BIOINFROMATICS AND INTELLECTUAL PROPERTY

Tereza Trencheva[1], Ivan Trenchev[2], Peter Milanov[2], Ivan Mirchev[2]

*State University of Library Studies and Information Technologies[1],
Sofia, Bulgaria, e-mail: e-mail: tereza_trencheva@abv.bg
South-West University "Neofit Rilski"[2], Blagoevgrad, Bulgaria*

Abstract: 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 Patent law, Copyright law and where protection may be available and practical.

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

1. INTRODUCTION

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.**

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 a 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 analysing 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.

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 four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T)[1]. These four bases are attached to the sugar/phosphate to form the complete nucleotide *(O'Neil, 2000).*
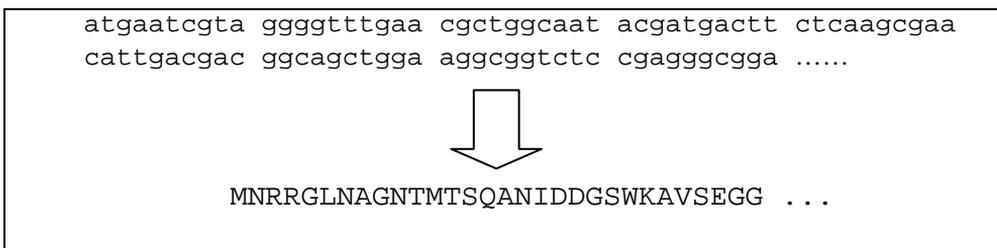
Like DNA, RNA is made up of a long chain of components called nucleotides. Each nucleotide consists of a nucleobase, a ribose sugar, and a phosphate group. The sequence of nucleotides allows RNA to encode genetic information[2]. The chemical structure of RNA is very similar to that of DNA, with two differences-(a) RNA contains the **sugar ribose** while DNA contains the slightly different sugar deoxyribose, and (b) RNA has the

---

[1] Definition of DNA according Wikipedia: http://en.wikipedia.org/wiki/DNA

[2] Definition of RNA according Wikipedia: http://en.wikipedia.org/wiki/RNA

nucleobase **uracil** while DNA contains thymine. The protein is a sequence of 20 standard amino acids.

```
atgaatcgta ggggtttgaa cgctggcaat acgatgactt ctcaagcgaa
cattgacgac ggcagctgga aggcggtctc cgagggcgga ......


         MNRRGLNAGNTMTSQANIDDGSWKAVSEGG ...
```

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

3. BIOINFORMATICS DATABASES.

Bioinformatics Databases play a key role in bioinformatics for the collection, storage and maintenance of biological data *(D'Souza, 2004 : 35).* 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 *(Chang, Zhu, 2010 : 447-454)*

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 meteonin - 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, tRNK, 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. *(McBride, 2002 : 13)*

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

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 HPfolding.

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

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 he owner a monopoly right for manufacture and sale the technology *(Muralidharan, 2005 : 43).*

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 bioinformatic tool involved.

In Bulgaria Copyright can be used to protect bioinformatics-related materials such as scientific articles, books, software, compilation of facts (databases), manuals and etc.

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".

Trade secrets can be used to protect Bioinformatics-related intellectual Property such as software code, manuals, databases, formulas and processes.

Trademarks can be used to protect trade names, product names, domain names, and service marks/slogans for bioinformatics companies.

6. CONCLUSION

Bioinformatics comprises a wide array of components, and it follows that a wide array of protection might be available, depending on the particular nature of the bioinformatic component and its intended use. In Bulgaria the IP protection of bioinformatics is still in the begging 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.

7. ACKNOWLEDGEMENTS

REFERENCE

1. **Chang**,J., Zhu,X., Bioinformatics Databases: Intellectual Property Protection Strategy, 2010, Journal of Intellectual property rights, Vol.15, p. 447-454
2. **D'Souza**, S. Gene meets machine: Intellectual Property issues in Bioinformatics, 2004, Health Law Review,Volume 12, Number 2, p. 34 – 43
3. **McBride**, M. Bioinformatics and Intellectual property protection, 2002, Berkley technology law journal, Vol 17:1, p.33
4. **Muralidharan**, R. Intellectual property rights in bioinformatics and biotechnology, 2005, Krishna & Saurastri Assosiates, p. 43-49
5. **O'Neil**, S.A., 2000. Mapping the human genome. Hillards Lyons Equity Research.
6. **Закон** за авторското право и сродните му права. // Държ. вестник, бр. 56, 1993; изм. Държ. вестник, бр. 32 от 28.04.2009.

7. **Закон** за патентите и регистрацията на полезните модели. // Загл. Изм. Държ. вестник, бр. 64, 2006; изм. Държ. вестник, бр. 36 от 04.04.2008.