Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks

2017 Laser Phys. 27 025203

(http://iopscience.iop.org/1555-6611/27/2/025203)

IOP Publishing | Astro Ltd Laser Physics

Laser Phys. **27** (2017) 025203 (8pp) doi:10.1088/1555-6611/aa51a7

Improvement of reliability of molecular DNA computing: solution of inverse problem of Raman spectroscopy using artificial neural networks

T A Dolenko[1,2,3], S A Burikov[1,2], E N Vervald[1], A O Efitorov[2,3], K A Laptinskiy[1,2], O E Sarmanova[1] and S A Dolenko[2]

[1] Department of Physics, Lomonosov Moscow State University, Leninskie Gory 1/2, Moscow 119991, Russia
[2] Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Leninsky Gory 1/2, Moscow 119991, Russia
[3] National Research Nuclear University MEPhI, Moscow 115409 Russia

E-mail: tdolenko@mail.ru

Received 30 November 2016
Accepted for publication 1 December 2016
Published 11 January 2017

Abstract

Elaboration of methods for the control of biochemical reactions with deoxyribonucleic acid (DNA) strands is necessary for the solution of one of the basic problems in the creation of biocomputers—improvement in the reliability of molecular DNA computing. In this paper, the results of the solution of the four-parameter inverse problem of laser Raman spectroscopy—the determination of the type and concentration of each of the DNA nitrogenous bases in multi-component solutions—are presented.

Keywords: molecular computations, nitrogenous bases, Raman spectroscopy, artificial neural networks

(Some figures may appear in colour only in the online journal)

1. Introduction

Achievements in biochemistry in recent years have provided significant success in the understanding of natural means of data storage, processing and transmission of information in biological systems. Intensive development of modern molecular and information technologies has allowed for the elaboration of various means of introducing information into biological media and its readout. This promoted the elaboration of the principles of the operation of biocomputers [1, 2] which are characterized by massive parallelism of operation, high energetic efficiency and high information density in unit volume. Deoxyribonucleic acid (DNA) is responsible for the storage and transmission of genetic information from one generation to another and our ability to genetically program live cell growth and functions. This is why the use of DNA molecules as the basis of a new generation of computing devices is so widespread.

1.1. Problem of reliability of DNA computing

The ability to use DNA for the solution of computational problems was demonstrated for the first time by Adleman in 1994 [3]. In his study, the combinatorial 'traveling salesman problem'—the search for the optimal path through a directed graph—was solved with the help of DNA strands. The solution to the problem consisted of the encoded data of the graph vertices and edges using DNA strands, conducting a series of biochemical reactions and the selection of the DNA strand containing the encoded answer to the problem.

Further this approach was developed in a series of studies devoted to the solution of optimization problems: scheduling

of work on elevators in multistory buildings [4, 5] and laying cable in the trench (CTP—cable trench problem) [6]. With the help of DNA strands the random number generator was created [7], data encoding and decoding algorithms were made [8] and the search of structural errors in rule-based systems were elaborated upon [9]. With the help of DNA, Shu *et al* [10] solved the problem of calculating the optimal route. For perspective, Shu *et al* [10] planned to use the suggested approach as an additional element to the existing navigation systems.

The basic advantage of using DNA computing over conventional methods to solve these problems is the high parallelism of computations. According to Rajaee *et al* [11], one drop of DNA can substitute 15 trillion powerful computers executing 1 billion operations per second. This means that DNA computer performance can be estimated at approximately 10^{21} operations per second. For comparison, the most powerful modern supercomputer, the Sunway TaihuLight System, can execute up to 125×10^{15} operations per second [12, 13].

At the present time, the use of DNA molecules as basic elements of computing devices—molecular electronics—is being developed the most intensively. The replacement of conventional silicon-based logic elements with molecular complexes containing DNA molecules has been proposed due to the opportunity to increase the density of logic elements (and therefore the productivity of computing device) on a chip of conventional computer are close to their limits. Today, technologies are available that can produce a transistor with a size of about 10–20 nm. A computer with a molecular transistor (which is two orders of magnitude smaller than most miniature silicon transistors) will have a great advantage in computing power, as the performance of a computer is proportional to the number of transistors per unit square [14].

In reference [15], Carell suggested creating new logic gates based on DNA and metal ions. Introducing metal ions (for example, mercury Hg^{+} and silver Ag^{+}) into the solution with DNA molecules breaks the connection patterns of the DNA strands: instead of the connection of complementary oligonucleotides, the connection of the same nucleotides (for example, thymine–Hg^{+}–thymine) becomes possible. Due to different combinations of connected oligonucleotides, Carell [15] was able to obtain DNA-based logic gates with three inputs: 'Yes', 'And', 'Or'. One can now mark some significant achievements in molecular electronics: the creation of programmable DNA controllers [16], implementation of basic logic (Boolean) operations using DNA [17] and the creation of schemes for analog calculations with the help of DNA [18].

The main problems with computational modules built based on DNA molecules, which do not allow them to compete with conventional computers right now, are the possibility of errors occurring during biochemical reactions and the complexity of reading out information in a form suitable for the user. It is noteworthy that the elaboration of a method that allows for the creation of a suitable interface for reading out the results of DNA computation is performed by some groups with the help of operations using DNA strands. So, the authors of references [19, 20] use a system that can show the result of the multiplication of two numbers encoded by DNA strands and fed to its two inputs, in the form of a number on the display. Unlike the conventional calculator, the answer is obtained not as the result of logic operations but as the result of the selection from a library of solutions.

Molecular computations are obviously very prospective in terms of future computers, and there are a number of problems in this field, the first of which is the problem of improving the reliability of the calculations. Computing using biological structures consists of many different biochemical reactions involving DNA strands with different lengths and concentrations in the solution, and reactions are executed in different buffer solutions, with different enzymes and primers, under different temperature conditions etc.

Errors can occur during many of these operations. They can be related to loss of a 'working substance' during a reaction or, for example, due to sticking of DNA molecules to the walls of the container, random point mutations or damage to the DNA molecules and violation of the conditions of the reaction etc. [21, 22] In real calculations, the loss of DNA molecules can be as high as 10–15% of the initial amount, while the elimination of even 1–2% of the DNA molecules from the calculation process leads to the wrong solution to the problem. Thus, minimization and control of errors in molecular calculations is one of the main problems in biocomputing.

To solve this problem, it is first necessary to control the concentration of the 'working substance' and the parameters of the flow of the biochemical reactions simultaneously throughout the computation time. As genetic information is encoded in the concentration and sequence of nitrogenous bases—adenine (A), guanine (G), cytosine (C), and thymine (T), loss of information can be controlled by changing the concentration of each nitrogenous base. The change in the total concentration of DNA molecules cannot provide information about damage in DNA strands. Thus, it is necessary to elaborate a method that allows for the simultaneous determination and monitoring of as many parameters of media as possible, including the measurement of the concentration of each nitrogenous base in the solution in the presence of the other three nitrogenous bases.

In addition, this method of control must obviously be nondestructive and express, and it must work in a real-time mode. Laser Raman spectroscopy possesses these properties. It is successfully used for the identification of various compounds and for the determination of their concentrations in solutions [23–27], including the identification and determination of the concentration of DNA molecules [28, 29].

The authors of this publication have demonstrated that Raman spectroscopy allowed us to determine (in non-contact real-time mode) both the total concentration of DNA molecules in the solution and the concentration of individual nitrogenous bases in their single-component solutions, using the calibration dependencies of Raman marker bands of nitrogenous bases (figure 1) on their concentration in the range of 0 to 9 g l^{-1} [30, 31]. Also it was shown that the use of laser Raman spectroscopy for monitoring biochemical reactions provided for the detection and control of the processes of renaturation and denaturation of DNA strands, determination of the state of the strands, and the control of possible mutations in DNA molecules [32].

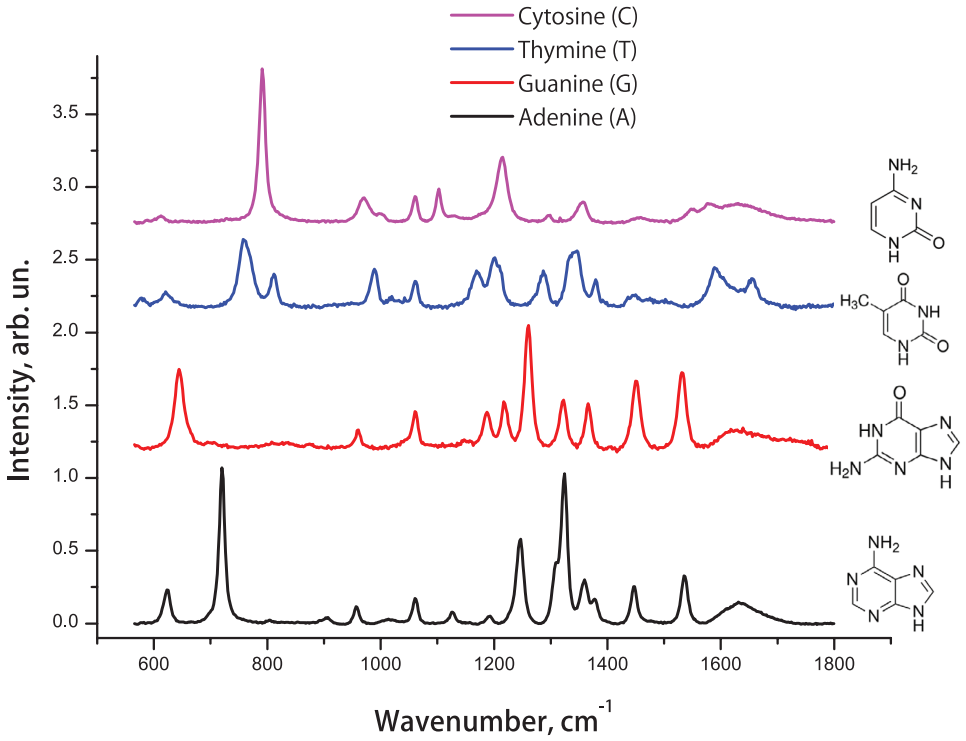

Figure 1. Raman spectra of single-component water solution of DNA nitrogenous bases. The concentration of the nitrogenous bases in solutions are 9 g l^{-1}.

As mentioned above, in order to control for possible errors in the process of molecular DNA computing, it is necessary to control the concentration of each of the four nitrogenous bases in the DNA solutions simultaneously, in the presence of all the other nitrogenous bases. While the selection of band markers is not difficult in Raman spectra of single-component solutions of nitrogenous bases (figure 1), in a spectrum of a solution containing all four nitrogenous bases the selection of the characteristic markers of each nitrogenous base is a very complicated task. As can be seen in figure 2, most of the marker bands of different nitrogenous bases overlap with each other, and it is impossible to construct any calibration dependences of the intensity of individual markers on the concentration of the corresponding nitrogenous base.

In addition, for a DNA concentration range of 0–9 g l^{-1} [30], the calibration dependence of the intensity of the nitrogenous base markers on its concentration could be approximated really well by a straight line. However, according to the literature, for a concrete problem the maximal concentration of DNA can vary within rather wide limits—from 3.32 g l^{-1} [33] up to 32.5 g l^{-1} [34]. At DNA concentrations higher than 9 g l^{-1} the dependence of the intensity of the marker bands on the concentration of the nitrogenous bases becomes significantly non-linear because of intensifying non-linear interactions between the molecules in the solution.

In order to overcome the difficulties described above, the stated multi-parameter inverse problem of laser Raman spectroscopy of DNA solutions—the identification and determination of the concentration of each nitrogenous base in a buffer solution of all four nitrogenous bases in the 'working

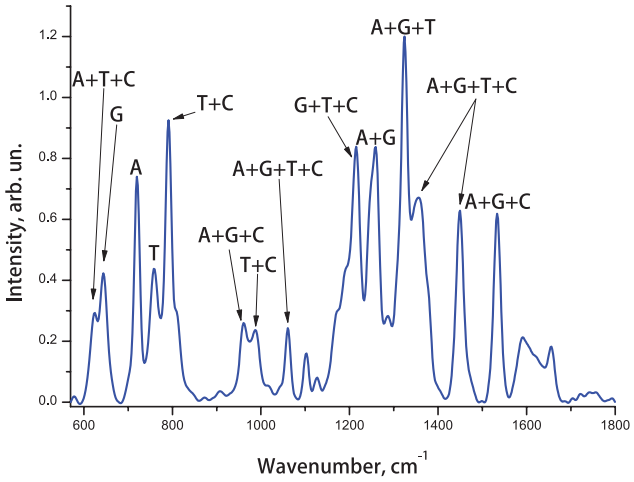

Figure 2. Raman spectrum of four-component water solution of DNA nitrogenous bases. The concentration of each nitrogenous base is 19 g l^{-1}.

range' of concentrations for DNA calculations—was solved using artificial neural networks (ANN) [35]. Properties such as training on real samples, resilience to noise and controversial data allow ANN to exceed traditional methods of the solution of ill-posed inverse problems by efficiency [36].

The authors of the present study have successfully used ANN before for the determination of the type and concentration of dissolved salts and ions in multi-component water solutions by their Raman spectra [37]. In reference [38], two two-parameter inverse problems were successfully solved using ANN: identification and determination of the concentration of each DNA nitrogenous base by Raman

spectra of two-component solutions—(adenine + cytosine) and (guanine + cytosine).

To solve the problems described above, three methods were used: (1) calibration dependencies of the intensity of nitrogenous base markers on the concentration of the corresponding bases; (2) method of projection to latent structures (PLS); (3) ANN. Concentrations of DNA ranged from 0 to 20 g l^{-1}. Comparative analysis demonstrated that adaptive methods—PLS and ANN—provide better accuracy of the determination of the concentration of each nitrogenous base by Raman spectra: 0.2–0.4 g l^{-1}. It is about 1–2% of the weight of DNA molecules taking part in biochemical reactions.

In the present study, the four-parameter inverse problem of the determination of the concentration of each of the four DNA nitrogenous bases dissolved in a buffer by Raman spectra is solved by ANN.

1.2. Artificial neural networks

ANNs are a class of mathematical methods that have proven to be highly efficient in solving problems involving intellectual data analysis (data mining)—problems of approximation, prediction, estimation, classification and pattern recognition etc [35].

The simplest element (node) of the network is called a neuron; it has several inputs and one output. A neuron sums up the values at its inputs with some weights and then non-linearly transforms the result, i.e. it is a non-linear weighted adder.

An ANN is determined by the topology of the connection and by the characteristics of its elements (neurons), and by the algorithm of training. The most widespread topology is called a multi-layer perceptron. Figure 3 presents a multi-layer perceptron with a single hidden layer; for some problems, perceptrons with a greater number of hidden layers, connected according to the same principle (output of each neuron of a previous layer is connected with the input of each neuron of the next layer), are used.

The first stage of work with an ANN is its training on known data samples. Training consists of changing the weight coefficients of the network in such a way that the error on the output of the network is decreased. The most widespread algorithm of perceptron training is the error back propagation algorithm. One training event consists of feeding a signal to network inputs, obtaining output values, comparison with the expected result for calculation of error, and sending the error back to the network for the correction of the weight coefficients. Initially all the weight coefficients are set to random values. The number of training events, which is equal to the number of patterns in the training set of the data, is called an epoch.

During network training, the degree of success of the process is estimated regularly using a separate validation data set. The validation data set is composed in such a way that its patterns do not intersect with patterns from the training set. Training of the network is stopped when during a predefined number of epochs the mean error of the network on the validation set does not decrease. This is necessary in order to prevent the so-called 'overtraining' of the network, when it starts to memorize noise contained in the training data set. To estimate the accuracy of the network response at different levels of noise in the input data, test sets with the addition of noise at a corresponding level are used. These test sets are given to the trained network.

Thus, to work with an ANN three data sets are required: training set for ANN training (in the process of weight correction), validation set to prevent network overtraining, and test data set to estimate the quality of the ANN training and to calculate the errors of determination of the desired parameters using independent data.

ANN can be used to solve the inverse problems of optical spectroscopy within the frameworks of three approaches: 'model-based', 'experiment-based', 'quasi-model' [39].

In the 'model-based' approach, to obtain data for ANN training, some available analytical or computational model of a solution of the direct problem is used. With such a model it becomes possible to provide the required representativity of all the necessary data sets for ANN training. However, the quality of the solution in this case depends directly on the adequacy of the model used. In situations where the construction of an adequate model is impossible because of the complexity of the required description of the object, this approach cannot be used.

In the 'experiment-based' approach, the ANN is trained on experimental data. The disadvantage of this approach is low representativity of data sets, since obtaining a large amount of experimental material is rather laborious. The main advantages of this approach are the following: when an ANN is trained using real experimental curves, all the molecular interactions are taken into account; the network is trained with real experimental noise, which increases the accuracy of the solution of the inverse problems.

In the 'quasi-model' approach, to obtain representative data sets, model spectra are used. In contrast with the 'model-based' approach (where an analytical expression describing spectra is known), in the 'quasi-model' approach, a parametrical 'quasi-model' describing spectra on the basis of a small set of experimental data, is constructed first. Then the data array is computed with the help of this 'quasi-model'. Obviously, in this case one can get a sufficient quantity of data samples and provide good representativity of data sets for ANN training, in contrast with the 'experiment-based' approach. However, the accuracy of the solution of an inverse problem in this case heavily depends on the compliance of the chosen or constructed numerical 'quasi-model' with reality, and on the difference of noise in the calculated curves and real noise in experimental data.

This paper continues the series of articles regarding the use of laser Raman spectroscopy to improve the accuracy of molecular computations. In this paper, the results of the solution of the inverse problem of the identification and determination of the concentration of each of the DNA nitrogenous bases in multi-component solutions by their Raman spectra with the help of an ANN are presented. The successful solution of this problem provides qualitative and quantitative control of the 'working substance' in biochemical reactions during DNA computing in a non-contact real-time mode.

2. Experiment

2.1. Preparation of water solutions of DNA nitrogenous bases

To form the sets of Raman spectra of water solutions of DNA nitrogenous bases, one- and four-component solutions of all nitrogenous bases—adenine, cytosine, guanine, thymine—were prepared in a sodium buffer (solution of NaOH in water, concentration 2 M). The concentration of each component in all solutions changed in the range of 10 to 19 g l^{-1} in increments of 1.5 g l^{-1}. Two- and three-component solutions were not considered, as in molecular computing all nitrogenous bases are present in the solution in any case. The solutions were prepared by dilution of the initial solutions of the individual nitrogenous bases at a concentration of 80 g l^{-1} with a buffer.

2.2. Raman spectrometer

Raman spectra of water solutions of DNA nitrogenous bases were obtained using a spectrometer consisting of an argon laser (wavelength 488 nm, power on the sample 400 mW) and a system of registration (monochromator Acton with focal length 500 mm, grade 1800 grooves mm^{-1} and charge-coupled device (CCD)-camera Horiba Jobin-Yvon, Synapse BIUV). Registration of spectra was performed in 90° geometry of experiment in the spectral range 500–4000 cm^{-1} with a practical resolution of 2 cm^{-1}. The recording of one spectrum was 10 s. The thermal stabilization system attached to the base of the cryostat KRIO-VT-01 maintained the solutions at room temperature (24 °C) within an accuracy of 0.1 °C.

Processing of the spectra consisted of the following: (1) correction to spectral and channel sensitivity of the detector, (2) subtraction of pedestal, (3) normalization to the area of water Raman valence band to eliminate the effect of the instability of the laser power on the measured spectra.

In total, 2414 Raman spectra of one- and four-component solutions of DNA nitrogenous bases were obtained. In figure 4, one can see the obtained Raman spectra of DNA nitrogenous bases with a concentration of each component of 19 g l^{-1}. Basic Raman bands of water molecules and of DNA hydroxyl groups are valence bands in the range 3000—3600 cm^{-1} and a deformation band with the maximum near 1625 cm^{-1}. The water Raman valence band has a peculiarity near 3600 cm^{-1} caused by vibrations of free OH-groups in the buffer solution. In the range 2800–3000 cm^{-1} one can see Raman bands of valence vibrations of CH-groups of nitrogenous bases. All characteristic Raman bands of nitrogenous bases are situated in the range 500–1800 cm^{-1} (figures 2 and 4). The total number of spectral channels in the range of 565–4000 cm^{-1} was 1961.

3. Results

3.1. Method of application of ANN

To train neural networks within the framework of the 'experiment-based' approach, all the obtained arrays of Raman spectra of one- and four-component solutions of nitrogenous bases was divided into three subsets: training, validation and test ones, as described above. The training was stopped when the average training error on the validation data set stopped decreasing for 500 consecutive epochs; the network, corresponding to the minimum error on the validation set was considered as trained and used further. (One can say that with further training the neural network begins to describe local peculiarities of the samples of the training set and not the common relationship of spectra and concentrations.) Of the total number of samples, 20% were randomly selected to be put into the test set, and 20% out of the remaining samples—into the validation set.

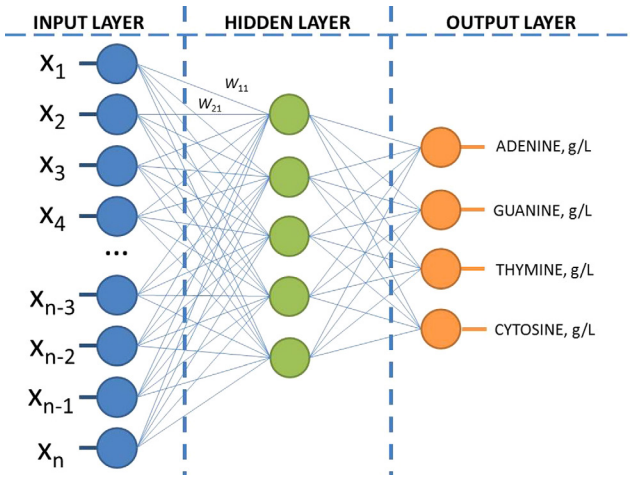

Figure 3. Scheme of a multi-layer perceptron with one hidden layer.

As the main indicator of the quality of the neural networks, we used the mean relative error (MRE) on the test data set, calculated according to the formula:

$$\mathrm{MRE} = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{|y_i - d_i|}{d_i} \cdot 100\%. \qquad (1)$$

Here N is the number of samples in the test data set, y_i is the network response to the ith sample of the test set, d_i—is the desired response to the ith sample of the test set. This indicator is calculated separately for each determined parameter, i.e. separately for the concentration of each nitrogenous base.

To eliminate the effect of partitioning into subsets on the results, as well as to eliminate the influence of weight initialization on the point of convergence of the training process of neural networks, the training was repeated three times for each of the architectures of the neural network used, with different partitioning into subsets and with different initial weights; the statistical indicators of the three trained networks were averaged. The indicator values presented below are in all cases values averaged over the three splits.

3.2. Determination of the concentration of each nitrogenous base in multi-component DNA solutions by full spectra

Because the quality of the solution can strongly depend on the complexity of the model used, neural networks of several architectures were trained—perceptrons with one and two

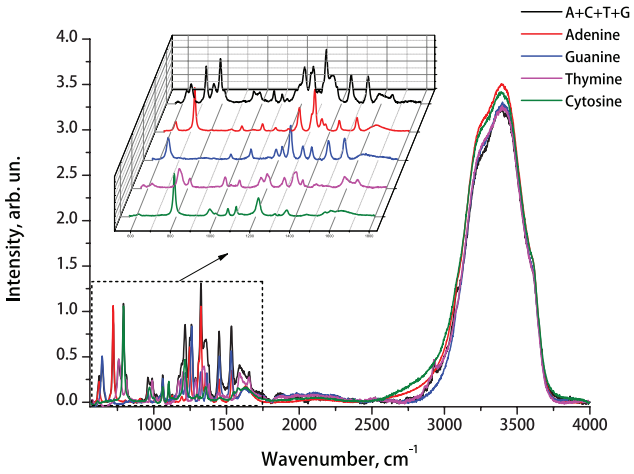

Figure 4. Raman spectra of single-component nitrogenous base solutions and of a four-component solution. The concentration of each nitrogenous base in all solutions is 19 g l^{-1}.

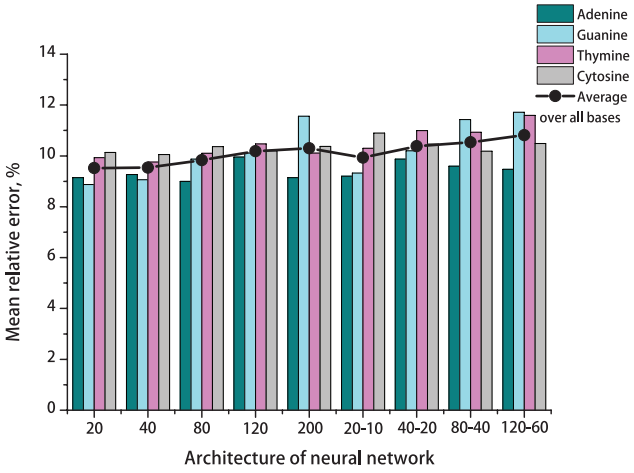

Figure 5. MRE on the test data set, averaged over three splits, for various nitrogenous bases and neural network architectures, obtained from full spectra; with the same averaged over all nitrogenous bases.

hidden layers, with various numbers of neurons in the hidden layers: 20, 40, 80, 120, 200, 20 + 10, 40 + 20, 80 + 40, 120 + 60. Figure 5 shows the MRE for the test data set, averaged over the three splits, for each nitrogenous base, obtained when the neural networks were fed the full spectra, consisting of the 1961 values of the intensities in the channels and the values are averaged over all four nitrogenous bases.

The best result on full spectra (MRE of 9.52%) was demonstrated by the neural network architecture having 20 neurons in the single hidden layer.

3.3. Application of feature selection methods

As mentioned above, the full spectrum consists of 1961 values of the intensities of the spectral channels. However, it is clear that in view of the presence of the characteristic spectral lines of each of nitrogenous base, not all channels of the spectrum are equally informative in determining the concentrations of the nitrogenous bases, even when taking into account the nonlinear interaction between the spectra of the individual nitrogenous bases. On the other hand, even the best neural network architecture has nearly 40 000 weight coefficients; this is obviously a lot with the total number of spectra equal to 2414. This means that reducing the number of spectral channels used may lead to an improvement in the neural network approximations of the desired dependencies, due to simplifying of the approximating function (the neural network), and as a consequence to a decrease in the errors in the determination of the concentration of each nitrogenous base. However, significant input features should be selected in an objective manner and not manually.

Earlier, the authors of this article used the same technology for the selection of significant input features for another spectroscopic problem, which improved the quality of its solution [40]. In this study, the technology was reproduced. The following algorithms of the selection of significant spectral channels (input features of the problem) were tested: (a) by the value of cross-correlation (Corr) of the intensity of a channel of the spectrum with the determined concentration of nitrogenous base; (b) by the value of cross-entropy (Cre) of the intensity of a channel of the spectrum with the determined concentration of nitrogenous base; (c) by the value of standard deviation (Std) of the intensity values of a channel of the spectrum; (d) by the results of analysis of the weights (Wa) of a trained neural network—a perceptron with a single hidden layer. The selection of the algorithms is described in more detail in reference [40].

For each of the four selection algorithms, three values of a parameter of the algorithm were chosen in such a way that the amount of the extracted significant channels was about 1400, 900, and 200. Next, for each selected set of significant channels, three neural networks were trained in the same way as was done above (on the same data sets); however, the neural network architecture was fixed (the one that showed the best results for the full spectra—having one hidden layer with 20 neurons). The obtained results are presented in figure 6.

The best result for this architecture (MRE of 5.17%) was demonstrated by the neural network, which was fed by the intensity values in 202 significant spectral channels selected with the method of neural network weight analysis.

As the number of spectral channels at the input of the neural network has been reduced almost tenfold, it should be checked whether a different architecture of the neural network could turn out to be optimal. For this purpose, neural networks of several architectures were re-trained—perceptrons with one and two hidden layers, with various numbers of neurons in the hidden layers: 10, 20, 40, 10 + 5, 20 + 10, 40 + 20. Figure 7 shows the average over the three splits of the MRE value on the test data set, separately for each nitrogenous base, obtained when the neural networks were fed with the intensity values in 202 channels selected with the method of ANN weight analysis, as described above. Figure 7 shows the values averaged over all four nitrogenous bases. It can be seen that the neural network with a single layer of 20 neurons remains the best architecture. Comparison of figures 5 and 7 also demonstrates well that reducing the number of input channels led to the decrease in the average relative errors for all the neural network architectures.

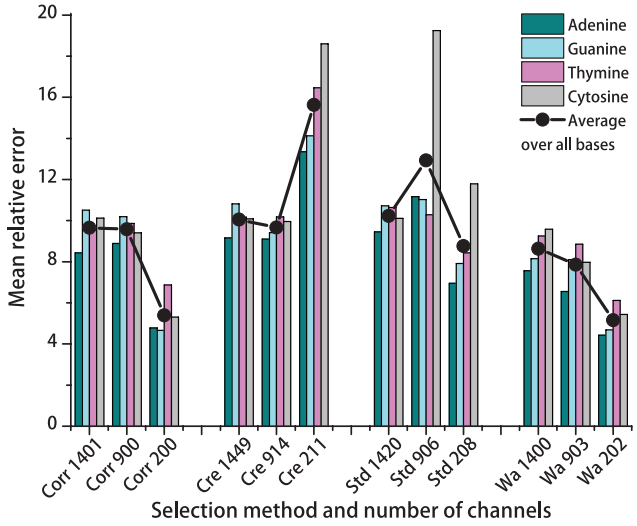

Figure 6. MRE on the test data set, averaged over three splits, for various nitrogenous bases and methods of the selection of significant spectral channels, architecture with 20 neurons in the single hidden layer, with the same averaged over all nitrogenous bases.

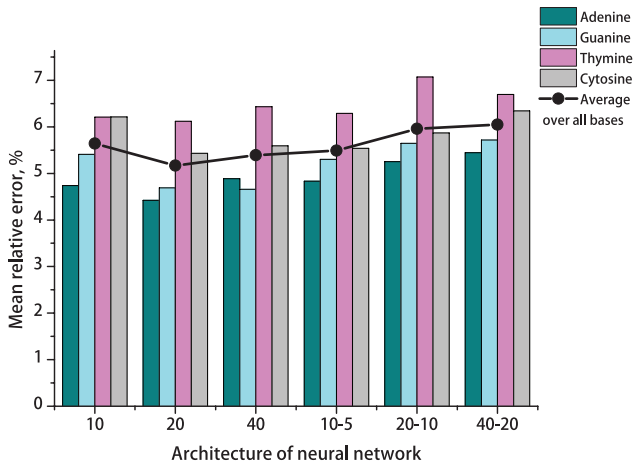

Figure 7. MRE on the test data set, averaged over three splits, for various nitrogenous bases and neural network architectures, obtained from 202 spectral channels selected by the method of neural network weight analysis, with the same averaged over all nitrogenous bases.

It is also interesting to compare how significant channels, objectively selected using the adaptive algorithm, coincide with the main spectral lines of the nitrogenous bases. In figure 8, in the spectrum of the solution corresponding to the maximum concentrations of all four nitrogenous bases (19 g l^{-1}), red color marks 202 significant channels selected by the method of weight analysis, on which the neural network training provided the best results. It can be seen that the significant channels include most of the major spectral lines of the nitrogenous bases; however, in addition, neural networks also require some information from some other spectral areas that are most sensitive to the concentrations of the individual nitrogenous bases.

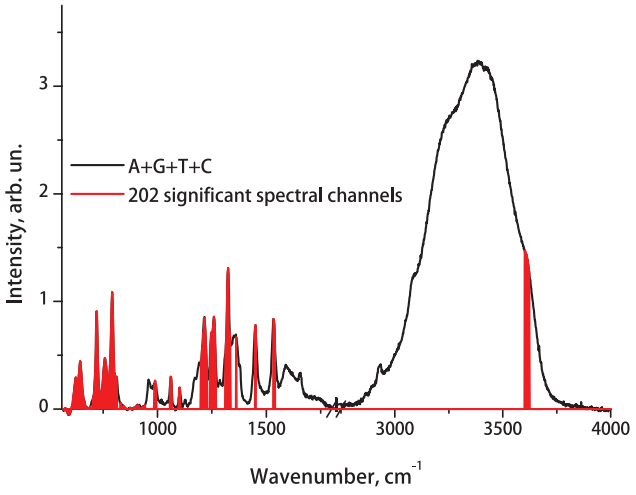

Figure 8. 202 significant spectral channels selected by the method of neural network weight analysis (filled).

Thus, in this study it has been shown that the application of ANN for the determination of the concentration of nitrogenous bases in four-component solutions with the concentrations of the components ranging from 10 to 19 g l^{-1} using the selection of the most significant input channels by the method of analysis of the weights of the neural network allows one to determine the concentration of each of the nitrogenous bases with a MRE of 5.17%. Thus, the average absolute error of determination of the concentration is 0.75 g l^{-1}.

4. Conclusions

The accuracy of determining the loss of the 'working substance' of molecular DNA computation obtained in this study has the same order as the maximum allowable loss for proper operation of computing biostructure (see the Introduction). Thus, the obtained results demonstrate the principle possibility and efficiency of the application of Raman spectroscopy in combination with adaptive data processing methods for non-invasive express control of the concentration of individual nitrogenous bases in the process of molecular DNA computing.

Acknowledgments

This study has been funded by the Russian Science Foundation grant no. 14-11-00579 (SAD, AOE—methods of application and implementation of ANN algorithms) and the Russian Foundation for Basic Research grant no. 14-02-00710_a (TAD, SAB, KAL—conduction of experiments, pre-processing of Raman spectra). This work was partially supported by the MEPhI Academic Excellence Project (contract No. 02.a03.21.0005, 27.08.2013) (TAD). The authors would like to express their deep appreciation to S P Kulik, Professor of Physical Department of MSU, who initiated and supported the research in the direction used.

References

[1] Paun G 2004 *DNA-Computer. New Computing Paradigms* (Moscow: Mir)
Moe-Behrens G H G 2013 The biological microprocessor, or how to build a computer with biological parts *Comput. Struct. Biotechnol. J.* **7** 1–18

[2] Xu J and Tan G 2007 A review on DNA computing models *J. Comput. Theor. Nanosci.* **4** 1219–30

[3] Adleman L M 1994 Molecular computation of solutions to combinatorial problems *Science* **266** 1021–4

[4] Muhammad M S, Ueda S, Ono O, Watada J and Khalid M 2005 Solving elevator scheduling problem using DNA computing approach *Adv. Soft Comput.* **29** 359–70

[5] Watada J, Kojima S, Ueda S and Ono O 2006 DNA computing approach to optimal decision problems *Int J. Innov. Comput.* I **2** 273–82

[6] Jeng D J-F, Kim I and Watada J 2006 DNA-based evolutionary algorithm for cable trench problem *Lecture Notes Artif. Intell.* **4253** 922–9

[7] Saman H, Nazri K and Suriayati C 2014 A sticker-based model using DNA computing for generating real random numbers *Int. J. Secur. Appl.* **8** 113–22

[8] Rahman N H U, Balamurugan C and Mariappan R 2015 A novel DNA computing based encryption and decryption algorithm *Proc. Comput. Sci.* **46** 463–75

[9] Madahian B, Salighehdar A and Amini R 2015 Applying DNA computation to error detection problem in rule-based systems *J. Intell. Learn. Syst. Appl.* **7** 21–36

[10] Shu J-J, Wang Q-W, Yong K-Y, Shao F and Lee K J 2015 Programmable DNA-mediated multitasking processor *J. Phys. Chem.* B **119** 5639 – 44

[11] Rajaee N, Zulkharnain A, Sallehin A A and Hussaini A 2016 Basic architecture and applications of DNA computing *Trans. Sci. Technol.* **3** 277–82

[12] Fu H H *et al* 2016 The Sunway TaihuLight supercomputer: system and applications *Sci China Inf. Sci.* **59** 1–16

[13] Dongarra J 2016 Sunway TaihuLight supercomputer makes its appearance *Natl Sci. Rev.* **3** 264–5

[14] Minkin B 2004 Molecular computers *Himiya i zhizn'* **2** 13–7

[15] Carell T 2011 DNA as a logic operator *Nature* **469** 45–6

[16] Chen Y-J, Dalchau N, Srinivas N, Phillips A, Cardelli L, Soloveichik D and Seelig G 2013 Programmable chemical controllers made from DNA *Nat. Nanotechnol.* **8** 755–62

[17] Wang F, Lu H-H and Willner I 2014 From cascaded catalytic nucleic acids to enzyme–DNA nanostructures: controlling reactivity, sensing, logic operations, and assembly of complex structures *Chem. Rev.* **114** 2881–941

[18] Song T, Garg S, Mokhtar R, Bui H and Reif J 2016 Analog computation by DNA strand displacement circuits *ACS Synth. Biol.* **5** 898–912

[19] Poje J E *et al* 2014 Visual displays that directly interface and provide read-outs of molecular states via molecular graphics processing units *Angew. Chem., Int. Ed. Engl.* **53** 9222–5

[20] Liu H, Wang J, Song S, Fan C and Gothelf K V 2015 A DNA-based system for selecting and displaying the combined result of two input variables *Nat. Commun.* **6** 1–7

[21] Kari L, Losseva E and Sosik P 2005 DNA computing and errors: a computer science perspective *Molecular Computational Models: Unconventional Approaches* (Hershey, PA: Idea Group Publishing)

[22] Faulhammer D, Lipton R J and Landweber L F 2000 Fidelity of enzymatic ligation for DNA computing *J. Comput. Biol.* **7** 839–48

[23] Durickovic I, Marchetti M and Claverie R, Fontana M D 2010 Experimental study of NaCl aqueous solutions by Raman spectroscopy: towards a new optical sensor *Appl. Spectrosc.* **64** 853–7

[24] Durickovic I 2016 Using Raman spectroscopy for characterization of aqueous media and quantification of species in aqueous solution *World's Largest Science, Technology & Medicine* pp 405–28 ch 19 (Rijeka: InTech)

[25] Li Z, Deen M J, Kumar S and Selvaganapathy P R 2014 Mint: Raman spectroscopy for in-line water quality monitoring—instrumentation and potential *Sensors* **14** 17275–303

[26] Burikov S, Dolenko T, Velikotnyi P, Sugonyaev A, and Fadeev V 2005 The effect of hydration of ions of inorganic salts on the shape of the Raman stretching band of water *Opt. Spectrosc.* **98** 235–9

[27] Burikov S, Dolenko T, Fadeev V and Sugonyaev A 2005 New opportunities in the determination of inorganic compounds in water by the method of laser Raman spectroscopy *Laser Phys.* **15** 1–5

[28] Palma B, Ferrari A, Bitar R, Cardoso A, Martin A and Martinho H 2008 DNA extraction systematics for spectroscopic studies *Sensors* **8** 3624–2

[29] Anokhin A, Gorelik V, Dovbeshko G, Pyatyshev A and Yuzyuk Y 2015 Difference Raman spectroscopy of DNA molecules *J. Phys. Conf. Ser.* **584** 1–6

[30] Dolenko T, Burikov S, Laptinskiy K, Moskovtsev A, Mesitov M and Kubatiev A 2015 Improvement of fidelity of molecular DNA computing using laser spectroscopy *Laser Phys.* **25** 1–10

[31] Laptinskiy K A, Burikov S A, Dolenko T A 2014 Determination of type and concentration of DNA nitrogenous bases by Raman spectroscopy *Proc. SPIE* **9448** 1–8

[32] Dolenko T, Burikov S, Laptinskiy K, Sarmanova O 2016 Improvement of fidelity of molecular DNA computations: control of DNA duplex melting using Raman spectroscopy *Laser Phys.* **26** 1–9

[33] Lee J Y, Shin S-Y, Park T H and Zhang B-T 2004 Solving traveling salesman problems with DNA molecules encoding numerical values *Biosystems* **78** 39–47

[34] James K D, Boles A R, Henckel D and Ellington A D 1998 The fidelity of template-directed oligonucleotide ligation and its relevance to DNA computation *Nucleic Acid Research* **26** 5203–11

[35] Hassoun M 1995 *Fundamentals of Artificial Neural Networks* (Cambridge, MA: MIT Press)

[36] Sandham W and Leggett M 2003 *Geophysical Applications of Artificial Neural Networks and Fuzzy Logic* (Dordrecht: Kluwer)

[37] Dolenko S, Burikov S, Dolenko T and Persiantsev I G 2012 Adaptive methods for solving inverse problems in Laser Raman spectroscopy of multi-component solutions *Pattern Recognit. Image Anal.* **22** 551–8

[38] Dolenko T, Burikov S, Efitorov A, Laptinsky K, Sarmanova O and Dolenko S 2016 Adaptive methods of solving inverse problems for improvement of fidelity of molecular DNA computations *Opt. Mem. Neur. Netw.* **25** 16–24

[39] Gerdova I V, Dolenko S A, Dolenko T A, Churina I V and Fadeev V V 2002 New opportunity solutions to inverse problems in laser spectroscopy involving artificial neural networks *Izv. Akad. Nauk Ser. Fiz.* **66** 1116–24

[40] Efitorov A, Burikov S, Dolenko T, Laptinskiy K and Dolenko S 2015 Significant feature selection in neural network solution of an inverse problem in spectroscopy *Proc. Comput. Sci.* **66** 93–102