DOI 10.1002/pmic.200600467 *Proteomics* 2007, *7*, 4–9

VIEWPOINT

AFM fishing nanotechnology is the way to reverse the Avogadro number in proteomics

Alexander Ivanovich Archakov, Yurii Dmitrievich Ivanov, Andrey Valerevich Lisitsa, Victor Gavrilovich Zgoda

Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Pogodinskaya Street, Moscow, Russia

Received: July 7, 2006
Revised: September 21, 2006
Accepted: September 21, 2006

Future development of proteomics may be hindered by limitations in the concentration sensitivity of widespread technological approaches. The concentration sensitivity limit (CSL) of currently used approaches, like 2-DE/LC separation coupled with MS detection, *etc.*, varies from 10^{-9} to 10^{-12} M. Therefore, proteomic technologies enable detection of up to 20% of the protein species present in the plasma. New technologies, like atomic force microscopy (AFM molecular detector), enable the counting of single molecules, whereas biospecific fishing can be used to capture these molecules from the biomaterial. At the same time, fishing also has thermodynamic limitations due to the reversibility of the binding. In cases where the fishing becomes irreversible, its combination with an AFM detector enables the registration of single protein molecules, and that opens up a way to lower the CSL down to the reverse Avogadro number.

Keywords:
Atomic-force microscopy / Biospecific fishing / Molecular detectors / Nanobiotechnology

Proteomics enables the identification and analysis of proteins with a high throughput. Methodically, proteomics is based on a combination of knowledge about the genome and the latest achievements in mass spectrometric technology; this combination allows for rapid identification of proteins in a biological sample [1].

Unfortunately, despite the impressive advances made, primarily, within the framework of the international project "Human Proteome" [2], proteomics has not justified many hopes it had initially inspired. The current situation is perfectly characterized as "we began to run before we could walk" [3]. Although the situation in proteomics has very much in common with the period of implementation of the "Human Genome Project" in 1991–1993, there is a fundamental difference in the proteomic, as opposed to the genomic approach.

The successful development of genomics (decoding of the human and some other genomes) became possible owing to PCR, which allows replication of nucleic acid molecules [4]. The ability to replicate molecules makes it possible to circumvent the CSL for the RNA and DNA molecules.

Proteomics lacks a method for making multiple copies of individual molecules similar to the PCR-based one [5]. As a result, proteomics is limited by the concentration sensitivity of existing technologies that are unable, despite the successful development of MS, 2-D-electrophoresis, multidimensional chromatography, and a variety of other methods, to overcome the CSL of 10^{-12} M (Table 1) [6–13].

Moreover, little attention has so far been given even to the existence of the CSL as such. For the most part, the efforts of investigators engaged in proteomics are directed towards overcoming the dynamic concentration barrier, *i.e.*, the problem of identification of low-abundant proteins ($<10^{-10}$ M) against a background of high-abundant ones

Correspondence: Professor Alexander Archakov, Institute of Biomedical Chemistry, Pogodinskaya Street 10, Moscow 119121, Russia
E-mail: archak@ibmc.msk.ru
Fax: +7(495)-2450857

Abbreviations: AFM, atomic force microscopy; **CSL,** concentration sensitivity limit

(>10^{-6} M) [10, 14]. The dynamic range of protein concentrations in biomaterials is very broad and corresponds to values upwards of 10^{12}. Its narrowing in the serum is achieved at the expense of technologies based on the depletion of major protein fractions; however, there is no guarantee that low-abundant proteins are not similarly subjected to depletion [15]. In our opinion, the existence of the CSL is an even more important problem in proteomics; indeed, the further development of proteomics as a scientific discipline is largely dependent on successfully overcoming the concentration limitations in proteomic technologies.

Why is the CSL so important for proteomics? To answer this question, it is necessary to analyze the dependence of the number of different proteins found on their concentrations in biological material. It has been proposed to investigate the plasma concentrations of various proteins [16] (Fig. 1). One can see that only several tens of different proteins occur in the plasma at concentrations of 10^{-3}–10^{-4} M. The lower the concentration, the greater is the diversity of proteins present in the plasma; there are more than 1000 proteins present at a concentration of 10^{-15} M.

The dependence in Fig. 1 can be extrapolated by the equation $N = 2.158 \times C^{-0.22}$, where N is the number of protein types occurring in the plasma with respective molar concentrations C. This equation shows that over 50 000 proteins are present in the concentration region 10^{-18} M. However, none of the current methods allow us as yet to "look into" this concentration region. As seen in Table 1, the methods of 2-D-electrophoresis and LC with mass spectrometric detection have a CSL in plasma assays of about 10^{-8} M; the CSL of microcantilever methods is about 10^{-14}–10^{-15} M, that of ELISA is about 10^{-12}–10^{-15} M and that of optical biosensors is 10^{-12} M; the CSLs of nanowire and nanopore electrochemical detectors are about 10^{-14}–10^{-15} M. Thus, the existing proteomic technologies enable detection and identification of up to 10–20% of the various protein species present in plasma.

A similar situation is observed with cellular biological material. Assuming that the copy number of a specific molecule is ten, *e.g.*, that ten signal molecules or receptors are to be found in the cell [17], the conversion of this number to concentration units (with typical volumes of eukaryotic cells of 10^{-10}–10^{-14} L), gives the values 10^{-16}–10^{-19} M. It is impossible to detect the presence of such molecules by the use of currently adopted proteomic technologies.

Figure 2a represents the results of studies of the CSL for one of the typical methods of proteomic analysis, 2-D-LC-MS/MS. The ghost preparation of mouse liver microsomes [18] was analyzed, and the protein concentration was calculated based on the average molecular weight of the proteins in the 50 kDa fraction; the total protein content was determined by the method of Lowry *et al.* [6] At concentrations below 10^{-7} M not a single protein was identified. Furthermore, by increasing the concentration from 10^{-7} to 10^{-5} M, the number of identified proteins at first rapidly increased, and then–with the subsequent increase in protein amount in the sample – it reached a plateaue. A polynomial approximation of the experimental data shows that an increase in protein concentration in the sample upwards of 10^{-4} M would not lead to a substantial increase in the number of identified proteins. In an extreme case, upon increasing the protein concentration to infinity, the number of various protein molecules detected by this method approaches 300 (Fig. 2a). In the inset of Fig. 2a, the dependence of the myoglobin fragments detected on the concentration of protein loaded onto the column is illustrated. One can see that, upon loading onto the column material with a protein content of 100 ng (5×10^{-6} M), 8 out of 17 possible peptide fragments of myoglobin proteolysis are revealed, while at levels lower than 1.0 ng (5×10^{-8} M) no fragments are registered at all. It is noteworthy that in the case of pure protein, the CSL for LC-MS/MS becomes 10^{-8} M (1 ng/μL, see the inset in Fig. 2a). Using the 2-D-LC-MS/MS method, we could not identify

Table 1. Sensitivity of current methods for protein detection

Name of the method	Sensitivity limit (M)	Reference
Lowry	10^{-3}–10^{-4}	[6]
2-DE	5×10^{-7}–5×10^{-9} (for silver staining)	[7]
2-DE-MALDI-PMF	10^{-7}–10^{-8}	[8]
2-DE + labels	10^{-6}–10^{-7}	BioRad Instruction Manual (SYPRO Ruby)
LC + MS/MS (ESI)	10^{-7}–10^{-8}	Our observations
SELDI-TOF-MS	10^{-6}–10^{-8}	[9]
ELISA	10^{-9}–10^{-14}	[10]
Microcantilever detectors	10^{-12}	[11]
Optical biosensors	10^{-11}–10^{-12}	BIAcore 3000 specification[12]
Electrochemical nanowire detectors	10^{-14}–10^{-15}	[13]

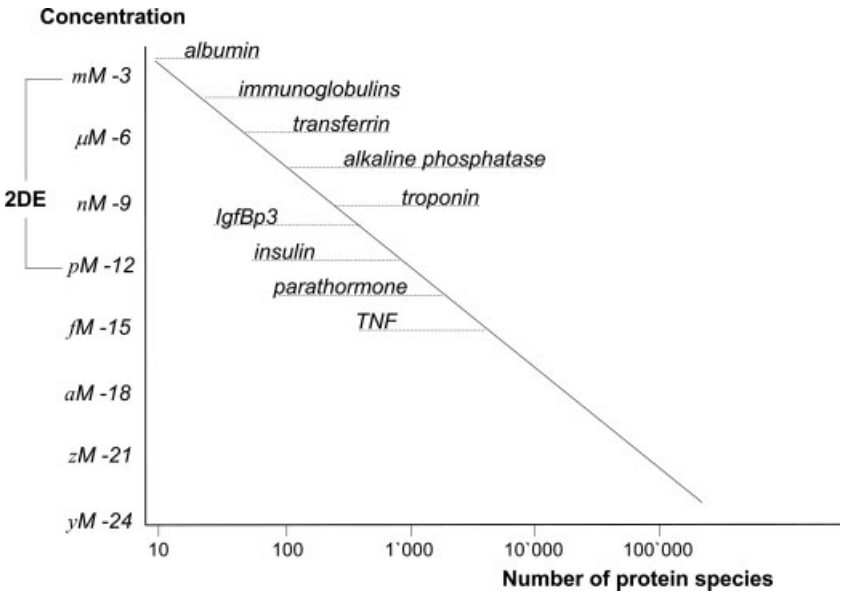

Figure 1. Plasma concentration dependence of the number of protein types [16].

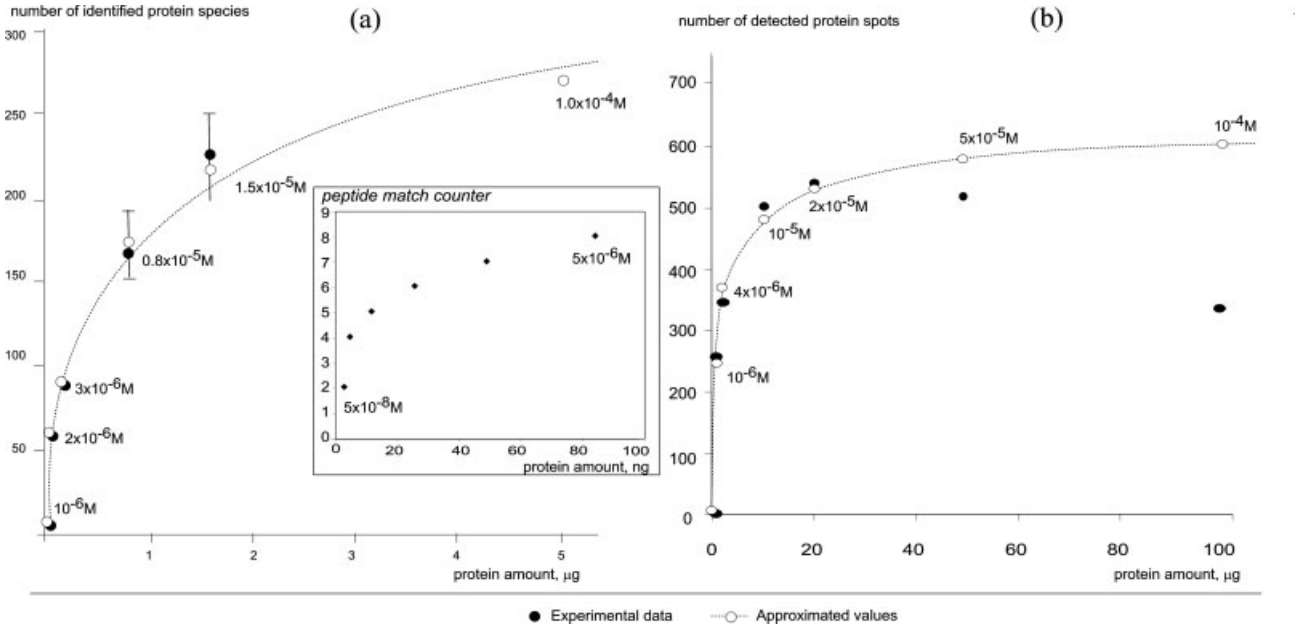

Figure 2. Dependence of the number of identified proteins or detected protein spots on the concentration. (a) -●- Identification by 2-D LC-MS/MS and (b) -●- separation 2-DE (silver-stained). -o-o- Approximation to the polynomial equation $(a_0 + a_1/x + a_1/x^2 + a_1/x^3)^{-1}$. Inset: dependence of the number of proteolytic peptides identified by LC-MS/MS on the myoglobin concentration in the sample.
Experimental conditions: Mice liver microsomal ghosts were prepared as described in [19]. 2-D-LC-MS/MS analysis was performed on an Agilent Proteomic Workstation with a LC/MSD IT detector according to the regimen described in [20]. Microsomal ghosts lysates were separated by small format IPG-2-DE (7 × 6 cm) and visualized by silver staining. Each 2-DE and 2-D-LC-MS/MS analysis was replicated three times; SDs are shown in the figure.

more than 250 proteins in the microsomal ghost material, while on the 2-D-electrophoregram of the preparation more than 860 spots were revealed [19].

Figure 2b summarizes data obtained upon the separation of the microsomal ghost fraction by 2-DE. Following the separation of various protein amounts in the concentration range 10^{-4} to 10^{-6} M, the number of protein spots determined in automatic mode by the Melanie III program (Swiss Institute of Bioinformatics, SWISS) was calculated. As 2-DE becomes overloaded when the protein amount exceeds 20 μg *per* probe (*i.e.,* 1 μg/μL or 2×10^{-5} M), the number of protein spots decreases. The results of approximation show that the

number of protein spots upon 2-DE separation is limited to 600 under our conditions. This number is 2.5-fold higher than the number of proteins identified by LC-MS/MS. However, as a rule, only one-third of spots visualized on the electrophoregram may be identified. Thus, neither of the two currently adopted procedures for proteomic analysis (2-D-LC-MS/MS and 2-DE) enables identification of more than 200–300 proteins, which corresponds to a sensitivity limit of 10^{-6} M (see Fig. 1).

Thus, it would be expedient to indicate, while compiling a proteomic map, at which concentration a particular proteome is obtained. It may well be that such a practice will help exclude discrepancies in the literature data [2] and, in our opinion, will compel the contributors to adhere to the criteria used by the editors of PROTEOMICS when assessing submissions [3].

The cause of the above-indicated limitations is the absence of a technological basis, solid enough to register single molecules, in proteomics. That is why the situation in proteomics is much more complicated than that in genomics, where PCR is widely adopted for the replication of low-abundant molecules. It appears that decisive successes in proteomics will only be gained when the CSL is lowered sufficiently to allow the investigator to detect one molecule in 1 L of biomaterial (10^{-24} M) and thus approach the reverse Avogadro number $(6 \times 10^{23}\ \text{molecules} \cdot \text{mol}^{-1})^{-1}$.

Are there any real prerequisites for the creation of technologies enabling work at protein concentrations close to the reverse Avogadro number? The required sensitivity (*i.e.,* the ability to register a single molecule in 1 L of biological material) may be achieved by the use of molecular detectors for the analysis of surfaces, which have been enriched by molecules from biological material [21]. The technology involves two stages: biospecific fishing onto the chip surface and the registration of "fished-out" objects by the use of molecular detectors. The "fishing" stage is realized by fixation, onto the chip surface, of specific probe molecules, *e.g.,* antibodies or aptamers. In the course of incubation of the chip in the sample, the interaction of chip-fixed molecules with partner molecules in solution occurs. The interaction may be either reversible or irreversible. The chip is incubated in the sample, and in the course of incubation the molecules captured from solution and placed onto the affine surface begin to interact with molecules fixed on the support. Then the chip is washed with the aim of minimizing the nonspecific adsorption of molecules, and analyzed by a molecular detector that is capable, by its registered characteristics, of distinguishing single molecules and their complexes.

A most promising molecular detector is one based on the atomic force microscope [21]. At the same time, other detecting devices cannot be excluded such as other types of molecular scanning microscopes [22] or electrochemical detectors on nanowires [23] and nanopores [24].

The general principle of how an AFM-based molecular detector functions is as follows: the cantilever moves across the smooth surface and meets various objects on its way such as individual molecules or their complexes with partner proteins. The heights (or, alternatively, volumes) of molecular objects are registered by the deflection of the cantilever. Analysis of the scanned images makes it possible to calculate the ratio of the number of single molecules to the number of molecular complexes. If antibodies or aptamers to certain proteins are used as immobilized molecules, we may identify their complexes with appropriate antigens and, hence, estimate the total number of antigen molecules in solution [21].

The major obstacle of such technology is a the reversible nature of the complex formation reaction proceeding on the biochip surface. Indeed, complex formation between ligand A and chip-immobilized ligate B can be expressed as:

$$A + B \Leftrightarrow AB$$

The dissociation constant of this reaction K_d is given by the following equation:

$$K_d = \frac{[A][B]}{[AB]} = \frac{([A_0] - [AB])([B_0] - [AB])}{[AB]}$$

where [A], [B], and [AB] are concentrations of the ligand, ligate and their complex, respectively, whereas $[A_0]$ and $[B_0]$ denote the initial concentrations of ligand and ligate.

The CSL can be taken as the minimal initial ligand concentration that could be detected by the registration device:

$$CSL = [A_0] = K_d \frac{[AB]}{[B_0] - [AB]} + [AB] = K_d[AB]([B_0] - [AB])^{-1} +$$

$$+[AB] = K_d\left(\frac{B_0}{[AB]} - 1\right)^{-1} + [AB] \quad (1)$$

The concentrations of the complex [AB] and ligate $[B_0]$ are expressed using Avogadro's number (N_A) and the volume of the probe V as follows:

$$[AB] = \frac{N_{AB}}{N_A V};\ [B_0] = \frac{N_{B_0}}{N_A V}, \quad (2)$$

where N_{B_0} and N_{AB} stand for the number of initial chip-immobilized molecules of the ligate and the number of complexes on the chip, respectively. Therefore, from Eq. (1) the CSL is expressed by the formula:

$$CSL = K_d\left(\frac{N_{B_0}}{N_A V}\frac{N_A V}{N_{AB}} - 1\right)^{-1} + \frac{N_{AB}}{N_A V} = K_d\left(\frac{N_{B_0}}{N_{AB}} - 1\right)^{-1} + \frac{N_{AB}}{N_A V} \quad (3)$$

For our AFM biochip containing $N_{B_0} \approx 15\,000$ immobilized molecules on an area of 400 μm^2 (with a sample volume V of 1 mL and the average affinity for antigen/antibody complexes $K_d \sim 10^{-12}$ M) and assuming that an N_{AB} of ten complexes is sufficient for registration, the CSL becomes 10^{-15} M:

$$CSL \approx \left[10^{-12}\left(\frac{1.5 \times 10^4}{10} - 1\right)^{-1} + \frac{10}{(6.02 \times 10^{23})10^{-3}}\right] \approx 10^{-15}\text{M} \quad (4)$$

In Eq. (3) and (4), the second term is negligible, whereas the first one determines the CSL at the level of 10^{-15} M.

However, in the case of irreversible binding, K_d equals zero, making the first term of the equation zero as well. In such a case, the CSL will be

$$\mathrm{CSL} \approx \frac{N_{AB}}{N_A V} \tag{5}$$

For the aforementioned particular AFM biochip

$$\mathrm{CSL} \approx \left(\frac{10}{(6.02 \times 10^{23})10^{-3}}\right) = 1.6 \times 10^{-20}\mathrm{M} \tag{6}$$

In the case of reversible binding, K_d primarily determines the CSL, whereas for irreversible binding the CSL is primarily a function of the probe volume. It is assumed that, if even a single target molecule is present in the given volume, it will be captured on the chip. Formally, the CSL can be defined as the concentration of the ligand at which no complexes are registered by the molecular detector, *i.e.,* $[C] \rightarrow \mathrm{CLS}$, if $N_{AB} \rightarrow 0$:

$$N_{AB} = N_A V[C] \tag{7}$$

In the experiment illustrated in Fig. 3, the dependence of the number of antigen/antibody complexes — HCV core antigens/HCV antibodies, formed on the AFM support after the fishing procedure — on the concentration in solution of the target HCV core antigen molecule is shown. The results obtained confirm the theoretically calculated limitation of reversible binding: in our experimental conditions — at the antigen concentration of 10^{-12} M — no complexes are registered on the biochip surface (Fig. 3A). In Fig. 3, the theoretical curves illustrate the reversible binding according to Eq. 3 (Fig. 3C) and irreversible binding according to Eq. 7 (Fig. 3C).

To convert the reversible antigen/antibody binding into an irreversible binding, the antibodies were treated (prior to their immobilization on the chip) with a photochemically activated label, *N*-5-azido-2-nitrobenzoyl oxysuccinimide [www.pirsenet.com]. The efficiency of antibody immobilization was substantially decreased, so that no more than 3000 molecules on the area of 400 μm^2 were available — with a further increase in antibody concentration the formation of aggregates was observable. However, the CSL was lowered by several orders of magnitude and reached the value of 10^{-17} M (see Fig. 3B).

In summary, it would be worthwhile to emphasize the following points. Overcoming the CSL at the expense of nanotechnologies is a necessary step toward further devel-

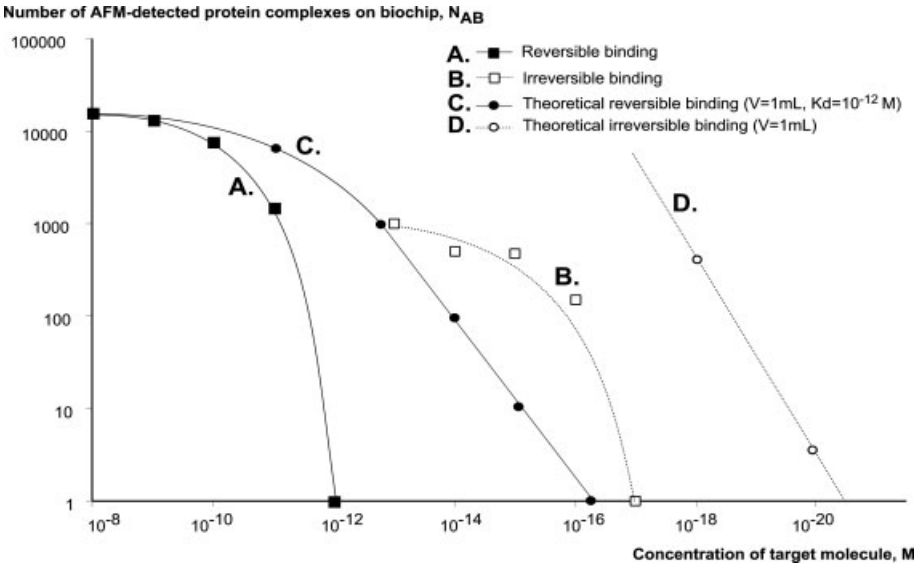

Figure 3. The dependence of the number of antigen/antibody complexes on the concentration of antigen (HCVcore) in solution. Experimental conditions: antiHCVcore antibodies were immobilized on the surface of aminosilanized mica by use of 1-ethyl-3(3-dimethylaminopropyl) carbodiimide (EDC)/N-hydroxysuccinimide (NHS) [25]. The surface concentration of the antibodies was 15 000 molecules *per* 400 μm^2 for the case of reversible binding (A) or 3000 molecules for the case of irreversible binding (B). To achieve irreversibility the anti-HCVcore was modified, prior to immobilization, using the photocrosslinker ANB-NOS [26]. The biochips were incubated in 1 mL of HCVcore antigen solution in 50 mM PBS buffer (10 mM Na-phosphate buffer, pH 7.4; 138 mM NaCl and 2.7 mM KCl) for 30 min at 37°C. (B) Incubation mixture was irradiated at 300 nm for 15 min to achieve covalent antigen/antibody binding. After the incubation the biochip was washed and the AFM scanning of its surface was carried out. The AFM measurements were made in air using tapping mode on a multimode "NTEGRA" atomic force microscope (NT-MDT, Moscow, Russia) as described in [21]. Cantilevers NSG 10 supplied by NT MDT were used. The resonant frequency of the cantilevers was 190–325 kHz and the force constant was about 5.5–22.5 N/m. The calibration of the microscope by height was carried out on a calibration *TGZ 1* grating NT-MDT with a step height 23 $\pm$ 1 nm and period 3 $\pm$ 0.05 μm. At least 1000 objects were measured in each experimental series, and at least 40 height values were obtained for each object. Theoretical reversible binding is assessed by Eq. (3), taking 1 mL as the volume of the sample and $K_d = 10^{-12}$ M (C). Theoretical irreversible binding (D) is assessed by Eq. (7) for the same sample volume.

opment of this area of research. As soon as we gain the capacity to register single protein molecules and their complexes, we shall be able to address the problem of compiling proteomic maps in any biological material on a single-molecule level. The alternative approach in this research area, which will probably find wide application in the future, lies in the development of methods enabling replication of protein molecules [27] or the use of PCR for monitoring the antigen/antibody interaction reactions (immuno-PCR [28], immuno-RCA [29]), and the method for proximal coupling of aptamers [30]). However, insufficient development of these methods does not yet allow us to consider them as a basis for highly productive proteomic investigations.

Concluding the discussion on CSL — an apparent "bottleneck" for present-day proteomics — we would like to emphasize that, in view of the above calculations and experimental data, the proposed technology does not look fantastic and may be realized even now. CSL-free proteomics will enable us to resolve important applied problems of medical diagnostics and accelerate the creation of a comprehensive atlas of proteins from the organs and tissues. It may well be suggested that overcoming the concentration barrier will equalize proteomics in its analytical potential to genomics and transcriptomics, thereby creating a real technological basis for the system biology of the future. For the time being — in order to eliminate the discrepancies in the particular characteristics of proteomic maps derived from the same biomaterial but from different laboratories — it would be worthwhile to indicate the CSL at which these characteristics were obtained upon their conversion to molecules *per* volume of biomaterial.

This article partially includes the results obtained with the support of the following funding sources: RFBR 050448690, RFBR-0604-08057-ofr, FASI ##02435.11.3010/02.434.11. 1013, Program of the Russian Academy of Medical Sciences "Proteomics in medicine and biotechnology", and Scientific School Funding (*02.445.11.7276*).

References

[1] Aebersold, R., Mann, M., *Nature* 2003, *422*, 198–207.

[2] Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W. *et al.*, *Proteomics* 2005, *5*, 3226–3245.

[3] Wilkins, M. R., Appel, R. D., Van Eyk, J. E., Chung, M. C. *et al.*, *Proteomics* 2006, *6*, 4–8.

[4] US 4,683,202, 1985, Mullis, K.

[5] Cutler, P., *Proteomics* 2003, *3*, 3–18.

[6] Lowry, O. H., Rosebrough, N. J., Farr, A. L., Randall, R. J., *J. Biol. Chem.* 1951, *193*, 265–275.

[7] White, I. R., Pickford, R., Wood, J., Skehel, J. M. *et al.*, *Electrophoresis* 2004, *25*, 3048–3054.

[8] Kuster, B., Mann, M., *Curr. Opin. Struct. Biol.* 1998, *8*, 393–400.

[9] Vorderwülbecke, S., Clereverley, S., Weinberger, S. R., Wiesner, A., *Syst. Nat. Methods* 2005, *2*, 393–395.

[10] Anderson, N. L., Anderson, N. G., *Mol. Cell Prot.* 2002, *1*, 845–867.

[11] Wu, G., Datar, R. H., Hansen, K. M., Thundat, T. *et al.*, *Nat. Biotechnol.* 2001, *19*, 856–860.

[12] Sonksen, C. P., Nordhoff, E., Jansson, O., Malmqvist, M., Roepstorff, P., *Anal. Chem.* 1998, *70*, 2731–2736.

[13] Zheng, G., Patolsky, F., Cui, Y., Wang, W. U., Lieber, C. M., *Nat. Biotechnol.* 2005, *23*, 1294–1301.

[14] Wiese, R., Belosludtsev, Y., Powdrill, T., Thompson, P., Hogan, M., *Clin. Chem.* 2001, *47*, 1451–1457.

[15] Bjorhall, K., Miliotis, T., Davidsson, P., *Proteomics* 2005, *5*, 307–317.

[16] Omenn, G. S., The HUPO Plasma Proteome Project Pilot Phase: Technology Platforms, Reference Specimens, DataBase. *Working Summary March* 2004.

[17] Albrecht-Buehler, G., *Int. Rev. Cytol.* 1990, *120*, 191–241.

[18] Kanaeva, I. P., Petushkova, N. A., Lisitsa, A. V., Lokhov, P. G. *et al.*, *Toxicol. In Vitro* 2005, *19*, 805–812.

[19] Zgoda, V. G., Tikhonova, O. V., Serebryakova, M. V., Lisitsa, A. V., Archakov, A. I., *Proteomics* 2006, *6*, 4662–4670.

[20] Xiang, R., Shi, Y., Dillon, D. A., Negin, B. *et al.*, *J. Proteome Res.* 2004, *3*, 1278–1283.

[21] Ivanov, Yu. D., Govorun, V. M., Bykov, V. A., Archakov, A. I., *Proteomics* 2006, *6*, 1399–1414.

[22] Garcia-Parajo, M. F., Veerman, J. A., Segers-Nolten, G. M. J., de Grooth, B. G. *et al.*, Greve, J., *Cytometry* 1999, *36*, 239–246.

[23] Patolsky, F., Zheng, G., Hayden, O., Lakadamyali, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 14017–14022.

[24] Fologea, D., Gershow, M., Ledden, B., McNabb, D. S. *et al.*, *Nano. Lett.* 2005, *5*, 1905–1909.

[25] Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B. *et al.*, *Bio Techniques* 1991, *11*, 620–627.

[26] Levis, R. V., Roberts, M. F., Dennis, E. A., Allison, W. S., *Biochemistry* 1977, *16*, 5650–5654.

[27] Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., Ghadri, M. R., *Nature* 1996, *382*, 525–528.

[28] Ruzicka, V., Marz, W., Russ, A., Gross, W. *et al.*, *Science* 1993, *260*, 688–689.

[29] Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2000, *97*, 10113–10119.

[30] Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C. *et al.*, *Nat. Biotechnol.* 2002, *20*, 473–477.

