minimum degree of similarity required. For example, the first claim requires only 80 or 85% sequence identity, a second claim requires 90% identity, and a third claim requires 95% identity. This strategy admits the possibility that a claim to sequences that are only 80% identical might be invalid.¹³¹

Claims to a nucleotide sequence “*encoding protein A*” that has “at least X% similarity to sequence S” were common in the surveyed patents. They are potentially problematic, though, because they do not explicitly require that the claimed structurally similar sequences have the same function as the isolated sequence or sequences.¹³² Such functional similarity could be inferred by the reference to the protein by its name. However, several patents claimed a nucleotide sequence “encoding a protein *having the activity of protein A*” that has “at least X% similarity to sequence S.”¹³³ They thereby restricting the claimed set of structurally similar nucleic acids to those that have a particular biochemical function.

c. Use of Structural Variants Having Similar Function

A more complex but potentially more accurate way to define a set of nucleic acids that vary structurally but not functionally considers the effect of amino acid substitutions on the structure and function of a molecule. Many amino acids may be replaced with other amino acids without changing the structure or function of the molecule. Information about the substitutability of amino acids can therefore be used to describe a set of nucleic acid sequences encoding a set of functionally similar polypeptides.

specific about the methods to be used when it claims “[a]n isolated polynucleotide having at least 95% sequence identity to nucleotides 43-3327 of the sequence of SEQ ID NO: 14, wherein % identity is calculated using the LALIGN program found in the FASTA Version 2.0 suit of programs using default parameters with the BLOSUM50 matrix, a ktup of 2 and a gap penalty of -12/ -2 . . .” Patent 6,194,152

¹³¹ A patent typically has many claims, which vary in scope from very broad to very narrow. The broadest claims are most likely to be found invalid by a court, but the narrowest claims are unlikely to be infringed because they are easy to work around. The use of a series of claims of decreasing scope is a strategy to ensure the broadest possible valid claim. This strategy was used in most of the DuPont patents that I read; it was not used, for example, in Patent Number 6,242,256.

¹³² Compare such a claim to Example 14 in the Written Description Training Materials, *supra* note 112; *see also* text accompanying note 123.

¹³³ For example, Patent Number 6,262,345 claimed “[a]n isolated polynucleotide comprising . . . a nucleotide sequence *encoding a polypeptide having glycogen synthase kinase activity*, . . . wherein the amino acid sequence of the polypeptide and the amino acid sequence of [sequence 1, 2, . . . X] have at least 90% identity based on the Clustal alignment method. . .”. A claim in Patent Number 6,274,379 is similar. Patent Number 6,277,568 claimed “[a]n isolated and purified polynucleotide having at least 90% sequence identity . . . [to] the polypeptide of SEQ ID NO: 2, and which *encodes a polypeptide that retains ubiquitin-conjugating activity*.”).

Amino acids may differ, for example, “in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues.”¹³⁴ If the substituted amino acid has similar characteristics, the change is “conservative” and is unlikely to change the structure or function of the protein. Substitutions involving amino acids with very different attributes are “non-conservative” and may produce “[s]ubstantial changes in function or immunological identity. . . For example, substitutions may be made which more significantly affect the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure, the charge or hydrophobicity of the molecule at the target site, or the bulk of the side chain.”¹³⁵

The USPTO training materials do not discuss the use of methods of amino acid substitution to describe a genus of nucleic acid. But many inventors discuss “variants” of a polypeptide in the specification of the patent made by either conservative or non-conservative substitutions.¹³⁶ They often indicate that conservative variants are within the scope of the claimed invention¹³⁷ and may specify methods for determining conservative substitutions.¹³⁸

D. Satisfying the Utility Requirement

The USPTO is clearly issuing patents that rely on computational methods to *describe* a genus or set of nucleic acid sequences. To assess whether, and if so, how computational methods are being used to establish the *utility* of patents for partial cDNAs or ESTs, I examined the USPTO Utility Training Materials¹³⁹ and the same twenty-three patents that I used to assess whether and how computational methods are being used to address the Written Description Requirement. As in the last section, I review the legal criteria and then assess the patents in light of the law.

1. Synopsis of Legal Criteria

An invention must be useful to qualify for a patent. The utility requirement is set forth in Section 101 of the Patent Act, but its application to gene patents has

¹³⁴ Patent Number 6,277,568.

¹³⁵ Patent Number 6,194,152.

¹³⁶ A “variant” . . . may have an amino acid sequence that is different by one or more amino acid “substitutions”. The variant may have “conservative” changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have “nonconservative” changes, e.g., replacement of a glycine with a tryptophan.” Patent Number 5, 840,544.

¹³⁷ For example, one inventor indicate that “[d]eliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of EXADH is retained.” Patent Number 20010010913.

¹³⁸ *E.g.* “Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.” Patent Number 5,840,544.

¹³⁹ Utility Training Materials, *supra* note 97.

not yet been addressed by the CAFC; nonetheless, the USPTO published guidelines in January 2001 explaining how to apply the requirement to various claims in biotech patents, including patents on ESTs.¹⁴⁰ I briefly review the requirement here.

The Supreme Court held in 1966 that an invention must have a real world, practical utility.¹⁴¹ In that case, it found that a process for making a chemical that was used only in research lacked such utility.¹⁴² Various appellate court cases since then have held, in addition, that an invention must have a “specific and substantial” utility.¹⁴³ And prior to 1996, the USPTO required its examiners to determine whether an invention had a “credible” or well-established utility.

The USPTO’s new Utility Guidelines require that all claimed inventions have a “specific, substantial, and credible utility.”¹⁴⁴ A “credible” utility is logically consistent with the asserted facts. For example, since at least some nucleic acids can be used as probes or chromosome markers, it is credible that any particular DNA can be used in this way. A “substantial” utility is a real-world use. For example, a claim that a nucleic acid is useful as a dietary protein supplement is insufficient; it is a “throw-away” use that lacks substance. A “specific” utility is particular to the subject matter claimed. For example, if a nucleic acid is claimed to be useful as a gene probe or chromosomal marker, then the specific DNA target must be disclosed.

The new Utility Guidelines raised the bar on utility because inventions must now have a substantial and specific use—not just a credible one. However, the procedural requirements for evaluating utility clearly favor the patent applicant. USPTO personnel must presume that statements by applicants are true, and they must allow applicants to rebut any prima facie finding of no utility.¹⁴⁵

Despite the wishes of many commentators, the new Utility Guidelines do not create a “per se” rule against homology-based assertions of utility. The PTO said there is no “scientific evidence that homology-based assertions of utility are inherently unbelievable or involve implausible scientific principles.”¹⁴⁶ Instead of an across-the-board rule, the PTO declared that assessments of utility would be “fact dependent” and determinations would be made “on the basis of scientific evidence.”¹⁴⁷

2. Observed Uses of Computational Methods

¹⁴⁰ See Part II.B.1 further discussion of the CAFC’s rulings and Part II.B.3 for further discussion of the USPTO’s guidelines.

¹⁴¹ 383 U.S. 519 (1966).

¹⁴² 383 U.S. 519 (1966).

¹⁴³ need to get some examples or summary citations here.

¹⁴⁴ The guidelines also discuss a “well-established” utility test, but even well-established utilities must be specific, substantial, and credible. However, if the utility is well-established, it need not be asserted explicitly in the patent. For an excellent review and critique of the Utility Guidelines, see Worrall, *supra* note 96, 132.

¹⁴⁵ Utility Training Materials, *supra* note 97; Worrall, *supra* note 96, at 132.

¹⁴⁶ Utility Guidelines, *supra* note 21, at 1096.

¹⁴⁷ *Id.*

Claims to nucleic acid sequences as compositions of matter must assert a credible and specific practical utility for the sequence. A nucleic acid may be useful because it encodes a particular known and useful protein, or because it can be used as a probe to identify or locate the full-length nucleic acid encoding a specific known and useful protein. Even if the function of the encoded protein is unknown, a nucleic acid that is transcribed in some cells but not others may be useful as a diagnostic tool—if its presence is correlated, for example, with a particular disease.¹⁴⁸

The utility of a nucleic acid thus often (but not always!) requires information about the biological function of the particular encoded polypeptide.¹⁴⁹ Such information may be obtained directly and experimentally in the laboratory. It may also be inferred from comparison to sequences whose function has already been directly and experimentally determined in the laboratory. The latter technique requires computational methods of sequence alignment and is the more contentious method for establishing the utility of a sequence.¹⁵⁰

In short and despite the debate, computational methods may be used to establish the utility of ESTs by comparing the partial or complete cDNA sequences to full length sequences encoding proteins of known function, and then inferring the function of the protein partially or completely encoded by the cDNA sequence.¹⁵¹ The patents that I examined used computation methods in precisely this fashion.

a. To Identify the Polypeptide Encoded by a Sequence

¹⁴⁸ “[T]he utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have a specific and substantial utility because, e.g., it hybridizes near a disease-associated gene or it has a gene-regulating activity.” Utility Guidelines, *supra* note 21, at 1095.

¹⁴⁹ As demonstrated in Example 9 of the Training Materials, a set of cDNAs is not useful merely because they encode part of some protein and can be used as probes to identify the full length nucleic acid encoding that protein; the particular protein that they encode must be determined and specified. Utility Training Materials, *supra* note 97, at 50-53. However, establishing the function of the encoded polypeptide is only one way to establish real world utility. Real-world utility and the function of the gene are frequently but inaccurately treated as synonyms. For example, the statement that “[p]atent applications that do not specify exactly what a gene or gene fragment is, or what its function is, will not be considered for approval, according to the new guidelines” confuse real-world utility and gene function. *Updated Guidelines from Patent Office Similar to Old Ones*, BIOTECHNOLOGY NEWSWATCH at 9 (Feb. 5, 2001).

¹⁵⁰ Experimental evidence is typically considered more reliable than the “hypotheses” or “theoretical results” resulting from the analysis of genomic databases. For example, one author noted that “[o]pen reading frames vary widely in the degree to which their encoded proteins assert a credible specific and substantial utility” and then explained that “[a]t one extreme, DNA sequences encoding proteins having experimentally verified function and use satisfy the utility requirement. At the other extreme, the function of an unknown protein can be *hypothesized* based on sequence similarity, or homology, to known sequences with known function.” Worrall, *supra* note 96, at 139. See also notes 81-83, *infra*, and accompanying text.

¹⁵¹ “[W]hen a patent application claiming a nucleic acid asserts a specific, substantial, and credible utility, and bases the assertion upon homology to existing nucleic acids or proteins having an accepted utility, the asserted utility must be accepted by the examiner unless the Office has sufficient evidence or sound scientific reasoning to rebut such an assertion”. Utility Guidelines, *supra* note 21, at 1096.

The USPTO training materials provide an example of the use of computational methods to assess the structure and function of the protein encoded by a full open reading frame, and thereby satisfy the utility requirement.¹⁵² In the example, a cDNA library is prepared, clones are sequenced, and their open reading frames are identified. The nucleic acid sequence is found to be similar to various known ligases, presumably by doing sequence alignments. The amino acid sequence that it encodes is compared to a consensus sequence of the known ligases, and “reveals a similarity score of 95%.” The nucleic acid sequence also has a “high homology” to DNA Ligase encoding nucleic acids, and has only 50% “homology” to the next most similar sequence. The Training Materials indicate that these disclosures are sufficient to establish that the claimed sequence encodes a DNA ligase and, since DNA ligases have “a well-established use in the molecular biology art,” the disclosure establishes a utility for the claimed sequence.

This basic method was used in the patent that Incyte claims was the first EST patent to issue.¹⁵³ The patent describes 44 partial cDNAs that were isolated from various cDNA libraries. According to the specification, each nucleotide and its corresponding amino acid sequence was compared to sequences in GenBank using a proprietary search algorithm,¹⁵⁴ and homologous regions were identified. The specification does not provide any statistics or results from the analysis of the described sequences.¹⁵⁵ It does, however, note that “protein kinases are associated with basic cellular processes such as cell proliferation, differentiation and cell signaling” and asserts that “[k]inase nucleotide sequences are [therefore] useful in diagnostic assays used to evaluate the role of a specific kinase in normal, diseased, or therapeutically treated cells”. A patent issued soon thereafter is very similar.¹⁵⁶

The same approach was used in two recently issued patents that were assigned to Incyte. In a patent on human ubiquitin-conjugating enzymes, Incyte took clones from a prostate cDNA library and then used BLAST to ascertain that one of them had “chemical and structural similarity with *Arabidopsis thaliana* [a plant] ubiquitin-conjugating enzyme (GI 1707021).”¹⁵⁷ The threshold for the BLAST was given as 10^{-25} for nucleotides and 10^{-8} for polypeptides. Thus, the probability that the newly discovered polypeptide sequence was the same as the *Arabidopsis* gene purely by chance was less than 10^{-8} , and the new sequence was inferred to be a ubiquitin-conjugating enzyme. Because ubiquitin is part of a pathway for

¹⁵² Utility Training Materials, *supra* note 97, at 53-55.

¹⁵³ Patent Number 5,817,479

¹⁵⁴ The search algorithm was “developed by Applied Biosystems and incorporated into the INHERIT TM 670 Sequence Analysis System” and used “Pattern Specification Language (TRW Inc, Los Angeles, Calif.)” to determine regions of homology.

¹⁵⁵ It merely explains in general that dot matrix plots were used “to distinguish regions of homology from chance matches” and Smith-Waterman alignments were used “to display the results of the homology search”. The specification also explains that BLAST could also be used to find High-scoring Segment Pairs, whose probability score meets a predetermined threshold level of significance.

¹⁵⁶ Patent Number 5,840,544.

¹⁵⁷ Patent Number 6,277,568.

selective protein degradation, the claimed sequence was asserted to be “useful in the diagnosis, treatment, and prevention of cancer, autoimmune disorders, and neuronal disorders.”

In a second patent issuing recently and assigned to Incyte, this one on human extracellular adhesive proteins,¹⁵⁸ Incyte used more seemingly more comprehensive but only vaguely described methods to identify homologs of known function.¹⁵⁹ And the specification merely asserts that the sequences are useful for diagnosing, treating or preventing disorders associated with expression of the proteins.¹⁶⁰

Eleven EST patents that recently issued and were assigned to DuPont were very similar in structure and approach.¹⁶¹ They all claimed partial cDNAs from plants, and the functions of the claimed sequences were usually determined by finding homologs of known function from humans or other animals. Most of them relied on BLAST to compare the isolated cDNAs to sequences in various databases.¹⁶² And most of them included p-values for the comparison of each described sequence and the homolog used to infer its function, as well as the percent identity of the described sequence and its homolog.¹⁶³ P-values were typically smaller than 10^{-25} , and the claimed cDNAs were usually more than 70% similar to the sequences of known function.¹⁶⁴

The patents assigned to DuPont were clearly distinguishable from the patents assigned to Incyte. They relied fairly exclusively on the described findings of

¹⁵⁸ Patent Number 20010010913.

¹⁵⁹ “The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.” Patent Number 20010010913.

¹⁶⁰ However, it recites a list of potentially treatable diseases that is 31 lines (>300 words) long!

¹⁶¹ See note 107, *infra*.

¹⁶² Typically, the cDNA sequences were “analyzed for similarity to all publicly available DNA sequences contained in the “nr” database using the BLASTN algorithm, [and] . . . [t]he DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the “nr” database using the BLASTX algorithm” Patent Number 6,255,090. Slightly different language is used in Patent Number 6,255,114.

¹⁶³ The specifications explained that “the P-value[s] (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as “pLog” values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST “hit” represent homologous proteins.”

¹⁶⁴ In two patents claiming transcription coactivators from plants by homology to mouse and human proteins, the claimed sequences were only 19-46% identical to the sequences of known function; the p-values were all less than 10^{-20} . Patent Number 6,255,090 and Patent Number 6,271,441.

homology to establish utility; that is, they typically did not include any additional laboratory work on the claimed sequences. The asserted utility of the cDNAs claimed in the DuPont patents also tended to be less explicit and more general in nature than the utility asserted in the Incyte patents. In general, the DuPont patents relied on sequence comparisons to claim sequences identified in an early stage of research, whereas the Incyte patents used sequence comparisons in combination with a variety of laboratory findings to justify their claims to such sequences.

b. To Show that the Polypeptide is Unknown

An expressed nucleic acid sequence may be useful even if the biological function of the protein that it encodes is unknown. The utility arises from knowledge of factors that are correlated with the expression of the sequence. For example, many sequences are expressed only in cancerous cells; these sequences are therefore useful as indicators of cancer. Several recently issued EST patents use computational methods to demonstrate that sequences are novel, and then assert utility based on their specificity to particular types of tumor or cancer cells.¹⁶⁵

Computational methods may also be used to establish whether or not a sequence is known or has known homologs so that research and patents can be designed appropriately. For example, if there are no known homologs of a sequence, its function cannot be inferred from the analysis of genomic databases but additional research may be advantageous. If the exact sequence is already described, additional research is unnecessary and the sequence itself cannot be claimed. However, it is possible that the sequence can be claimed as an indicator of disease.¹⁶⁶

E. Discussion and Critique

Randall Scott of Incyte asserts that “there are many, many families [of genes] now for which the function can be reasonably predicted from the structure, and [our ability to predict function from structure gets] better and better . . . every year.”¹⁶⁷ He was presenting testimony to a Congressional Hearing on Genomic Inventions, arguing for the patentability of ESTs whose utility was established

¹⁶⁵ For example, computational methods were used to establish that sequences specific to human prostate tumor cells were novel. Patent Number 6,194,152.

¹⁶⁶ A patent assigned to Incyte for consensus sequences from cancer cells reports whether or not each sequence has a known homolog; if it does, then the specification adds that the sequence has now been observed from a cancer cell. Similarly, two patents assigned to Bayer are careful to distinguish “1) matches to known human genes, 2) matches to human EST sequences, and 3) no significant match to either 1 or 2, and therefore a potentially novel human sequence.” Patent Number 6,262,333 and Patent Number 6,262,334. Patent Number 5,932,442.

¹⁶⁷ Dr. Randal W. Scott, President And Chief Scientific Officer, Incyte Genomics. Prepared Statement at Congressional Hearing on Genomic Inventions, *supra* note 1.

from comparison to sequences of known function. His comments reflect both legal and scientific problems in inferring function from structure.

Patent law requires that every invention be adequately described and have a practical utility. The courts have made it clear that a nucleotide sequence can only be patented when its “structure” is adequately described¹⁶⁸. However, it must also have an asserted utility, which is often only possible when the function of the encoded protein is known. Thus, to patent a gene sequence or set of gene sequences, one must usually know both its structure and the function of the encoded protein or proteins.

Discussions of about the patentability of genes, especially partial cDNAs or ESTs, have tended to focus on the utility requirement. However, the utility requirement and the written description requirement are flip sides of same coin, because both create issues about the use of computational methods to translate between structure and function.

The genetic code provides one biological reality that has required an adjustment to the idea that a nucleic acid must be structurally described in order to be patented. It allows one structure (i.e. an amino acid sequence) to be reliably translated into another (i.e. a nucleotide sequence), and vice versa if the reading frame is known. The legal world struggled to distinguish a claim to a “theoretical” genus of nucleotide sequences from a claim to a naturally occurring nucleotide sequence, but the basic idea is simple and sound. All the nucleic acids that encode a polypeptide can be patented if the amino acid sequence of the polypeptide is known, because all those sequences code for the same polypeptide structure.

The description of all the nucleotides that encode a set of “similar” amino acid sequences (or a set of “similar” nucleotides) by measures of percent identity is more problematic because structural similarity does not correlate exactly with functional similarity. Some differences in some amino acids are more important than others. Definition of sequences by their percent similarity is computationally simple and it provides a bright-line test for deciding whether two sequences are similar or not. However, unless the definition is extremely rigid, so that only very similar sequences are considered the same, it will probably include sequences that encode polypeptides with other functions—however slightly.

Ideally the definition of a set of sequences will clearly distinguish those that are functionally similar and those that are not, and that threshold can be accurately determined. In other words, the receiver operator curve for the method will have a sharp transition, indicating a clear separation between true positives and false positives.

The USPTO addresses this problem of identifying functionally similar sequences by proposing the definition of a set of nucleotides that share some degree of structural similarity *and have the same activity as the given sequence*. However, this technique poses legal problems because the CAFC seems to have asserted that functional attributes cannot be used to define a claimed structure. The court could distinguish this technique by noting that it merely limits a set of

¹⁶⁸ Case law forbids the use of functional attributes to describe a claimed composition. See Part II.B.1.

structurally similar sequences, but that seems to push the structural definition rule beyond the bounds of legal or scientific reason.

It is, however, well-known in the art that some methods for comparing sequences or defining sets of sequences are better than others in identifying sequences of similar function based upon their sequence similarity. For example, a gapped Blast may provide a more functionally accurate analysis of sequence similarity than an ungapped Blast if there are many insertions or deletions. A multiple alignment that uses an appropriately selected substitution matrix results in fewer false positives than one that assumes all substitutions are equally likely. Hidden Markov Models may provide a better model for identifying functionally similar sequences than, for example, a simple gapped BLAST search.

It is also well known in the art that more exhaustive, more sensitive methods tend to be slower and are often more complex than others. In many cases, a simple, approximate method is sufficient to identify all functionally similar sequences; in other cases, it may not. The sufficiency of a method for assessing the similarity of sequences or defining a set of sequences will be case-specific, depending on the actual sequence landscape and the extent of clustering within that landscape. All else equal, simpler methods are probably preferable.

Patent applicants have addressed the problem of identifying functionally similar sequences by discussing the difference between conservative and non-conservative changes and appealing to the knowledge of one skilled in the art. This approach may avoid the legal problem of using function to define a structure, since the approach is based on inferences of functional equivalency of parts of the polypeptide rather than functional equivalency of the entire protein. It is philosophical related to the use of substitution matrices, but more flexible.

In short, the use of any method for assessing sequence similarity is potentially problematic when the measure of similarity is used to infer function. Methods that account for the greater likelihood of particular amino acid substitutions assume that such changes will not affect the proteins function, and may permit more accurate inferences of function from structure. Percent identity for sequences aligned with a model that uses reasonable parameters is probably a good and simple rule of thumb for describing a set of sequences that are likely to have similar function. The adequacy of the threshold may vary with the protein, though; for example, stricter thresholds may be necessary when function varies greatly with small changes in structure. Similarly, methods that assess the probability that a sequence is structurally similar to a protein of known function can probably often be used reliably, especially when the sequences are very similar.

V. CONCLUSION

The USPTO is issuing large number of patents on ESTs whose utility is often established by comparing them to sequences of known function, and allowing claims to sequences that share some critical but arbitrary percentage of identical nucleotides or amino acids. The methods used to infer utility and describe a

claimed set of sequences appear scientifically sound and will likely produce reliable results in most cases. Sequences that encode proteins with different functions are best excluded by reference to their difference in function.

The recent guidelines issued by the USPTO have clarified their position with respect to a number of issues: “A DNA sequence per se is not patentable. Isolated genes can be patented. The entire gene sequence doesn’t have to be disclosed. The gene must have a use. An EST must have a use. The applicant only has to disclose one use for the gene. The gene’s function doesn’t have to be known in order for the DNA to be useful.”¹⁶⁹ The guidelines are a declaration that “the patenting of genomic inventions is consistent with our law and with our practice.”¹⁷⁰

However, the CAFC has not ruled on either of the two more contentious issues involving the use of computational methods in describing partial cDNAs and identifying their utility. It is possible that the court will view these issues quite differently than the USPTO and scientists. The simultaneous failure of politicians to appreciate the scientific validity of genomic methods and the sophistication of patent applications with claims to ESTs is remarkable.

The business world is likely to have more effect on the issuance of patents than the courts and the USPTO. The USPTO says that it is seeing more “generation three” EST patents—patents whose utility is supported by more than “mere homology,” and fewer “generation two” EST patents, whose utility is supported only by homology. However, my reading of several recent patents suggests that there are differences in the patenting strategies of companies in the human gene business and companies in the plant gene business. These strategies may reflect differences in publicity and political pressure.

In sum, the use of computational methods to identify the utility of ESTs and describes claims to similar sequences is probably scientifically and legally feasible—although not without problems on either account. How the issuance of such patent affects the progress of research and the development of industries that rely on genetic information is another issue.

¹⁶⁹ Van Brunt, *supra* note 24.

¹⁷⁰ Todd Dickinson. Statement at Congressional Hearing on Genomic Inventions, *supra* note 1.