ELSEVIER

Journal of Chromatography A, 691 (1995) 67–79

JOURNAL OF CHROMATOGRAPHY A

Multivariate visualization in the size-exclusion chromatography and pattern recognition of biological samples

R.D. Ricker[a,*], L.A. Sandoval[a], J.D. Justice[b], F.O. Geiser[c]

[a]*Rockland Technologies, Inc., 538 First State Boulevard, Newport, DE 19804, USA*
[b]*Justice Innovations, Inc., 1240 L'Avenida Avenue, Suite A, Mountain View, CA 94043, USA*
[c]*Geiser Scientific, Inc., 287-A Wilson Avenue, Glen Mills, PA 19342, USA*

Abstract

A principal-component technique that maps highly correlated data by the use of non-orthogonal vectors has been described by Geiser et al. [*J. Chromatogr.* 631 (1993) 1–13]. As many as 35 chromatographic variables were simultaneously visualized in a single three-dimensional map, using cosines of the correlation coefficients as vector angles. Applications included the validation of system-suitability parameters and the elucidation of biomolecular structure–activity relationships. In this paper, the mapping technique was extended to the chemometric analysis of data for the size-exclusion chromatography of biological samples. The study was to characterize protein–packing interactions for 25 molecules at nine mobile phase concentrations. Using this chemometric technique, the relationships between all of the data points could be viewed in a single graph. Analysis of the graph revealed mapping regions for proteins demonstrating hydrophobic characteristics, ionic characteristics, a combination of hydrophobic and ionic characteristics, or little interaction with the packing. The optimal mobile phase concentration for ideal size-exclusion chromatography in this system could also be identified from the graph.

1. Introduction

The trend in high-performance liquid chromatography (HPLC) is towards faster and improved chromatography. This is being accomplished in several ways. First, highly purified, low-acidity silica is being used to provide better peak shapes and higher resolution [1,2]. In addition, highly stable bonded phases have been developed that can withstand high-temperature and low-pH operation, reducing separation time by sharpening peaks, and lowering viscosity [3]. Carefully controlled particle-size distributions allow the use of small-particle-size columns having shorter lengths, with no loss in efficiency [4]. Improvements in the unattended operation of HPLC instrumentation and data collection have also increased the number of chromatographic runs that can be performed. Using these advancements, one can easily generate more data than can be readily analyzed and interpreted. With the resulting increase in ability to generate data, and strong demands for faster analyses, chemometrics is becoming an increasingly popular tool [5–9]. Chemometrics is designed to reduce the dimensionality of a multivariate dataset so that patterns in the data might be more easily viewed and interpreted by the human mind. While many of the variables of chromatographic data are highly correlated, the information becomes very complex with the introduction of biological sam-

* Corresponding author.

0021-9673/95/$09.50
SSDI 0021-9673(94)01136-2

ples, different mobile phases and different stationary phases.

We have used a unique chemometric program (DataMax) for visualization and partial quantitation of differences in the interactions between 21 proteins and the column packing during size-exclusion chromatography (SEC) on the Zorbax (registered trademark of DuPont) BioSeries GF-250. Interpretation of these interaction data is especially complicated in SEC because the chromatography is normally carried out under conditions such that the sample is in its native state. The chemometric software (DataMax) maps the correlations of all the attributes (mobile phase concentrations) and all the objects (proteins) of this system, in a single, non-orthogonal, three-dimensional graph (MultiGraf; registered trademark of Echo Data). Reports of the actual statistical values and mapping coordinates are also available. DataMax allowed us to determine the best of our mobile phase concentrations for performing ideal SEC in this system. The technique also provided a means of grouping proteins by their relative hydrophobicities and ionic character, as determined by the extent of their interaction with the column-packing surface.

2. Experimental

2.1. Instrumentation

HPLC was carried out using a Ti-Series 1050 liquid chromatograph equipped with a four-solvent pump, autosampler and variable-wavelength UV detector, all from Hewlett-Packard (Valley Forge, PA, USA). HPLC was also carried out with an LKB 2150 pump (Pharmacia LKB, Piscataway, NJ, USA), DuPont 8800 Series multiwavelength detector (DuPont, Wilmington, DE, USA), and a 9125 injector from Rheodyne (Cotati, CA, USA). Data were collected with an IBM-compatible 80486 computer running ChromPerfect Direct 4I chromatography analysis software (Justice Innovations, Mountain View, CA, USA), using a DT2804 data-acquisition card, all from Geiser Scientific (Glen Mills, PA, USA). Multivariate maps of log M_r vs. relative retention volume were prepared using DataMax software (ECHO Data, Orem, UT, USA). The final graphics were exported as HPGL files and enhanced using CorelDraw 4.0 (Corel, Ottawa, Canada).

2.2. Columns

The columns used were Zorbax BioSeries GF-250, obtained from Mac-Mod Analytical (Chadds Ford, PA, USA). The columns had dimensions of 250 × 9.4 mm, with 5 μm particle diameter and 150 Å pore size.

2.3. Chemicals and proteins

All proteins and low-molecular-mass markers used in this study were of 90% purity or better, purchased from Sigma (St. Louis, MO, USA). Sodium phosphate, monobasic, ultrapure bioreagent, and analytical-grade sodium hydroxide were obtained from VWR Scientific (Bridgeport, NJ, USA). Samples and buffers were made with de-ionized water filtered through a 0.22-μm hydrophilic filter (Millipore, Bedford, MA, USA).

2.4. Procedures

Protein stock solutions were prepared individually at 10 mg/ml. The stock solutions were prepared in de-ionized prefiltered water, with the exception of bovine insulin (INS), which contained 0.05% trifluoroacetic acid (HPLC Spectro-grade, Pierce, Rockford, IL, USA) for its solubilization. Sample mixtures containing 1–3 proteins were prepared for injection by diluting the stock solutions. This was typically a 10-fold dilution with 0.20 *M* sodium phosphate adjusted to a pH of 7.0 with 10 *M* NaOH. In cases where the relative retention volume (K_d) was greater than 1, individual protein concentrations were increased to 2–3 mg/ml to facilitate detection. Sample mixtures were prepared fresh daily. Detection was carried out at 230 nm with a flow-rate of 2 ml/min and ambient temperature.

2.5. Theoretical

Relative retention volume (K_d)

Retention times, t_R, for each protein were determined in mobile phase containing sodium phosphate buffer (pH 7.0) at concentrations from 50 to 1000 m*M*. To reduce experiment time, flow-rates were 2 ml/min. V_R values were used to calculate K_d, defined here as $K_d = \{V_R / V_m\}$, where V_R is the retention volume of the solute and V_m is the column volume of the mobile phase, or total permeation volume of the column (determined as the retention volume of the solute sodium azide at 200 m*M* sodium phosphate, pH 7.0). This use of K_d is simply $k' + 1$, and is useful in this context because it indicates whether a particular solute elutes within the SEC separation range of the column ($V_R < 1$).

Log K_d and regression analysis

For the approximation of missing retention values, regression analyses of log K_d values were carried out using a Hewlett-Packard HP-42S calculator. By restricting the log K_d values used to those obtained at lower mobile phase concentrations, correlation values were kept greater than 0.95 in each case. The resulting extrapolated log K_d values were then converted back to standard K_d.

3. Results and discussion

3.1. Solute–packing interactions in SEC

SEC is one of the few chromatographic methods in which biomolecules may be separated in their native state. However, all columns have surface characteristics that can adversely affect the retention of solute molecules [10,11]. This is especially true in SEC, where the mobile phase is often designed to provide a good environment for the protein and not necessarily for the best performance of the column. A considerable number of column types and mobile phases have been used to better understand and predict protein–packing interactions, and achieve ideal size-exclusion behavior [12–16]. The column used in this study was silica-based and is surface-stabilized with zirconia to withstand operation at higher pH (up to pH 8). Zorbax silica particles have extreme physical stability, allowing high flow-rates and operational pressures. The particles are coated with a bonded monolayer of silane containing a diol functionality. The result is a highly consistent coverage of the particle with the bonded phase. This surface is very hydrophilic and some interaction of proteins with unreacted silanol groups may occur because of this hydrophilic nature. Other SEC supports are made by coating the entire surface of a silica particle with a polymeric bonded phase. While solutes are completely shielded from the silanol groups of these column packings, this surface is not likely to be as reproducible as a monolayer coating and can have more hydrophobic character. Additionally, size-exclusion columns can be made from cross-linked polysaccharide particles. While being very hydrophilic, these packings are relatively soft and can undergo electrostatic interactions with some solutes. Because of this wide variability in synthesis of size-exclusion columns, the chromatographer should have an understanding of solute–column packing interactions before using a particular column type. (See [17] for a recent review of SEC.)

This study reports this type of information for the Zorbax BioSeries GF-250 column. The current results are only one set of a large number of experiments that might be performed to better understand chromatography on a column packing. We intend to complete a more rigorous treatment of this topic, that will include the use of different mobile phase types buffered to higher and lower pH values. These parameters are known to have a dramatic effect on the interaction of proteins with the packing surface. However, with the addition of just four different buffers and three different pH values for these 21 proteins, conclusions would require the analysis of more than 2000 data points. We have, therefore, spent time looking for a tool that would allow rapid visualization of these results, and the work became more focused on learning about the non-orthogonal vector-plotting charac-

teristics of DataMax and introducing it as a valuable tool for this kind of study.

Once all of the data has been entered into the DataMax table, versatile analysis is allowed. Provisions are made for selecting and de-selecting any of the attributes and objects, and making comparisons with external data. A corresponding change is made to the statistical report. The DataMax MultiGraf may be rotated to any orientation to simplify determination of the alignment between any of its objects. In these experiments, it was useful to orient the MultiGraf so that the log M_r axis was vertical and neither angled forward nor backward. This provided a fixed axis for comparison of other attributes to those of "ideal" SEC.

3.2. Extrapolation of K_d values

Data are entered into DataMax using a standard table format —the columns are the attributes (mobile phase concentration) and the rows are individual objects (proteins). Because of the statistical calculations involved, the data table must be completely filled. In this type of study, as in many others, this is not always possible. Many of the proteins were simply not detectable at very low mobile phase concentration. The missing numbers can be determined by extrapolation using linear or curvilinear regression, in systems where this is meaningful. A very similar MultiGraf is created when the estimated values are removed (50, 70 and 80 mM data). Once an initial trend is generated by the data, DataMax seems to properly interpret and represent that trend.

In the case of electrostatic and hydrophobic interactions, a great deal of theory is known. In both cases, linear maps are produced when the logarithm of the retention factor is plotted against logarithm of salt concentration or reciprocal salt concentration.

Specifically, for ion exchange [18,19],

$$\log k = 2z \log \left[1/(X^{+u}Y^{-v})\right] + \log K_z$$

where k = capacity factor; X^{+u} = concentration of displacing ion; Y^{-v} = concentration of counter ion; z = the number of charges interacting between the solute and the surface; K_z = a constant.

For hydrophobic interactions [20],

$$\log k = (1/2.303)Sm + \log K_w$$

or

$$\ln k = Sm + \ln K_w$$

where m = the molality of the salt; $K_w = k$ when water is the mobile phase; S = parameter that varies with ion and solute type.

When these treatments are applied to SEC K_d values, the data should be linearized, simplifying the approximation of missing values. In DataMax, when data increase logarithmically, the position of data points with smaller values becomes compressed to the center of the graph. In preliminary application of the logarithm calculations, an important additional benefit was found. The MultiGraf "opened up"; many data points were moved out from the center of the map to group more closely with other objects of similar type, and the amount of information visualized was markedly increased. It should be kept in mind that linearization should only be applied to the attributes for which it is valid. In this case, log K_d and log (1/[sodium phosphate]) should be used for the lower-ionic-strength data. The log K_d and log [sodium phosphate] should be applied only to the higher-ionic-strength data. These topics will be discussed further in a following paper.

3.3. Raw SEC data

The characterization of solute interactions with the surface of a size-exclusion column packing is often carried out by an analysis of retention volumes of the sample molecules over a wide range of ionic strengths, pH, and mobile phase type [10–16]. These interactions were characterized for proteins eluting from a Zorbax BioSeries GF-250 column using sodium phosphate, pH 7.0 as a mobile phase. Sodium phosphate is a very common biological buffer and is the mobile phase currently suggested for use by the column manufacturer. The relative retention volumes (K_d) for the proteins at various mobile phase concentrations are shown in Table 1, along

Table 1
Relative retention volumes (K_d) for the 21 proteins and 4 low-molecular-mass markers

Sample name	p*I*	M_r	K_d of sample at listed concentration of sodium phosphate, pH 7.0								
			50 m*M*	70 m*M*	80 m*M*	100 m*M*	150 m*M*	200 m*M*	400 m*M*	600 m*M*	1000 m*M*
Thyroglobulin	5.1	669 000	0.54	0.54	0.54	0.54	0.54	0.53	0.55	0.55	0.60
Ferritin	4.2	440 000	0.67	0.63	0.64	0.65	0.64	0.64	0.67	0.73	0.88
Catalase	5.4	250 000	0.68	0.68	0.68	0.69	0.68	0.68	0.71	0.74	**1.54**
β-Amylase	4.8	200 000	0.64	0.63	0.63	0.64	0.64	0.64	0.65	0.66	0.78
Alcohol dehydrogenase	6.8	150 000	0.66	0.66	0.66	0.67	0.67	0.67	0.67	0.68	0.72
BSA dimer	5.4	132 000	0.64	0.63	0.64	0.64	0.64	0.64	0.65	0.66	0.72
BSA	5.1	66 430	0.70	0.70	0.70	0.70	0.70	0.70	0.72	0.73	0.79
Ovalbumin	4.6	44 000	0.75	0.74	0.75	0.75	0.75	0.75	0.78	0.79	0.89
Peroxidase	9.0	44 000	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.75	0.88
Pepsin	2.0	35 000	0.92	0.82	0.81	0.80	0.81	0.81	0.92	**1.16**	**1.70**
Carbonic anhydrase	5.9	29 000	0.91	0.85	0.85	0.83	0.82	0.82	0.83	0.85	0.97
Chymotrypsinogen A	9.3	25 000	**2.51**	**1.16**	**1.00**	0.88	0.84	0.83	0.83	0.90	**1.23**
Trypsinogen	9.3	23 977	0.87	0.84	0.83	0.82	0.80	0.82	0.83	0.85	0.89
Trypsin	10.5	23 280	0.93[a]	0.90	0.87	0.85	0.83	0.83	0.85	0.86	0.95
α-Chymotrypsin	8.8	21 600	**2.10**	**1.09**	0.96	0.86	0.82	0.82	0.83	0.84	0.92
Myoglobin	7.2	17 600	0.82	0.82	0.82	0.82	0.82	0.82	0.83	0.83	0.87
Lysozyme	10.0	14 300	**2.10**[a]	**1.59**	**1.33**	**1.02**	0.94	0.95	0.95	**1.00**	**1.38**
Ribonuclease A	7.8	13 700	**4.25**[a]	**2.74**	**1.82**	**1.78**	0.96	0.85	0.86	0.87	0.91
Cytochrome *c*	9.6	12 400	**6.47**[a]	**4.90**[a]	**4.26**[a]	**3.72**	**1.20**	0.92	0.85	0.84	0.85
Aprotinin	10.5	6 500	**2.75**[a]	**2.18**	**1.72**	**1.11**	0.96	0.94	0.93	0.96	**1.10**
Insulin	5.7	6 000	0.97	0.97	0.97	0.95	0.96	0.97	**1.09**	**1.34**	**2.92**
Vitamin B_{12}	–	1 355	0.95	0.96	0.95	0.95	0.96	0.99	**1.04**	**1.19**	**1.44**
Uridine	–	244	0.95	0.99	0.95	0.94	0.95	0.96	0.97	0.97	**1.03**
Uracil	–	122	0.99	0.99	0.99	0.98	0.99	**1.00**	**1.01**	**1.01**	**1.02**
Sodium azide	–	67	0.99	0.99	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.02**	**1.05**

Protein samples (1 mg/ml) were chromatographed on a GF-250 size-exclusion column at 2 ml/min. Numbers in bold type are retention volumes greater than V_R for sodium azide at 200 m*M* sodium phosphate, pH 7.0. All p*I* values and molecular masses were obtained from Sigma and Refs. [22] and [23].

[a] Values approximated by linear regression as described under Experimental.

with values for the low-molecular-mass markers. Although only one mobile phase type was used, the number of data points needed to perform an adequate characterization was quite large. However, by careful analysis of the resulting K_d values (Table 1), certain patterns can be recognized. At mobile phase concentrations of 70 to 100 m*M*, several proteins had a $K_d > 1$. In fact, trypsin, lysozyme, RNase A, cytochrome *c* and aprotinin were not detected, even at run times of 40 min. These proteins are known to have relatively high isoelectric points (p*I*) (p$I > 8$) and carry an overall positive charge at the pH of the mobile phase (pH 7.0). It is also possible to pick out proteins that elute with $K_d > 1$ at very high mobile phase concentrations —notably, catalase, pepsin, chymotrypsinogen A, lysozyme, aprotinin and insulin. Table 1 shows that the proteins elute at different retention volumes that change at dramatically different rates towards the extremes of high or low mobile phase ionic strength. K_d values were, therefore, estimated using regression analysis. These estimated values appear in bold type, followed with a superscript "a".

Although various characteristics can be surmised from the table, it is difficult to visualize, all-at-once, how they interrelate. If one views the data using traditional x–y plots of K_d versus mobile phase concentration, or as log M_r versus K_d, the result is numerous, slightly differing curves plotted over one another. Placing them in separate graphs makes comparison and presentation difficult. Individual graphs of log M_r plotted

against K_d (traditional calibration curves) are not quantitative and limit the comparison of data for different mobile phase strengths.

3.4. The DataMax statistical report

Table 2 is a statistical report exported directly from DataMax using the data in Table 1. The report first shows the amount of variance in the data accounted for in one, two and three dimensions of the MultiGraf. In this case, 97.729% of the total variance was accounted for in the three-dimensional map (Fig. 1). This indicates that the angles between the attributes (correlations between mobile phase, ionic strength data) add up to somewhat more than what can be represented in a three-dimensional sphere. The more correlated the attributes, the smaller the angles between the axes; hence, more axes (attributes) can be mapped in the same three-dimensional sphere.

In the second section of its report, DataMax lists the correlations between the attributes. This is reported as a matrix of correlations; one for each attribute calculated against every other. In SEC, one expects the elution volume to increase as log M_r decreases (negative correlation). Therefore, the mobile phase concentration with the greatest negative value represents the best mobile phase in this study for ideal SEC. It can be seen in the table "Correlations between attributes" section of Table 2 that 200 mM sodium phosphate has the highest negative correlation (-0.88914) with log M_r. Correlation values close to zero indicate attributes that are not correlated to each other.

The DataMax report continues with a table of the actual x, y and z coordinates (normalized between $+1$ and -1) for mapping the high end of the axes (attributes) on the sphere. The mean of each data set is placed at the center of the MultiGraf. At the right middle of the DataMax report, communalities are listed for the first three factors of the factor analysis. These values indicate the variance of each attribute accounted for in each of the 1st, 2nd and 3rd dimensions of the graph. At the far right appear the totals of variance accounted for in all three dimensions; as these totals approach 1, the MultiGraf better depicts the entire data set.

Each of the factors represents a characteristic of the system. It can be seen in this section of Table 2 that the middle mobile phase concentrations (150, 200 and 400 mM) make up a large percentage of the first factor. The second factor is largely accounted for by characteristics of the 50, 70, 80, 100, 400 and 600 mM mobile phase concentrations. The third factor is almost exclusively made up by the characteristics of the 1000 mM mobile phase concentration. More than 97% of the system characteristics were attributable to these first three factors. SEC systems are well defined; therefore, it seems appropriate to assign actual characteristics to these three factors. The first factor is most likely related to separation of the molecules by size (hydrodynamic volume); it occurred at medium ion strengths and correlated most negatively with log M_r. The second factor is best described by ion exchange of solute with the column packing; it occurred at high and low ionic strengths (except where SEC dominated). The third factor is strongly associated with hydrophobic interaction of solute with the column packing; it occurred only at the highest ionic strength used.

At the end of the report, DataMax lists the actual x, y and z coordinates for placement of the objects within the three-dimensional sphere (MultiGraf). When these objects are mapped, they appear perpendicular from their proper, relative position on each axis. As a result of this type plotting, the attributes and objects are pushed out into three dimensions, and group into different regions of the sphere.

3.5. The DataMax MultiGraf

The DataMax MultiGraf of chromatographic data from Table 1 is shown in Fig. 1. The MultiGraf may be rotated in any direction. In Fig. 1, the sphere graph has been rotated so that the log M_r axis runs vertically through the center of the sphere. The label for the attribute is always near the high end of the axis. (Larger molecular mass values are toward the top of the graph, and lower masses are near the bottom.) It

Table 2
DataMax statistics and plotting parameters generated from Table 1

Correlations between attributes:

	50 m*M*	70 m*M*	80 m*M*	100 m*M*	150 m*M*	200 m*M*	400 m*M*	600 m*M*	1000 m*M*	Log M_r
50 m*M*	1.0000	0.9721	0.9385	0.9188	0.6676	0.3569	0.2078	0.1059	−0.0494	−0.1413
70 m*M*	0.9721	1.0000	0.9866	0.9684	0.7156	0.3979	0.2472	0.1361	−0.0351	−0.1939
80 m*M*	0.9385	0.9866	1.0000	0.9852	0.7279	0.4103	0.2588	0.1505	−0.0191	−0.2151
100 m*M*	0.9188	0.9684	0.9852	1.0000	0.7183	0.3897	0.2475	0.1429	−0.0247	−0.2250
150 m*M*	0.6676	0.7156	0.7279	0.7183	1.0000	0.9191	0.8335	0.6888	0.3257	−0.7708
200 m*M*	0.3569	0.3979	0.4103	0.3897	0.9191	1.0000	0.9662	0.8358	0.4489	−0.8894
400 m*M*	0.2078	0.2472	0.2588	0.2475	0.8335	0.9662	1.0000	0.9406	0.6174	−0.8488
600 m*M*	0.1059	0.1361	0.1505	0.1429	0.6888	0.8358	0.9406	1.0000	0.7971	−0.6695
1000 m*M*	−0.0494	−0.0351	−0.0191	−0.0247	0.3257	0.4489	0.6174	0.7971	1.0000	−0.2277
Log M_r	−0.1413	−0.1939	−0.2151	−0.2250	−0.7708	−0.8894	−0.8488	−0.6695	−0.2277	1.0000

Attribute coordinates: *Communalities:*

Attribute name	*X*	*Y*	*Z*	1 D	2 D	3 D	Total
50 m*M*	0.13	0.79	−0.56	0.518	0.417	0.011	0.9461
70 m*M*	0.18	0.80	−0.57	0.583	0.402	0.007	0.9926
80 m*M*	0.19	0.79	−0.57	0.597	0.381	0.007	0.9844
100 m*M*	0.19	0.79	−0.55	0.580	0.379	0.005	0.9635
150 m*M*	0.78	0.36	−0.51	0.980	0.002	0.014	0.9963
200 m*M*	0.92	0.03	−0.37	0.777	0.169	0.037	0.9829
400 m*M*	0.88	−0.19	−0.43	0.653	0.337	0.002	0.9928
600 m*M*	0.70	−0.39	−0.58	0.487	0.431	0.052	0.9701
1000 m*M*	0.22	−0.61	−0.74	0.146	0.365	0.468	0.9783
Log M_r	−0.98	−0.02	0.01	0.492	0.220	0.254	0.9657

Object coordinates:

Data point name	*X*	*Y*	*Z*
Thyroglobulin	−0.414	−0.031	0.211
Ferritin	−0.290	0.050	0.072
Catalase	−0.292	0.165	−0.107
β-Amylase	−0.260	0.004	0.152
Alcohol dehydrogenase	−0.211	−0.012	0.169
BSA dimer	−0.236	−0.012	0.183
BSA	−0.141	0.004	0.154
Ovalbumin	−0.071	0.028	0.112
Peroxidase	−0.106	0.014	0.132
Pepsin	−0.021	0.280	−0.246
Carbonic anhydrase	0.008	0.032	0.069
α-Chymotrypsin	−0.025	0.010	−0.085
Trypsin	0.020	0.021	0.101
Trypsinogen	0.033	0.025	0.073
Chymotrypsinogen A	0.005	−0.043	0.045
Myoglobin	0.039	0.010	0.126
Lysozyme	0.113	0.027	−0.166
Ribobuclease A	0.016	−0.329	−0.169
Cytochrome *c*	−0.002	−0.860	−0.511
Aprotinin	0.143	−0.131	−0.114
Insulin	0.114	0.521	−0.615
Vitamin B_{12}	0.310	0.181	−0.086
Uridine	0.362	0.017	0.158
Uracil	0.439	0.014	0.167
Sodium azide	0.466	0.014	0.175

Total variance accounted for: in one dimension: 58.134%; in two dimensions: 89.158%; in three dimensions: 97.729%.

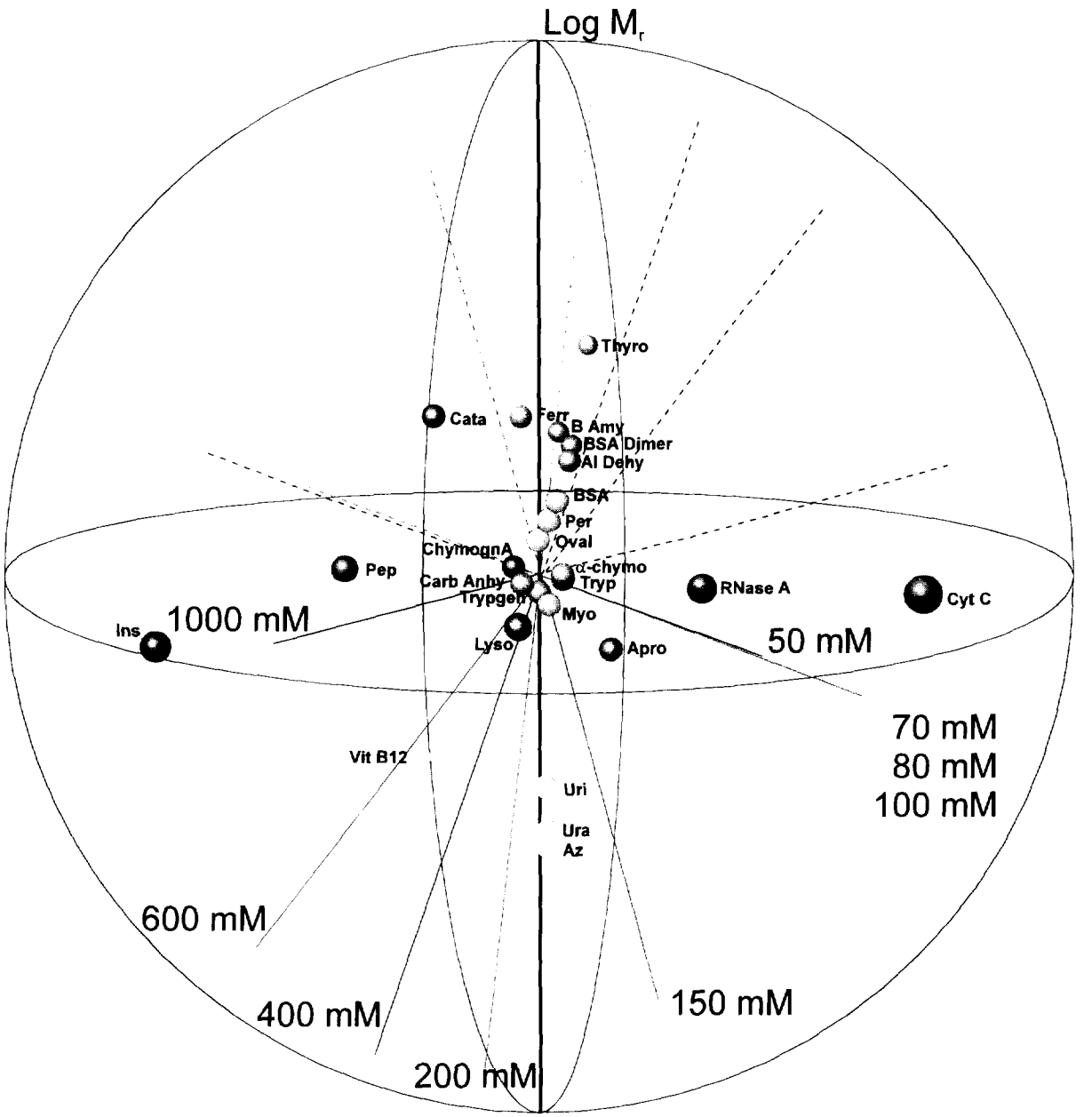

Fig. 1. Analysis of protein–packing interactions in SEC by multivariate non-orthogonal plotting. DataMax map of all relative retention volumes (K_d) for 21 proteins and 4 small molecules determined at 9 sodium phosphate concentrations (pH 7.0). Column: Zorbax GF-250 (250 × 9.4 mm), flow-rate 2 ml/min.

can be seen that the axis of the 200 m*M* mobile phase concentration lies nearly anti-parallel to that of the log M_r —with its low-value end tilted slightly back. Solid lines indicate that the end of the axis extends towards the front of the graph, while dashed lines indicate extension towards the back. This MultiGraf shows that each of the mobile phase concentrations are consistently

correlated with each other, from low concentrations, which are grouped tightly at the right, to the 1000 m*M* data on the left. The DataMax report was used to reduce these data for a graph that adequately represents only the three principal factors. The mobile phase concentrations chosen, by their correlations and communalities, were 100, 200 and 1000 m*M*. The resulting MultiGraf appears in Fig. 2. In the front view (Fig. 2a), it is seen that the majority of the proteins lie along the 200 m*M* and log M_r axes. In the side view (Fig. 2b) it is possible to see the deviations of the objects from these two axes. To account for the correlation between the attributes, these axes are placed at nearly 90° from each other and those of the 200 m*M* and log M_r. Again, data values that deviate from their relative position on the 200 m*M* axis are pushed out into the sphere at a point perpendicular to their relative value on each axis. The high retention times for cytochrome *c* and RNase A may be readily visualized in the MultiGraf. These large K_d values are attributable to electrostatic interactions. The non-ideal retention of insulin is also readily apparent, being mapped at the high end of the 1000 m*M* axis, indicating hydrophobic interactions. Two proteins, lysozyme and aprotinin, have a tendency towards both ionic and hydrophobic interactions. These objects appear between the 100 and 1000 m*M* axes. By comparing the front and side views of the MultiGraf (Fig. 2a and b), it can be seen that the majority of the proteins fall along the log M_r axis and are undergoing ideal size separation at these various conditions. It appears that larger molecules cannot sample as much of the packing surface as a small molecule. As a result, these larger molecules should have less interaction with the packing and exhibit little deviation from their proper elution volume.

3.6. Chromatographic examples

Through the use of the DataMax MultiGraf, three objects (proteins) were chosen to demonstrate their different interactions with the column packing at various mobile phase concentrations. By showing the actual chromatograms, one can more easily see the dramatic differences that change of mobile phase has on the chromatography of these proteins. Cytochrome *c* was positioned in Figs. 1 and 2 at the high end of the 100 m*M* axis, at the edge of the MultiGraf. This protein was chosen to show the elution characteristics of a molecule that undergoes electrostatic interactions with the column packing (Fig. 3). Cytochrome *c* has a high p*I* (Table 1) and is known to have positive charges at its surface under native conditions and pH 7.0 [21]. At low mobile phase ionic strengths, these cations can undergo ion exchange with silanol groups on the surface, causing the peak to be retained and broadened. Higher mobile phase ionic strengths tend to shield these interactions. Fig. 4 shows the elution characteristics of a relatively hydrophobic protein, insulin. This protein is very sensitive to hydrophobic interactions and maps on the edge of the MultiGraf at the high end of the 1000 m*M* axis. It is retained past the total permeation volume at mobile phase concentrations as low as 400 m*M* sodium phosphate, pH 7.0. Extensive band broadening is observed at this mobile phase concentration. Finally, Fig. 5 shows the elution profile for lysozyme, a protein that exhibits both electrostatic and hydrophobic tendencies [18]. This duality is made possible by discrete domains of the protein (e.g., certain folding, amino acids, and conformational changes, allow two adjacent areas of the protein surface to be completely different). It should be noted that the extents of these protein interactions with the column packing are relative. At lower ionic strengths, the interactions of all the proteins with the packing becomes exaggerated. In addition, these proteins can change their conformation as a result of mobile phase changes, and may consequently mask or display different surface characteristics.

4. Conclusions

DataMax allowed visualization of large amounts of SEC data, through principal-component analysis and non-orthogonal-plotting techniques (a MultiGraf). The relationships between log M_r and the retention volumes for all

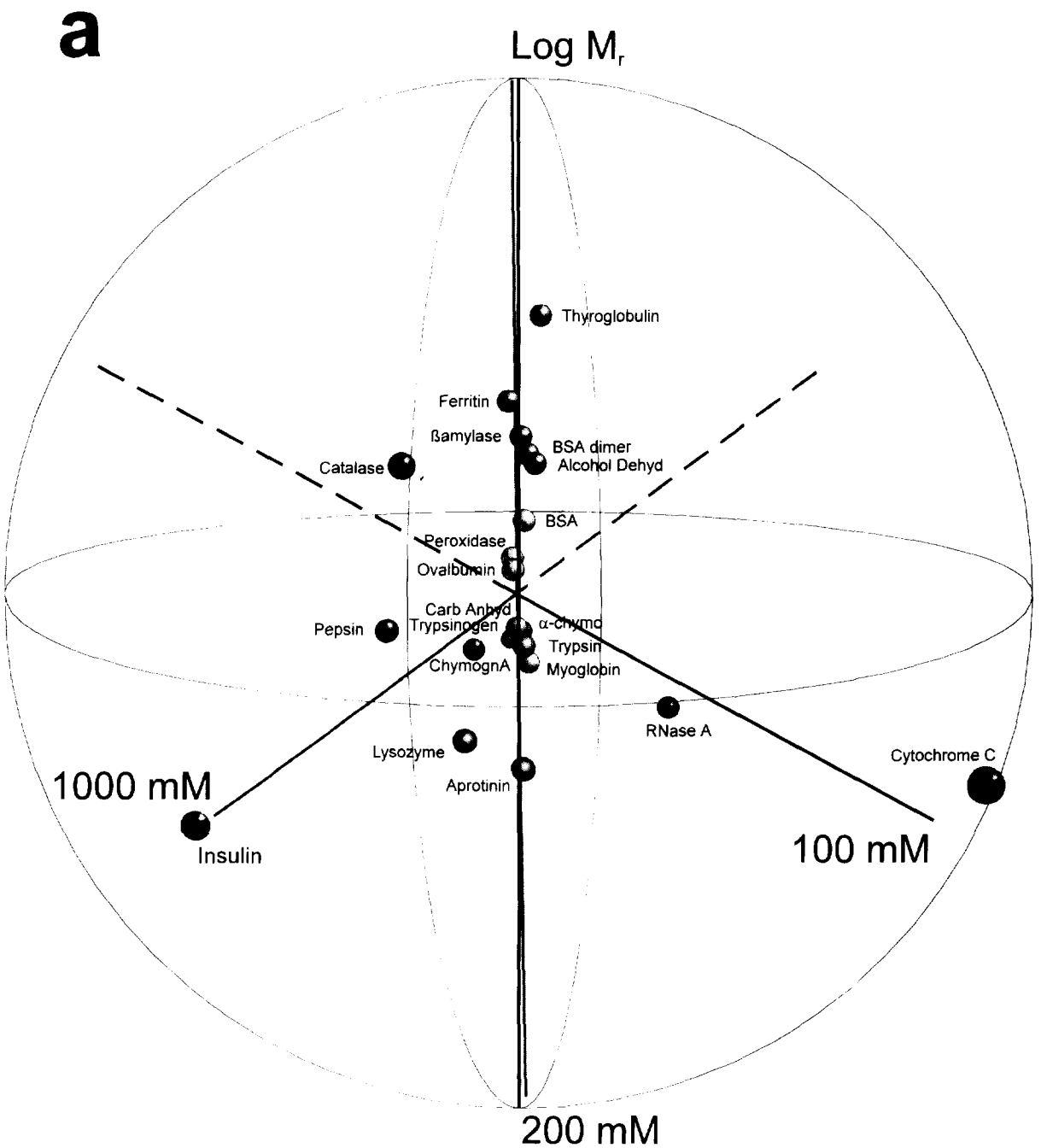

Fig. 2.

the proteins at different mobile phase concentrations were visualized in a single graph. As a result, a mobile phase could be selected for achieving the most-ideal SEC in this system —200 m*M* sodium phosphate (pH 7.0). DataMax also provided a means of viewing the grouping proteins by their relative hydrophobicities and ionic character, as determined by the

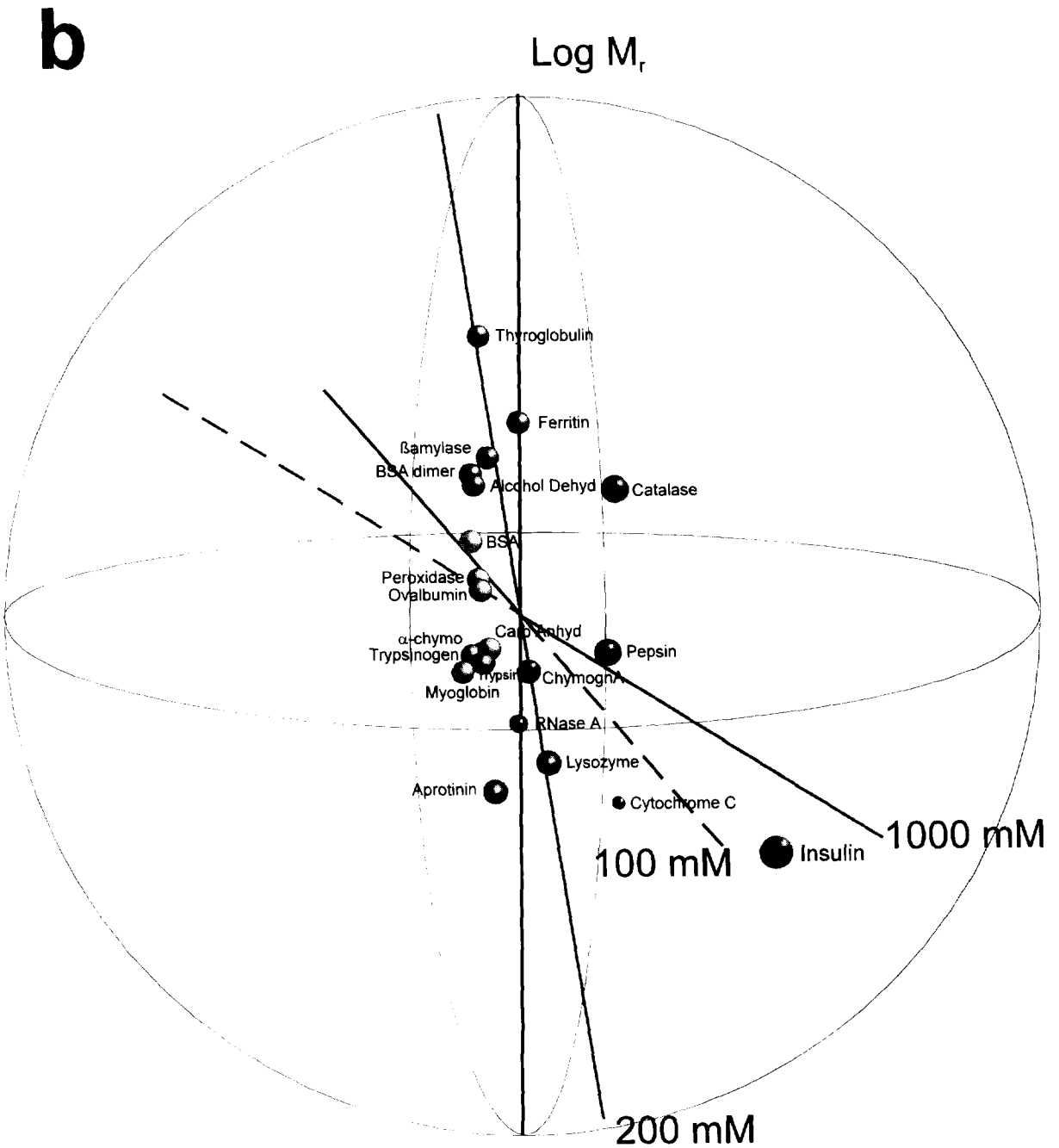

Fig. 2. Analysis of protein–packing interactions in SEC by multivariate non-orthogonal plotting. Reduction of data in Fig. 1 to emphasize the major attributes (mobile phase concentration) and objects (proteins). (a) Front view; (b) is (a) spun 90° around its log M_r axis.

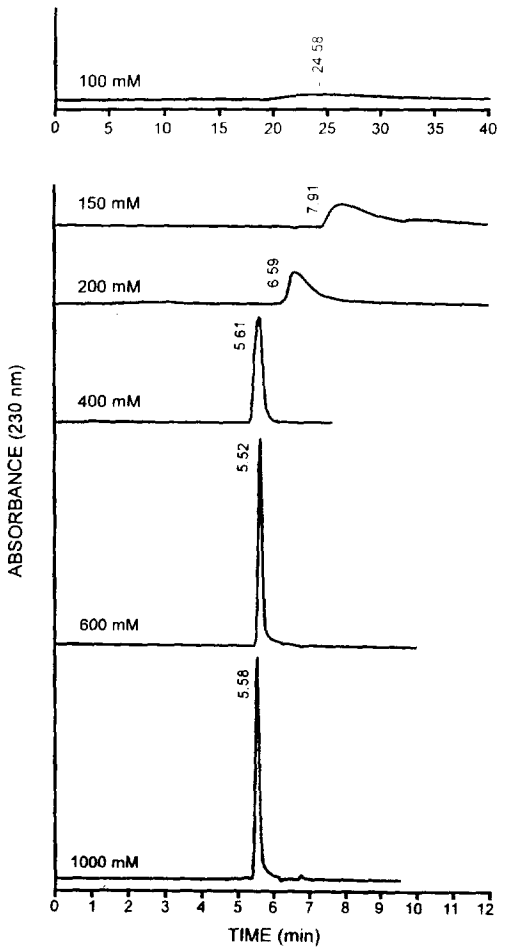

Fig. 3. Characteristic elution profiles of cytochrome *c*, identified in Figs. 1 and 2 to undergo electrostatic-type interactions. Actual chromatograms showing change in retention volume and peak shape. Conditions as in Fig. 1.

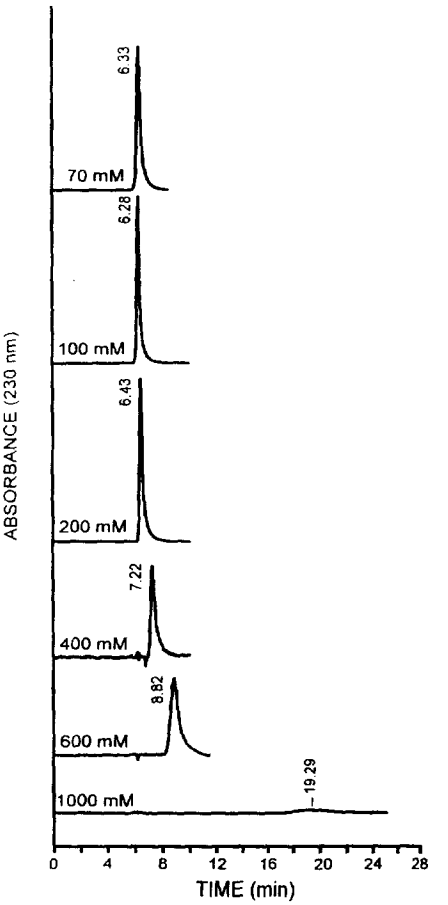

Fig. 4. Characteristic elution profiles of insulin, identified in Figs. 1 and 2 to undergo hydrophobic interactions. Actual chromatograms showing change in retention volume and peak shape. Conditions as in Fig. 1.

extent of their interaction with the column-packing surface. Proteins, including cytochrome *c* and ribonuclease A, grouped in a region of the MultiGraf that indicates electrostatic interactions as a major characteristic. Catalase, pepsin, chymotrypsinogen A, insulin and vitamin B_{12} mapped in an area indicating that they undergo hydrophobic interaction with the column packing. Aprotinin and lysozyme had a tendency towards both types of interaction, and were mapped between the hydrophobic and electrostatic objects on the MultiGraf. Principal-component analysis has been used in a similar way to study reversed-phase systems [24], but the MultiGraf allows improved visualization of the data. DataMax would be a useful aid in demonstrating the hydrophobic and ionic characteristics of an unknown protein. Alternatively, optimal mobile phase conditions could be suggested for a protein of known characteristics. We plan to extend these analyses in the future, providing a more rigorous analysis of non-ideal SEC. It is clear, however, that the ability to visualize *all* of the data from an experiment, through non-orthogonal plotting and techniques such as linearizing data through their logarithms, make DataMax a powerful chemometric tool, especially in SEC.

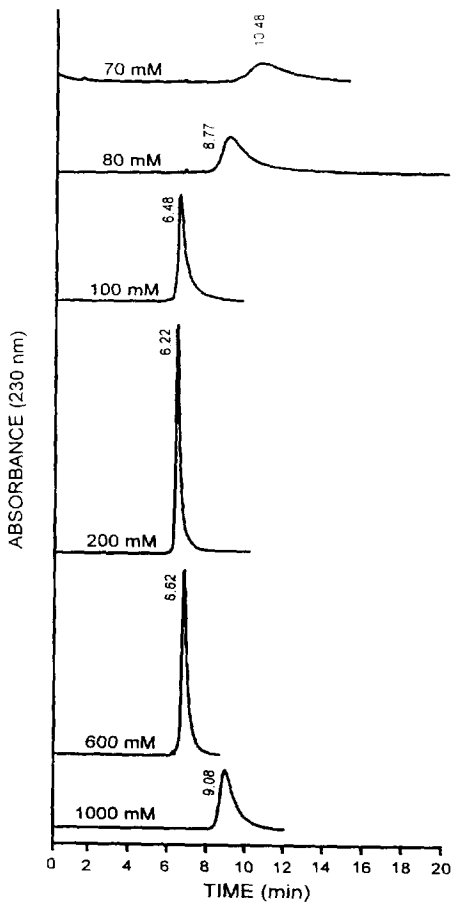

Fig. 5. Characteristic elution profiles of lysozyme. Lysozyme was identified in Figs. 1 and 2 to undergo both electrostatic and hydrophobic interactions. Actual chromatograms showing change in retention volume and peak shape. Conditions as in Fig. 1.

Acknowledgement

The authors would like to thank J.J. DeStefano for many helpful discussions and critical reading of the manuscript.

References

[1] J.J. Kirkland, C.H. Dilks, Jr. and J.J. DeStefano, *J. Chromatogr.*, 635 (1993) 19.
[2] J.J. Kirkland, C.H. Dilks, Jr. and J.E. Henderson, *LC·GC*, 11 (1993) 290.
[3] B.E. Boyes and J.J. Kirkland, *Peptide Res.*, 6 (1993) 249.
[4] J.J. Kirkland, *J. Chromatogr. Sci.*, 31 (1993) 493.
[5] L. Morin-Allory and B. Herbreteau, *J. Chromatogr.*, 590 (1992) 203.
[6] P.J. Bailey and B.G. Rohrback, *Food Technol.*, 48, No. 4 (1994) 69.
[7] F.O. Geiser, C. Golt, L. Kung, Jr., J.D. Justice and B.L. Brown, *J. Chromatogr.*, 631 (1993) 1.
[8] S.J. Schmitz, H. Zwanziger and H. Engelhardt, *J. Chromatogr.*, 544 (1991) 381.
[9] P. Karsnas and T. Lindblom, *J. Chromatogr.*, 599 (1992) 131.
[10] Y. Kato and T. Hashimoto, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 6 (1983) 324.
[11] W. Kopaciewicz and F.E. Regnier, *Anal. Biochem.*, 126 (1982) 8.
[12] M. Potschka, *J. Chromatogr.*, 441 (1988) 239.
[13] F. Hefti, *Anal. Biochem.*, 121 (1982) 378.
[14] M. LeMaire, A. Ghazi, J.V. Moller and L.P. Aggerbeck, *Biochem. J.*, 243 (1987) 399.
[15] M. LeMaire, L.P. Aggerbeck, C. Monteilhet, J.P. Andersen and J.V. Moller, *Anal. Biochem.*, 154 (1986) 525.
[16] P.L. Dubin and J.M. Principi, *Macromolecules*, 22 (1989) 1891.
[17] H.G. Barth, B.E. Boyes and C. Jackson, *Anal. Chem.*, 66 (1994) 595R.
[18] R.W. Stout, S.I. Sivakoff, R.D. Ricker, H.C. Palmer, M.A. Jackson and T.J. Odiorne, *J. Chromatogr.*, 352 (1986) 381.
[19] W. Kopaciewicz, M.A. Rounds, J. Fausnaugh and F.E. Regnier, *J. Chromatogr.*, 266 (1983) 3.
[20] G. Rippel and L. Szepesy, *J. Chromatogr. A*, 664 (1994) 27
[21] O.-W. Reif, V. Nier, U. Bahr and R. Freitag, *J. Chromatogr. A*, 664 (1994) 13.
[22] P.G. Righetti and C. Tiziana, *J. Chromatogr.*, 127 (1976) 1.
[23] D.W. Darnell, in G.D. Fasman (Editor), *CRC Practical Handbook of Biochemistry and Molecular Biology*, CRC Press, Boca Raton, FL, 1989, p. 158.
[24] M.C. Pietrogrande, M.I. Turnes Carou and F. Dondi, *Analusis*, 20 (1992) 111.