"Prostate Cancer Proteomics" Database

S. S. Shishkin[1*], L. I. Kovalyov [1], M. A. Kovalyova [1], K. V. Lisitskaya [1], L. S. Eremina[1], A. V. Ivanov [1], E. V. Gerasimov[1], E. G. Sadykhov[1], N. Y. Ulasova[2], O. S. Sokolova[2], I. Y. Toropygin[3], V. O. Popov[1]

[1] Bach Institute of Biochemistry, Russian Academy of Sciences
[2] Lomonosov Moscow State University
[3] Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences
*E-mail: shishkin@inbi.ras.ru
Received 28.09.2010

ABSTRACT **A database of Prostate Cancer Proteomics has been created by using the results of a proteomic study of human prostate carcinoma and benign hyperplasia tissues, and of some human-cultured cell lines (PCP, http://ef.inbi.ras.ru). PCP consists of 7 interrelated modules, each containing four levels of proteomic and biomedical data on the proteins in corresponding tissues or cells. The first data level, onto which each module is based, is a 2DE proteomic reference map where proteins separated by 2D electrophoresis, and subsequently identified by mass-spectrometry, are marked. The results of proteomic experiments form the second data level. The third level contains protein data from published articles and existing databases. The fourth level is formed with direct Internet links to the information on corresponding proteins in the NCBI and UniProt databases. PCP contains data on 359 proteins in total, including 17 potential biomarkers of prostate cancer, particularly AGR2, annexins, S100 proteins, PRO2675, and PRO2044. The database will be useful in a wide range of applications, including studies of molecular mechanisms of the aetiology and pathogenesis of prostate diseases, finding new diagnostic markers, etc.**
KEYWORDS proteomics, prostate cancer, digital database
ABBREVIATIONS 2DE—two-dimensional gel electrophoresis, BPH—benign prostate hyperplasia, PCa—prostate cancer, PCP—prostate cancer proteomics

INTRODUCTION

The first decade of the post-genome era was marked by a rapid development in the field of bioinformatics, the extension of major databases (such as NCBI and UniProt), and the creation of specialised information resources for biomedical research in many countries [1−4]. The impressive resources created in Ireland (UCD-2DPAGE, http://proteomics-portal.ucd.ie:8082/cgi-bin/2d/2d.cgi) [2] and India (Human Proteinpedia, www.humanproteinpedia.org) [3] make the state of things in Russia pale in comparison.

Currently, one of the most important tasks for biomedical research is to find efficient prostate cancer (PCa) biomarkers which would enable new diagnostic methods [5−8]. The fact that in recent years the PCa incidence rate has dramatically increased worldwide [9, 10], and particularly in Russia, making PCa the most frequent male oncological disease in some countries [11, 12], is reason enough to pay close attention to this disease. In early diagnostics of PCa at the moment it is important to establish the presence of one of the most studied biomarkers, the so-called Prostate-Specific Antigens (PSA), in the blood. The test, however, is known to produce a significant number of false-positive and false-negative results, leading to the wrong clinical and financial outcomes [5, 7]. Therefore, in the U.S. and in other Western countries, new PCa biomarkers are being actively sought, an initiative recently stimulated by the development of proteomic and other post-genome technologies [6, 8, 13].

Since 2005, the Bach Institute of Biochemistry, in collaboration with other research and medical institutions, has been researching new PCa biomarkers by utilising various proteomic technologies [14, 15]. In 2009, the "Prostate Cancer Proteomics" (PCP, http://ef.inbi.ras.ru/) national database was created in order to facilitate this research, summarising experimental and referenced published data and providing links to several other biomedical Internet databases. This paper describes the structure and capabilities of the new, extended PCP version.

MATERIALS AND METHODS

Biomaterials— Biopsy and surgical samples of prostate tissues from patients with PCa ($n = 72$) and benign prostatic hyperplasia (BPH, $n = 69$) were provided by staff members of the Urology Department of the Botkin Clinical Hospital (Moscow). Diagnosis
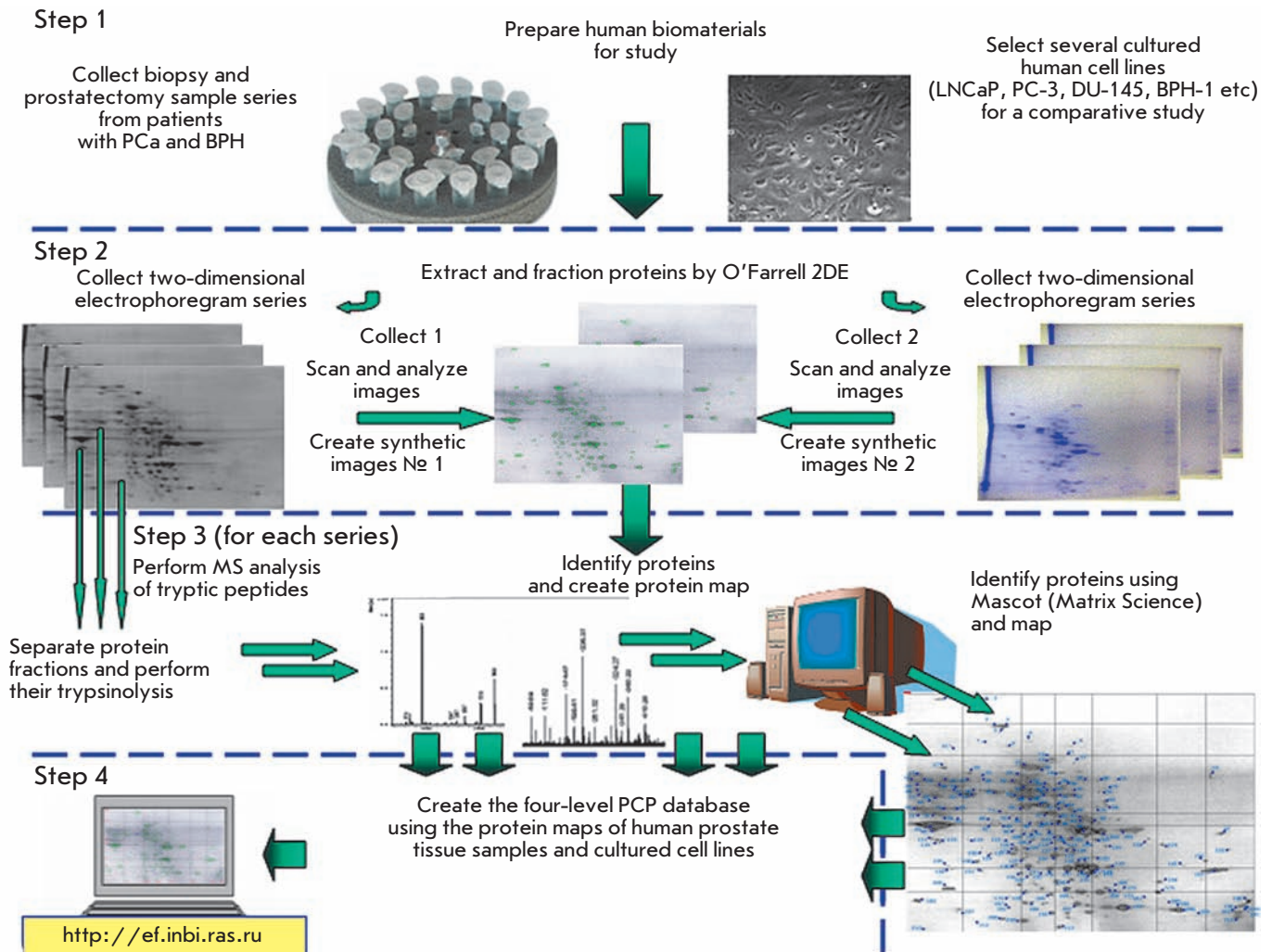
Fig. 1. Main steps in the proteomic study of prostate tissue samples from patients with malignant and benign tumors, as well as from several cultured human cell lines.

was performed using clinical, histological, and immunochemical (PSA level) tests. Histological verification was performed via U.S.-controlled transrectal multifocal needle biopsy; up to 18 tissue samples from various prostate zones per patient were taken [16, 17]. All PCa cases were found to be adenocarcinoma. Gleason score was determined by following the standard procedure [16, 17].

In parallel tests, we analysed the proteins of the PC-3 (ACC 465), DU-145 (ACC 261), and BPH-1 (ACC 143) cell cultures purchased from the German Collection of Microorganisms and Cell Cultures, as well as the proteins of cultured cells of the LNCaP line provided by Dr. I. G. Shemyakin (Obolensk National Science Centre for Applied Microbiology and Biotechnology). The cells were cultured in the RPMI-1640 medium with HEPES, sodium pyruvate, gentamicine and 20% fetal bovine serum (FBS) [18], using cell culture plastic (Costar, USA and Nunc, Denmark) in a CO_2-incubator (Sanyo, Japan). In addition, we studied proteins from the cultured cells of two lines of human rhabdomyosarcoma (A-204 and RD) purchased from the Ivanovsky Virology Institute, RAMS, and proteins from the cultured normal human myoblasts kindly provided by Dr T. B. Krohina [19].

The preparation of protein extracts, their O'Farrell 2DE fractioning, Coomassie Blue R-250 and silver nitrate staining, and 2DE analysis were performed following the techniques described in [20, 21]. In addition, we used a 2DE procedure with isoelectric focusing using IPG-PAGE and Ettan IPGphor 3 kit (GE Healthcare), according to the manufacturer's protocol. Pro-
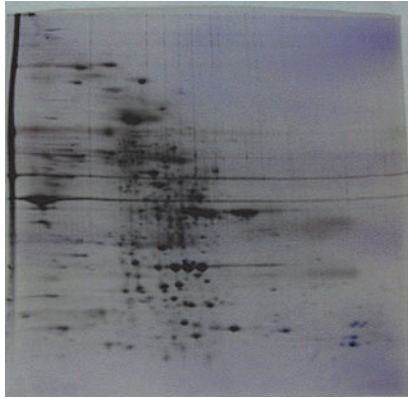
Fig. 2. Results of proteomic study of malignant prostate tissue samples. A. Typical 2D electrophoregram of proteins from prostate tissue biopsy samples with PCa. B. Synthetic 2DE image of proteins from prostate tissue biopsy samples with PCa; reference fractions are marked with red ovals. Vertical axis: molecular weights. Horizontal axis: isoelectric points. C. 2D protein map constructed from synthetic 2DE image of proteins from prostate tissue biopsy samples with PCa; fractions characterized with 2DE are marked with blue arrows.

a

b

c
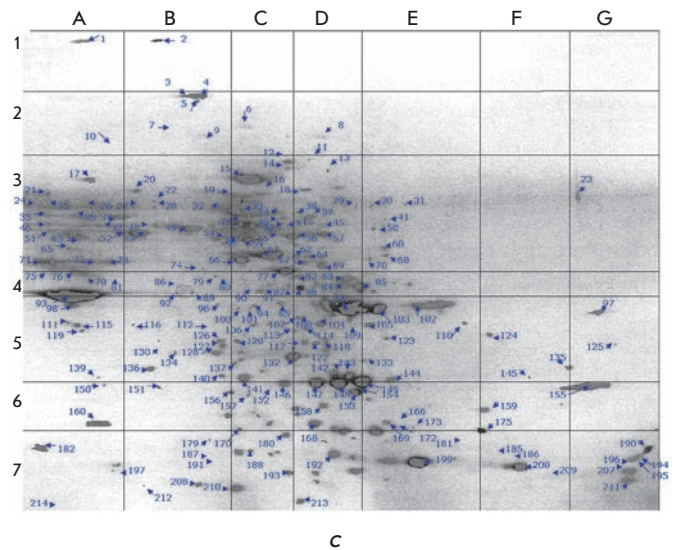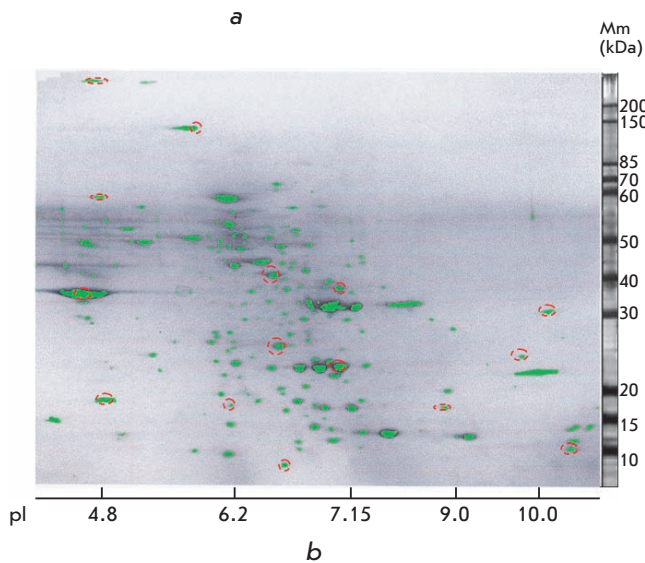
teins were identified with MALDI-TOF MS and MS/MS using an Ultraflex instrument (Bruker) at a 336-nm UV laser beam in a 500−8000 Da cation mode calibrated using reference trypsin autolysis peaks and processed with Mascot software, Peptide Fingerprint option (Matrix Science, USA) [21, 22]. The proteins were identified by matching experimental masses with the masses of proteins listed in the NCBI Protein and SwissProt/TrEMBL databases. The accuracy of monoisotopic masses measured in the reflection mode calibrated with autolytic trypsin peaks was 0.005%, and the accuracy of the fragment masses was ±1 Da. Hypothetical proteins identified with MALDI-TOF MS corresponding to fragments of the full-size proteins, which are products of corresponding genes, were revealed with MS/MS. The molecular masses of protein fractions were determined using the ultrapure recombinant protein sets SM0661 (10−200 kDa) and SM0671 (10−170 kDa) (Fermentas). The measurement of the optical density of 2DE images and/or their fragments was performed following scanning (E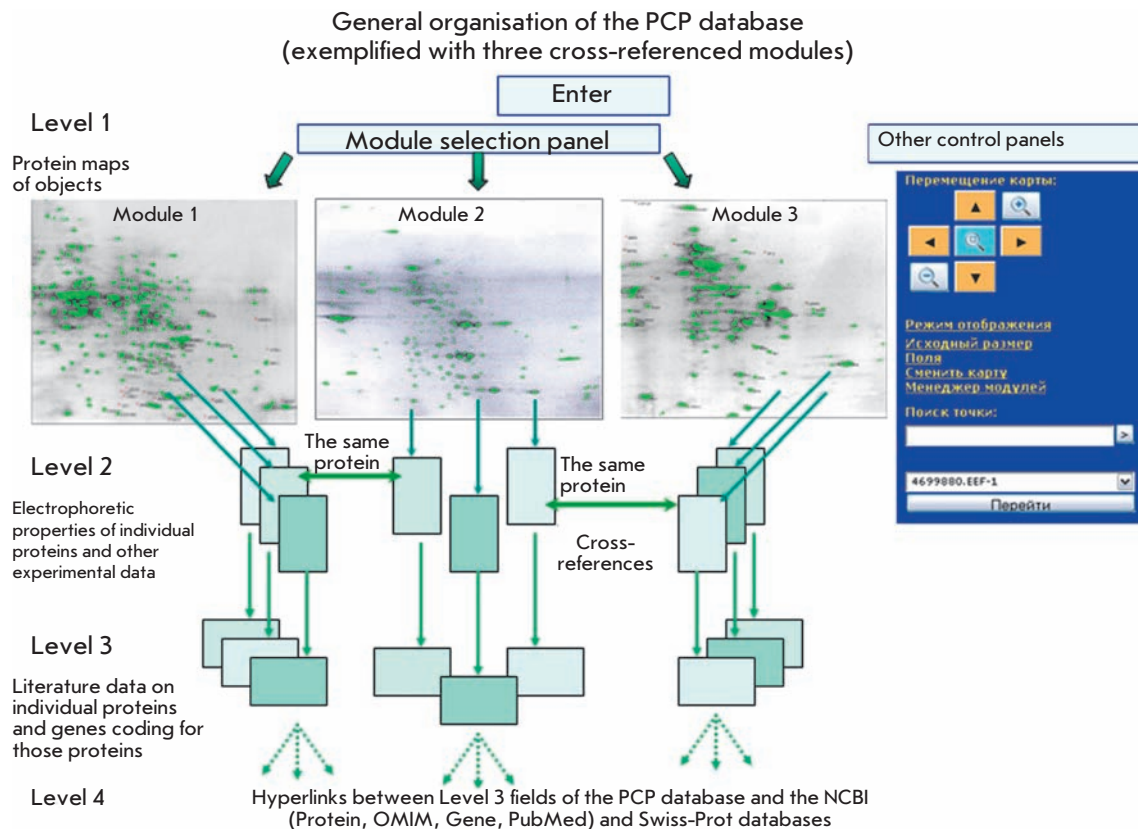pson expression 1680) or digital photography (Nikon 2500 or Canon PowerShot A1000 IS). Digital image processing with densitometry of the protein fractions was performed with Melanie Image-Master, versions 6 and 7 (Genebio).

Data logging and processing for the Prostate Cancer Proteomics multilevel database were done with various software packages, including MapThis!, Molly Penguin Software, Mozilla Firefox, and some Microsoft Office applications. A MySQL-based interactive database was used which could be updated and modified online using any computer with Internet connection. The BIOSTAT and Microsoft Office Excel 2003 software packages were used for statistical analysis.

RESULTS AND DISCUSSION

According to the conventional proteomics strategy developed in the late 20th century, the national PCP database was created in several consecutive steps which involved systematic characterisation of proteins in prostate tissue samples obtained from benign and malignant tumors (Fig. 1) ([23, 24]). Proteins from several

General organisation of the PCP database
(exemplified with three cross-referenced modules)

Enter

Level 1

Protein maps of objects

Module selection panel

Other control panels

Module 1   Module 2   Module 3

Level 2

Electrophoretic properties of individual proteins and other experimental data

The same protein

The same protein

Cross-references

Level 3

Literature data on individual proteins and genes coding for those proteins

Level 4

Hyperlinks between Level 3 fields of the PCP database and the NCBI (Protein, OMIM, Gene, PubMed) and Swiss-Prot databases

cultured human cell lines were studied in parallel experiments (Fig. 1).

The first step was to make series of 2DE protein samples (50 or more) by fractioning dozens of bioptates or prostate tissue samples (from 30 or more patients). Figure 2A illustrates a typical 2DE of the PCa prostate tissue proteins. The 2DE series for cell line proteins were created with 20 2DE, taking into account the homogeneity of the analyte.

The distribution of 2DE protein fractions was registered and stored as graphic *.tif files. The images of the entire 2DE and (in some cases) their segments were produced by scanning and/or digital photography. The relevance of the selected 2DE was assessed by comparing protein fractioning results using digital image matching [23, 24].

The second step was to construct synthetic 2D maps of the proteins. 2DE images from each series were standardised with Melanie ImageMaster software using 15 selected reference points corresponding to easily identifiable major protein fractions. Figure 2B shows reference points on the 2DE images of proteins from prostate tissue samples.

Each image was analysed using the Cummings technique [25] with some modifications [20, 24]. The analysis was based on dividing the images into 49 rectangular fragments, the sides of which formed six horizontal and six vertical standard lines and the sides of the 2DE image itself. The points for plotting the horizontal lines were determined by special molecular-weight-marker proteins, which were placed on each gel plate before fractioning in the second direction (SDS-PAGE). Thus, the protein fractions located on the corresponding horizontal lines have identical molecular weights. For plotting the vertical lines, protein markers with previously measured pI values were used [20, 24]. As a result, each image analyzed consisted of 49 rectangular fragments, each usually containing no more than 10 protein fractions (only 4 fragments contained more than 20 fractions). Image fragmenting significantly simplified the image comparison and construction of synthetic 2D maps.

The described procedure was performed with the 60 best 2DE images of proteins from BHP samples and with 70 2DE images of proteins from PCa samples. The comparison of standardised BHP and PCa 2DE images showed that the coordinates of more than 95% of protein fractions were constant. Quantitative or qualitative variations in the coordinates were observed for less than 5% of the fractions. The variations could be caused

Table 1. Description of the PCP modules and the number of identified proteins by modules.

Modules—synthetic maps of proteins in specific objects (fractioning technique)	Proteins identified
Prostate biopsy samples, PCa and BPH (IEF-PAGE*)	165
LNCaP (IEF-PAGE*)	60
LNCaP (IPG-PAGE)	18
PC-3 (IEF-PAGE*)	25
BPH-1 (IEF-PAGE*)	24
Rhabdomyosarcoma (IEF-PAGE*)	29
Normal human fibroblasts (IEF-PAGE*)	38
* modification [20]	

by genetic factors (e.g. single nucleotide polymorphism) or a different expression of corresponding genes, as well as differences in tissue composition of the samples and the pathology's intensity.

Having made sure that the positions of the majority of the protein fractions were constant, we were able to construct the 2D maps of prostate tissue proteins from patients with BHP and PCa. After that, we performed a fragment-by-fragment comparison of the 2D maps constructed. The fraction patterns in BHP and PCa maps were found to be quite similar, the difference being that about 20 fractions present in the PCa map were either present in a much smaller amount in the BHP map or absent altogether. We paid particular attention to those fractions in further study, as described below. In general, as a result of our analysis, an integrated synthetic 2D map of human prostate proteins was constructed that contains more than 200 protein fractions in the ranges of Mw 8.5−450 kDa and pI 4.5−11.5 (Fig. 2C). Each fraction was assigned a unique seven-digit number; the first four digits representing the logarithm of the fraction's molecular weight, and the next three digits representing the value of the isoelectric point expressed in units described in [20, 24].

The same procedure was applied to construct other synthetic maps of proteins from cultured human cell lines, although much fewer 2DE images were used for those maps.

Thus, each of the constructed maps contained data on the electrophoretic properties of the protein fractions (represented by their coordinates) in the corresponding object. These maps were in *.jpg format (with a resolution of at least 300 dpi), constituting Level 1 of the database. Further studies and analysis of data on proteins were based on the information contained in the Level 1 maps. Therefore, the synthetic maps represented original modules enabling the characterizing and formalizing of the biochemical properties of the proteins studied. There are currently seven modules in PCP (Table 1). There is a special panel enabling navigation among the modules. The 2D maps are scaleable, and the user can mark certain proteins on the maps and create links (buttons) for accessing other levels− second, third, and fourth−that contain data on the proteins. The database also automatically displays 2D coordinates (along the two fractioning axes) as the cursor moves around the map. The general organisation of the PCP database is presented in Fig. 3.

The third step in the proteomic study was to identify individual protein fractions. The proteins were mainly identified by mass-spectrometry: the results are presented in Table 1.

As Table 1 shows, there is a total of 359 identified proteins in PCP. Among them there are many well-known proteins, such as the enzymes responsible for glycolysis (glyceraldehyde-3-phosphate dehydrogenase, triose-phosphate isomerase, etc.) and other metabolic processes, as well as cytoskeletal (actin, transgelins, etc.) and mitochondrial (porins, superoxide dismutase, etc.) proteins. Some of the identified proteins, for instance transgelins [21, 22], were represented with several isoforms.

We paid particular attention to the identification of the protein fractions which differed qualitatively or quantitatively in the prostate tissue samples from patients with BHP and PCa. We previously reported on the results of identification of two potential PCa biomarkers, the proteins AGR2 [14] and Dj-1 [26]. In total, we succeeded in identifying 17 potential PCa biomarkers, some of which are new. Table 2 provides a short description of the potential PCa biomarkers. For example, Fig.4 shows the results of MS identification of one of the new biomarkers, protein PRO2675, which contains an albumin domain in its primary structure.

For each protein identified (and marked with a "button" on the 2D map), the second information Level was formed, comprising a standardized system of 15 fields for the entry of text and graphical data obtained during characterization of the corresponding protein fraction. In four fields, general information about the protein is entered, in the next six fields the identification results are entered, and in the other five fields additional in-

Table 2. Potential PCa biomarkers in the "Prostate biopsy samples, PCa and BPH" module and other PCP modules

Unique identifier*	Protein (synonyms and symbol in PCP)	Numbers in NCBI** and *Swiss-Prot*	Additional information in PCP and references***
5653580	**Ferritin light chain complex (K-(L)F)**	182516, *P02792*	[15]
4785508 (4799550)	Chaperonin (HSPD1)	31542947, NP_002147, 118190, *P10809*	Found in rhabdomyosarcoma cells; {Bindukumar B. et al. 2008, 18646040}
4716560 (4756612)	Protein-disulphide isomerase (ER60)	7437388, *P30101*	[15]
4531685	N-acetylneuraminat phosphate synthase (NANS)	12652539, AAH00008, NP_061819, 605202, Q9NR45	Found in rhabdomyosarcoma cells; [15]
4502675	Annexin 2, isophorm 2 (ANXA2-i2)	4757756, NP_004030, 151740, *P07355*	{Shiozawa Y. et al. 2008, 18636554; Hastie C. et al. 2008, 18211896}
4454692	**Unknown protein PRO2675 containing the albumin domain (PRO2675)**	7770217	[15]
4447605	Protein 29 of endoplasmic reticulum, isophorm 1 (Erp29)	5803013, NP_006808, 602287, *P30040*	{Myung J.K. et al. 2004, 15598346}
4352630 (4342620)	Dj-1 protein (Dj-1)	50513593, 1SOA_A, 606324, *Q99497*	{Bindukumar B. et al. 2008, 18646040}
4356607 (4344615)	**Dj-1 protein, electrophoretic isophorm (Dj-1-ei)**	31543380	
4336712 (4301795)	Prostatic binding protein (neuropolypeptide h3, PEBP1)	21410340, AAH31102, 604591, *P30086*	[15]; {Li et al. 2008, 18161940; Woods Ignatoski K.M. et al. 2008, 18722266}
4286750 (4290620)	NM23B-protein, nucleoside phosphate kinase B	4505409, NP_002503, 156491, *P22392*	{Johansson B. et al. 2006, 16705742}
4255880	**Unnamed protein (NEDO human cDNA sequencing project, tissue type = "testis") (NEDO)**	21758704, BAC05360	
4204630	Fatty acid binding protein, isophorm 5 (E-FABP)	30583737, AAP36117, 605168, *Q01469*	[15]; {Morgan E.A. et al. 2008, 18360704}
4279900	AGR2 (AGR2)	37183136, AAQ89368, 606358. *Q4JM47*	[14, 15]; {Zhang J.S. et al. 2005, 15834940; weitzig D.R. et al. 2007, 17694278}
41811130	Hystone H3, 3A family (H3f3a)	55665435	[15]
4161675	**Unknown protein PRO2044 containing the albumin domain (PRO2044)**	6650826	[15]
4021610	S100 calcium binding protein A11 (S100A11)	12655117, AAH01410, 603114, *P31949*	{Rehman I. et al. 2004, 15668896; Schaefer K.L. et al. 2004, 15150091}

* Numbers in the "LNCaP (IEF-2DE Modification)" module are given in parenthesis
** Numbers from the NCBI databases are given in the following order: Protein, Genbank and/or Nucleotide, OMIM
*** Publications listed in the References section of this article are given in square brackets; publications from the PubMed database are given in curly brackets.
Note: a new potential biomarkers are shown in bold

formation is entered. The filled Level 2 fields for one of the potential PCa biomarkers, protein NANS (N-acetylneuraminic acid phosphate synthase), are shown in Fig. 5.

The same protein could be present in more than one object. Therefore, within Level 2, one can use the control panel to create cross-links between identical proteins in different modules. An example of such cross-referencing for protein Dj-1 is presented in Fig. 6.

The majority of the 359 identified fractions were well-known proteins (and/or their electrophoretic isophorms) described in the literature and various databases. Some information on those proteins, relating to the PCP scope, constituted the third information Level of the database. Level 3 is a standardized system of 23 fields for the entry of text and graphical data. In twelve fields information about the protein is entered, in the next six fields information about the gene coding for

A1

Start-end	Experimental Mw	Calculated Mw	Peptide sequence
1 – 16	1706.77	1705.82	MPADLPSLAADFVESK + Oxidation (M)
27 – 39	1639.82	1638.79	DVFLGMFLYEYAR + Oxidation (M)
76 – 92	2045.05	2044.09	VFDEFKPLVEEPQNLIK
106 – 113	960.52	959.56	FQNALLVR
118 – 131	1511.74	1510.84	VPQVSTPTLVEVSR
249 – 263	1763.73	1762.77	AVMDDFAAFVEKCCK + Oxidation (M)

A2

B1
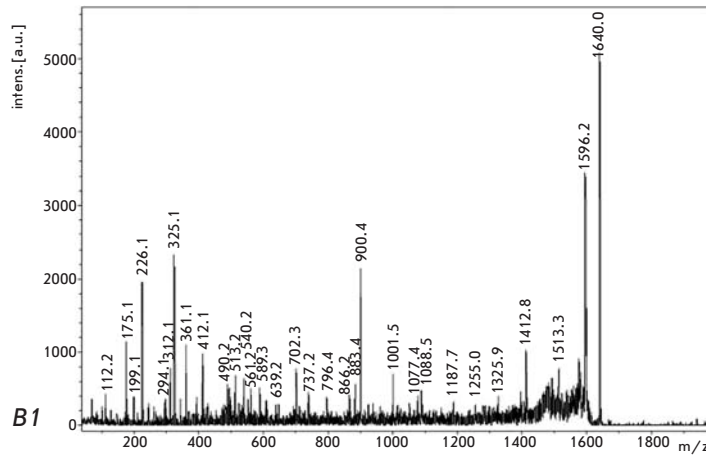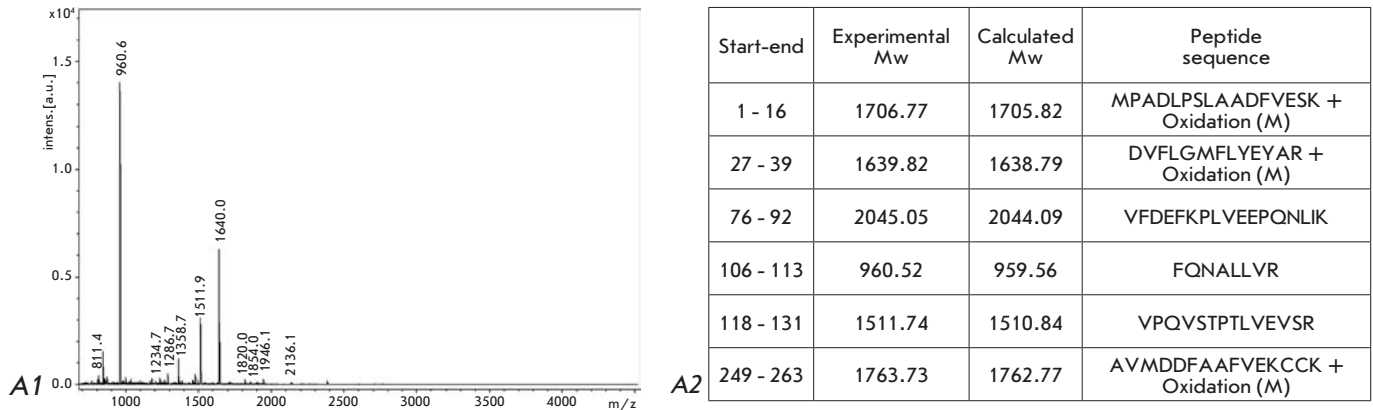
#	Immon.	a	a*	a⁰	b	b*	b⁰	Seq	v	w	w*	y	y*	y⁰	#
1	101.11	101.11	84.08		129.10	112.08		K					1511.84	1494.82	1493.83	15
2	72.08	200.18	183.15		228.17	211.14		V	1467.78	1480.80		1511.84	1494.82	1493.83	14
3	70.07	297.23	280.20		325.22	308.20		P	1370.73	1369.73		1412.77	1395.75	1394.76	13
4	101.07	425.29	408.26		453.28	436.26		Q	1242.67	1241.67		1315.72	1298.70	1297.71	12
5	72.08	524.36	507.33		552.35	535.32		V	1143.60	1156.62		1187.66	1170.64	1169.65	11
6	60.04	611.39	594.36	593.38	639.38	622.36	621.37	S	1056.57	1055.57		1088.59	1071.57	1070.58	10
7	74.06	712.44	695.41	694.42	740.43	723.40	722.42	T	955.52	968.54	970.52	1001.56	984.54	983.55	9
8	70.07	809.49	792.46	791.48	837.48	820.46	819.47	P	858.47	857.47		900.51	883.49	882.50	8
9	74.06	910.54	893.51	892.53	938.53	921.50	920.52	T	757.42	770.44	772.42	803.46	786.44	785.45	7
10	86.10	1023.62	1006.59	1005.61	1051.61	1034.59	1033.60	L	644.34	643.34		702.41	685.39	684.40	6
11	72.08	1122.69	1105.66	1104.68	1150.68	1133.66	1132.67	V	545.27	558.29		589.33	572.30	571.32	5
12	102.05	1251.73	1234.70	1233.72	1279.73	1262.70	1261.72	E	416.23	415.23		490.26	473.24	472.25	4
13	72.08	1350.80	1333.77	1332.79	1378.79	1361.77	1360.78	V	317.16	330.18		361.22	344.19	343.21	3
14	60.04	1437.83	1420.80	1419.82	1465.83	1448.80	1447.82	S	230.12	229.13		262.15	245.12	244.14	2
15	129.11							R	74.02	73.03		175.12	158.09		1

B2

```
  1 MPADLPSLAA DFVESKDVCK NYAEAKDVFL GMFLYEYARR HPDYSVVLLL RLAKTYETTL

 61 EKCCAAADPH ECYAKVFDEF KPLVEEPQNL IKQNCELFEQ LGEYKFQNAL LVRYTKKVPQ

121 VSTPTLVEVS RNLGKVGSKC CKHPEAKRMP CAEDYLSVVL NQLCVLHEKT PVSDRVTKCC

181 TESLVNRRPC FSALEVDETY VPKEFNAETF TFHADICTLS EKERQIKKQT ALVELVKHKP

241 KATKEQLKAV MDDFAAFVEK CCKADDKETC FAEEGKKLVA ASQAALGL
```

C

Fig. 4. Results of mass-spectrometric identification of the PRO 2675 protein.
A. Mass-spectrum of tryptic peptides. (1) MALDI-TOF MS data. (2) Peptide identification by Mascot.
B. Mass-spectrum of one of the tryptic peptides (1) MALDI-TOF MS data. (2) Peptide identification by Mascot.
C. Amino acid sequence of the PRO 2675 protein (records AAF69644.1 GI:7770217 in the NCBI Protein database); amino acid residues of the tryptic peptides are printed in red; the peptide's whose sequence was identified by MALDI-TOF MS/MS and is highlighted in grey; the sites corresponding to the albumin domain are underlined.

Experimental data (Level 2)

General properties	
Fraction (Protein)	N-acetylneuraminic acid phosphate synthase, NANS Edit
Localization, number	D4, 4531685 Edit
Mw, kDa	38.0 Edit
pl (experimental)	6.85 Edit
Identification	
Method	MALDI TOF MS Edit
Sequence coverage, %	58% Edit
Mascot (Matrix Science, USA) search results	(see peptide table below)

```
Start-End    Observed   Mr(expt) Mr(calc) ppm   Miss            Peptide
 66 – 74     1064.56   1063.55  1063.57  –0.02   0  ALERPYTSK
 87 – 96     1331.59   1330.58  1330.61  –0.03   0  HLEFSHDQYR
132 – 145    1540.68   1539.68  1539.72  –0.04   0  VGSGDTNNFPYLEK
150 – 166    1919.79   1918.78  1918.83  –0.04   0  GRPMVISSGMQSMDTMK  4 Oxidation (M)
199 – 205     866.44    865.43   865.45  –0.03   0  VISEYQK
206 – 233    2767.42   2766.41  2766.50  –0.08   0  LFPDIPIGYSGHETGIAISVAAVALGAK
247 – 264    1866.88   1865.87  1865.91  –0.04   0  GSDHSASLEPGELAELVR
301 – 315    1693.83   1692.82  1692.87  –0.05   0  IPEGTILTMDMLTVK  2 Oxidation (M)
321 – 333    1448.71   1447.70  1447.73  –0.04   0  GYPPEDIFNLVGK
335 – 355    2401.08   2400.07  2400.14  –0.06   0  VLVTVEEDDTIMEELVDNHGK  Oxidation (M)
```
Edit

Figure	Edit
Amino acid sequence (without signal peptides)	(see sequence below)

```
0001 MPLELELCPG RWVGGQHPCF IIAEIGQNHQ GDLDVAKRMI RMAKECGADC
0051 AKFQKSELEF KFNRKALERP YTSKHSWGKT YGEHKRHLEF SHDQYRELQR
0101 YAEEVGIFFT ASGMDEMAVE FLHELNVPFF KVGSGDTNNF PYLEKTAKKG
0151 RPMVISSGMQ SMDTMKQVYQ IVKPLNPNFC FLQCTSAYPL QPEDVNLRVI
0201 SEYQKLFPDI PIGYSGHETG IAISVAAVAL GAKVLERHIT LDKTWKGSDH
0251 SASLEPGELA ELVRSVRLVE RALGSPTKQL LPCEMACNEK LGKSVVAKVK
0301 IPEGTILTMD MLTVKVGEPK GYPPEDIFNL VGKKVLVTVE EDDTIMEELV
0351 DNHGKKIKS
```
Edit

Comments to amino acid sequence	Amino acid sequence of the NANS protein. Matched peptides shown in Red. Edit
Supplementary identification information	
Supplementary information	MS/MS Fragmentation of VGSGDTNNFPYLEK Found in gi|8453156, N-acetylneuraminic acid phosphate synthase [Homo sapiens] Match to Query 16: 1539.695159 from(1540.702435,1+) Edit
Tandem mass-spectrometry	Monoisotopic mass of neutral peptide Mr(calc): 1539.72 Ions Score: 22 Expect: 8.7

Fig. 5. Level 2 fields for the NANS protein.

Fig. 6. Control panel with cross-references for the Dj-1 protein from the "Prostate biopsy samples, PCa and BPH" module.

Identical proteins in other modules			
Module	Point	Link	Disconnect
LNCaP (IPG-2DE)	4301630.Dj-1	Link	Disconnect
LNCaP (IEF-2DE modification)	4342635.Dj-1	Link	Disconnect
Rhabdomyosarcoma	4342690.DJ-1	Link	Disconnect
PC3	4342630.DJ-1	Link	Disconnect
Add			

the protein is entered, and in the next three fields information about the protein's polymorphism is entered. In the other fields, selected references to publications about the general properties of the proteins, as well as its oncological properties, are entered (Fig. 7).

The Level 3 text fields can contain hyperlinks to various Internet databases, such as NCBI's Protein, OMIM and PubMed, and SwissProt. This feature allowed the creation of the fourth information Level, providing the user with prompt access to international databases containing, in particular, the results of human genome sequencing.

The PCP database is an interactive MySQL-based web resource located at http://ef.inbi.ras.ru and can

Fig. 7. Level 3 fields for specially selected references to publications on the AGR2 protein.

References	
General	**PMID: 9790916** 1. Thompson DA, Weigel RJ. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell. Biochem. Biophys. Res. Commun. 1998. V.251(1), P.111-116. **PMID: 12975309** 2. Clark HF, Gurney AL, Abaya E, Baker K, Baldwin D, Brush J, Chen J, Chow B, Chui C, Crowley C, Currell B, Deuel B, Dowd P, Eaton D, Foster J, Grimaldi C, Gu Q, Hass PE, Heldens S, Huang A, Kim HS, Klimowski L, Jin Y, Johnson S, Lee J, Lewis L, Liao D, Mark M, Robbie E, Sanchez C, Schoenfeld J, Seshagiri S, Simmons L, Singh J, Smith V, Stinson J, Vagts A, Vandlen R, Watanabe C, Wieand D, Woods K, Xie MH, Yansura D, Yi S, Yu G, Yuan J, Zhang M, Zhang Z, Goddard A, Wood WI, Godowski P, Gray A. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res. 2003. V.13(10), P.2265-2270. **PMID: 15340161** 3. Zhang Z, Henzel WJ. Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci. 2004. V.13(10). P.2819-2824. Edit
Oncological	**PMID: 12592373** 1. Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, Daviet L, Legrain P, Parekh R, Harris AL, Terrett JA. hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. Br. J. Cancer. 2003. V.88(4). P.579-585. **PMID: 15532095** 2. Kristiansen G, Pilarsky C, Wissmann C, Kaiser S, Bruemmendorf T, Roepcke S, Dahl E, Hinzmann B, Specht T, Pervan J, Stephan C, Loening S, Dietel M, Rosenthal A. Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol. 2005. V.205(3). P.359-376. **PMID: 15867376** 3. Liu D, Rudland P.S, Sibson DR, Platt-Higgins A, Barraclough R. Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. Cancer Res. 2005. V.65(9). P.3796-3805. **PMID: 15958538** 4. Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, Meropol NJ, Weiner LM, Cohen SJ, Moreno JG, Connelly MC, Terstappen LW, O'Hara SM. Global Gene Expression Profiling of Circulating Tumor Cells. Cancer Res. 2005. V.65. P.4993-4997. **PMID: 15834940** 5. Zhang JS, Gong A, Cheville JC, Smith DI, Young CY. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. Genes Chromosomes Cancer. 2005. V.43(3) P.249-259. **PMID: 17022460** 6. Kovalev LI, Shishkin SS, Khasigov PZ, Dzeranov NK, Kazachenko AV, Toropygin Ilu, Mamykina SV. Identification of AGR2 protein, a novel potential cancer marker, using proteomics technologies. Prikl Biokhim Mikrobiol. 2006. V.42(4). P.480-484. **PMID: 18829536** 7. Ramachandran V, Arumugam T, Wang H, Logsdon CD. Anterior gradient 2 is expressed and secreted during the development of pancreatic cancer and promotes cancer cell survival. Cancer Res. 2008. V.68(19). P.7811-7818. Edit

be accessed from any computer connected to the Internet using the Mozilla Firefox and Microsoft Internet Explorer browsers. There are three access permission categories: "Guest," "Manager," and "Administrator," each giving certain rights for working with PCP. In particular, users with "Manager" access permission can make entries into and correct the Level 2 and 3 fields, while users with the "Administrator" category access permission, have the ability to expand the database by creating new modules and new functional elements. Users with "Guest" access can browse all fields but cannot edit them.

In conclusion, our work resulted in the creation of an original multi-module national database entitled "Prostate Cancer Proteomics," which summarizes data on the proteins in prostate tissue collected from patients with BHP and PCa, as well as on proteins in several human cell lines. This is very promising in the further use of proteomic and other biochemical data. We are hopeful that the PCP database will be useful to biochemists and other biomedical scientists, making their research on PCa more efficient. ●

REFERENCES

1. Gottlieb B., Beitel L.K., Wu J.H., Trifiro M. // Hum. Mutat. 2004. V. 23. P. 527−533.
2. Westbrook J.A., Wheeler J.X., Wait R., Welson S.Y., Dunn M.J. // Electrophoresis. 2006. V. 27. P. 1547−1555.
3. Kandasamy K., Keerthikumar S., Goel R., Mathivanan S., Patankar N., Shafreen B., Renuse S., Pawar H., Ramachandra Y.L., Acharya P.K., Ranganathan P., Chaerkady R., Keshava Prasad T.S., Pandey A. // Nucleic Acids Res. 2009. V. 37 (Database issue). P. D773−D781.
4. Vizcaino J.A., Cote R., Reisinger F., Barsnes H., Foster J.M., Rameseder J., Hermjakob H., Martens L. // Nucleic Acids Res. 2010. V. 38 (Database issue). P. D736−742.
5. Stamey T.A., Caldwell M., McNeal J.E., Nolley R., Hemenez M., Downs J. // J. Urol. 2004. V. 172. P. 1297−1301.
6. Zhang J.S., Gong A., Cheville J.C., Smith D.I., Young C.Y. // Genes Chromosomes Cancer. 2005. V. 43. № 3. P. 249−259.
7. Lim L.S., Sherin K. // Am. J. Prev. Med. 2008. V. 34. № 2. P. 164−170.
8. Leman E.S., Getzenberg R.H. // J. Cell Biochem. 2009. V. 108. № 1. P. 3−9.
9. Zlokachestvennie novoobrazovania v Rossii v 2008. Zabolevaemost i smertnost (Malignant Tumors in Russia in 2008. Morbidity and Mortality) / ed. Chissov V.I., Starinsky V.V., Petrova G.V. M.: «MEDpress-inform», 2008. P. 18.
10. Zlokachestvennie novoobrazovania v Rossii i SNG v 2002 (Malignant Tumors in Russia and CIS in 2002). N. N. Blokhin Russian Cancer Research Center of Russian Academy of Medical Sciences / ed. Davidov M.I., Aksel E.M. M.: «MIA», 2004. 256 p.
11. Jemal A., Siegel R., Ward E., Murray T., Xu J., Smigal C., Thun M.J. // CA Cancer J. Clin. 2007. V. 57. № 1. P. 43−66.
12. Maddams J., Brewster D., Gavin A., Steward J., Elliott J., Utley M,. Muller H. // Br. J. Cancer. 2009. V. 101. № 3. P. 541−547.
13. Primrose S.B., Twyman R.M. Genomika. Rol v Medicine (Genomics. Applications in Human Biology). M.: «BINOM. Laboratoria znaniy, 2008. 277 p.
14. Kovalyov L.I., Shishkin S.S., Hasigov P.Z., Dzeranov N.K., Kazachenko A.V., Kovalyova M.A., Toropygin I.Y., Mamikina S.V. // Prikladnaya biokhimia i mikrobiologia (Applied Biochemistry and Microbiology). 2006. V.42. №4. P.480-484.
15. Shishkin S.S., Dzeranov N.K., Totrov К.I., Kazachenko A.V., Kovalyov L.I., Eremina L.S., Kovalyova M.A., Toropygin I.Y. // Urologiia (Urology). 2009. V. 1. P. 56-58.
16. Kogan M.I., Loran O.B., Petrov S.B. Radikalnaya khirurgia raka predstatelnoy zhelezi (Radical Surgery of Prostate Cancer). M.: «Geotar-Media», 2006. 392 p.
17. Shishkin S.S., Kovalyov L.I.., Kovalyova M.A., Krakhmaleva I.N., Eremina L.S., Makarov A.A., Lisitskaya K.V., Loran O.B., Veliev E.I., Okhrits V.E. Problemi ranney diagnostiki raka prostati i vosmozhnosti primenenia novih potencialnih biomarkerov (The Problems of Early Diagnostics of Prostate Cancer and Possibilities of Using of New Potential Biomarkers. (Informational-methodical letter)). M.: OOO "Originalnaya compania", 2009. 45 p.
18. Chernikov V.G., Terehov S.M., Krohina T.B., Shishkin S.S., Smirnova T.D., Lunga I.N., Adnoral N.V., Rebrov L.B., Denisov-Nikolsky Y.I., Bikov V.A. // Bull. eksp. biol. med. (Bulletin of Experimental Biology and Medicine). 2001. V. 131. № 6. P. 680-682.
19. Krohina T.B., Shishkin S.S., Raevskaya G.B., Kovalyov L.I.., Ershova E.S., Chernikov V.G, Mirochnik V.V., Bubnova E.N., Kucharenko V.I. // Bull. eksp. biol. med. (Bulletin of Experimental Biology and Medicine). 1996. V. 122. № 9. P.314-317.
20. Kovalyov L.I., Shishkin S. S., Efimochkin A.S., Kovalyova M.A., Ershova E.S., Egorov T.A., Musalyamov A.K. // Electrophoresis. 1995. V. 16. P. 1160−1169.
21. Eremina L.S., Kovalyov L.I., Shishkin S.S., Toropygin I.Y., Burakova M.I., Kovalyova M.A., Makarov A.A., Dzeranov N.K., Kazachenko A.V., Totrov К.I., Kononkov I.V., Loran O.B. // Vopr. biol. med. farm. khimii (Problems of Biological, Medical, and Pharmaceutical Chemistry). 2007. V.3. P.49-52.
22. Kovalyova M.A., Kovalyov L.I., Eremina L.S., Makarov A.A., Burakova M.I., Toropygin I.Y., Serebryakova M.V., Shishkin S.S., Archakov A.I. // Biomedicinskaya khimia (Biomedical Chemestry). 2008. V. 54. № 4. P. 420-434.
23. Anderson N.G., Anderson L. // Electrophoresis. 1996. V. 17. P. 443−453.
24. Shishkin S.S., Kovalyov L.I., Gromov P.S. / Mnogolikost sovremennoy genetiki cheloveka (The Variety of Contemporary Human Genetics). M.-Ufa: «Gilem», 2000. P. 17-50.
25. Cumings D. // Clin. Chem. 1982. V. 28. P. 782−789.
26. Loran O.B., Veliev E.I., Okhrizts V.E., Lisitskaya K.V., Eremina L.S., Kovalyov L.I., Kovalyova M.A., Shishkin S.S. // Eur. Urol. Suppl. 2010. V. 9. № 2. P. 309.