ISSN 1060-992X, Optical Memory and Neural Networks (Information Optics), 2016, Vol. 25, No. 1, pp. 16–24. © Allerton Press, Inc., 2016.

Adaptive Methods of Solving Inverse Problems for Improvement of Fidelity of Molecular DNA Computations

T. A. Dolenko[a, b], S. A. Burikov[a, b], A. O. Efitorov[a, b], K. A. Laptinsky[a, b], O. E. Sarmanova[b], and S. A. Dolenko[a]

[a] *Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Russia*

[b] *Physics Department, Lomonosov Moscow State University, Russia*

e-mail: tdolenko@lid.phys.msu.ru, burikov@lid.phys.msu.ru, dolenko@srd.sinp.msu.ru

Received November 30, 2015; in final form, December 7, 2015

Abstract—Elaboration of methods of monitoring of biochemical reactions with DNA strands is necessary to solve one of the main problems in creation of biocomputers—improvement of fidelity of molecular DNA computations. In this paper, the results of solution of inverse two-parameter problems of laser Raman spectroscopy on determination of the types and concentration of DNA nitrogenous bases in multicomponent solutions are presented. Comparative analysis of the three used methods of solving these problems has demonstrated convincing advantages of artificial neural networks and of the method of projection to latent structures. Use of adaptive methods allowed achieving the accuracy of determining the concentration of each base in two-component solutions about 0.2–0.4 g/L.

Keywords: inverse problems, molecular DNA computations, neural networks, projection to latent structures, laser Raman spectroscopy

DOI: 10.3103/S1060992X16010021

INTRODUCTION

Discoveries of the last decades in biochemistry allowed us to understand the natural ways of storing, processing and information transfer in biological systems. In turn, modern molecular and information technologies have facilitated the development of different ways to put information into biomolecular environment and its reading. This gave impetus to the development of the concepts and practical creation of biocomputers [1, 2].

It is known that in living cells genetic information is encoded in the molecule of deoxyribonucleic acid (DNA). DNA provides storage, transfer from generation to generation and realization of genetic program of development and functioning of living organisms [3, 4]. These unique properties of DNA are the basis of molecular calculations. For the first time, DNA molecules were used to solve the problem of passage of a directed graph by Adleman [5, 6]. The information about the edges and the vertexes of the graph was encoded using DNA strands, and the process of discovering a solution to the problem—a Hamiltonian path through the graph—was to carry out a series of biochemical reactions and to find the "right" DNA chain by means of isolation, sorting, and sequencing the resulting chains [6]. The "computing device" was a test-tube with an aqueous solution of DNA chains. The suggested approach has been developed in the works of Lipton for hacking one of the systems of encryption—DES (Data Encryption Standard System), with 256 different ways to encode information [7]. The algorithms for solving problems with DNA chains proposed by Adleman and Lipton [5–7] have found their development in the works of many scientific groups.

In [8, 9], molecular DNA computing was used to solve the problem of scheduling first for 2 lifts in a 6-storey building, and then for multiple lifts in a high-rise building. The authors of [10] developed the algorithm to solve the problem of the theory of graphs (routing in communication and transport systems)—optimal cable trench problem (CTP). Using biochemical reactions (polymerase chain reaction, denaturation, renaturation, etc.) and special temperature conditions in solutions of long DNA chains, the authors of [11] have created a random number generator. In [12], molecular calculations were used to find structural errors in Rule-Based Systems (server of business rules execution). The authors of [13] proposed an algorithm for encryption and decryption, based on molecular DNA computing.

Now it is evident that molecular computing is very promising for the development of computer technology. However, research in this area encountered a number of difficulties, primarily, the problem of the fidelity of calculations. Molecular computing is a set of biochemical reactions involving the DNA chains of various lengths, with different concentrations in the solution. These reactions are carried out at various temperatures, in various buffer solutions, etc. In process of such complex operations, errors associated with loss of substance during the reaction, random point mutations in DNA molecules, violation of the conditions of the reaction, are possible. In real experiments, for example, loss of DNA molecules can reach 10% of the initial mass, while elimination of even 1% of the DNA molecules from the computation process leads to incorrect solution of the problem. This means that the problem of monitoring errors in molecular calculations is extremely important: it is necessary to monitor the concentration of DNA molecules and the conditions of the reactions during the whole time of the biochemical reactions. It is obvious that the methods of such monitoring should be non-invasive, express, operated in real-time. In addition, it is necessary to develop an approach that could simultaneously identify and control the largest number of parameters of DNA solutions. This study is devoted to elaboration of such methods.

It is known that laser Raman spectroscopy can provide a method that is noninvasive, remote, rapid, and able to work in real time. Raman spectroscopy has been successfully used to study the structure and properties of DNA molecules [14, 15]. In Raman spectra of DNA molecules the lines corresponding to vibrations of the backbone of DNA and nitrogenous bases (adenine, guanine, thymine, cytosine) are identified [14], the markers of each base are identified too [15]. In [16, 17] it is shown that Raman spectroscopy is the basis of the non-invasive express method for determination of the total concentration of DNA molecules in a solution and the concentration of certain nitrogenous bases in single-component solutions by the calibration dependences of the intensity of the Raman markers on the concentration.

In the context of solution of the problem of monitoring eventual errors in process of molecular calculations, determination of concentration of individual nitrogenous bases in DNA solutions (i.e. in presence of the other nitrogenous bases) is the most important problem. This is due to the fact that information is encoded in the sequence of nitrogenous bases, so one can conclude about loss of information when the concentration of at least one nitrogenous base changes. Thus, it is necessary to create a method providing the ability of measurement and monitoring of the concentration of every nitrogenous base in DNA solution in presence of the three other bases. This is the multi-parameter inverse problem of Raman spectroscopy of DNA. In solving such problems, the technique of artificial neural networks (ANN) proved to be successful [18–21]. For many ill-posed inverse problems, such properties as training by real examples, stability to noise, stability to contradictory data, allow ANN to overcome traditional methods of solving inverse problems by efficiency [20]. Earlier, the authors of this paper successfully applied ANN to solve such problems of laser spectroscopy as extraction of a weak signal against a strong fluorescent background [22], identification of the type and determination of the concentration of dissolved salts in multi-component aqueous solutions by Raman valence band of water and by full Raman spectra of the solutions [21].

In this study, the results of the solution of two-parameter inverse problems of identification of nitrogenous DNA bases and of determination of their concentration by Raman spectra of DNA solutions are presented. The solution of such problems is necessary to improve the fidelity of molecular DNA calculations, i.e. to create methods of monitoring the reactions with DNA strands in remote mode in real-time.

MATERIALS AND EXPERIMENT

In DNA molecular calculations, the value of concentration of DNA molecules—the "working substance" in the solution—is one of the most important parameters. According to the literature data, this value can vary in a wide range—from units [23] to tens [24] g/L depending on the problem. Such content of DNA molecules in the solution provides the opportunity of use of Raman spectroscopy in its traditional variant, when Raman signal is collected from the bulk of the solution. Such variant allows monitoring of practically the whole volume of the "working substance".

The following solutions of adenine, guanine, cytosine in buffer (1 M NaOH in water) were prepared: (1) single-component solutions of all nitrogenous bases in the concentration range 0–20 g/L with increment 1 g/L; (2) two-component solutions of (guanine + cytosine) and (adenine + cytosine) in the concentration range of each base 0–20 g/L with increment 1 g/L. In one of the two-components solutions the bases were complementary to each other (guanine + cytosine), and in the other one—non-complementary (adenine + cytosine).

Raman spectra of DNA solutions were obtained using a spectrometer consisting of an argon laser (wavelength 488 nm, power 250 mW) and the system of registration composed of a monochromator (Acton) and a CCD-camera (Horiba Jobin-Yvon). Spectra were measured in the 90-degree scattering

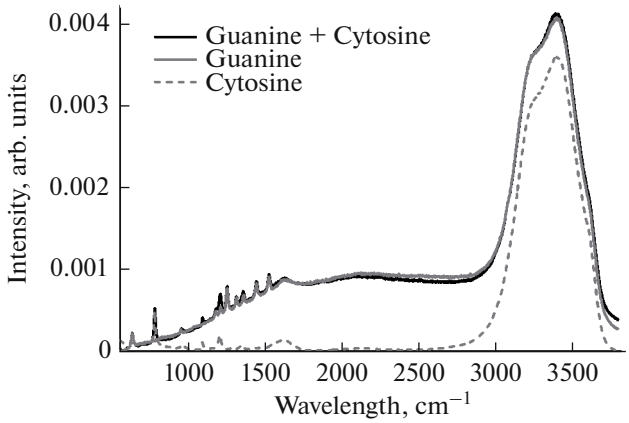

Fig. 1. Raman spectra of single-component solutions of guanine and cytosine and of two-component solution (guanine + cytosine). The concentration of each base in all solutions is 10 g/L.

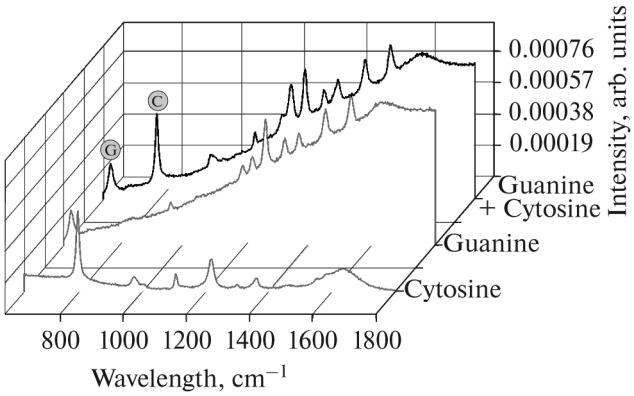

Fig. 2. Low-frequency region of Raman spectra of single-component solutions of guanine and cytosine and of their two-component solution. The concentration of each base in all solutions is 10 g/L.

geometry with practical resolution 1 cm^{-1} in the range 200–4000 cm^{-1}. Spectra processing included subtraction of a fluorescent pedestal and normalization by the area of water Raman valence band (in order to eliminate the influence of instability of laser power on the obtained spectra).

Methods of Determination of the Concentration of the Nitrogenous Bases by Raman Spectra of DNA Solutions

For 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)—total 465 spectra with different concentration of the bases—was obtained. In Fig. 1, the spectra of single-component solutions of guanine and cytosine and of two-component solution of (guanine + cytosine) with the concentration of each base 10 g/L are presented. The broad high-intensity band near 3400 cm^{-1} and the band near 1600 cm^{-1} are Raman valence and bending bands of water, respectively. The narrow Raman peaks of own vibrations of the nitrogenous bases are situated in the region 600–1600 cm^{-1} [14, 15].

The main characteristic Raman bands of all DNA bases are situated in the low-frequency region of Raman spectrum. This region of the spectrum is shown in Figs. 2 and 3.

As it can be seen in Figs. 2 and 3, in the region 500–1600 cm^{-1} each of the nitrogenous bases has a set of Raman spectral lines, which substantially overlap with each other. This complicates analysis of the spectrum, but, nevertheless, it is possible to select the marker lines that are specific for each of the bases.

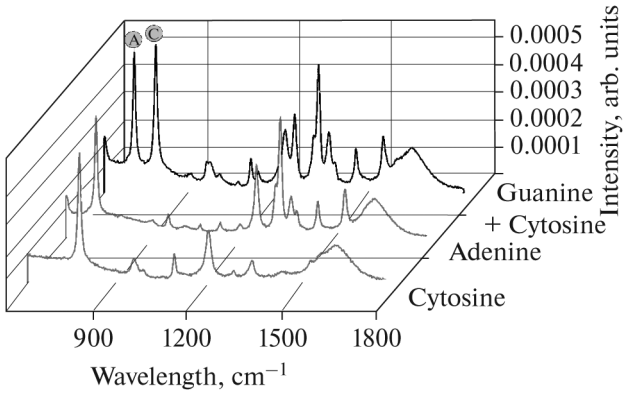

Fig. 3. Low-frequency region of Raman spectra of single-component solutions of adenine and cytosine and of their two-component solution. The concentration of each base in all solutions is 10 g/L.

Guanine has the marker line with maximum at 650 cm^{-1}, adenine has the marker with maximum at 720 cm^{-1}, cytosine has the marker with maximum at 790 cm^{-1}.

Method 1. Determination of DNA nitrogenous bases concentration using the calibration dependencies of intensity of marker spectral lines. The intensity of spontaneous Raman bands of a substance is directly proportional to its concentration in the solution. If experimentally obtained calibration dependences of the intensity of Raman markers of nitrogenous bases on their concentration in the solution are available, one can solve the inverse problem—determine the concentration of the nitrogenous base in the solution using the intensity of its marker and the corresponding calibration dependence [16, 17]. For multi-component solutions, the problem is how the accuracy of measuring the concentration of a single base is affected by the presence of other bases in the solution. In addition, as it can be seen in Fig. 1, the Raman spectrum of guanine solution has a broad band of fluorescence, which complicates the determination of the intensity of bases markers by DNA solution Raman spectra.

Method 2. Method of projections to latent structures. As it can be seen in Figs. 1–3, Raman spectra of single- and two-component solutions of DNA nitrogenous bases have a set of own lines, many of which overlap, are poorly resolved, and are combined into a common band. It is obvious that correct selection of marker lines of the bases for the use of method 1 is extremely difficult. In addition, fluorescence of certain bases is possible. This also prevents extraction of markers of the bases. These difficulties prompted application of the technique of adaptive algorithms, using the "experiment-based" approach [20], to solve the inverse problems of molecular DNA computing. In the "experiment-based" approach, a regression model is formed directly on the experimental data, what provides taking into account interactions in the solution and, accordingly, the characteristics of the Raman spectra of multicomponent solutions of DNA bases. One more advantage of this approach is that the model is formed taking into account real experimental noise.

Projection methods [25–27] are widely used for solving different problems of spectroscopy [28, 29]. Earlier, the authors have demonstrated the advantages of using the method of projections to latent structures, or partial least squares (PLS), to determine the concentrations of inorganic salts in aqueous solutions by Raman spectra [30, 31]. The essence of the PLS method is interdependent decomposition of matrices X (the matrix of spectra) and Y (the matrix of concentrations) according to the principle of principal component analysis (PCA). The difference is in the fact that in the calculation of each principal component of matrices X and Y, a common matrix of scores (coordinates of samples in the space of principal components) is used. This provides taking into account the interconnection of the matrices X and Y during formation of the vectors of loadings (the basis of PC space). On the basis of the latter, the matrix of regression coefficients $Y = XB$, which provides solution of the inverse problem, is created.

Method 3. Application of artificial neural networks. A serious disadvantage of the method of PLS is that when a regression model is created, only linear interconnections in the data are used. In the presence of nonlinearities this leads to a serious deterioration in the quality of solution of the inverse problem [28]. Therefore, to construct a regression model taking into account possible nonlinearities, ANN were used, whose training was also carried out within the framework of the "experiment-based" approach.

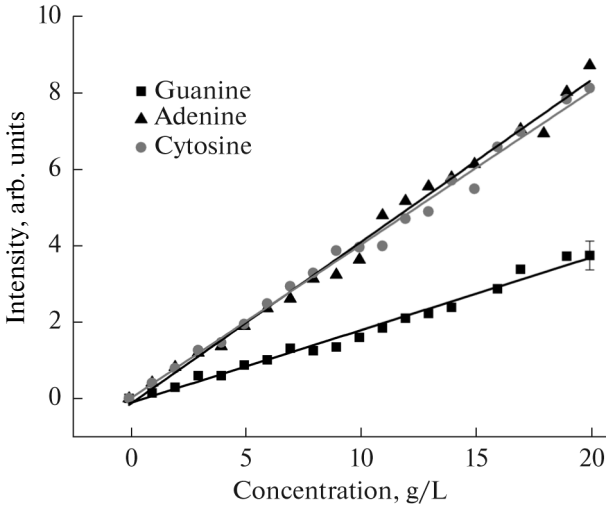

Fig. 4. Calibration dependences of Raman intensity of marker lines of nitrogenous bases on the concentration of each of the bases in the corresponding single-component solutions.

RESULTS AND DISCUSSION

Determination of DNA Nitrogenous Bases Concentration by the Calibration Dependencies of Intensity of Marker Spectral Lines (Method 1)

For single-component solutions of all DNA nitrogenous bases, the experimental dependencies of the intensity of Raman signal in maxima of their marker spectral lines on the concentration of the corresponding base were obtained. These dependencies are approximated by straight lines (Fig. 4). Using the constructed calibration dependencies, the concentration of each nitrogenous base in two-component solutions of DNA bases was determined. To reduce the influence of fluorescent background when measuring the intensity of the Raman marker line, this pedestal was approximated by a polynomial curve and subtracted. For guanine, the marker line with maximum at 646 cm^{-1} was used, for adenine, that with maximum at 722 cm^{-1} was used, and for cytosine, the marker line with maximum at 793 cm^{-1} was used (Fig. 4).

The calculations showed that the accuracy of determining the concentration of each base in two-component solutions by the obtained calibration dependencies was from 0.6 to 1.2 g/L. The results are presented in Table 1.

Determination of DNA Nitrogenous Bases Concentration Using the Method of Projections to Latent Structures (Method 2)

In this study, two 2-parameter inverse problems of determining the concentrations of the DNA nitrogenous bases for each two-component solution (adenine + cytosine) and (guanine + cytosine) were solved.

Table 1. The errors of determination of bases concentration in two-component solutions obtained using methods 1 and 2

Method of determination of concentration	Adenine (with cytosine) ΔC, g/L	Cytosine (with adenine) ΔC, g/L	Guanine (with cytosine) ΔC, g/L	Cytosine (with guanine) ΔC, g/L
Method 1 By the intensity at the maximum of the marker line	0.64	1.14	1.22	0.52
Method 2 PLS	0.18	0.23	0.37	0.23

Table 2. The errors of determination of bases concentration in two-component solutions obtained using method 3

Method 3	Adenine (+ cytosine) ΔC, g/L	Cytosine (+ adenine) ΔC, g/L	Guanine (+ cytosine) ΔC, g/L	Cytosine (+ guanine) ΔC, g/L
Architecture (1824) – 16 – 8 – 2				
Without selection	0.32	0.31	0.38	0.29
Selection by StD 74*/1472**	0.27	0.27	0.38	0.29
Selection by CC 417/1426	0.34	0.30	0.45	0.30
Selection by CE 262/1371	0.25	0.24	0.47	0.31
Selection by PLS 1450/1415	0.34	0.27	0.41	0.29
Selection by PLS 218/891	0.31	0.22	0.45	0.30
Architecture (1824) – 8 – 6 – 4 – 2				
Without selection	0.29	0.21	0.36	0.31
Selection by StD 436/1472	0.27	0.20	0.36	0.33
Selection by CC 975/1426	0.36	0.27	0.49	0.27
Architecture (1824) – 16 – 8 – 4 – 2				
Without selection	0.30	0.27	0.35	0.27

* The number of selected input features in the problem (adenine + cytosine).

** The number of selected input features in the problem (guanine + cytosine).

To implement the "experiment-based" approach, the whole obtained array of Raman spectra of single- and two-component solutions for each problem was randomly divided into training, test and examination sets according to ratio 70 : 20 : 10. To maintain due representativity of data, each of the data sets included samples with both boundary and middle values of bases concentrations. For the problem (adenine + cytosine), the ratio of the spectra in the pointed data sets was 164 : 46 : 23 (total 233 spectra), for the problem (guanine + cytosine) it was 163 : 46 : 23 (total 232 spectra).

In solving the problem by the method of PLS, the formation of the regression model with knowingly excessive number of components was performed on the training set. Further, based on the values of mean absolute error of determination of nitrogenous bases concentrations on the test set, the choice of the optimal number of PC was performed. For out-of-sample estimation of the quality of the inverse problem solution, the examination set was used. The results are presented in Table 1. It should be noted that small discrepancy between the results on test set and examination set demonstrates the adequacy of the generated regression model.

The calculations showed that the highest accuracy of determining the concentration of each base in the two-component solutions, provided by the PLS method, was from 0.2 to 0.4 g/L.

Determination of DNA Nitrogenous Bases Concentration Using Artificial Neural Networks (Method 3)

ANN training was performed using the training set. The test set was used to prevent noise memorization by the network and network overtraining. The accuracy of solution of the inverse problem was evaluated using the out-of-sample examination set.

Several architectures of neural networks were used—perceptrons with 2 and 3 hidden layers. To eliminate the influence of weights initialization on the point of convergence of the training process of neural networks, training was repeated 5 times for each architecture, with different initial weight values. The best results of the application of various architectures are presented in Table 2. In the architecture designation, the numbers of neurons in its layers are indicated. For the input layer, the number of input features is shown in parentheses, because it changes as a result of compression of the input data (see below). For all architectures, the following values of parameters of ANN and its training were used: gradient-based learning method, logistic transfer function in the hidden layers, linear in the output layer, learning rate 0.01, moment 0.5. Training stopped after 500 epochs after minimum of error on the test dataset.

Working with spectroscopic data, one often meets the following problem: the number of samples for training is much less than that of spectral channels (inputs of the network). At the same time, it is obvious that not all points of the spectrum are equally informative. The number of spectral channels in the obtained Raman spectra was 1824. This means that the inverse problem of determination of concentration by Raman spectra is an ill-posed inverse problem with high input dimensionality. One of the ways to increase the accuracy of determination of the desired parameters is reduction of dimensionality of the problem using selection of significant features [30].

In this study, the selection of the most significant features had been implemented in several ways: selection by the absolute value of the standard deviation (StD) of the intensity in each channel, selection by the values of cross correlation (CC) or by the values of cross entropy (CE) between the input and output features, and the selection based on the evaluation of weighted loadings of the PLS model.

In the first case, the selection is based on the fact that the amount of information contained in each input feature (i.e., spectrum channel), is proportional to the entropy of this channel, which, in its turn, is proportional to the standard deviation of the values in the channel. However, some of the information contained in the channel can be not related to the solved problem.

The values of CC and CE evaluate the relationships that are relevant to the solved problem. However, the values of CC takes into account only linear relationships; and the accuracy of the estimation of the CE in this case is low because of the small number of samples available.

The absolute values of the weighted loadings indicate to what extent the information of channels was used in the formation of the PC vector. The final regression model includes more than a dozen of PC, and thus all of them should be used to evaluate the significance of channels. It should be noted that different PC have different significance for the model, and their positive impact decreases with the increase of the number of PC [32]. Eventually, the weighted sum of the values of weighted loadings for each channel, normalized to its maximum value, was used to evaluate the significance of this channel. As weights in this weighted sum, the values of mean squared error of the inverse problem solution using the PLS model with the given number of components were used. Thus, the weights for the first PC were equal to one, and for subsequent PC they decreased with the same dynamics as the values of StD in solving the problem.

The results (average absolute errors) of the inverse problem solution using neural networks trained on the complete set of input features and using various techniques of selection of input features are presented in Table 2. The presented results show that the highest accuracy of determining the concentration of each base in two-component solutions, provided by ANN, is from 0.2 to 0.4 g/L.

COMPARATIVE ANALYSIS OF THE USED METHODS

The results in Tables 1 and 2 show that the adaptive methods provide appreciably more precise solution for the two-parameter 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. The improvement in the precision is 2–3-fold on the average.

The methods of PLS and ANN provide almost the same accuracy of measurement of the concentration of DNA nitrogenous bases in two-component solutions, which is 0.2–0.4 g/L, with some advantage of the method of PLS. A similar trend was already found by the authors in applying PLS and ANN techniques for solving five-parameter inverse problem of laser spectroscopy of multi-component aqueous solutions of salts [33]. In the case of weakly nonlinear interactions and of small number of components in solutions, the higher accuracy of problem solution is provided by the method of PLS. The advantages of this method also include a much lower computational cost and the ability to work with a small amount of data.

When the number of components increases and, accordingly, the interactions in the solutions strengthen, the ANN method provides better results [33]. It can be explained by the fact that ANN trained on real experimental spectra can take into account all types of molecular interactions and processes occurring in the solutions, including all non-linear effects. For calibration dependencies of intensity of markers on the concentration of the corresponding bases (method 1), these interactions are factors that can impair the accuracy of determining the concentration of substances.

When ANN (method 3) was used to solve the problem for (guanine + cytosine), none of the methods of input feature selection could provide increase in the accuracy of the problem solution. The selection of input features by the absolute value of the standard deviation, i.e. with the method that preserves the largest amount of information, provided a small improvement in the solution of the problem for (adenine + cytosine). It can be explained by the fact that each of the spectral bands of the complex Raman spectrum of DNA solutions may contain some useful information about the characteristics and concentration of the

nitrogenous bases in solutions, and simultaneous taking into accounting of this information allows reducing the error in determining the concentrations. Selection of input features by the CE value and by the method of PLS also succeeded in improving the solution of the problem (Table 2).

CONCLUSIONS

This paper presents the results of solving two 2-parameter inverse problems of laser Raman spectroscopy on identification of DNA nitrogenous bases—adenine, guanine and cytosine—and determination of their concentrations in two-component solutions. To solve these problems, the following three methods were used: (1) using calibration dependencies of intensities of marker spectral lines on the concentration of the corresponding bases; (2) the method of PLS—projections to latent structures (partial least squares); (3) artificial neural networks (ANN).

The comparative analysis has shown that the adaptive methods—PLS and ANN—provide higher accuracy in determining the concentration of each of the three bases by Raman spectra: 0.2–0.4 g/L, which is 1–2% by the weight of the amount of DNA molecules involved in biochemical reactions. This accuracy of determining the loss of the "working substance" in molecular DNA computations has the same order as the maximum acceptable loss for the correct operation of computing biodevice (see Introduction). Thus, the obtained results demonstrate the principle possibility and efficiency of application of Raman spectroscopy in combination with adaptive methods for non-invasive express monitoring of the concentration of individual nitrogenous bases in the process of molecular DNA computing. The perspective of this study is solution of 4–5 parameter inverse problems using adaptive methods.

ACKNOWLEDGMENTS

This work was supported by the grant no. 14-02-00710-a of the Russian Foundation for Basic Research (T.A.D., K.A.L., S.A.B., conducting experiment, implementation of method 1) and by the grant no. 14-11-00579 of the Russian Science Foundation (A.O.E., S.A.D., implementation of methods 2 and 3).

REFERENCES

1. Paun, G., *DNA-Computer. New Computing Paradigms*, M.: Mir, 2004, p. 528.
2. Gerd, HG.M.-B., The biological microprocessor, or how to build a computer with biological parts, *Comput. Struct. Biotechnol. J.*, 2013, vol. 7, no. 8, pp. 1–18.
3. Watson, J. and Crick, F., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature*, 1953, vol. 171, no. 4356, pp. 737–738.
4. Bruce, A., *Molecular Biology of the Cell*, Fifth Ed., N.Y.; L.: Garland Science, 2007, p. 1392.
5. Adleman, L.M., Molecular computation of solutions to combinatorial problems, *Science*, 1994, vol. 266, pp. 1021–1024.
6. Adleman, L.M., On constructing a molecular computer, *DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, 1996, vol. 27, pp. 1–21.
7. Lipton, R.J., DNA solution of hard computational problems, *Sci. New Ser.*, 1995, vol. 268, no. 5210, pp. 542–545.
8. Muhammad, M.S., Ueda, S., Ono, O., et al., Solving elevator scheduling problem using DNA computing approach, *Adv. Soft Comput.*, 2005, pp. 359–370.
9. Watada, J., Kojima, S., Ueda, S., and Ono, O., DNA computing approach to optimal decision problems, *Int. J. Innovative Comput., Inform. Control*, 2006, vol. 2, no. 1, pp. 273–282.
10. Jeng, D.J-F., Kim, I., and Watada, J., DNA-based evolutionary algorithm for cable trench problem, *Lect. Notes Artificial Intelligence*, 2006, vol. 4253, pp. 922–929.
11. Saman, H., Nazri, K., and Suriayati, C., A sticker-based model using DNA computing for generating real random numbers, *Int. J. Security Its Appl.*, 2014, vol. 8, no. 4, pp. 113–122.
12. Madahian, B., Salighehdar, A., and Amini, R., Applying DNA computation to error detection problem in rule-based systems, *J. Intelligent Learning Systems Appl.*, 2015, vol. 7, pp. 21–36.
13. Rahman, N.H.U., Balamurugan, C., and Mariappan, R., A Novel DNA computing based encryption and decryption algorithm, *Procedia Comput. Sci.*, 2015, vol. 46, pp. 463–475.
14. Prescott, B., Steinmetz, W., and Thomas, G.J., Jr., Characterization of DNA structures by Laser Raman Spectroscopy, *Biopolymers*, 1984, vol. 23, pp. 235–256.
15. Shanmugasundaram, M. and Puranik, M., Vibrational markers of structural distortion in adenine nucleobases upon DNA damage, *Phys. Chem. Chem. Phys.*, 2011, vol. 13, pp. 3851–3862.

16. Dolenko, T.A., Burikov, S.A., Laptinskiy, K.A., Moskovtsev, A.A., Mesitov, M.V., and Kubatiev, A.A., Improvement of fidelity of molecular DNA computing using laser spectroscopy, *Laser Phys.*, 2015, vol. 25, p. 035202 1–10.
17. Laptinskiy, K., Burikov, S., and Dolenko, T., Determination of type and concentration of DNA nitrogenous bases by Raman Spectroscopy, *Proceedings of SPIE—The International Society for Optical Engineering*, 2015, vol. 9448, pp. 94480 1–8.
18. Hassoun, M.H., *Fundamentals of Artificial Neural Networks*. Massachusetts, Cambridge: MIT Press, 1995, p. 511.
19. Yang, H. and Xu, M., Solving inverse bimodular problems via artificial neural network, *Inverse Problems Sci. Eng.*, 2009, vol. 17, no. 8, pp. 999–1017.
20. Gerdova, I.V., Dolenko, S.A., Dolenko, T.A., Churina, I.V., and Fadeev, V.V., New opportunity solutions to inverse problems in laser spectroscopy involving artificial neural networks, *Izvestiya Akademii Nauk, Seriya Fizicheskaya*, 2002, vol. 66, no. 8, pp. 1116–1124.
21. Burikov, S.A., Dolenko, S.A., Dolenko, T.A., and Persiantsev, I.G., Application of artificial neural networks to solve problems of identification and determination of concentration of salts in multi-component water solutions by Raman spectra, *Opt. Mem. Neural Networks (Inform. Opt.)*, 2010, vol. 19, no. 2, pp. 140–148.
22. Burikov, S.A., Vervald, A.M., Vlasov, I.I., Dolenko, S.A., Laptinskiy, K.A., and Dolenko, T.A., Use of neural network algorithms for elaboration of fluorescent biosensors on the base of nanoparticles, *Opt. Mem. Neural Networks (Inform. Opt.)*, 2013, vol. 22, no. 3, pp. 156–165.
23. Lee, J.Y., Shin, S.-Y., Park, T.H., and Zhang, B.-T., Solving traveling salesman problems with DNA molecules encoding numerical values, *Biosystems*, 2004, vol. 78, pp. 39–47.
24. 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, *Nucl. Acid Res.*, 1998, vol. 26, no. 22, pp. 5203–5211.
25. Sharaf, M., Illman, D., and Kowalski, B., *Chemometrics*, N.Y.: Wiley, 1986, p. 1931.
26. Brown, S.D., Tauler, R., and Walczak, B., *Comprehensive Chemometrics. Chemical and Biochemical Data Analysis*, vol. 4, Amsterdam: Elsevier, 2009, p. 2896.
27. Esbensen, K.H., *Multivariate Data Analysis—In Practice, an Introduction to Multivariate Data Analysis and Experimental Design*, 5th Ed., 2006, p. 598.
28. Kessler, W. and Kessler, R.W., Multivariate curve resolution: a method of evaluating the kinetics of biotechnological reactions, *Analyt. Bioanalyt. Chem.*, 2006, vol. 384, no. 5, pp. 1087–1095.
29. Dolenko, T., Burikov, S., Dolenko, S., Efitorov, A., Plastinin, I., Yuzhakov, V., and Patsaeva, S., Raman spectroscopy of water-ethanol solutions: the estimation of hydrogen bonding energy and the evidence of clathrate-like structures, *J. Phys. Chem. A*, 2015, vol. 119, no. 44, pp. 10806–10815.
30. Efitorov, A.O., Burikov, S.A., Dolenko, T.A., Persiantsev, I.G., and Dolenko, S.A., Comparison of the quality of solving the inverse problems of spectroscopy of multi-component solutions with neural network methods and with the method of projection to latent structures, *Opt. Mem. Neural Networks (Inform. Opt.)*, 2015, vol. 24, no. 2, pp. 93–101.
31. Dolenko, S., Efitorov, A., Burikov, S., Dolenko, T., Laptinskiy, K., and Persiantsev, I., Neural network approaches to solution of the inverse problem of identification and determination of the ionic composition of multi-component water solutions, *Commun. Comput. Inform. Sci. (CCIS)*, 2015, vol. 517, pp. 109–118.
32. Wehrens, R., *Chemometrics with R.*, Heidelberg, Springer-Verlag, 2011, p. 286.
33. Dolenko, S., Dolenko, T., Burikov, S., Fadeev, V., Sabirov, A., and Persiantsev, I., Comparison of input data compression methods in neural network solution of inverse problem in Laser Raman Spectroscopy of natural waters, *Lect. Notes Comput. Sci.*, 2012, vol. 7553, pp. 443–450.