Superresolution Chromatography*

E.L.Kosarev[a] and K.O.Muranov[b]

[a]P.L.Kapitza Institute for Physical Problems,
Russian Academy of Sciences,
ul.Kosygina, 2, Moscow 117334, Russia

[b]N.M.Emanuel Institute of Biochemical Physics,
Russian Academy of Sciences,
ul.Kosygina, 2, Moscow 117977, Russia

Abstract

A method for improving the resolution of the chromatographic analysis based on deriving the point-spread function of a chromatographic column, i.e. a chromatogram of an individual compound, is described. A system of two functions - a chromatogram of the substance analyzed and a point-spread function of the chromatographic column - in combination with noise statistics enables the application of a RECOVERY signal-reconstruction software package in order to obtain a superresolution chromatography. The superresolution means a better resolution than that determined by the width of the point-spread function. The proposed method is tested with a size exclusive chromatography of bovine serum albumin. The resolution obtained exceeds one of high-performance liquid chromatography (with a lower cost of the instrument system by a factor of 15-20).

1 Introduction

Chromatography is a physicochemical method for the investigation of substances. This method is based on the separation of components moving through a chromatographic column [1], which is a calibrated tube 10 cm (high performance liquid chromatography (HPLC)) to 30 m (gas chromatography) long. Size exclusion chromatography or gel filtration clearly demonstrates the process of separation. The column bed is the gel formed by twisted polymer molecules. The polymer (agarose, dextran, polyacrylamide, etc.) is formed as ~ 100 μm granules. When a sample moves through the column, large molecules crossing the space between the granules move together with the solvent flow. Smaller molecules

[1]Preprint version of the paper published in "Instruments and Experimental Techniques", Vol.44, No.5, pp.638-643, 2001. Translated from "Pribory i Tekhnika Eksperimenta, No.5, pp.74-79, 2001.

penetrating into the granules are delayed and exit later. Thus, the sample is divided into fractions, depending on the molecular weight. The retention time t of a substance (which is equal to the time necessary for the molecule passage through the chromatographic column) is connected with its molecular weight M by the formula

$$\log M = a - bt \tag{1}$$

(see [2]). The constants a and $b > 0$ are determined by calibration measurements of samples with known molecular weights.

A chromatogram is often represented as the sum of peaks, many of which overlap, making an accurate analysis difficult. To increase the resolving power of the method, both the new sorbents, which perform a finer separation, are developed and calculation methods are used. The calculation methods are based on the assumption that the peak shape of an individual substance can be described analytically, for example, with the use of the Gauss function [2]. However, this is not often true. For example, a gel bed has some adsorption activity. Therefore, one kind of separation (for example, size exclusion chromatography) is imposed on another kind of chromatography (adsorption chromatography), resulting in distortion of the peak shape. In this case, because of irreversible adsorption, the properties of the column can vary. This leads to the fact that the peak shape of an individual compound becomes not only asymmetric but also time-dependent.

Therefore, the parametric technique used in many standard chromatography data processing packages cannot give reliable results in spite of a variety of used basis functions. The shape of the peak observed in experiments does not often belong to a vast set of standard package functions.

We suppose that the problem of overlapping peak's separation could be accurately solved with the use of a nonparametric method. The peak's shape in this method does not described by number of parameters (position of peak maximum, peak width, e.t.c.) but the peak's shape is determined directly from measurements by using the chromatography of an individual compound.

A chromatogram of a substance mixture is the superposition of individual peaks. Every peak is an individual substance. If the shapes of these peaks are the same throughout the entire working range of the device, then chromatogram is the convolution of a sought for distribution with the peak of this shape. In this case (gel-filtration), the molecular-weight distribution is studied. The peak's shape is determined directly from the chromatography of an individual compound. The peak's shape can be called, by analogy with optics, the point spread function of a chromatographic column.

The inverse problem of decomposing a complex spectrum into the same components is achieved by solving an integral convolution equation. For this equation, the input data is the chromatogram, and the convolution operator kernel is the point spread function of the chromatograph column. Solutions to similar problems can be found in [3-6].

2 RECOVERY software package and superresolution

To solve this problem, we use a RECOVERY software package restoring signals from the noisy data [5,6] that was developed specially to restore nonnegative functions. The desired

spectra are always such functions, in this case representing the molecular weight distribution for the substance under study. In the general case, the spectrum is an intermolecular interaction distribution of the studied compound, accounting for the separation.

The RECOVERY software package is based on the maximum likelihood principle (MLP), since this principle provides the maximum possible restoring efficiency [7]. According to the MLP it is necessary to determine the likelihood function:

$$L = \mathcal{P},$$

where

$$\mathcal{P} = \mathcal{P}(F|f_0)$$

is the conditional probability of observing a set of experimental data, coinciding with the actually obtained set

$$F(x_i), \quad i = 1, 2, ..., n,$$

on the condition that the desired solution is equal to the function f_0. We can consider many unknown values of the function $f_0(t_j),\ j = 1, 2, \ldots, m$ as a vector in m-dimensional space of solutions. Each point in the space corresponds to one of possible solutions. Further, the maximum of the likelihood function is derived in this m-dimensional space of solutions on the condition that some necessary restrictions on the solution are met.

The likelihood function for experimental data with the Gaussian statistics reduces to a negative value of a quadratic deviation between the experimental data $\{F\}$ and their approximation: $\{\hat{F}_0\}$

$$L = -\frac{1}{2}\|F - \hat{F}_0\|^2.$$

The work [6] gives iteration formulas for localization of the likelihood function maximum. These iteration formulas implement the nonlinear signal restoration method that, unlike the linear methods, is capable of expanding the bandwidth of the restored signal in comparison with that of the input signal. The nonlinear method is capable to determine thinner details in the output signal, even invisible in the input data and, therefore, increase the resolution.

The natural measure of the resolution of a chromatograph is equal to the width of its point spread function Δ. The value of width is determined by the diffusion processes eroding the peak. If we use the RECOVERY software package, it is possible to obtain a resolution δ, which is better than this width $\delta < \Delta$ and, in this way, obtain a superresolution

$$SR = \frac{\Delta}{\delta} > 1.$$

It was shown in [3,4] that the limiting value of the achievable superresolution is limited by the input data noise and can be calculated with the formula

$$SR < SR_{\max} = \frac{1}{3}\log_2(1 + E_s/E_n), \tag{2}$$

where

$$E_s = \int_{t_0}^{t_1} F^2(t)\,dt,$$

is the signal energy; $E_n = n\sigma^2$ is the noise energy; $F(t)$ is the value of the obtained chromatogram at the time t; t_1 and t_2 are the initial and final times in the chromatogram; n is the number of data in the chromatogram; and σ^2 is the noise variance at each point. If we express the signal-to-noise ratio in (2) in decibels,

$$dB = 10 \lg_{10}(E_s/E_n),$$

it is possible to write a very easy-to-use approximate expression for the limiting superresolution

$$SR_{\max} \simeq dB/10.$$

The signal-to-noise ratio for the data below on the chromatographic analysis of bovine serum albumin (BSA) (Fig.1a) is 47 dB. This means that the quality of the data makes it possible to obtain a five times higher resolution than the width of the point spread function of the chromatography column. The proposed technique yielded good results in spectroscopy [8-10] and nuclear physics [11]. In this work, we use it to solve chromatographic problems.

3 Materials and methods

The experimental technique was as follows: a protein mixture of known composition (bovine serum albumin (BSA) monomer, dimer, and trimer) was separated by gel filtration. However, the used separating gel, as it had been known earlier, could not in principle separate the mixture. So, a chromatogram with heavily overlapping peaks was obtained. The data was processed with the use of the RECOVERY software package, and the result was compared with the finer separation data obtained with the use of the HPLC.

The bovine serum albumin (fraction No. 5, Sigma, USA) was dissolved in a phosphate buffer (NaCl - 100 mM, NaH_2PO_4 40 mM, ethylenediaminetetraacetic acid 1 mM, pH = 6.8) and, after ultra- filtration (Millipore, 0.2 μm), was applied to a Fractogel TSK HW 55 gel (TOYO SODA, Japan). The separation was carried out at a solvent current rate of 1.5 ml/min, column (25 x 900 mm). The column eluent was monitored for absorbance at a wavelength $\lambda = 280$ nm (Uvicord S 2138, LKB, Sweden). The output signal from the detector was inputted to a computer via a special interface [12]. The protein purity was checked by electrophoresis [13] in a polyacrylamide gel in the presence of sodium dodecylsulfate and mercaptoethanol (SDS-PAAG) both in denaturing (boiling during 5 min) and native conditions. The high performance liquid chromatography was performed in an TSK G2000 SW Spherogel column, 10 μm, 7.5 x 600 mm (LKB, Sweden) at a solvent current rate of 0.5 ml/min, Beckman-Altex 334 chromatography system (USA). The columns were calibrated with the use of the Pharmacia Gel Filtration Standard. The peak area was measured calculations with the use of the MicroCal Origin 4.1 software package. The calculations were made on IBM-compatible personal computers with MS-DOS and Windows NT operating systems.

Of the entire RECOVERY software package, in this work, we used **Dconv_** and **Dconv2_** programs designed to solve an integral convolution equation with a Gaussian noise distribution with an increased amount of input data in the chromatogram: up to 4096 and 2048, respectively, for the above programs. Complete listings of a standard version of the

RECOVERY software package written in Fortran-77 are given in [5] and in a computer program library at http://cpc.cs.qub.ac.uk under the name ACLJ [6]. Additional data on this software package can be obtained at http://kapitza.ras.ru/people/kosarev/recovery.html

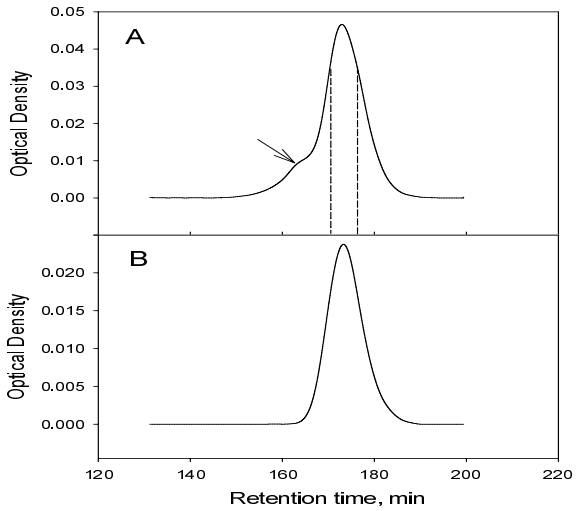

Figure 1: Determination of the point-spread function of a column with TSK gel: (a) chromatogram of the BSA initial mixture (30 mg). The arrow indicates an unresolved peak; dashed lines mark the time interval when the BSA monomer fraction was collected; (b) chromatogram of the collected BSA fraction. Attention should be paid to the absence of the unresolved peak and slight visible asymmetry.

4 BSA chromatography and point spread function determination

Figure 1a illustrates a BSA chromatogram. It can be seen that the substance consists of at least two components (the second component is indicated by an arrow in the figure). The resolution of the used gel bed is insufficient for the separation of these components. This is no surprise, since the used column is designed for a wide molecular weight range of 1-700 kD [14]. Therefore, it is incapable to separate a BSA monomer and dimer with molecular weights of 67 and 134 kD, respectively.

Dashed lines indicate the time interval when a fraction containing the BSA monomer was collected. The obtained fraction was repeatedly chromatographed on the same column. The elution profile shown in Fig.1b shows the relative purity (absence of the BSA dimer or trimer) of the obtained fraction. The obtained fraction was also studied with the use of electrophoresis in SDS-PAAG. The results indicate that the studied fraction contains the BSA monomer, $\sim$15% low molecular weight impurities, and trace quantities of the BSA dimer or trimer. The observed slight peak asymmetry can be caused by the impurities of low molecular weight substances (products of the BSA proteolysis) and by the above mentioned extraneous causes: an irreversible adsorption, hydrophobic interactions, etc.

This peak asymmetry does not prevent using the proposed method in order to reach a

superresolution at the chromatographic separation. The curve shown in Fig.1b was used in this work as a point spread function.

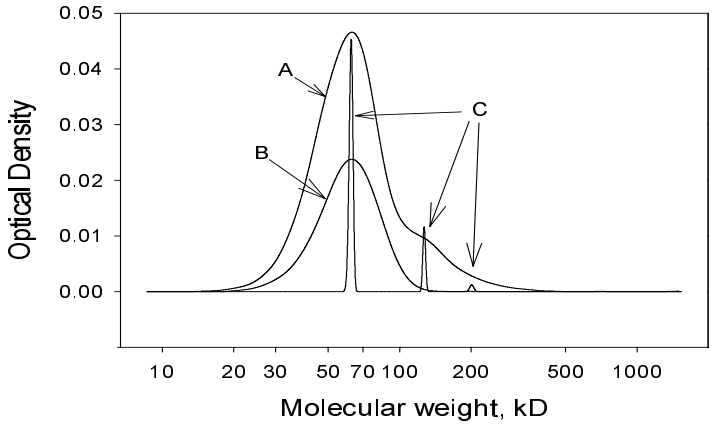

Figure 2: The use of the RECOVERY software package for the initial data given in Fig.1: (A) the BSA chromatogram obtained for a TSK-gel column; (B) point-spread function (BSA monomer chromatogram); (C) recovering result of the chromatographic separation data with the use of the **Dconv2_** program from the RECOVERY software package.

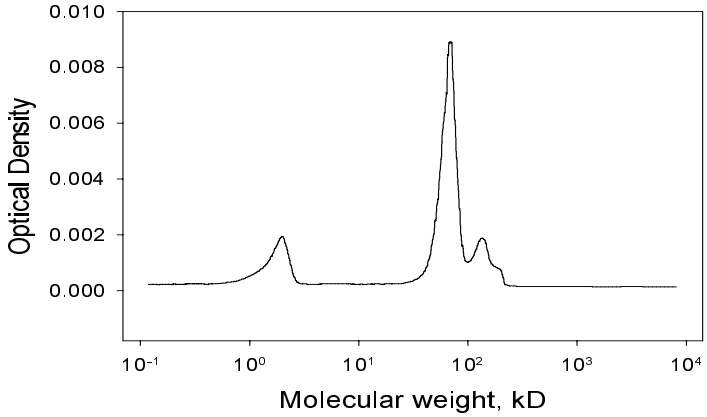

Figure 3: High-performance liquid chromatogram of the BSA compound; the column is TSK G2000 SW Spherogel, 10 μm, 7.5 x 600 mm (LKB); the solvent current rate is 0.5 ml/min.

Figure 2 illustrates the BSA chromatogram (curve A) and BSA monomer fraction used as a point spread function (curve B). The narrow peaks are the result of recovering the chromatographic separation data by using the **Dconv2_** program. Along the X-axis, we did not plot the substance retention time, as was done in Fig.1, but the molecular weights calculated by formula (1). The recovery results in a set of three peaks, which correspond to the BSA monomer, dimer, and trimer.

Figure 3 illustrates the HPLC separation of the same BSA sample. It is common knowledge that this method allows high performance separation due to using the finely dispersed gel beds. We used a HPLC column with 10 μm granules.

This small granule size allows one to increase significantly the separation efficiency. However, this significantly increases the operating pressure required for pumping a solvent through the column, and, hence, special pumps, high quality compounds, and special materials should be used. All these factors appreciably increase the cost of the analysis. The HPLC separation of a mixture is much more effective than TSK gel chromatography. It can be seen from the figure that the compound contains four components, namely, a low molecular weight impurity (of a 2 kD molecular weight), monomer, and dimer with an unresolved BSA trimer. The results given in this figure coincide with the data in the literature [15], which indicate that the BSA fraction No.5 used in our work contains monomer, dimer and trimer molecules; some quantity of low molecular weight impurities; and trace quantities of a BSA tetramer.

A peak with a small molecular weight, which can be observed in Fig.3 and is absent in Fig.2, lies off the molecular weight range of the column used. At the same time, the molecular weight range (67-201 kD), which is not completely separated by the HPLC, is excellently separated by our method. If we evaluate the superresolution obtained with the use of the RECOVERY software package by the data given in Fig.2 as the ratio of a width of the point spread function to a distance between the BSA dimer and trimer, we obtain a value of 1.1. This already justifies the use of the term *superresolution* in the title of our work. As mentioned above, the quality of the data allows one to improve this value as much as five times. Below, the molecular weights of compounds (kD) obtained with the use of the standard TSK size exclusion chromatography, HPLC, and the method described in this work are given in comparison with data from the Sigma-Aldrich catalogue [16].

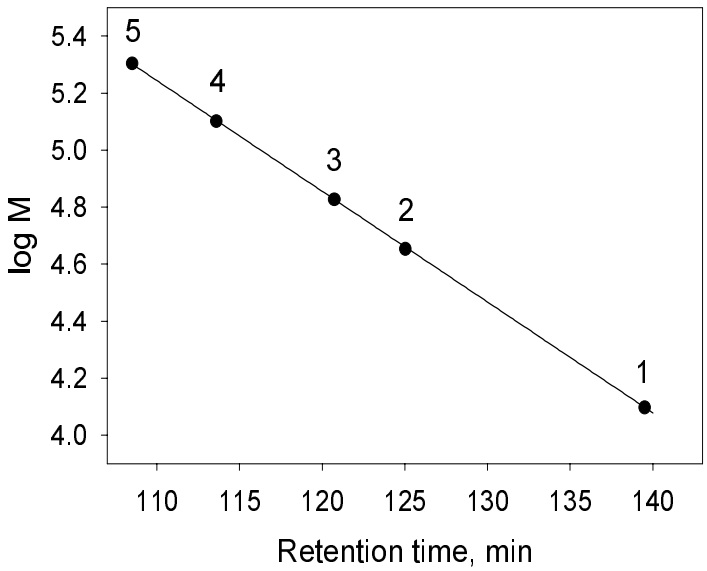

Figure 4: Calibration straight line for determining the constants a and b from formula (1) obtained in the separation of standard proteins: (1) cytochrome C (12.5 kD), (2) ovalbumin (45 kD), and (3) BSA (67 kD). The decimal logarithm of the protein molecular weight in [D] is plotted along the Y-axis. The approximating straight line $y(x) = -0.0385x + 9.47$ is plotted by the least squares technique with the use of points Nos.1, 2, and 3. Points Nos.4 and 5 represent BSA dimer (127 kD) and trimer (200 kD).

Table 1. Molecular weights of BSA components (kD), determined by various methods

Peak number	0	1	2	3
Table value		67 ± 1	134 ± 1.4	201 ± 1.7
TSK-gel	-	67	-	-
HPLC	2	68	135	190
RECOVERY	-	66.4 ± 0.53	127 ± 1.62	200 ± 5.78

The line denoted RECOVERY contains data averaged over five independent measurements with a probable error of $\pm 1\sigma$. Peaks are numbered in increased order of the molecular weights of the compounds. The peak No.0 corresponds to a low molecular weight impurity and peaks Nos.1, 2, and 3 correspond to the BSA monomer, dimer, and trimer. Figure 4 illustrates a calibration curve for determining the constants a and b from formula (1). The proposed method demonstrates a good agreement with the published data with respect to the determined molecular weights, namely, better than 5%, and demonstrates an important improvement in the resolution in comparison with the HPLC. Note that it proved possible to reach a 6% error with the use of standard methods only through the simultaneous use of three detectors, namely, light scattering, viscosimetric, and refractive ones [17], while this work uses only one detector.

Table 2. The percentage of the components of a separated mixture

Peak number	0	1	2	3
TSK-gel	-	100	-	-
HPLC	17.3	66.7	13.7	2.3
RECOVERY	-	83 ± 6	14.7 ± 6	2.3 ± 1

The reasons behind the remaining slight discrepancy between the determined molecular weights can be divided into two groups: the first one is related to the determination of the molecular weights by using the calibration curve, and the second, to the use of the nonparametric method. Let us consider these reasons one after the other. The albumin monomer molecule has a shape close to an ellipsoid of revolution with the 140 x 40 Å principal and transverse axes. At the same time, a dimer is an aggregate of two such molecules with a 50% overlap [18]. The hydrodynamic properties of the dimer and, naturally, its capability to penetrate into the separating gel significantly differ from the monomer properties. That is, possibly, the reasons why a dimer molecule gives a decreased molecular weight value.

The second reason is the special features of the method used. As indicated above, we use the nonparametric method without fixing in advance the shape of the point spread function of the chromatographic column. The point spread function is determined experimentally, not assuming in advance the shape of the molecular weight spectrum of the substance analyzed as a set of discrete lines. We have to pay for these less stringent requirements by decreasing the accuracy of the spectrum restoration in comparison to the parametric method. The least possible error in determining the parameters at the MLP application is derived from the Cramer-Rao inequality [7], and this error varies as a power dependence of an input data error. At the same time, the limiting possible resolution of the nonparametric method depends logarithmically on the input data error [3, 4]. This resolution is generally worse than that obtained with the parametric method.

The optimal way is, at first, to measure the sufficiently reliable data on the shape of the point spread function and recovered spectra using the nonparametric method, and then the parametric method should be applied. An example of a successful use of this approach can be found in [11].

These data demonstrate good agreement of the component percentage determined by the well-known and new methods. It should be noted that, although the peak of a low molecular weight impurity is not visible in experiments with the TSK gel column, its presence is reflected in the percentage of the main peak No.1. At the same time, a contribution of the low molecular weight impurity into the total balance (peak No.1) coincides, with good accuracy, with the sum of the peaks Nos.0 and 1 determined with the HPLC technique.

Thus, we have shown that the proposed method fundamentally improves the quality of the chromatographic separation. Quantitatively measure of overlapping peaks is achieved. Note that these new possibilities are achieved through reasonable processing of the measured data with no complication in the instrumentation.

The use of the RECOVERY software package for the gel filtration data significantly increased the resolution of this method and exceeded the quality of the separation obtained with the HPLC technique. Note that the cost of the used instrument complex for gel filtration ($1000. - US dollars) is roughly 15-20 times lower than that for the HPLC setup ($20, 000. - US dollars).

Acknowledgments

We are grateful to S.N.Bubenchikova from the Belozerskii Research Institute for Physics and Chemical Biology, Moscow State University, for her interest in this work and assistance in performing the HPLC.

This work was supported by the Russian Foundation for Basic Research and Research Program of the Higher Education Ministry of the Russian Federation and Moscow State University "Fundamental Studies of Higher School in the Area of Natural and Humanitarian Sciences. Russia's Universities, " project Nos. 00-04-48364 and 992582, respectively.

References

[1] Sakodynskii K.I., Brazhnikov V.V., Volkov S.A., et al. *Analiticheskaya khromatografiya (Analytical Chromatography),* Moscow: Khimiya, 1993, pp. 16-21. (in Russian)

[2] Friefelder D. *Physical Biochemistry.* San Francisco: Freeman, 1976. Translated into Russian under the title Fizicheskaya biokhimiya, Moscow: Mir, 1980, pp. 200-201.

[3] Kosarev E.L. The Shannon limit of superresolution and its attainment in signal restoration. *Instruments and Experimental Techniques,* Vol.32, No.4, part 1, pp.827-831, 1989

[4] Kosarev E.L. Shannon's superresolution limit for signal recovery. *Inverse Problems,* Vol.6, No.1, pp.55-76, 1990

[5] Gelfgat V.I., Kosarev E.L., and Podolyak E.R., Programs for signal recovery from noisy data by maxumun likelihood method. *Instruments and Experimental Techniques,* Vol.34, No.5, part 1, pp.1070-1075, 1991, available from VINITI, 1991, Moscow, No. 2635-V91.

[6] E.L.Kosarev E.L., Gelfgat V.I., Podolyak E.R. Programs for signal recovery from noisy data using the maximum likelihood principle. 1: General description. *Computer Physics Communications,* Vol.74, No.3, pp.335-348, 1993. 2: Program implementation. ibid, pp.349-357, 1993

[7] Linnik, Yu.V., *Metod naimen'shikh kvadratov i osnovy teorii obrabotki nablyudenii (Least Squares Method and Principles of Data Processing),* Moscow: Fizmatgiz, 1962, ch.3, sect.3. (in Russian)

[8] Kosarev E.L., Peskov V.D., and Podolyak E.R., High resolution soft X-ray spectrum reconstruction by MWPC attenuation measurements. *Nucl.Instrum. and Methods,* Vol.208, pp.637-645, 1983

[9] Korotkikh V.L., Kosarev E.L., Ormont A.B., and Korotkikh A.V. Improving the energy resolution of photoelectron spectra by digitally correcting for the instrumental function. *Instruments and Experimental Techniques,* Vol.37, No.6, part 2, pp.721-725, 1994

[10] Kaminskii A.S., Kosarev E.L., E.V.Lavrov E.V. Using comb-like instrumental function in high-resolution spectroscopy. Meas.Sci.Technol. Vol.8, No.8, pp.864-870, 1997

[11] Kolganova E.A., Kosarev E.L., Ososkov G.A. Superresolution algorithms for data analysis in nuclear physics. *Nuclear Instruments and Methods in Physics Research*, Vol. A443, pp.464-477, 2000

[12] Kirillov V.I. Computer interface for digital instruments using the HP-IB standard. *Prib. Tekh. Eksp.*,, No.4, p.89, 1981 (in Russian)

[13] U.K.Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* Vol.227, No.259, pp.680-685, Aug.15, 1970

[14] TSK-Gel Toyopearl, *TOYO SODA manufactoring Co., LTD*

[15] T.Arakawa, Y.Kita. Protection of Bovine Serum Albumin from Aggregation by Tween 80. *Journ. Pharmaceutical Sci.*, Vol.89, pp.646-651, 2000

[16] Sigma-Aldrich 2000-2001 Catalog, Sigma-Aldrich Co.

[17] S.Ross. Size exclusion chromatography helps improve production of silicone fluids, *LabPlus International*, v.14, No.4, p.26, 2000

[18] P.G.Squire, P.O.Moser, C.T.Konski. The Hydrodynamic properties of Bovine Serum Albumin Monomer and Dimer. *Biochemistry*, v.7, No.12, p.4261-4272, 1968