



The Role of Intellectual Property in Bioinformatics in Bulgaria

Tereza Trencheva¹, Ivan Trenchev²

¹State University of Library Studies and Information Technologies, SULSIT (Bulgaria)

²South-West University "Neofit Rilski" (Bulgaria)

¹t.trencheva@unibit.bg, ²trenchev@swu.bg

Abstract

Intellectual Property (IP) rights are essential and very important in today's technology-driven world. In fact IP is directly related to the information, which contains in the objects themselves. In other words, the IP is the ownership of the information, which intellectual products contain themselves, and their creators and legal possessors have full ownership of them. Bioinformatics in its nature refers to the intellectual, high-tech science section, where the results obtained largely depend on the development of the creative potential of scientists, but also influenced by technical support. It is relatively new discipline that has gained much recognition in the past few years. Basically, bioinformatics is the convergence of analytical and computational tools with the discipline of biological research. Scientific interest in the problems of Intellectual Property and its relation to the different parts of Bioinformatics in the digital realm takes an important research position in the last few years. The article aims to describe the nature of bioinformatics and its relation to intellectual property. Bioinformatics is presented by its three main categories: first of all biological sequences such as DNA, RNA and protein sequences; second of all databases in which these sequences are organized; and third of all software and hardware designed to access, organize and analyze information contained within this sequences and databases. The article analyzes how these three categories of bioinformatics are related to the Patent and Utility Models Law, Copyright Law, Trade Marks and Geographical Indications Law and where protection may be available and practical. The overview of the investigated problem is very important especially for the interdisciplinary character because it makes a bridge between the Bulgarian legislative system and Bioinformatics.

Keywords: bioinformatics, DNA, RNA, intellectual property, copyright, patents, databases, trade marks

1. Introduction. Brief description of Bioinformatics

It is said that 'need is the mother of all inventions'. The world is changing very fast with the inventions of new technologies and sophisticated gadgets to comfort the human life [1]. Bioinformatics is an exciting new area of interdisciplinary science. While many definitions exist, in general bioinformatics blends technologies from computing, mathematics and statistics to help solve biological problems. Bioinformatics is about creating biological information and knowledge by the acquisition, archiving, integration, analysis and interpretation of biological data. Biotechnology, drug development, medicine, cancer research, agriculture and plant science are just a few of the many areas in which bioinformatics is having a massive impact, and in which the demand for skilled bioinformaticians is rising rapidly [2,8]. Bioinformatics is a relatively new discipline which has gained much recognition in the last decade. In fact Bioinformatics is a quite new interdisciplinary scientific area, which operates at the intersection of Biology (Molecular Biology, Biotechnology, Genetic Engineering), Chemistry (Biochemistry), Mathematics, Engineering, Computer Systems and Computational Biology. It can be defined in simple terms as the use of informatics for the investigation of biological databases [9].

It can be hard to describe what bioinformatics is, it is somewhat easy to explain what bioinformatics is not - it is not just using computers to look at biological sequence data. Bioinformatics has very fast become established as a main branch of modern bioscience, with an assortment of sophisticated tools for analyzing proteins and genes, in silico, in vitro, and in vivo.

In common, we can say that bioinformatics is concerned, but not limited to:

- modeling of biological systems and functions;
- analysis of laboratory data;
- generation models based on accumulated data from experiments;
- study of new data using mathematical models;
- recognition motifs in the experimental data;
- predicting functions of genes and proteins.



- in silico experiments.

Using a computer in bioinformatics is ideal for the job of analyzing biological complex data sets, such as sequence data for proteins and nucleic acids. This often involves a series of computing procedures, each of which may be comparatively simple in isolation. The calculation may need to be repeated millions of times; therefore, it is momentous that this is achieved quickly and exactly.

Before one can understand intellectual property protection for bioinformatics, it's necessary to understand the nature of the various components that comprise the field of bioinformatics.

For the purpose of this article, in general bioinformatics is presented by its three main categories: first of all biological sequences such as DNA, RNA and protein sequences; second of all databases in which these sequences are organized; and third of all software and hardware designed to access, organize and analyze the information contained within this sequences and databases.

Another focus on this paper is that it is discussed the current scenario of legal protection in bioinformatics inventions and products, which may vary country to country but the focus here is placed on the Bulgarian Legislation, where the area of bioinformatics is restricted to Patents, Copyright and Trade secrets and Trade Marks of intellectual property rights.

2. Biological Sequences - DNA, RNA and PROTEIN SEQUENCES

A biological sequence is a single, continuous molecule of nucleic acid or protein. Biological sequences can be presented as follows:

- DNA (nucleotides, 4 types): storage of the genetic information;
- RNA (nucleotides, 4 types): bridge from DNA to protein;
- Protein (amino acids, 20 types): active molecules.
- Genetic code: deciphering genetic information.

The main role of DNA molecules is the long-term storage of genetic information. The human genome or our genetic blueprint is written in four alphabet of chemical compounds called adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T) . These four bases are attached to the sugar/phosphate to form the complete nucleotide [7,8,9]. Like DNA, RNA is made up of a long chain of components called nucleotides. Each nucleotide consists of a nucleobase, a ribose sugar, and a phosphate group. The sequence of nucleotides allows RNA to encode genetic information . The chemical structure of RNA is very similar to that of DNA, with two differences-(a) RNA contains the sugar ribose while DNA contains the slightly different sugar de-oxyribose, and (b) RNA has the nucleobase uracil while DNA contains thymine. The protein is a sequence of 20 standard amino acids.

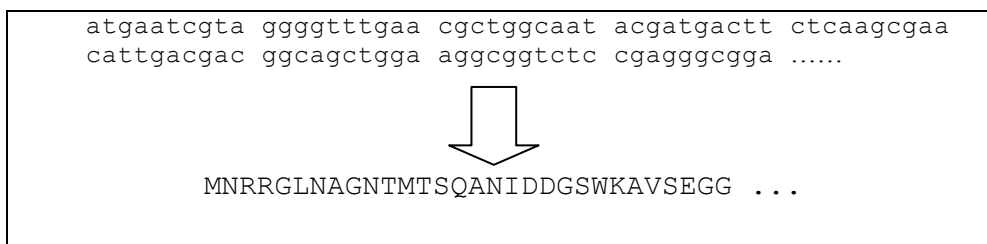


Table 1 Scheme of the translation of the genetic information from DNA to protein

3. Bioinformatics Databases

According one definition of S. Suryawanshi and I. Jacob - Bioinformatics Database “is the storage or collection of information. In scientific perspective, these are libraries of life science information, collected from scientific experiment, published literature, and high throughput screening and computational analysis. It includes experimental study of genomics, proteomics, metabolomics and microarray gene expression and etc.” [1].

Bioinformatics Databases play a key role in bioinformatics for the collection, storage and maintenance of biological data [3]. To search for information on biological, physicochemical, biochemical and others. Characteristics of different nucleotide sequences and amino acids and the sites on which they are obtained it using biological databases.



Bioinformatics database is a combined product of biotechnology and information technology and plays a vital role in accelerating modern life science research [4]

Basics elements of these databases are:

- **Sequence** - a key element showing the sequence of nucleotides or amino acids, it is continuous, as found in nature, if the sequence has not been investigated fully, then each fragment is a separate record, nucleotides are marked with four Latin letters A, G, T, C, amino acids with a three-letter code, for example, amino acid series are marked by Ser, amino acid methionin - Met; nucleotides in the database is marked with four small Latin letters a, c, g and t, and AK with a big Latin letter as serine is denoted with S;
- **Length** - the number of nucleotides or the number of amino acids involved in sequencing sequence;
- **The type of nucleic acid** - DNA, mRNA, tRNA, rRNA;
- **Location** - Defines the position of the 5' end of the nucleotide chain or the location of the genetic map;
- **Segment** - defines the relationship between adjacent, disjoint sections of the sequences if the distance between two or more sequences known to be marked and the number of amino acids or nucleotide [5].

There are various open sources (BIOPERL EMBOSS, BIOJAVA) and web services software (MSA, SEQUENCE SEARCH SERVER) which are available. These are online tools to make protein models and docking etc. [1]

Some of the most popular database in the field of bioinformatics are ExpPASy, KEGG and NCBY. They are freely available, as the oldest of them is ExpPASy, established in 1993 in Geneva.

4. Software for Biological Databases

Software for access, organization and analyses of biological sequence and databases can be freely available or commercial. Here are some examples for commercial software.

- **Matlab** is a software package, built on a modular principle which has its own language, called the M language, which is similar to the language C + +. The advantage of Matlab is the ability to visualize data in 3D graphics.
- **CIPLEX** is an optimization package to IBM, which can perform both integer optimization, and continuous. This package solves very successfully problems which examine HP folding.
- **Tomlab** is a software package for optimization that runs in the middle of Matlab, but is installed separately. It has more options than optimization program Matlab.
- **Hyperchem** is a software product that provides good opportunities for establishing a chemical molecular models, spatial monitoring of the established models and tools to optimize them. During the creation of models can be set and changed certain lengths of chemical bonds, valent and torsion angles and atomic charges. Moreover, the program has databases of amino acids and nucleosides, which makes it easy to be modeled natural polymers - protein and nucleic acids.

5. Main Types of Intellectual Property in Bioinformatics in Bulgaria

Intellectual property protection in bioinformatics is a really important process. As it not only afford the owner, the right to exclude others from using the protected technology, but can also potentially provide the owner a monopoly right for manufacture and sale the technology [6].

Intellectual property protection for bioinformatics is a largely untested area in Bulgarian legislation. However, it's clear that ability to obtain intellectual property protection in any given area of bioinformatics depends on the type of bioinformatics tool involved. In terms of bioinformatics according the Bulgarian legislative system there are four types of intellectual property rights are generally considered.

Copyright

In Bulgaria Copyright can be used to protect bioinformatics-related materials such as scientific articles, books, software, compilation of facts (databases), manuals and etc. The term period for database protection is 15 years.

Patents

According to the Law of patents and registration of utility models by a patent can be protected: "an element isolated from the human body or otherwise obtained through the technical process, including the sequence or partial sequence of the gene may constitute a patentable invention, even if the



structure of this element is identical to that of a natural element". Patents are the exclusive right granted by government to the original inventor, developer or researcher of invention, which prevents other from making, using or selling the invention. The term period of patent is 20 years. For example, patent for a new drug.

Trade Secrets

Trade secrets can be used to protect Bioinformatics-related Intellectual Property such as software code, manuals, databases, formulas and processes. Trade secrets refer to proprietary information having commercial value and application. The matter of undisclosed information comes under civil courts.

Trade Marks

Trade Marks can be used to protect trade names, product names, domain names, and service marks/slogans for bioinformatics companies. The duration of the protection according the Law is 10 years and can be renewable every 10 years.

6. Conclusion

The scope of intellectual property rights varies country to country, bioinformatics inventions can be protected under intellectual property rights. Intellectual property rights are the negative rights, which prevent others from using your invention. IPRs have made revolution in the field of science and technology and assured a protection for your idea or innovation, and motivates the researchers but Indian patent law does not categorize bioinformatics inventions as patents, only copyrights and trade secrets rights are rewarded for an inventor, hence law should get amendment. The state and through it the global IP regime seeks to strike a balance between the interest of the creator of IP and the interest of other entrepreneurs and the society in general [1].

Bioinformatics comprises a wide array of components, and it follows that a wide array of protection might be available, depending on the particular nature of the bioinformatics component and its intended use. In Bulgaria the IP protection of bioinformatics is still in the beginning and there will be a lot of discussions on this hot point of innovation. The article just aimed to outline the framework of main types intellectual property forms of bioinformatics according the Bulgarian laws.

Intellectual property rights have become crucial for getting an idea patented otherwise; your efforts of making new inventions will be useless. If you have an idea, get it patented before anybody copies.

7. Acknowledgments

The paper and the survey is part of the scientific projects: 1) SULSIT - "Analysis of the Common Practices in the Use of Products of Intellectual Property in University Information Environment" (2012-2014), financed by National Science Fund of the Bulgarian Ministry of Education and Science, Contract № DMU 03/3-19.06.2012 in competition for "Young Scientists", led by Assoc. Prof. Tereza Trencheva PhD; 2) SWU "Neofit Rilski" - "Bioinformatics Research: Protein Folding, Docking and Prediction of Biological Activity (2014-2016) financed by National Science Fund of the Bulgarian Ministry of Education and Science, Contract № NSF - I02/16 - 12.12.2014, led by Assoc. Prof. Ivan Trenchev PhD.

References

- [1] Suryawanshi, S., Jacob, I. Bioinformatics in the World of Intellectual Property Rights. International Journal of Research in Pharmaceutical and Biomedical Sciences. Vol. 2 (3) Jul – Sep 2011, pp. 930-933
- [2] School of Computing Science <http://www.ncl.ac.uk/computing/study/postgrad/taught/5198/>
- [3] D'Souza, S. Gene meets machine: Intellectual Property issues in Bioinformatics, Health Law Review, Volume 12, Number 2, 2004, pp. 34 – 43
- [4] Chang, J., Zhu, X., Bioinformatics Databases: Intellectual Property Protection Strategy, Journal of Intellectual property rights, Vol.15, 2010, pp. 447-454
- [5] McBride, M. Bioinformatics and Intellectual property protection, 2002, Berkley technology law journal, Vol 17:1, p.33
- [6] Muralidharan, R. Intellectual property rights in bioinformatics and biotechnology, 2005, Krishna & Saurastri Associates, p. 43-49
- [7] O'Neil, S.A., 2000. Mapping the human genome. Hillards Lyons Equity Research.
- [8] Todorova T., Traykov, M., Tadjer, A., Velkov, Zh. Structure of flavones and flavonols. Part I: Role of substituents on the planarity of the system. Computational and Theoretical Chemistry, Volume 1017, 2013, pp. 85–90.



[9] Mavrevski R., Selection and comparison of regression models: estimation of torque-angle relationships. C. R. Acad. Bulg. Sci., 2014, 67(10), pp.1345-1354.