METHOD OF PATTERN RECOGNITION OF BIOCHIPS IN GENETIC ENGINEERING

Jaromír VOLF[1], Viktor NOVÁK[1], Vladimír RYŽENKO[1]

[1]Department of Electrical Engineering and Automation, Czech University of Life Sciences Prague, Faculty of Engineering, Kamýcká 129, 165 21 Prague, Czech Republic

Abstract
The article describes the genome analysis based on method of automatic pattern recognition of biochips. The described method enables processing of multitude genetically modified DNA samples. This method can be used for analysing DNA samples in all areas of genetic engineering, particularly for analysing of samples of cattle or corn for agricultural purposes. The image of the samples is obtained by confocal microscopy; it is subsequently digitized and then processed by specific originally developed program Genocol, based on program code Borland Delphi. The program detects the degree of DNA modification of the samples and it represents it graphically by means of representative colour scale.

Key words: *genome; DNA; pattern recognition biochip; genetic engineering.*

INTRODUCTION

By genetic engineering we understand the artificial creation of genetically modified cells or complete organisms by direct manipulation of the organism´s genome on the level of genes. The genome is a set of the complete genetic material of a cell or more generally of an individual. The aim of the study of a genome is the analysis of the genetic structure and function of DNA sequences, detection of the organization of the genome, i.e. regularities of the arrangement of genes in a biological species or detection of similarities and dissimilarities between species and detection of differences between species and detection of the function of protein products which control the genome that creates biochip. Genetic engineering is a constituent of molecular biotechnology *(Bittner, 1999)*. The creating of biochips is described in *(Lyu, 2019)*, using of biochips in biochemistry procedure is described in *(Wan, 2019)* and other methods of analysis of biochips in *(Singla, 2019)*.
The amount of processed data is rapidly expanding and consequently the operation with data must be automated as much as possible, i.e. to apply computational technology in the pattern recognition and analysis and processing of data. The present paper deals with the concept of an appropriate method of the pattern recognition of biochips with individual DNA probes deposited on them. Confocal microscopy is used for the scanning of DNA on biochips. The biochips are placed in the microscope and the colour of the individual biochip DNA probes is scanned. Then the colour of the individual samples is stored in a file and processed by a computer.
The aim of study was automatic creation of biochips and their automatic pattern recognition because their number is about several thousand. This number of biochips is impossible to pattern recognition manually.

MATERIALS AND METHODS

The probe is a section of the DNA or RNA fibre (e.g. 1 gene) with a known sequence. Since it has an ability to be paired with a complementary sequence it can be used for the detection of a certain sequence in the DNA or RNA sample *(e.g. Pardatscher et al., 2016)*. For example if we know the normal sequence of a certain gene and the deviation from this sequence (mutation) which lead to the development of an illness, then by means of the probe we can find out if the gene is in order or mutated. For a known sequence we can even calculate the temperature of denaturation (ion concentration) of the probe. If we set the hybridization temperature of the probe with the investigated DNA exactly on the denaturation temperature of the probe we can even identify the difference between a normal gene (the probe binds) and a mutation (pairing is imperfect, the probe under the given condition does not bind). This is followed by washing when the incorrectly paired probes or absolutely unbound probes are removed. In order to be able to do this the investigated DNA must be fixed to a solid base.

Membranes which are by UV radiation charged positively and DNA charged negatively are used for this purpose.
Hybridization and repeated washing is followed by the detection of the labelled probe by various systems. Radioactive phosphorous is traditionally used for labelling so that each molecule emits β-radiation detected e.g. photographically. New systems are based on fluorescence, luminescence etc. Labelling of the probe is performed by e.g. biotin, fluorescein (fluorescent dye). In practice one of the nucleotides labelled in production with one of the mentioned labels is used in the preparation of the probe. In the case of fluorescein the detection of the labelled probe is performed by UV radiation. Biotin is detected by means of the protein streptavidin which binds to biotin. The producer of streptavidin already adds an enzyme. For the detection of the activity of the enzyme a substrate is added which can be chromogenic (the formed product is coloured, a coloured spot appears on the membrane in the place where the probe binds) or can excite light emission in the form of luminescence. The sensitivity of luminescent systems is equal to that of radioactive systems. In connection with the computational processing of images this makes possible a very fast and precise evaluation.
Biochips are then created by means of probes. From the viewpoint of their function biochips are actually membranes with deposited probes on which complementary sections of DNA (or RNA) are selected by means of hybridization. We can speak about a real biochip when the membrane (probe carrier) to which the DNA is firmly bound is replaced with a glass platelet about the size of a postage stamp and carries hundreds of thousands of functional points *(Ivkovic, N., 2016)*. At present biochips with up to 400 000 different oligonucleotide probes (400 000 functional points – size 1.28x1.28 cm) can be manufactured. The technology of their production is similar to procedures applied in electrical engineering industry.
By gradual and intermittent shading (and irradiating) further nucleotides can be added to the chain of the oligonucleotide probe to make the biochip contain all possible sequences for the given length. For example for an 8nucleitide sequence we need a biochip with 65 536 working spots (4 potential combinations of 4 various nucleotides in an 8nucleitide chain).
All hitherto mentioned biochips are based on the principle of hybridization (joining of complementary sections of nucleotide acids). The hybridization reaction is used for finding complementary sequences in the analysed samples, both for the identification of a particular similarity and for the specification of the sequence of nucleotides in the investigated nucleic acids.
A separate group are biochips the activity of which is not based on the hybridization reaction. Some of them are designed to sort cells according to their size and serve for obtaining DNA from captured leucocytes. In others a polymerase chain reaction (PCR) is executed. Some firms attempt to construct a so-called “mini-machine“, which could sort the cells, isolate DNA from them, multiply their corresponding sections (e.g. genes) compare them with the probes, detect the products and finally evaluate all.
For recognition in medicine in many applications it appears that the best method of scanning is confocal microscopy. Due to its great resolution this method is used in genetic engineering to recognize microscopic biochip patterns.
During observation in a confocal microscope the studied sample is illuminated by a point source of light, most frequently a laser beam focused on a diaphragm (so-called first confocal point diaphragm), which is then projected by the microscope objective lens into a point inside the sample. The same objective lens then collects the reflected and scattered light and perhaps even fluorescence. The image of the sample, owing to the space filtration of incident radiation on the detector, does not show any defocused background on extra focal areas of the sample.
Among the main advantages of confocal microscopy compared with optical microscopy is higher axial resolution and suppression of interfering radiation from the defocused areas of the studied sample. Confocal microscopes can image mounts with three-dimensional resolution. Biology and medical research benefit from the fact that confocal microscopy is actually a non-destructive method of study of the special arrangement of cells and tissues.
The benefit of confocal microscopy is that it can distinguish fluorescence from molecules inside and outside cells. By means of immunofluorescence methods which are based on the precise chemical uniqueness of antibodies labelled by fluorescence, not only the distribution of receptors in the membranes can be studied but for instance also the organization of the cycloskeleton. In many cases the

improvement of contrast is so dramatic, that cell structures which could absolutely not be observed with a classical microscope are revealed.
In our case the method of confocal microscopy was based on the simultaneous scanning of fluorescence images by three photomultipliers with spectral filters for the blue, green and red colour. By a recombination of three partial images in basic colours we obtain an optical section in real colours of the emitted fluorescence.
The image from a confocal microscope is then processed by pattern recognition methods, *(Masoud, 2000)*. Image converters with an accuracy of at least 10 bits are used for special photometric purposes. Methods of noise filtration (important for amplified fluorescence microscopy) or on the other hand detection and accentuation of intensity gradients and image segmentation belong to basic methods of computational image processing. Segmentation requires that individual objects must differ from each other by their brightness or colour. Objects labelled during image segmentation can then be counted or their dimensions and brightness statistically evaluated.
Ratio imaging is of basic significance in the biological application of fluorescence microscopy. This method makes it possible to measure the relative variations of fluorescence intensity in the cell compared with an appropriate initial state and to measure the ratios of intensities on two different excitation wavelengths. The measured ratios can then be imaged in imitated colours according to an appropriate colour code. Ratio imaging is used to convert the fluorescence images of cells coloured by indicators into quantitative images of the distribution of these physiologically significant factors in live cells.

RESULTS AND DISCUSSION

For the automatic evaluation of genetic operations a program called Genocol was created for the recognition of the colours of DNA probes located on the biochip and obtained by confocal microscopy.
First of all biochips are created according to the following procedure:

1. The investigated DNA is cloned.
2. Probes are produced from the DNA (DNA is split into smaller sections).
3. DNA is labelled (coloured) by fluorescence colours (green, red – e.g. Biotin).
4. By means of a micro manipulator (see Fig. 1) DNA is deposited on the biochip on which a physiological solution with a certain DNA sequence is located and also a membrane which fixes the DNA in the given position.
5. Now on each point of the biochip follow certain processes which serve for the joining of two fibres of a certain DNA sequence.
6. Complementary DNA sequences are joined and their bond is relatively strong. Partially joined fibres and those that are not joined are washed by the solution.
7. The biochip is completely scanned by a confocal microscope. In our case the colour of individual points on the biochip ranges from deep green to deep red according to which DNA probes joined.
8. The biochip with individual samples can be directly observed by the objective lens of the confocal microscope. A biochip usually has many thousands up to tens of thousands points and this is why most frequently the image is processed by a computer program. And this is the task of the Genocol program.

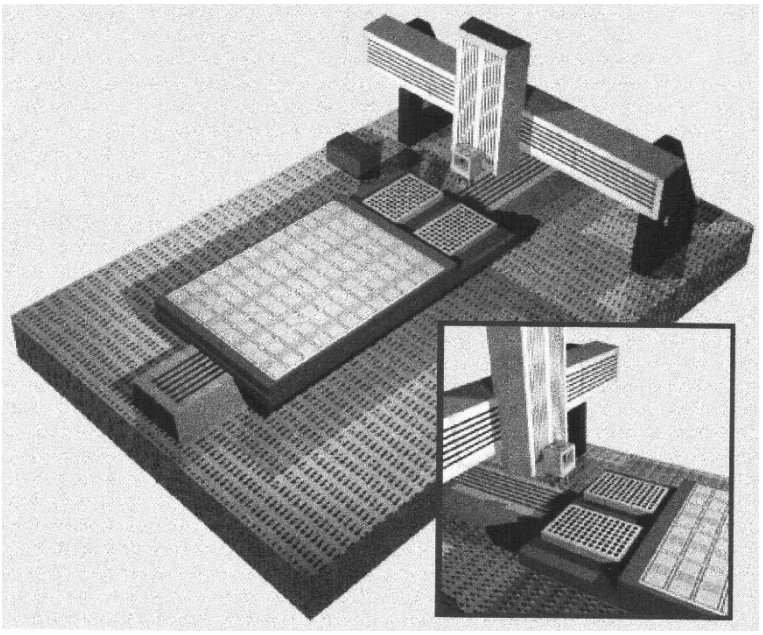

Fig. 1 Micro manipulator depositing the DNA samples.

The Genocol program *(Kolpek, 2003)*, with which the images with the samples of DNA probes are processed is written in the Borland-Delphi programming language. The image data are downloaded directly from the monitor and this is why it is necessary to set the colour deepness to 24 bits (True colour). This method is simpler than downloading data directly from the file because we need not know the exact structure and arrangement of data which are different in each format. The processed file is stored in the BMP file format (Microsoft Windows bitmap image file). This graphical file uses 24 bit imaging. 24 bit images use 3 bytes of colour information per pixel stored in the sequence red, green and blue channel (each 8 bite, i.e. a palette of 256 colours). The BMP format is not a packed format which is essential for dealing with this task since information on the real image in the packed format can be lost due to reduction of data. A section of the recognized bitmap file is in Fig. 2.

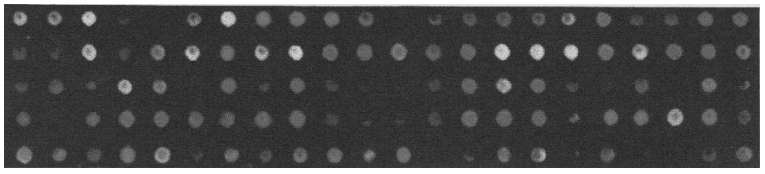

Fig. 2 Section of the recognized bitmap file.

The processed graphical file is a matrix of objects. The pixel which belongs to our searched sample (object) must have a zero or minimum magnitude of the blue colour component. This is due to the fact that DNA probes are labelled only with a red and green colour. However scanning may not be absolutely precise and this is why the tolerance of this colour must be set. Each object on our investigated image is composed of many pixels. This is why the quantity of the red and green colour is obtained by averaging. This means that first all values of the green and red colour are downloaded from one object and then divided by the number of downloaded pixels. This is how the average representation of the red and green colour in the object is obtained. This procedure is performed cyclically in all objects in the image. The result must be stored in two matrices. The coordinates are the same as in the objects on the image. The two matrices containing the values of the representation of the red and green colour are divided one by the other. Thus the resultant matrix is obtained giving the share of the red colour compared with that of the green one in each object (sample). The last operation consists in the ordering of the numbers of the resultant matrix of shares according to their magnitude. In order that the program could recognize which pixels belong to individual objects, grids are formed between

individual objects. The number of grids must be adjustable ("Number of samples, to the right, down") because the number of objects can vary. The program also makes it possible to set various tolerances of the blue colour since this colour is used for specifying if the pixel belongs to the object (sample) investigated by us. The setting of tolerance is also necessary due to the different quality of individual files. The main window with the individual settings is in Fig. 3.

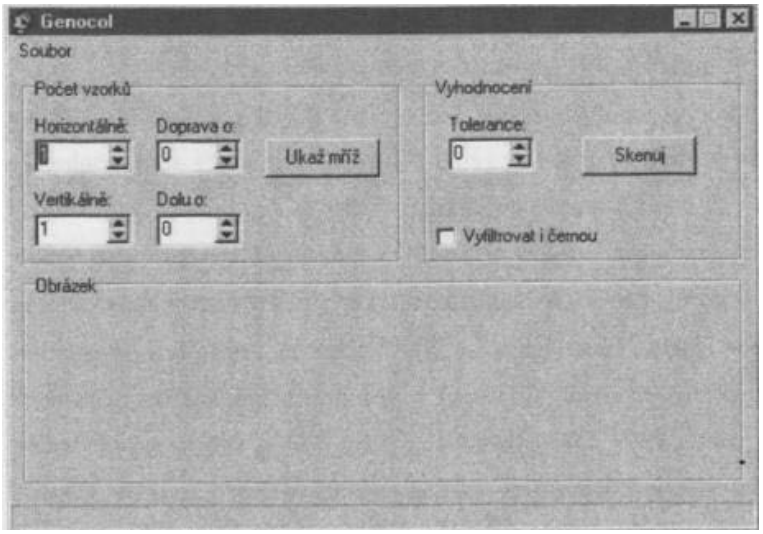

Fig. 3 The main interface window of the application.

The result shows the values of the representation of colours, their share and arrangement according to their magnitude. The program window with results is in Fig. 4.

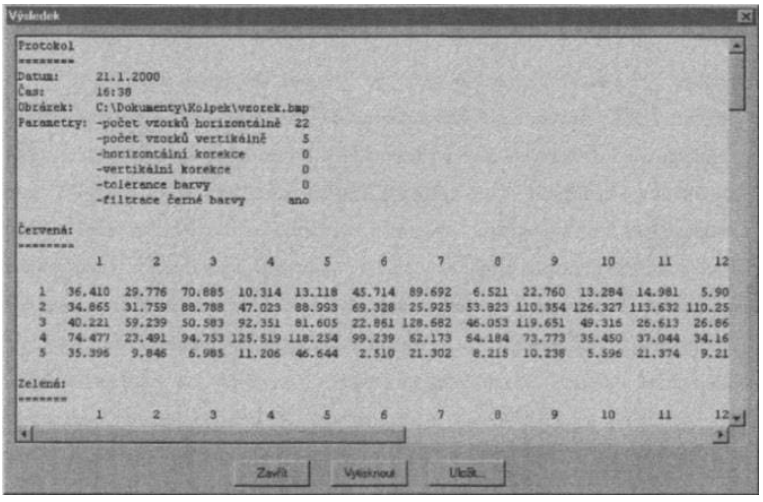

Fig. 4 Program window with results.

In this paper was shown originally methods of creating and automatically patern recognition of bio-chips used in genetics.
Other method of analysis of DMF biochips is described in (*Singla, 2019*).

CONCLUSIONS

The present paper deals with the issue of the recognition of biochip patterns obtained by fluorescence microscopy. The aim was to recognize DNA probes labelled by red and green colour implanted directly into the DNA. These probes were deposited on the biochip where certain processes proceeded, e.g. denaturation, by which the DNA probes were joined with the investigated DNA sections.

By means of this process we can determine the exact sequence of nucleotides in an unknown DNA because the sequence of nucleotides in the DNA probe is precisely known and must be absolutely identical to allow joining of two DNA fibres. The biochip was scanned by a confocal microscope which makes it possible to convert the scanned image into a digital form.
The Genocol program makes it possible to evaluate image information. The recognized file contained a matrix of samples deposited on a biochip. Each sample has a certain colour in dependence on how many DNA probes were complementary and hence were joined. The program screens the quantity of red and green colour in each pixel and sums up these values over the entire surface of the sample. Then it calculates the average representation of the red and green colour. The program distinguishes 256 shades of red and 256 bits of green (8 bite, 3 Byte). The Genocol program makes possible an output to the file and printing. The concept of the program is universal in order to allow recognition of files with a different magnitude and number of samples. The program can also set various tolerances of the blue colour because this colour is used for determining if the pixel belongs to the object (sample) investigated by us. The setting of tolerance is also necessary due to the different quality of individual files.

REFERENCES

1. Bittner, M. (1999). Nature genetics. *Nature Américaine, 21*, 10–14.
2. Ivkovic, N., Golub, M., & Jakobovic, D. (2016). Designing DNA Microarrays with Ant Colony Optimization. *Journal of Computers, 11*(6), 528–536.
3. Kolpek, V. (2000). *Using of Pattern Recognition Methods in Genome Engineering* (in Czech). Prague: Czech University of Life Sciences.
4. Lyu, Z.M., Zhou, X.L., Wang, X.N., Li, P., Xu, L., & Liu, E.H. (2019). Miniaturized electroluminescent biochip prepared on gold nanoparticles-loaded mesoporous silica film for visual detection of hydrogen peroxide released from living cells. *Sensors and Actuators, B: Chemical, 1. April 2019,* 437-443.
5. Masoud, M. (2000). *New Frontiers in Computational Intelligence and its Application*. Amsterdam: Press IOS.
6. Pardatscher, G., Bracha, D., Daube, S.S., Vonshak, O., Simmel, F.C., & Bar-Ziv, R.H. (2016). DNA Condensation in One Dimension. *Nature Nanotechnology, 11*, 1076–1081.
7. Singla, A., Agarwal, V., Roy, S., & Mondal, A. (2019). Reliability analysis of Mixture Preparation Using Digital Microfluidic Biochips. *IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems, 38*(4), 654-667.
8. Wan, C., Chen, X., & Liu, D. (2019). A multi-objective-driven placement technique for digital microfluidic biochips. *Journal of Circuits, Systems and Computers, 28*(5).

Corresponding author:
Prof. Ing. Jaromír Volf, DrSc., Department of Electrical Engineering and Automation, Faculty of Engineering, Czech University of Life Sciences Prague, Kamýcká 129, Praha 6, Prague, 16521, Czech Republic, phone: +420 22438 3203, e-mail: volf@tf.czu.cz