

Determination of type and concentration of DNA nitrogenous bases by Raman spectroscopy using artificial neural networks

Kirill A. Laptinskiy*^a, Sergey A. Burikov^a, Olga E. Sarmanova^a,
Sergey A. Dolenko^b, Tatiana A. Dolenko^a

^aMoscow State University, Department of Physics, 119991, Leninskie Gory 1 Bldg 2, Moscow, Russian Federation; ^bD.V.Skobeltsyn Institute of Nuclear Physics, 119991, Leninskie Gory Bldg 2, Moscow, Russian Federation

ABSTRACT

In this article the results of solution of two-parametrical inverse problems of laser Raman spectroscopy of identification and determination of concentration of DNA nitrogenous bases in two-component solutions are presented. Elaboration of methods of control of reactions with DNA strands in remote real-time mode is necessary for solution of one of the basic problems of creation of biocomputers – increase of reliability of molecular DNA-computations. The comparative analysis of two used methods of solution of stated problems has demonstrated convincing advantages of technique of artificial neural networks. Use of artificial neural networks allowed to reach the accuracy of determination of concentration of each base in two-component solutions 0.2-0.3 g/l.

Keywords: Molecular computations, nitrogenous DNA bases, laser Raman spectroscopy, artificial neural networks

1. INTRODUCTION

As a result of discoveries of last decades in biochemistry natural means of data storage, delivery and processing information in biological systems became clear. Development of the newest informational and molecular technologies provided the facilities of placement of information in biomolecular medium and read-out of information. It put elaboration of principles of work and practical realization of biocomputers on order of the day¹. Development of calculations using biological structures is very important not only for achievement of a new level of development of information technologies but for comprehension of functioning living matter on the molecular level.

As it is known, in living cells genetic information is encoded in molecule of deoxyribonucleic acid (DNA). DNA provides storage, delivery from one generation to another and realization of genetic program of functioning living organisms¹. These unique properties of DNA were used for solution of computation problem in². In this study the problem of going through directed graph was solved.

The information about the vertexes and the arcs of graph was encoded by DNA strands and the process of search of Hamiltonian path through the graph represented itself the set of biochemical reactions. The “Computing device” was the test-tube with water solution containing DNA strands.

Later the suggested approach was developed in a series of studies³⁻¹¹. In³ the opportunity of use of “biocomputer” based on DNA for solution of satisfaction problem of a Boolean expression was considered, in⁴ the molecular calculations were used for breaking one of encoding systems – so called DES (Data Encryption Standard System) having 256 different ways of information encoding. One should note that in this study the principal opportunity of breaking encoding system was demonstrated: it was supposed that there were no experimental errors duration realization of the algorithm.

The algorithms of solution of different problems using DNA strands suggested by Adleman and Lipton³⁻⁴ were developed in studies of many groups. Thus, in⁵ DNA molecules were used for solution of the problem of search of maximal clique in graph. In⁵ Adleman’s algorithm was used for solution of the problem of optimal delivery of mail. In⁶ it was suggested to multiply matrixes using DNA.

*laptinskij@physics.msu.ru; phone 7 495 939-1653; fax 7 495 939-1104; <http://rswater.phys.msu.ru/>

Well-known optimization “knapsack problem” – the problem of packing as more as possible useful things in knapsack under the condition that the total size or weight of all objects which can be packed is limited – was solved in⁷ using DNA computations. Authors of⁸ have developed the method of solution of map coloring problem using computations on DNA molecules. In⁹ the algorithm of encoding and decoding based on manipulations with DNA molecules was suggested.

The publication¹⁰ was devoted to the solution of sufficiently important problem of construction of true random number generator. It was demonstrated that using the biochemical reactions (polymerase chain reaction, denaturation, renaturation etc.) and the special temperature conditions in the solutions containing long DNA strands one can construct random number generator. In¹¹ the molecular computations were used for search of errors in Rule-Based Systems (Business Rule Management System). Suggested algorithm allowed to find all basic errors of Rule-Based Systems: redundancy, incompleteness, circularity, inconsistency.

Today it is obvious that molecular computations are very prospective way of development of computer technique. But the studies in this field faced with series of difficulties, firstly with the problem of reliability of molecular computations. Molecular computations regardless of the used algorithm consist of series of biochemical reactions with DNA strands of different length, with different concentration in the solution. Reactions are carried out at different temperature conditions, in different buffer solutions etc.

During such complicated operations the errors related with losses of substance during reaction, with accidental dot mutations in DNA molecules, with transgression of conditions of reaction etc. can occur. In real experiments, for example, the losses of DNA molecules can reach 10% from initial mass of working substance. At the same time elimination of even 1% DNA molecules leads to incorrect solution of problem. All this makes problem of control of errors during molecular computations very actual: it is necessary to control the concentration of DNA molecules and the parameters of reaction conditions during all time of carrying out the biochemical reactions.

It is obvious that methods of such control must be non-invasive, express, capable to work in real-time mode. In addition, it is necessary to elaborate the approach by which it could be possible to determine and control as more parameters of medium as possible. This study is devoted to elaboration of exactly such kind of methods.

Laser Raman spectroscopy provides non-invasive, express work in real-time mode. Vibrational spectroscopy including Raman spectroscopy (RS) is successfully used for a long time for study of structure and properties of DNA molecules¹²⁻¹⁵. In Raman spectra of DNA molecules peaks corresponding to vibrations of DNA backbone and nitrogenous bases (adenine, guanine, cytosine, thymine) being constituent of DNA¹²⁻¹⁴ were identified, markers of every base were determined¹⁵. In^{16,17} it was demonstrated that Raman spectroscopy allows to determine by non-contact means in real-time mode the total concentration of DNA molecules in the solution as well as the concentration of individual nitrogenous bases in one-component solutions using calibration dependencies of intensity of the Raman markers on the concentration.

In the context of solution of the problem of control of possible errors during molecular calculations the determination of concentration of individual nitrogenous bases in DNA solutions (i.e. in the presence of the other bases) is the most important problem. Information is encoded by the sequence of nitrogenous bases, that is why one can conclude about the loss of information when changing the concentration of at least one nitrogenous base. Thus, it is necessary to elaborate the method providing the ability of measurement and control of the concentration of every base in the solution in the presence of three other bases. This is the multi-parametrical inverse problem of Raman spectroscopy of DNA. To solve such problems the technique of artificial neural networks proved itself successfully¹⁸⁻²⁴.

Artificial neural networks (ANN) are the class of mathematical algorithms demonstrating high efficiency in solution of problems of pattern recognition, approximation, classification, prognosis etc.¹⁹⁻²⁴. In solution of many ill-posed inverse problems, such properties as training by real patterns, stability to noise, stability to contradictory data allow ANN to overcome the conventional methods of solution of inverse problems by efficiency²¹. Earlier the authors of the present work successfully used ANN for solution of such problems of laser spectroscopy as the extraction of weak useful signal from the intensive fluorescence background²², for the determination of the type and the concentration of dissolved salts in multi-component water solutions by water Raman valence band^{23, 24} and by full Raman spectra of the solutions^{25, 26}.

In this study the results of the solution of two-parametrical inverse problems of the identification and concentration determination of nitrogenous DNA bases using Raman spectroscopy are presented. The solution of such problems is

necessary for increase of reliability of molecular DNA calculations i.e. for elaboration of the methods of control of the reactions with DNA strands in remote real-time mode.

2. MATERIALS AND METHODS

In molecular calculations the value of concentration of DNA molecules in the solution is very important. According to the literature data this value of concentration can vary in rather wide limits – from units²⁷ to tens²⁸ g/l in depending on the problem. Such concentration of DNA in the solution provides the possibility of use of Raman spectroscopy in its conventional variant – the signal is collected from the bulk of the solution. It allows to control practically the whole bulk of “working substance”.

2.1. Preparation of solutions

The objects of research were the solutions of adenine, guanine, cytosine (Serva, Germany) in buffer solution (1M NaOH in water). For preparation of solutions the bidistilled water with specific electrical conductivity 0.1 $\mu\text{Sm/cm}$ was used. Single-component solutions of these bases and two-component (guanine+cytosine) and (adenine+cytosine) solutions in the concentration range 0 - 20 g/l with increment 1 g/l were prepared. The components of the solutions were chosen in such a way that in one of the solution the bases were complementary to each other (guanine+cytosine) and in another one – non-complementary (adenine+cytosine).

2.2. Raman spectrometer

Raman spectra were obtained using spectrometer consisted of argon laser (wavelength 488 nm, incident power 250 mW) and the system of registration composed of monochromator (Acton, focal length 500 mm, grades 900 and 1800 g/mm) and CCD-camera (Horiba Jobin-Yvon, Synapse 1024*128 BIUV-SYN). Spectra were measured in 90-degree scattering geometry with the practical resolution 2 cm^{-1} in the range 200-4000 cm^{-1} . In order to suppress the signal of the elastic scattering the edge-filter (Semrock) was used. Primary spectra processing consisted in subtracting pedestal and normalization by square of water Raman valence band (in order to eliminate the influence of instability of laser power on measured spectra).

2.3. Methods of determination of the concentration of the DNA bases by Raman spectra of their solutions

For the solutions of guanine, cytosine, and adenine the database of experimental Raman spectra of single-component solutions of each base and of two-component solutions (guanine+cytosine and adenine+cytosine) – the total 440 spectra with different concentration of the bases – was obtained.

In Fig. 1 and Fig. 2 the spectra of two-component solutions (guanine+cytosine) and (adenine+cytosine) with the concentration of each base 10 g/l are presented. The broad intensive band near 3400 cm^{-1} and the band near 1600 cm^{-1} are valence and bending bands of water, respectively. The narrow Raman peaks of nitrogenous bases are situated in region 600-1600 cm^{-1} as it is known^{17,18}.

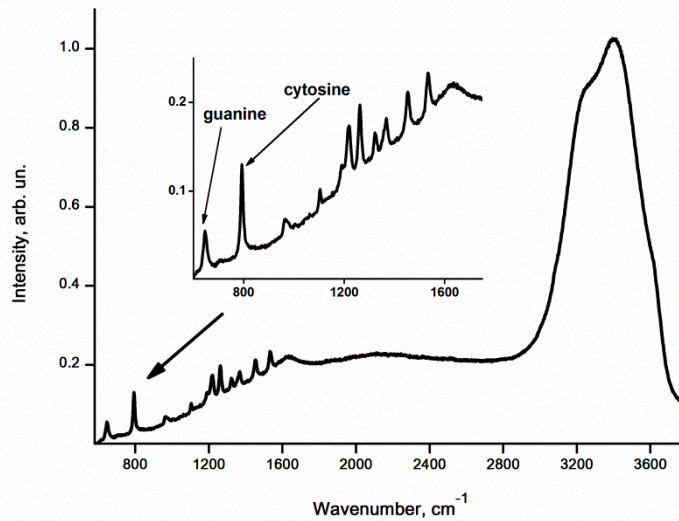


Figure 1. Raman spectrum of the solution (guanine+cytosine). The concentration of both components is 10g/l. In the inset the arrows indicate the markers of guanine and cytosine.

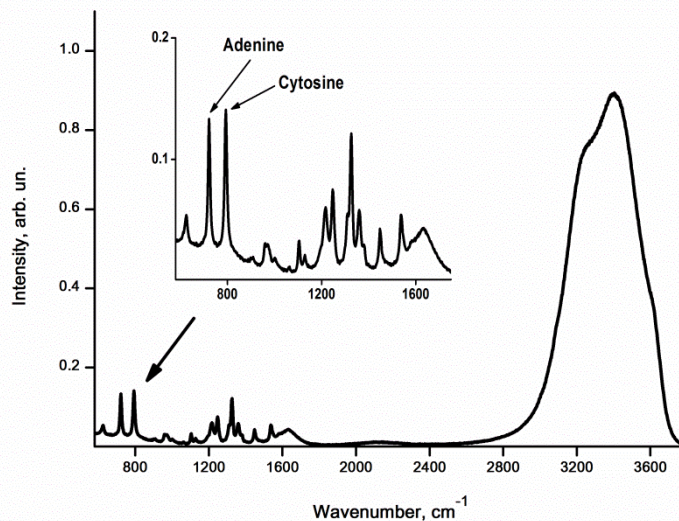


Figure 2. Raman spectrum of the solution (adenine+cytosine). The concentration of both components is 10g/l. In the inset the arrows indicate the markers of adenine and cytosine.

Every nitrogenous base has the series of Raman peaks in the region $500\text{-}1600\text{ cm}^{-1}$. These peaks overlap with each other to a large extent. It complicates the analysis of spectrum, but, nevertheless, one can single out the certain peaks-markers which are specific for each base. Guanine has the marker near 650 cm^{-1} , adenine – near 720 cm^{-1} , cytosine – near 790 cm^{-1} (Fig. 1 and Fig. 2).

2.3.1. Method 1. Determination of concentration of DNA nitrogenous bases using the calibration dependencies of intensity of spectral peaks-markers

The intensity of Raman bands of the substance depends on its concentration in the solution. After obtaining calibration dependence of the intensity of Raman markers of nitrogenous base on its concentration in the solution one can solve the inverse problem – to determine its concentration in the solution using intensity of peak-marker and corresponding calibration dependence¹⁷. For multi-component solutions the problem is how the accuracy of measuring the concentration of a single base is affected by the presence in solution of another base.

2.3.2. Method 2. Application of artificial neural networks

A unique feature of artificial neural networks (ANN) is their ability to learn by examples and not just to memorize but to generalize the presented information, to identify hidden regularities and to classify the presented information^{18,28}.

The simplest unit of the network – neuron – has several inputs and one output (Fig. 3). Neuron realizes weighed summation of inputs and then non-linearly converts result i.e. from mathematical point of view neuron is non-linear weighed summer

$$y = F(\sum_i w_i x_i) \quad (1)$$

where F — non-linear transmission function, w_i — weights of links.

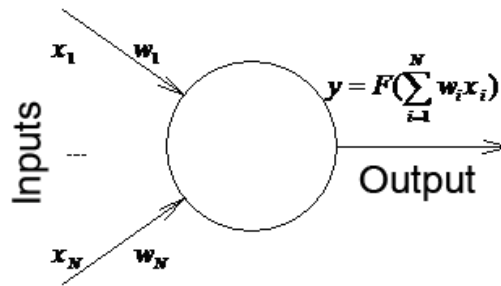


Figure 3. Artificial neuron.

ANN is defined by a topology of connections of neurons, characteristics of neurons and principles of training^{18,28}. The topology of neurons called 3-layer perceptron is widespread. It is demonstrated schematically in Fig. 4.

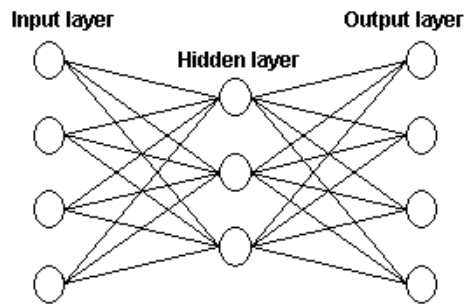


Figure 4. The scheme of 3-layer perceptron.

On the first stage of work the network is trained using known patterns – on the training set. Training consists of a change of weight coefficients of network in such a way that an output error of the network decreases. The most widespread algorithm of perceptron training is the algorithm of backward propagation of an error. One training event consists of applying a signal to inputs of network, obtaining output values y_i , comparison them with expected results y_i' for calculation of error

$$\varepsilon = (\sum_i (y_i' - y_i)^2) / n \quad (2)$$

and sending an error back to network for correction of weight coefficients. Initially the weights are initialized by random values. The main point of the method of backward propagation of an error is concluded in search of global minimum of an error in the space of weights by the method of gradient descent.

The estimation of degree of network training is carried out using a test set. Patterns of the test set do not intersect with patterns of the training set. This is necessary for prevention of memorization of noise by network and of its “overtraining”. “Overtrained” network is adapted to the training set well, but it loses ability to generalization. It is unacceptable. That is why if an error on the test set stops to decrease one should stop training.

The ANN method itself does not give an algorithm of estimation of accuracy of calculation of desired parameters that is why for estimation of accuracy of result one uses examination sets. These examination sets are applied to trained network.

3. RESULTS

3.1. Determination of the concentration of DNA bases using the calibration dependencies of intensity of the spectral peaks-markers

Determination of the concentration of nitrogenous base in the solutions was carried out using calibration dependence of intensity of its Raman peak-marker on the concentration of each base in single-component solution. For guanine the peak with the position of maximum at 646 cm^{-1} was used, for adenine – the peak with maximum at 722 cm^{-1} , for cytosine - the peak with maximum at 793 cm^{-1} (Fig. 1 and Fig. 2).

As it can be seen from Fig.1, background pedestal caused by fluorescence of the substance can be present at real experimental spectra. In order to decrease its influence on the accuracy of measurement of the concentration this pedestal was subtracted.

For single-component solutions of all nitrogenous bases the dependencies of the intensity of Raman signal in maximum of their spectral peaks-markers and the dependencies of integral intensity (square under the peak) on concentration of corresponding base were constructed. These dependencies can be good approximated by straight lines (Fig. 5, Fig. 6).

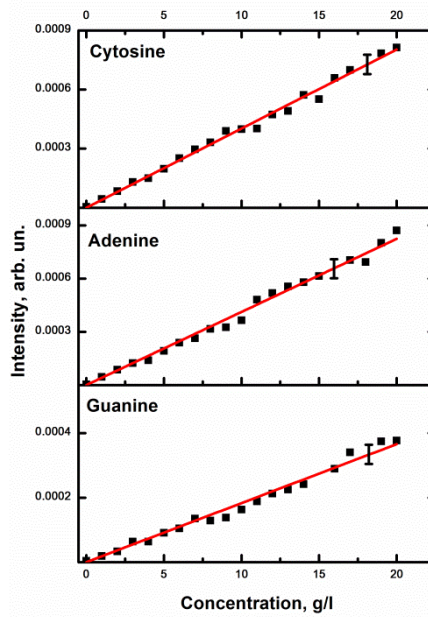


Figure 5. Calibration dependencies of the intensity of Raman markers of nitrogenous bases on concentration of each base in the corresponding single-component solutions.

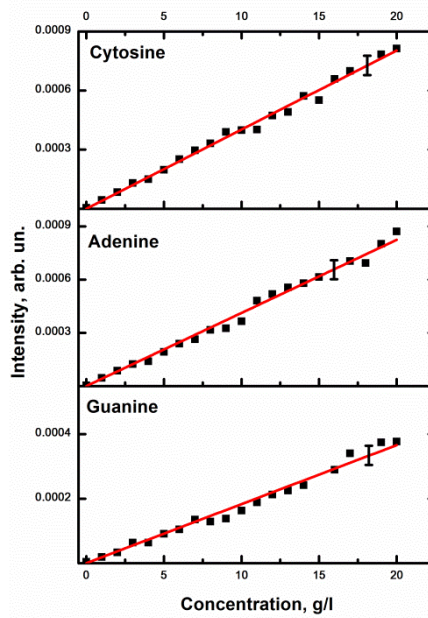


Figure 6. Calibration dependencies of integral intensity of Raman markers of nitrogenous bases on concentration of each base in the corresponding single-component solutions.

As a result of quantitative estimations of the accuracy of determination of the concentration of each base in two-component solutions using constructed calibration dependencies the best accuracy was achieved when the intensity of the Raman marker was used. The results are presented in Table 1.

Table 1. Accuracy of determination of the concentration of bases in two-component solutions obtained by different methods.

Method of determination of concentration	Guanine (with adenine) $\Delta C, \text{g/l} (\%)$	Adenine (with guanine) $\Delta C, \text{g/l} (\%)$	Cytosine (with guanine) $\Delta C, \text{g/l} (\%)$	Cytosine (with adenine) $\Delta C, \text{g/l} (\%)$
<i>Method 1</i>				
- by intensity of Raman marker	0.70 (3.5 %)*	0.70 (3.5 %)	0.65 (3.3%)	0.81 (4.1 %)
- by integral intensity of Raman marker	0.77 (3.9 %)	0.70 (3.5 %)	0.93 (4.7 %)	1.17 (5.9 %)
<i>Method 2</i> ANN	0.35 (1.8 %)	0.27 (1.35 %)	0.27 (1.4 %)	0.2 (1 %)

* The accuracy in percent was calculated relatively to the range of changes of bases concentration.

As one can see from Table 1, the accuracy of determination of the concentration of nitrogenous bases in two-component solutions by Raman markers is from 3% to 6 %.

3.2. Determination of the concentration of DNA bases using artificial neural networks

In this study two two-parametrical inverse problems of determination of the concentration of nitrogenous bases of DNA for each two-component solution were solved. During solution of the both problems the “experimental-based” approach of using ANN was used. I.e. ANN training was performed using real experimental spectra^{21,22}. Networks are trained on real experimental noise and it is advantage of this approach. Moreover this approach allows to take into account all interactions in the solution.

3.2.1 Two-parametrical inverse problem for the mixture (adenine+cytosine)

For realization of the “experimental-based” approach the whole array of 440 Raman spectra of single- and two-component solutions (adenine+cytosine) was randomly divided into training, test and examination sets according to ratio 70:20:10. ANN training was performed using training set. Test set was used for prevention of memorization of noise by network and “overtraining” network. The accuracy of solution of the problem was evaluated by independent examination set.

To train ANN two architectures of neural networks were used – 3-layer and 5-layer perceptrons. The best results were shown by 5-layer perceptron. When working with spectroscopic data one often faces with the following problem: number of desired parameters (outputs of the network) is much less than that of spectral channels (inputs of the network). At that it is obvious that not all points of spectrum are equally informative.

It means that the inverse problem of determination of concentration by Raman spectra is the ill-posed inverse problem with high input dimensionality. One of the ways for increase of an accuracy of determination of desired parameters is reduction of dimensionality of the problem by means of feature selection^{29,30}. In present study for selection of most significant features the analysis of weights was performed. According to the level (average value + standard deviation) from 1824 input features (spectral channels) 436 most significant features were selected.

Application of 5-layer perceptron to the array of Raman spectra with selected 436 most significant input features for single- and two-component solutions (adenine+cytosine) provided the accuracy of determination of concentration of adenine and cytosine in two-component solutions 0.27 and 0.20 g/l, respectively (Table 1).

3.2.2. Two-parametrical inverse problem for the mixture (guanine+cytosine)

The solution of two-parametrical inverse problem for the mixture (guanine+cytosine) using ANN was performed similarly but without compression of input data. Significant feature selection using the same algorithm as in the first problem gave no improvement of the result.

Application of 5-layer perceptron to the array of Raman spectra of single- and two-component solutions (guanine+cytosine) provided the accuracy of determination of the concentration of guanine and cytosine in two-component solutions 0.35 g/l and 0.27 g/l, respectively (Table 1).

Thus, the method ANN provides appreciably more precise solution – 2-3 times on the average – for the two-parametrical inverse problems of determination of the concentration of nitrogenous bases in two-component solutions in comparison with the method of determination of the concentration using the calibration dependencies of intensity of Raman markers of bases on their concentration. It can be explained by the fact that ANN is trained on real experimental spectra taking into account all types of molecular interactions and processes occurring in the solutions. For the calibration dependencies of intensity of markers on concentration of corresponding bases these interactions become a factor impairing an accuracy of determination of concentration of substances.

4. CONCLUSIONS

In present study the results of solution of two two-parametrical inverse problems of laser Raman spectroscopy – identification of DNA nitrogenous bases (adenine, guanine, cytosine) and determination of their concentration in two-component solutions are presented. For solution of the stated problems two methods were used: 1) using the calibration dependencies of intensity of own Raman markers of nitrogenous bases on the concentration of corresponding bases; 2) using artificial neural networks.

The comparative analysis has demonstrated that the method ANN provides more high accuracy of determination of the concentration of each of three bases by Raman spectra. Obtained results demonstrate the high efficiency of using Raman spectroscopy in combination with the method ANN for non-invasive express control of concentration of individual nitrogenous bases during molecular DNA calculations.

5. ACKNOWLEDGEMENTS

This work was supported by grant no.14-02-00710-a of Russian Foundation for Basic Research (T.A.D., K.A.L., S.A.B., conducting experiment, implementation of method 1) and by grant no.14-11-00579 of the Russian Scientific Foundation (S.A.D., implementation of method 2).

REFERENCES

- [1] Moe-Behrens, G.H.G., “The biological microprocessor, or how to build a computer with biological parts,” *Comput Struct Biotechnol J* 7(8), e201304003, (2013).
- [2] Watson, J., Crick, F., “Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid,” *Nature* 171(4356), 737–738 (1953).
- [3] Adleman, L.M., “Molecular computation of solutions to combinatorial problems,” *Science* 266, 1021-1024 (1994).
- [4] Lipton, R.J., “DNA solution of hard computational problems,” *Science, New series* 268(5210), 542-545 (1995).
- [5] Ouyang, Q., Kaplan, P., Liu, S., Libchaber, A., “DNA solution of the maximal clique problem,” *Science* 278(5337), 446-449 (1997).

- [6] Kari, L., Gloor, G., Yu, S., "Using DNA to solve the bounded post correspondence problem," *Theoretical Computer Science*, 23, 193-203 (2000).
- [7] Oliver, J., "Computation with DNA: matrix multiplication," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 44, 113-122 (1999).
- [8] Baum, E., Boneh, D., "Running dynamic programming algorithms on a DNA computer," *DNA Based Computers II*, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 44, 77-85 (1999).
- [9] Jonoska, N., Kar, I. S., "A molecular computation of the road coloring problem," *DNA Based Computers II*, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 44, 87-96 (1999).
- [10] Rahman, N.H.U., Balamurugan, C., Mariappan, R., "A novel DNA computing based encryption and decryption algorithm," *Procedia Computer Science* 46, 463 – 475 (2015).
- [11] Hedayatpour, S., Kama, N., Chuprat, S., "A sticker-based model using DNA computing for generating real random numbers," *International Journal of Security and Its Applications* 8(4), 113-122 (2014).
- [12] Madahian, B., Salighehdar, A., Amini, R., "Applying DNA computation to error detection problem in rule-based systems," *Journal of Intelligent Learning Systems and Applications* 7, 21-36 (2015).
- [13] Thomas, G.J., Jr., "Raman spectroscopy of protein and nucleic acid assemblies," *Annu. Rev. Biophys. Biomol. Struct.* 28, 1-27 (1999).
- [14] Prescott, B., Steinmetz W., Thomas, G.J., Jr., "Characterization of DNA structures by laser Raman spectroscopy," *Biopolymers* 23, 235-256 (1984).
- [15] Deng, H., Bloomfield, V.A., Benevides, J.M., Thomas, G.J.Jr., "Dependence of the Raman signature of genomic B-DNA on nucleotide base sequence," *Biopolymers* 50, 656–666 (1999).
- [16] Shanmugasundaram, M., Puranik, M., "Vibrational markers of structural distortion in adenine nucleobases upon DNA damage," *Phys. Chem. Chem. Phys.* 13, 3851–3862 (2011).
- [17] Dolenko, T.A., Burikov, S.A., Laptinskiy, K.A., Moskovtsev, A.A., Mesitov M.V., Kubatiev, A.A., "Improvement of fidelity of molecular DNA computing using laser spectroscopy," *Laser Phys.* 25, 035202 (10pp), (2015).
- [18] Laptinskiy, K.A.; Burikov, S.A.; Dolenko, T.A., "Determination of type and concentration of DNA nitrogenous bases by Raman spectroscopy," *Progress in Biomedical Optics and Imaging - Proceedings of SPIE* 9448, 94480J-1-8 (2014).
- [19] Hassoun, M.H., "Fundamentals of artificial neural networks," MIT Press, Cambridge, Massachusetts (1995)
- [20] Li, M., Verma, B., Fan, X., Tickle, K., "RBF neural networks for solving the inverse problem of backscattering spectra," *Neural Computing & Applications* 17(4), 391-397 (2008).
- [21] Yang, H., Xu, M., "Solving inverse bimodular problems via artificial neural network," *Inverse Problems in Science and Engineering* 17(8), 999-1017 (2009).
- [22] Gerdova, I.V., Dolenko, S.A., Dolenko, T.A., Churina, I.V., Fadeev, V.V., "New opportunity solutions to inverse problems in laser spectroscopy involving artificial neural networks," *Izvestiya Akademii Nauk Seriya Fizicheskaya.* 66(8), 1116-1124 (2002).
- [23] Dolenko, T.A., Burikov, S.A., Vervalde, A.M., Vlasov, I.I., Dolenko, S.A., Laptinskiy, K.A., Rosenholm, J.M., Shenderova, O.A., "Optical imaging of fluorescent carbon biomarkers using artificial neural networks," *J. Biomed. Opt.* 19(11), 117007 1-9 (2014).
- [24] Burikov, S.A., Dolenko, T.A., Fadeev, V.V., Sugonyaev, A.V., "New opportunities in the determination of inorganic compounds in water by the method of laser Raman spectroscopy," *Laser Physics* 15(8), 1-5 (2005).
- [25] Burikov, S.; Dolenko, S.; Dolenko, T.; et al., "Identification and determination of concentration of salts in natural waters by Raman spectroscopy using artificial neural networks," *AIP Conference Proceedings* 1267, 756-757 (2010).
- [26] Dolenko, S., Burikov, S., Dolenko, T., Efitov, A., Gushchin, K., Persiantsev, I., "Neural network approaches to solution of the inverse problem of identification and determination of partial concentrations of salts in multi-component water solutions," *Lecture Notes in Computer Science (LNCS)* 8681, 805–812 (2014).
- [27] Lee, J.Y., Shin, S.Y., Park, T.H., Zhang, B.T., "Solving traveling salesman problems with DNA molecules encoding numerical values," *Biosystems* 78, 39–47 (2004).
- [28] James K. D., Boles A. R., Henckel, D. and Ellington A. D., "The fidelity of template-directed oligonucleotide ligation and its relevance to DNA computation," *Nucleic Acid research* 26 (22), 5203-5211 (1998).
- [29] Specht, D., "A general regression neural network. *IEEE Trans.*" *Neural Networks* 2(6), 568-576 (1991).

- [30] Dolenko, S., Dolenko, T., Burikov, S., Fadeev, V., Sabirov, A., Persiantsev, I., "Comparison of input data compression methods in neural network solution of inverse problem in laser Raman spectroscopy of natural waters," *Lecture Notes in Computer Science* 7553, 443-450 (2012).
- [31] Gushchin, K.A., Burikov, S.A., Dolenko, T.A., Persiantsev, I.G., Dolenko, S.A., "Data dimensionality reduction and evaluation of clusterization quality in the problems of analysis of composition of multi-component solutions," *Optical Memory and Neural Networks (Information Optics)* 24(3), 186-192 (2015).