Ludmila Frolova*, Ivan Aponasevich, Arthur Husainov

Ecological estimation of fresh water resources using the markerproteins of indicators species of *Bacillariophyceae*

Еколошка процена ресурса свеже воде употребом протеинских маркера индикатора врста *bacillariophyceae*

Kazan Federal University, Russia
*e-mail: Lucie.Frolova@gmail.com

Abstract

In the article, we suggestedto use the rbcL protein as a marker for the identification of Bacillariophyceaein the bioindication method of the ecological estimation of fresh water resources.Here we perform the results of the Multiple alignments and 3D models respectively of 45and 27rbcL proteins indicator species of Bacillariophyceae from the V. Sladechek (1973) list,whichavailable in the international database of proteins - GenPept (NCBI) at present time. Short (6-35) amino- acid sequencesof 27 unique variable sites identified for 19 species of Bacillariophyceae, which normally used as bioindicators of water saprobity. Theseunique variable sites located on rbcL protein surface of Bacillariophyceae can be specifically find by antibodies in one water samples with use of the IFA-method. The use the rbcL protein as marker for identification of Bacillariophyceae in water samples gives best reliable estimation of the water ecological condition in a compare with the traditional identification of the organisms under the microscope.

Keywords: rbcLproteinmarker, Bacillariophyceae, algae, indicator species

Introduction

At present new methods of molecular genetics and bioinformatics are using in the fields of the biology and the ecology. For example, the Consortium for the Barcode of Life (CBOL) plant working group has recommended a standard barcode comprising ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit (rbcL) for the barcoding of all land plants [1,2]. And the mitochondrialcytochrome c oxidase subunit 1 (CO1) uses forthe barcoding of all animals [3, 4, 5]. The primary sequences of DNA and proteins of plants and animals are located in the international databasesGenBank and GenPept[6].

In our previously articles we shown that the DNA barcode with gene CO1 good work for animals and we can designed the unique variable sites for the antibodies [7,8]. Here we continue our investigations for Bacillariophyceae using the rbcL gene product.We took 45 primary sequences of proteins of the indicator species of Bacillariophyceaewhich presented in database GenPeptbeginning from 2004 till 2014, most of them – 23 records–are from USA, 9 records from Germany and 5 records from France, 2 records from Belgium, and others on 1 record from Chine, Republic of Korea, Spain, South Korea, Russia[6].Such approach based on molecular data increases the level of a confidence of the indicator species identification special in the Bioindicationmethods which use for the ecological estimation of water of fresh water resources.

At this stage, we conducted researches on use of rbcL protein as Bacillariophyceaeidentification marker which resulted in water samples, in more reliable assessment of reservoir ecological condition.We suggested to use the marker genes and their products for the identification of indicator species as a part of the method of Bioindication for the estimation of water environment. We used the marker genes and the proteins such as *COI*for animals and *rbcL*for plants. Our approach based on the way of search of the unique sites of marker genes and proteins for each organism using the methods of bioinformatics such as the Multiple alignments and Phylogenetic analysis, and 3D modelling for proteins.

Materials and methods

In this work indicator species of Bacillariophyceaefrom V. Sladechek's list are used [9], amino-acid sequences of rbcL protein of Bacillariophyceaeindicator species are received from the international database of proteins - GenPept database on the NCBI site [6].

Multiple alignment of rbcL protein amino-acid sequences of Bacillariophyceaeindicator species is performed with using the Clustal Omega program [10], the rbcL protein amino-acid sequences of Bacillariophyceaeare analysed in the Jalview program [11].

3D model operations of Bacillariophyceaeindicator speciesof rbcL proteins is carried out in the SWISS-MODEL program [12], and the analysis of these models in the PDB format is carried out in the Jalview program.

Table 1. Ecologic-and-genetic database of indicator species of Bacillariophyceae

№	Species	Saprobity*	Indicator weight	Accession number in the database GenPept
1	Achnanthescoarctata	x	0.10	AEB91214
2	Amphora normanii	x	0.10	CAM97884
3	Amphora ovalis	o-b	1.65	AIT92093
4	Asterionellaformosa	o-b	1.40	AEB91253
5	Caloneis amphisbaena	b-a	2.35	AIT92083
6	Cymbellaaspera	b	2.20	AIY31910
7	Cymbellacistula	b	1.80	AIY31913
8	Cymbella Helvetica	x-o	0.50	AIY31915
9	Cymbellalanceolata	b	1.90	AIY31917
10	Cymbellanaviculiformis	b	2.00	AIY31926
11	Didymospheniageminata	x	0.10	AIY31931
12	Epithemiaturgida	b	2.00	AEB39377
13	Eunotiapectinalis	x	0.20	AEB91256
14	Fragilariacapucina	o-b	1.60	AGG86633
15	Fragilariacrotonensis	o-b	1.40	CAN84680
16	Gomphonemaacuminatum	b	1.70	AIY31964
17	Gomphonemaangustatum	o	1.15	AIY31946
18	Gomphonemacapitatum	b	2.00	AAT78574
19	Gomphonemaclevei	x	0.30	AFV95053
20	Gyrosigmaacuminatum	b	2.20	AEB91218
21	Hantzschiaamphioxys	a	2.90	AEB39371
22	Melosiravarians	b	1.85	AGN91098
23	Naviculacryptocephala	a	2.70	AEB91223
24	Naviculapupula	b	2.20	AGT21406
25	Navicularadiosa	o-b	1.60	AIT92055
26	Navicularhynchocephala	a	2.70	AEZ56384
27	Nitzschiapalea	a	2.75	CBH19906
28	Nitzschiasigmoidea	b	2.00	CBH19911
29	Pinnularia borealis	x-o	0.40	AER42043
30	Pinnulariamesolepta	o	1.15	CAM97949
31	Pinnulariamicrostauron	o	0.80	CAM97908
32	Pinnulariasubcapitata	o	1.00	AER42030
33	Pinnulariaviridis	b	2.10	CAM97948
34	Rhizosoleniacurvata	b	1.85	AGC59927
35	Rhopalodiagibba	o	1.00	AEB39374
36	Stephanodiscushantzschii	a	2.70	AFO84271
37	Surirellabiseriata	b	2.00	AGE34629
38	Surirellacapronii	b	2.00	AGE34620
39	Surirellalinearis	b	2.20	AGE34603
40	Surirellaovata	b	1.85	AEB91278
41	Surirellaspiralis	o	1.00	AGE34584
42	Surirellatenera	b	2.10	AGE34632
43	Synedraparasitica	b-a	2.50	AEB91231
44	Synedratabulata	a	2.70	AEB91210
45	Tabellariafenestrata	o-b	1.40	AEB91204

* x–xenosaprobity; o–oligosaprobity; b – beta-mesosaprobity; a – alpha-mesosaprobity

As we can see from above table, 45 speciesofBacillariophyceae are indicators of different zones of fresh water reservoir saprobity: from „clear" (xenosaprobicwith weighing 0.1) to „dirty" (the alpha-mesosaprobicwith weighing 2.7).

We believe that Bacillariophyceae indicator organisms can be identified by means of rbcL marker genes products and thus approach will increase reliability of reservoir ecological condition estimation.

Identification of rbcL proteins epitopes of Bacillariophyceae

For identification of unique sites of rbcL proteins in BacillariophyceaeMultiple alignment of 45 amino-acid sequences of rbcL protein indicators is carried out (Fig. 1).

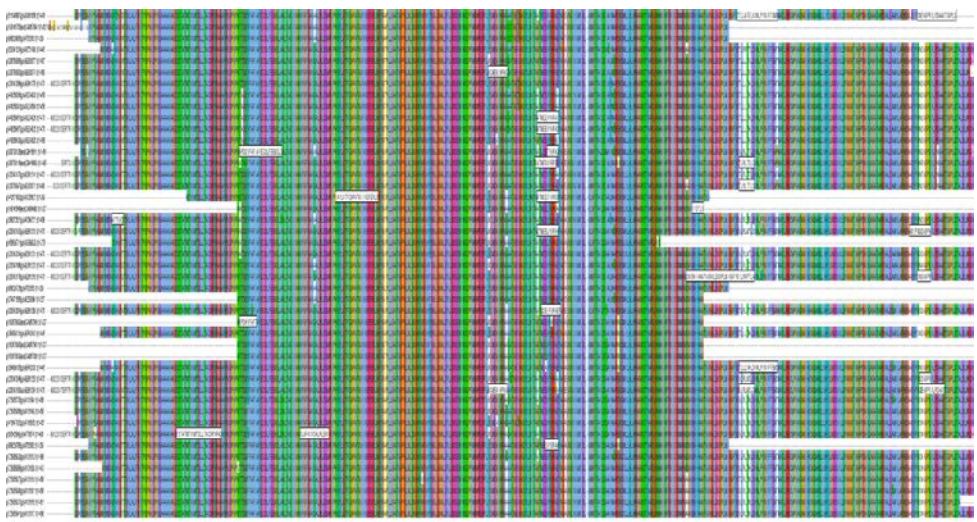


Figure 1. Multiple alignment of amino-acid sequences rbcLBacillariophyceae

Analysis of multiple amino-acid sequences alignment of rbcL protein in Bacillariophyceae organisms showed the existence of 27 unique sites for 19 species of Bacillariophyceae from 6 to 35 amino-acid (a.o.) with an optimum length from 6 to 20 a.o. for antibodies discernment (table 2).

Table 2. Unique variable sequences of the rbcL protein of indicator species of Bacillariophyceae

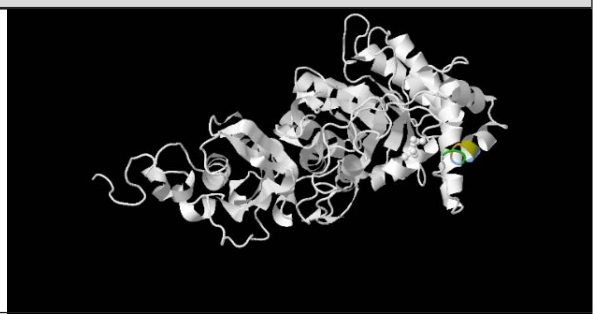
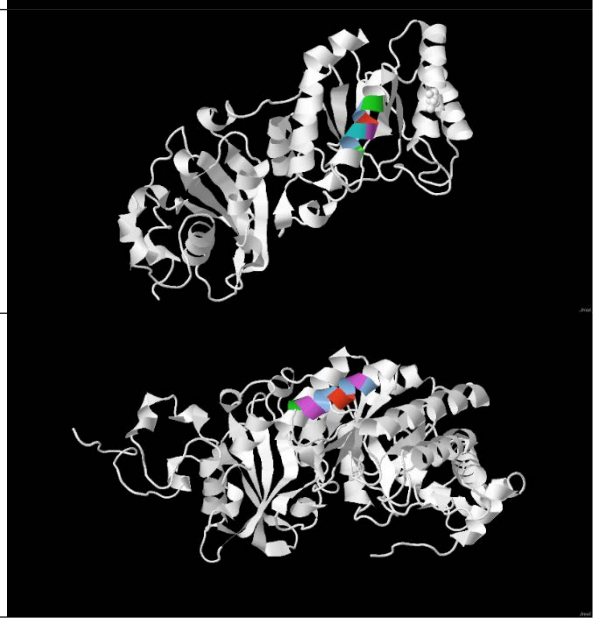
Species / Accession number in database GenPept	Saprobity*	Indicator weight	Unique fragments of amino acid sequences of the protein rbcL of indicators species of Bacillariophyceae with the positions in the alignment
Asterionellaformosa AEB91253.1	o-b	1.40	QQVGPV (443-448)
Caloneis amphisbaena AIT92083.1	b-a	2.35	QVYERAN (234-240)
Eunotiapectinalis AEB91256.1	x	0.20	QDEIFKRAEF (251-260)
Fragilariacrotonensis CAN84680.1	o-b	1.40	IISTLV (232-237)
Gomphonemacapitatum AAT78574.1	b	2.00	YSTATWTVVWTDLLTACDRYRAQ (65-87) XLGFKXVSALRLENI (128-142)
Gyrosigmaacuminatum AEB91218.1	b	2.20	CGVDHIHAGTVVGKLEGDPLMIKGFYDILRKPTLA (324-358) SNQVGPR (442-448)
Hantzschiaamphioxys AEB39371.1	a	2.90	TLRLTTLS (350-357)
Melosiravarians AGN91098.1	b	1.85	TTLLATELKVNLPYGIFFDMD (338-358) SNEVGPRILRDAAKTCGPLQ (430-449)
Nitzschiapalea CBH19906.1	a	2.75	GATMEDVYERS (243-253) TLRLTSLQ (346-353)
Nitzschiasigmoidea CBH19911.1	b	2.00	APDQYFAFIAYECDLFEEGSLS (82-103) TYKRAQ (239-244)
pinnulariasubcapitata AER42030.1	o	1.00	TLLLVHLDVNLPYGIFFEMS (326-345)
Pinnulariaviridis CAM97948.1	b	2.10	KPDHYFAFT (2-10)
Rhizosoleniacurvata AGC59927.1	b	1.85	IAYLKTFQGPATGIIVERERLD (77-98) GTMDECIKRGD (180-190)
Rhopalodiagibba AEB39374.1	o	1.00	NCMEGINRAC (212-221)
Stephanodiscushantzschii AFO84271.1	a	2.70	AYTVKD (20-25)
Surirellabiseriata AGE34629.1	b	2.00	ATMEEVYKRAI (249-259)
Surirellatenera AGE34632.1	b	2.10	GTMEEVYKRAS (249-259)
Synedratabulata AEB91210.1	a	2.70	ADYFNQQVGPA (438-448)
Tabellariafenestrata AEB91204.1	o-b	1.40	NCMEGINRAV (224-233) VLRLMSLQ (351-358) EVGPVILRQAAS (444-455)

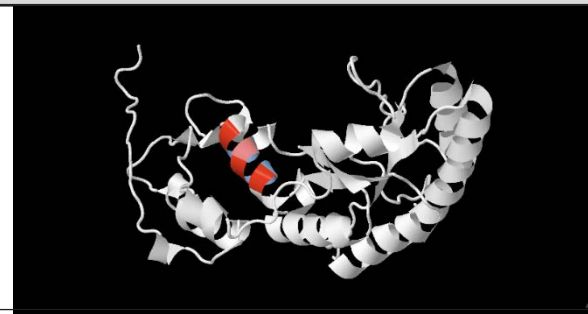
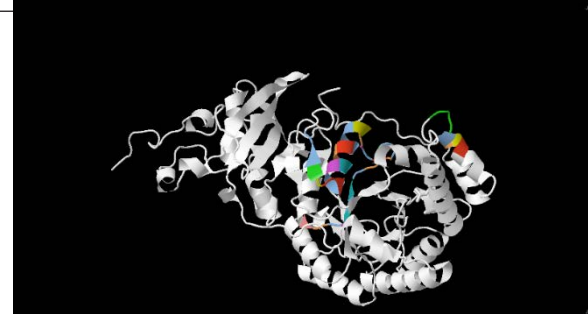
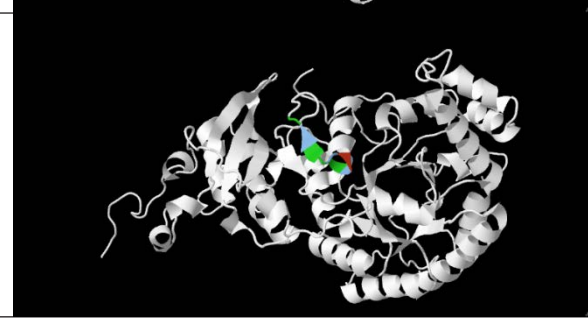
* x–xenosaprobity; o–oligosaprobity; b – beta-mesosaprobity; a – alpha-mesosaprobity

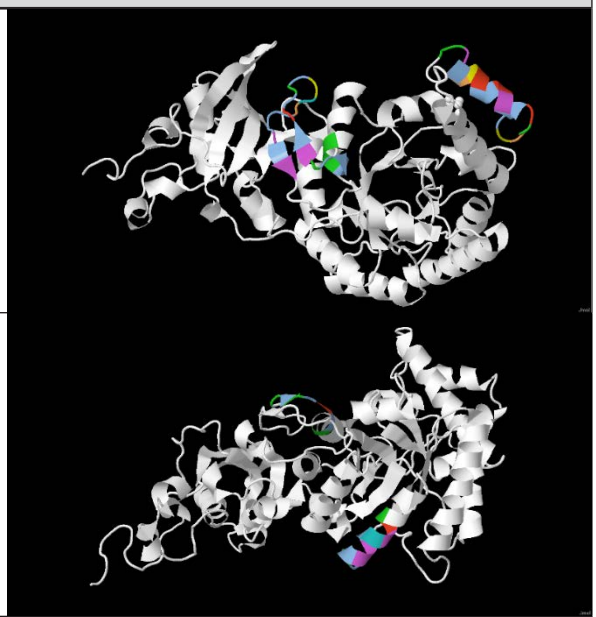
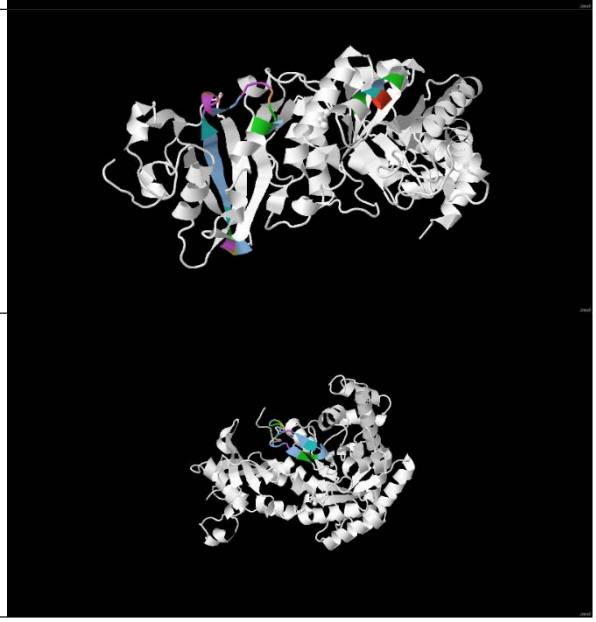
Verification of epitopes availability for antibodies discernment

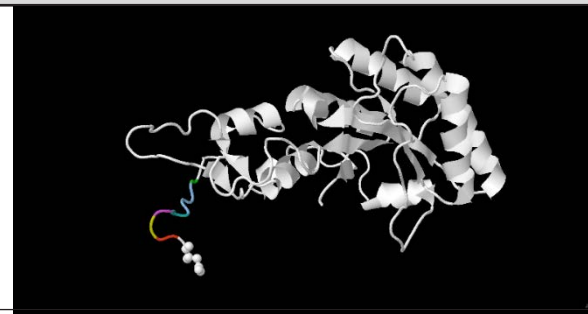
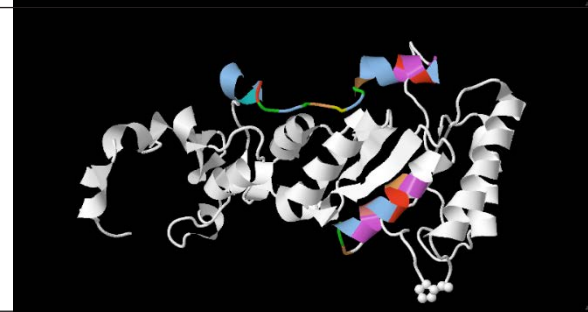
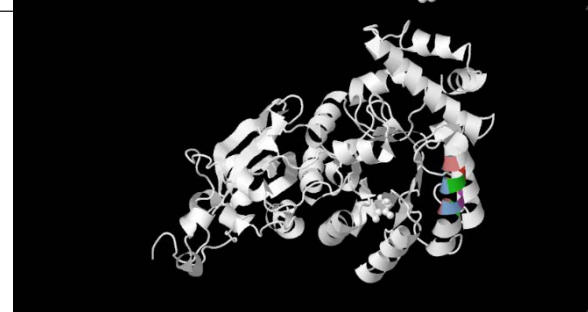
3D models are constructed for verification of 27 unique sites of rbcL proteins indicator in Bacillariophyceaeavailability for antibodies discernment (Tabl. 3).

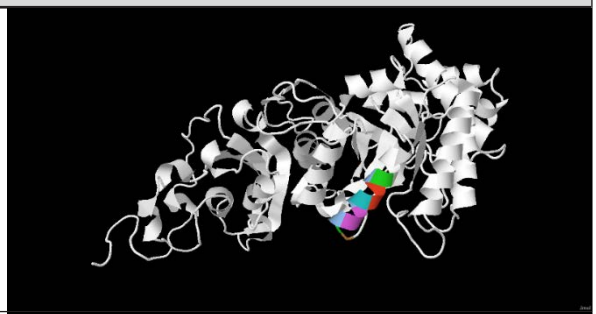
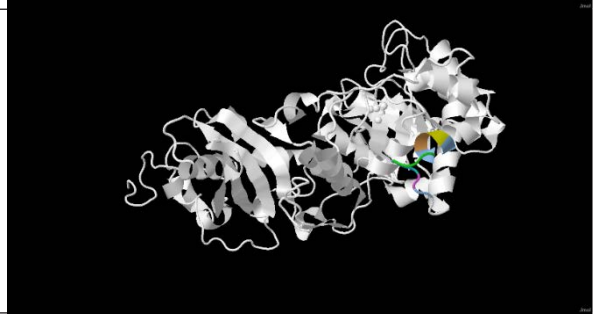
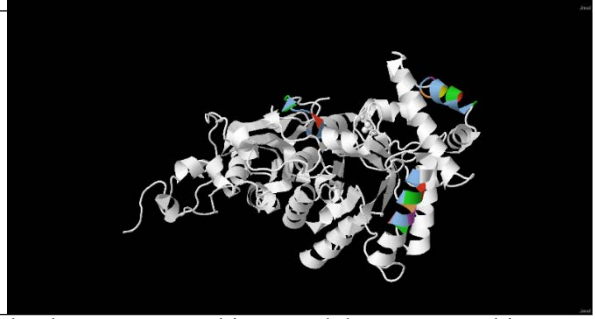
Table 3. 3D-models of amino-acid sequences rbcL of some indicator species of Bacillariophyceae

The indicator species of Bacillariophyceae / Accession number in the databank GenPept / saprobity*/ indicators weigh	3D-models of amino-acid sequences rbcL
Asterionellaformosa AEB91253 o-b/1.40	
Caloneis amphisbaena AIT92083.1 b-a/2.35	
Eunotiapectinalis AEB91256 x/0.20	

The indicator species of Bacillariophyceae / Accession number in the databank GenPept / saprobity*/ indicators weigh	3D-models of amino-acid sequences rbcL
Fragilariacrotonensis CAN84680 o-b/1.40	
Gomphonemacapitatum AAT78574.1 b/2.00	
Gyrosigmaacuminatum AEB91218 b/2.20	
Hantzschiaamphioxys AEB39371.1 a/2.90	

The indicator species of Bacillariophyceae / Accession number in the databank GenPept / saprobity*/ indicators weigh	3D-models of amino-acid sequences rbcL
Melosiravarians AGN91098 b/1.85	
Nitzschiapalea CBH19906 a/2.75	
Nitzschiasigmoidea CBH19911 b/2.00	
Pinnulariasubcapitata AER42030 o/1.00	

The indicator species of Bacillariophyceae / Accession number in the databank GenPept / saprobity*/ indicators weigh	3D-models of amino-acid sequences rbcL
Pinnulariaviridis CAM97948 b/2.10	
Rhizosoleniacurvata AGC59927 b/1.85	
Rhopalodiagibba AEB39374.1 o/1.00	
Surirellabiseriata AGE34629.1 b/2.00	

The indicator species of Bacillariophyceae / Accession number in the databank GenPept / saprobity*/ indicators weigh	3D-models of amino-acid sequences rbcL
Surirellatenera AGE34632.1 b/2.10	
Stephanodiscushantzschii AFO84271 a/2.70	
Synedratabulata AEB91210 a/2.70	
Tabellariafenestrata AEB91204 o-b/1.40	

* x–xenosaprobity; o–oligosaprobity; b – beta-mesosaprobity; a – alpha-mesosaprobity; p – polysaprobity

The analysis of 3D models on rbcL protein indicator for Bacillariophyceae organisms showed that all 27 unique sites are located on molecule surfaces. They are located on alpha spirals and can be therefore accessible to antibodies discernment. It should be noted that these Bacillariophyceae organisms are used as bioindicators for reservoir saprobity in the wide range, from xeno-saprobic zone with weighing 0.2, for example - *Eunotiapectinalis*, to an alpha saprobic zone with weighing 2.7, for example - *Stephanodiscushantzschii*.

Conclusion

The analysis of 45 rbcL proteins indicator of Bacillariophyceaeorganisms from the V. Sladechek (1973) list revealed existence of 27 unique sites of rbcL amino-acid sequences for 19 indicator species of Bacillariophyceaeorganisms. Unique variable sites are located on rbcL protein surface of Bacillariophyceaeorganisms, they do not coincide with 27 unique variable sites of indicator Bacillariophyceaeorganisms, and can be specifically distinguished by antibodies for precise identification of Bacillariophyceaespecies among organisms in one water sample with use of the IFA-method.

The studied organisms are indicators of fresh-water reservoirs saprobity in the wide range: from „clear" to „dirty", and, therefore, their identification on marker rbcL protein will allow to estimate reservoir ecological condition in shorter terms authentically.

Therefore, it is allows us to use the antibodies for direct qualification and quantification of the indicator Bacillariophyceae species in the samples of water using the rbclproteins and together with CO1 proteins of water indicator invertebrate species make the conclusion of ecological state of the water resources.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References:

CBOL Plant Working Group (2009): A DNA barcode for land plants. ProcNatlAcadSci U S A.106 (31): 12794-12797.

Hebert PDN, Ratnasingham S, deWaard JR, (2003): Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Roy SocLond B 270:96–99; doi: 10.1098/rspb.2003.0025.

Hebert PD, Cywinska A, Ball SL, de Waard JR, (2003): Biological identifications through DNA barcodes. Proc Roy SocLond B 270:313–321; doi: 10.1098/rspb.2002.2218.

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R., (1994): DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar BiolBiotechnol 3:294–299.

The National Center for Biotechnology Information – http://www.ncbi.nlm.nih.gov/.

Frolova LL., Kayumov AR., Husainov AM., (2015): Evaluation of the Ecological State Of Water Reservoirs Using the COI Protein of Zooplankton. RJPBCS 6(4): 82-94.

Frolova LL., Husainov AM., (2015): Evaluation of the Water Reservoirs Ecological State Using the COI Protein of Zoobenthos. RJPBCS 6(5): 1636-1649.

Sladechek V. (1973): System of water quality from the biological point of view. Arch. Hydrobiol. Ergeb. Limnol 179-191.

Sievers F., Wilm A., Dineen D., Gibson TJ., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson JD., Higgins DG. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539; doi: 10.1038/msb.2011.75 (http://www.ebi.ac.uk/Tools/msa/clustalo/).

Waterhouse AM., Procter JB., Martin DMA., Clamp M., Barton G J., (2009): Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. Bioinformatics 25 (9) 1189-1191; doi: 10.1093/bioinformatics/btp033 (http://www.jalview.org/).

Studer G., Schmidt T., Kiefer F., Cassarino TG., Bertoni M., Bordoli L., Schwede T. (2014): SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research; doi: 10.1093/nar/gku340 (http://swissmodel.expasy.org/).