Determination of type and concentration of DNA nitrogenous bases by Raman spectroscopy

Kirill A. Laptinskiy[*a], Sergey A. Burikov[a], Tatiana A. Dolenko[a]

[a]Moscow State University, Department of Physics, 119991, Leninskie Gory 1/2, Moscow, Russia.

ABSTRACT

In this paper results of application of laser Raman spectroscopy for increasing reliability of molecular DNA computations are presented. It is shown that elaborated method provides the accuracy of determination of concentration of individual nitrogenous bases 0.03 g/l and the accuracy of determination of total DNA concentration in the solutions 0.04 g/l by Raman spectra.

Keywords: DNA chain, molecular computations, nitrogenous bases, Raman spectroscopy

1. INTRODUCTION

Now it is obvious that in the near future development of conventional computing facilities (computers based on silicon elements) will come into collision with natural limitation in increase of quantity of elements per square unit of integral circuit. It leads to impossibility of increase of computational power. That is why the active search of alternative possibilities of development of computing facilities is carried out. Intrinsic interest of great amount of research groups is generated by native means of data storage, processing and transmission in biological systems. As it is known in living cells the genetic information is encoded by molecule of deoxyribonucleic acid – DNA. DNA is one of the two types of nucleic acids providing storage, transmission from generation to generation and realization of genetic program of development and functioning living organisms.

Molecular computations using DNA strands

The basic idea of computations using biological structures is the creation of new models and new algorithms on the basis of knowledge about DNA structure and functions and on the basis of operations on DNA molecules which are carried out in living cells by different ferments.

From the chemical point of view DNA is long polymeric molecule consisting of repeating proteins – nucleotides.[1, 2]. Every nucleotide consists of nitrogenous base (adenine (A), guanine (G), cytosine (C), thymine (T)), deoxiribose and phosphate group. Nucleotides are bonded by the phosphodiether bonds between deoxiribose and phosphate group. Thus, in these supermolecules the phosphate and the deoxiribose play role of supporting structure – backbone (they alternate in the strand). Information is encoded by the sequences of nitrogenous bases.

For elaboration computing structures based on DNA the specialists in the field of theory of computing and mathematical simulation describe the properties of DNA strands by formal machine languages.[3]. DNA molecules are represented by the double string with alphabet {A,T,G,C} in compliance with complementarity of nucleotides. Manipulations on the DNA molecules (elongation, contraction and cutting, modification, splicing, sequencing etc.) are as operations on these strings. With the use of theory of formal languages one can formalize the operations on DNA practically in full.

The principle opportunity of computing using DNA was demonstrated firstly by L. Adleman in 1994.[4] In his study the DNA molecules were used for solution of problem of evidence of existence of Hamiltonian way in graph using DNA-computer: if in the given directed graph (first and final vertexes are chosen) exists Hamiltonian way which passes through every vertex exactly one time. This problem was solved by the biochemical methods using specially synthesized

---

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

DNA strands. It took 7 days to solve this problem.[5] At that the most of efforts were spent not for synthesis of the initial data and the elaboration of algorithm but for filtration and extraction of already finished result in the convenient for human perception form.

Later on the approach suggested by Adleman was developed in the solution of number of another problems, for example, breaking one of encryption systems DES (Data Encryption Standard System)[6, 7], solution of some chess problems[8], for solution of post delivery with defined conditions and limitations[9], matrix multiplication[10] etc. Now on the basis of DNA-computations the algorithms of solution of some NP-full problems connected with logic, graph theory, computer networks, set theory, mathematical programming, optimization of processes and programs are theoretically elaborated[11].

1.2. Control of errors during molecular DNA-computations

In fact the computational process based on DNA strands represents a set of biochemical reactions. All operations with DNA molecules are rather complicated and laborious and that is why on the every stage of reactions the errors on molecular level are eventual[12]. The problem of last stage of reading of result – isolation, sorting and analysis of DNA strands using gel-electrophoresis is the most daunting problem. The modern methods do not provide sequencing strands with length of at least several thousands of bases at once. Solutions of the problem can be lost during the intermediate operations (molecules stick to walls of container), the point mutation can occur in DNA, DNA molecules tend to disintegrate as time goes by etc. The authors of many studies[13, 14, 15] note that the accuracy of computation depends strongly on concentration of biological molecules, time and temperature regime of biochemical reactions, type of buffer solution, values of pH of mixtures, type of PCR-primers. Thus one of the main problems in development of molecular DNA computations is the minimization of errors during biochemical reactions i.e. increase of reliability of molecular computations.

As it was mentioned above during molecular computations for the realization of different algorithms one has to deal with different DNA concentrations and different length of DNA strands, with different types and concentrations of buffer, different temperature profiles of reaction – i.e. with many changes of parameters of media influencing on the accuracy of solution of the problem. In real experiments the losses can achieve up to 10% of initial mass of molecules. At the same time the loss of even 1% of DNA molecules leads to incorrect solution of the problem. It means that it is necessary to control these parameters during full time of biochemical reactions. It is obvious that such methods of control must be non-invasive, express, and able to work in real time mode. Besides it is necessary to elaborate approach allowing simultaneous determination of as many parameters of media as possible. The methods of vibrational spectroscopy (mainly Raman spectroscopy) comply with these requirements.

The method of Raman spectroscopy provides the good results in solution of the problems of express monitoring of the state of natural water and aqua solutions[16], simultaneous determination of several parameters of aqua media[17], including the solution of multi-parametrical inverse problems of laser spectroscopy of aqua media by means of the modern methods of pattern recognition.[18, 19]

Laser Raman spectroscopy is widely used for study of the structure and properties of DNA molecules.[20, 21]. The bands corresponding to the vibrations of bonds in nitrogenous bases and the vibrations of DNA backbone are identified; the behavior of vibrational spectra of DNA in dependence on change of temperature, DNA conformation etc. is studied.[22, 23]. It provides an opportunity to use of this method for solution of the problem of control of state of reaction mixture during molecular computations.

This work is devoted to application of the method of Raman spectroscopy for increase of reliability of molecular computations using DNA strands – namely for the express and remote control and measurement of parameters of media where the biochemical reactions with DNA molecules occur. The results of the determination of concentration of individual nitrogenous bases and of the total concentration of DNA in solutions by Raman spectra are presented.

2. MATERIALS AND METHODS

For molecular DNA-computations (it is their peculiarity) one needs rather large amount of "working substance" – from units of g/l of DNA molecules (3.32 g/l, [24]) up to tens of g/l (32.5 g/l, [25]) – depending on the problem. Such concentration of DNA molecules in the solution provides the ability of using Raman spectroscopy in its conventional

variant – the signal is collected from the bulk of solution. It allows to control almost the entire bulk of "working substance".

2.1. Objects of research

As the objects of research 2 samples were used:

1) The Sample 1 – the solution of sodium deoxyribonucleate containing DNA extracted from salmon sperm. Maximal concentration was 20 g/l in the SSC buffer (3X) (water solution of 0.45 M NaCl+0.045M Na3C6H5O7*5.5H2O).

2)The Sample 2 – the solution containing specially synthesized and sequenced linear strands consisting of predetermined sequence of 100 nucleotides (83 adenine (A) bases and 17 (C) cytosine bases) in the SSC buffer (3X). Maximal concentration was 9g/l. This sample was synthesized and sequenced in the Institute of General Pathology and Pathophysiology of Russian Academy of Sciences.

The sequence of nucleotides was the following:

AAA-AAA-AAA-AAA-AAA-AAA-AAA-AAA-ACC-CAA-AAA-AAA-AAA-AAA-AAA-ACC-CAA-AAA-AAA-AAA-AAA-AAA-AAA-AAA-AAA-CCC-AAA-AAA-AAA-AAA-AAA-AAA-CCC-C.

2.2. Experiment

For the excitation of Raman signal the argon laser (wavelength 488 nm, output power 200 mW) was used. Spectra were registered in 90° geometry by CCD-camera in the range 500-4000 cm^{-1} with practical resolution 2 cm-1. In order to remove the background signal caused by elastic scattering the edge-filter was used. Spectra processing consisted of the subtraction of the fluorescence pedestal, the correction by spectral sensitivity of detector and accumulation time, normalizing by the square of water Raman valence band (the most intensive band near 3400 cm^{-1}). Such normalization allows to remove influence of instability of laser power.

In Fig. 1 the panoramic Raman spectrum of Sample 1 is presented. Raman bands of DNA are situated mostly in the low-frequency part of spectrum – from 500 up to 1600 cm^{-1}. The intensive band with the maximum near 3400 cm^{-1} is water Raman valence band, the band with the maximum near 1600 cm^{-1} – water bending band.
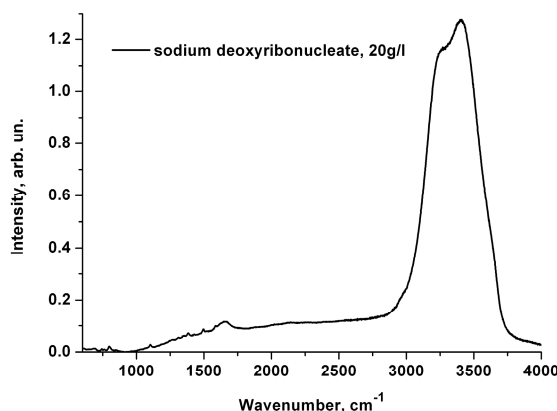

Fig.1 Raman spectrum of solution of the Sample 1.

2.3. Method of determination of concentration of substance by Raman spectra

The most important information about DNA-strands is contained in the low-frequency part of spectrum 1500-1600 cm^{-1} (Fig. 2). In this region there are the bands of vibrations of DNA backbone (vibrations of phosphate groups near 1100 cm^{-}

[1], vibrations of phosphodiether bonds near 840 cm-1) and the bands-markers of nitrogenous bases (adenine near 760, 1260, 1300, 1350 cm[-1], cytosine near 800 cm[-1], guanine near 685 cm[-1], thymine near 740 cm[-1]).
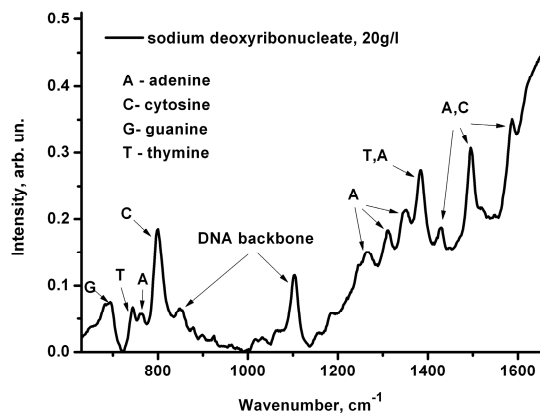

Fig.2 The low-frequency Raman spectrum of the solution of sodium deoxyribonucleate

Exactly these bands-markers provide the principal opportunity of determination of concentration of nitrogenous bases in the solutions: the nitrogenous bases can be identified by the position of bands and the concentration of corresponding nucleotide – by the intensity of bands-markers. At that the total concentration of DNA-molecules can be determined by the intensity of band near 1100 cm[-1] corresponding to vibrations of phosphate groups of DNA backbone. Moreover, for determination of the total DNA concentration one can use the band of vibrations of CH-groups of deoxyribose which is contained in backbone. This band is situated near 2900 cm[-1] and considerably overlaps with water Raman valence band (Fig. 3). But it is possible to extract this band by method of difference spectra – by subtraction of spectrum of solvent from the initial spectrum of solution (the inset in Fig. 3).
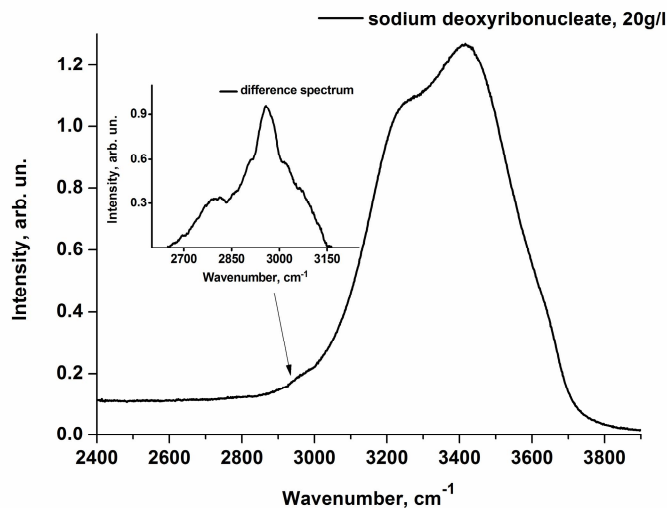

Fig. 3 Raman spectrum of solution of sodium deoxyribonucleate, the band of CH-groups is near 2900 cm[-1].

3. RESULTS

3.1. Determination of concentration of individual nitrogenous bases

During elaboration of the method of determination of concentration of individual bases the experimental dependence of Raman spectra of Sample 2 on the concentration of DNA (in the concentration region 0-9 g/l with increment 0.3 g/l) in the solution was obtained. The given sample represents the solution of linear DNA strands consisting of 100 nucleotides (83 adenine and 17 cytosine, see 2.1.)

For the extraction of useful Raman signal in the region 500-1500 cm^{-1} from the background caused by the more intensive elastic scattering method of so-called R-presentation was used.[26] In Fig. 4 the obtained spectra at different concentrations are presented.
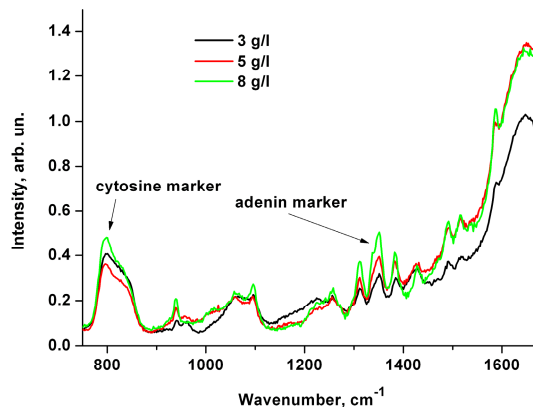

Fig. 4 Raman spectra of Sample 2 at different concentrations.

3.1.1. Determination of adenine concentration

For determination of adenine concentration the band of its vibrations in the region 1300-1400 cm^{-1} was used. For more correct extraction of these bands-markers the method of difference spectra was used (from spectra of solutions of DNA the spectrum of buffer solution was subtracted). In Fig. 5 one can see the dependence of integral intensity of adenine bands-markers in the region 1300-1400 cm^{-1} on its concentration.
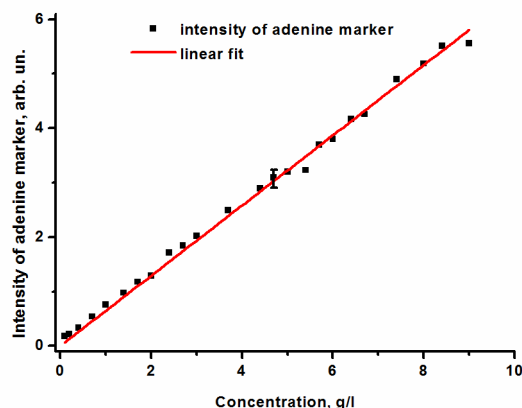

Fig. 5 The dependence of intensity of adenine marker on concentration of Sample 2.

The obtained concentration dependence can be approximated with good accuracy by linear function and it allows to determine the concentration of adenine with the accuracy 0.03 g/l.

3.1.2. Determination of cytosine concentration

The determination of concentration of cytosine is complicated by the fact that its marker – the band near 800 cm^{-1} - considerably overlaps with the band of vibrations of phosphordiether bond between nucleotides near 840 cm^{-1} (Fig.4). For extraction of band-marker of cytosine the spectra in this region (770-870 cm^{-1}) were decomposed into two Gaussians curves (Fig. 6). Decomposition of the band into Gaussian curves was performed by the least-squares method.
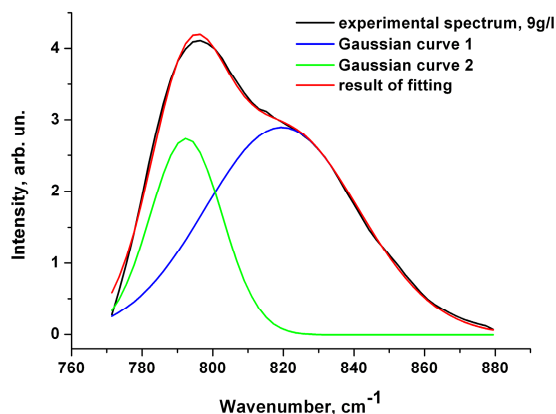

Fig. 6 Decomposition of spectra in the region near 800 cm^{-1} in two Gaussian curves.

For determination of cytosine concentration the square under the Gaussian curve with maximum near 795 cm^{-1} was used. The obtained concentration dependence of integral intensity of pointed Gaussian curve was also approximated by the linear function. It allows to determine the concentration of cytosine with the accuracy 0.03 g/l.

3.2. Determination of total DNA concentration

As it was mentioned in 2.3, for determination of the total DNA concentration in the solution one can use the bands of phosphate groups (1100 cm^{-1}) and the bands of valence vibrations of CH-groups (2900 cm^{-1}). In this experiment the better accuracy was achieved by using band of valence vibrations of CH-groups (they are more intensive in comparison with the band of phosphate groups). Because of the contribution of bands of SSC buffer in this part of Raman spectrum is large, the difference spectra were constructed for correct extraction of bands of CH-groups: in the region near 2800-3000 cm^{-1} the spectrum of SSC-buffer was subtracted from spectra of DNA solutions (similar difference spectrum is demonstrated in the inset in Fig. 3).

It turned out that the integral intensities of difference bands of vibrations of CH-groups in the region 2850-3000 cm^{-1} also linearly depend on DNA concentration. The obtained dependence allows to determine the total DNA concentration with the accuracy 0.04 g/l.

4.   CONCLUSIONS

As the result of realized experiments it was shown that the method of laser Raman spectroscopy provides an principal opportunity of determination of individual nitrogenous bases in DNA solutions and the total DNA concentration by Raman spectra. The suggested method is non-invasive, remote and can be used in real-time mode. It provides the accuracy of determination of concentration of individual nitrogenous bases 0.03 g/l and the accuracy of determination of total DNA concentration 0.04 g/l. The obtained results allow to used the elaborated method for monitoring of DNA solutions and control of errors during molecular computations using DNA strands.

5. ACKNOWLEDGMENTS

REFERENCES

[1] Watson, J., Crick, F., "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid," Nature 171(4356), 737–738 (1953).

[2] Bruce, A., [Molecular Biology of the Cell], Garland Science, New York & London, 1392 P (2007).

[3] Paun, Gh., Rozenberg G., Salomaa, A., [DNA Computing: New Computing Paradigms], Springer, Heidelberg (1998).

[4] Adleman, L.M., "Molecular computation of solutions to combinatorial problems," Science 266, 1021-1024 (1994).

[5] Adleman, L., "Computing with DNA," Scientific American 279(2), 54-61 (1998).

[6] Lipton, R.J., "DNA Solution of Hard Computational Problems," Science, New series 268(5210), 542-545 (1995).

[7] Boneh, D., Dunworth, C., Lipton, R., "Breaking DES using a molecular computer," Princeton university Technical report CS-TR-489-95, (1995).

[8] Faulhammer, D., Cukras, A.R., Lipton, R.J., and Landweber, L.F., "Molecular computation: RNA solutions to chess problems," Proc. National Acad. Sciences 97(4), 1385-1389 (2000).

[9] Kari, L., Gloor, G., Yu, S., "Using DNA to solve the Bounded Post Correspondence Problem," Theoretical Computer Science 231(2), 193–203 (2000).

[10] Oliver, J.S., "Computation with DNA: matrix multiplication," // DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Amer. Mathematical Society 44, 113–122 (1999).

[11] Cormen, Th. H.; Leiserson, C. E.; Rivest, R. L., [Introduction to Algorithms], MIT Press and McGraw-Hill, Cambridge MA, 1048 P (1990).

[12] Kari, L., Losseva, E., Sosik, P., [DNA Computing and Errors: A Computer Science Perspective, in Molecular Computational Models: Unconventional Approaches Ed. by M. Gheorhe], Idea Group Inc., Calgary, 56-77 (2005).

[13] Benenson, Y., Adar R., Paz-Elizur T., Livneh Z., and Shapiro, E., "DNA molecule provides a computing machine with both data and fuel," Proc. of the National Acad. of Sciences 100 (5), 2191-2196 (2003).

[14] Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E., "An autonomous molecular computer for logical control of gene expression", Nature 429 (6990), 423–429 (2004).

[15] Faulhammer D., Lipton R.J., Landweber L.F., "Fidelity of enzymatic ligation for DNA computing," Journal of computational biology: a journal of computational molecular cell biology 7 (6), 839–848 (2000).

[16] Dolenko, T.A.,Churina, I.V., Fadeev, V.V., Glushkov, S.M., "Valence band of liquid water Raman scattering: some peculiarities and applications in the diagnostics of water media," Journal of Raman Spectroscopy 31(8-9), 863-870 (2000).

[17] 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).

[18] 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).

[19] Dolenko, S., Burikov S., Dolenko, T., Efitorov, 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 8681, 805–812 (2014).

[20] Thomas, G.J., Jr., "Raman spectroscopy of protein and nucleic acid assembles," Annu. Rev. Biophys. Biomol. Struct. 28, 1-27 (1999).

[21] Prescott, B., Steinmetz W., Thomas, G.J., Jr., "Characterization of DNA Structures by Laser Raman Spectroscopy," Biopolymers 23, 235-256 (1984).

[22] Movileanu, L., Benevides J. M., Thomas, G. J., Jr., "Temperature Dependence of the Raman Spectrum of DNA. II. Raman Signatures of Premelting and Melting Transitions of Poly (dA)Poly(dT) and Comparison with Poly(dA-dT)Poly(dA-dT)," Biopolymers 63, 181–194 (2002).

[23] Martin, J.C., Wartell, R. M., "Changes in Raman Vibrational Bands of Calf Thymus DNA During the B-to-A Transition," Biopolymers 21, 499-512 (1982).

[24] 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).

[25] 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).

[26] Murphy, W. F., Brooker, M. H., Faurskov O., Praestgaard N.E., Bertie J.E., "Further assessment of reduction procedures for Raman spectra," Journal of Raman spectroscopy 20, 695-699 (1989).