Journal of Chromatography, 605 (1992) 205–213
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 222

Analytical and micropreparative high-performance gel chromatography of proteins with a short column

Determination of molecular size, rapid determination of ligand binding constant and purification of S-carboxymethylated proteins for microsequencing

Hiromasa Tojo

Department of Molecular Physiological Chemistry, Osaka University Medical School, 2–2 Yamadaoka, Suita, Osaka 565 (Japan)

Kihachiro Horiike, Tetsuo Ishida, Takehiro Kobayashi and Mitsuhiro Nozaki

Department of Biochemistry, Shiga University of Medical Science, Seta, Ohtsu, Shiga 520-21 (Japan)

Mitsuhiro Okamoto

Department of Molecular Physiological Chemistry, Osaka University Medical School, 2–2 Yamadaoka, Suita, Osaka 565 (Japan)

(First received January 10th, 1992; revised manuscript received March 30th, 1992)

ABSTRACT

A short (75-mm) column of TSK-GEL G3000SW (S-column) can be used for both analytical and micropreparative gel chromatography. The elution conditions for the S-column were optimized and thereafter the column was well calibrated under native and denaturing conditions in a manner similar to that with a 600-mm column. The error in the Stokes radii of proteins obtained with the S-column was treated statistically. This S-column has the advantage of the economy in the time taken for analysis and solvent exchange and in the sample amounts required. Hence, it is useful for studying protein–ligand interactions by Hummel–Dreyer gel chromatography and for purifying S-carboxymethylated proteins prepared for microsequencing. These were exemplified by the analysis of the binding of *o*-nitrophenol to catechol 2,3-dioxygenase and of the NH_2-terminal amino acid sequence of rat pancreatic prophospholipase A_2.

INTRODUCTION

High-performance gel chromatography with porous silica- and polymer-based supports has been used to determine molecular weights and Stokes radii, as under suitable conditions the column functions in the size-exclusion mode, and the advantages of higher resolution and speed compared with the usual soft gel columns, have been achieved [1–19]. Because of the superior resolving power of the column, it is possible to perform gel chromatography with a shorter column. Although the reduction in column length more or less sacrifices resolution, the use of a short column may be preferable,

Correspondence to: Dr. Hiromasa Tojo, Department of Molecular Physiological Chemistry, Osaka University Medical School, 2–2 Yamadaoka, Suita, Osaka 565, Japan.

0021-9673/92/$05.00

particularly when economy in time taken for analyses and solvent exchanges and in the sample and solvent amounts required is important, and when dilution of solutes due to axial dispersion should be minimized. These requirements are encountered, for example, in the analysis of protein–protein interactions by gel chromatography [15].

In this paper, we report that a short (75-mm) column of TSK-GEL G3000SW (S-column), which is a commercially available precolumn, has a resolving power sufficient for the determination of molecular weights and Stokes radii of purified proteins using native and denaturing solvent systems. We experimentally determined the random error in gel chromatography with an S-column and examined its effect on the calculated parameters on the basis of a statistical evaluation. Additional utility of the S-column was illustrated by the analysis of protein–ligand interactions by Hummel–Dreyer gel chromatography [20,21] and by the purification of small amounts of reduced and S-alkylated proteins.

EXPERIMENTAL

Materials

The following standards for molecular weight and size determinations were used: cytochrome *c* (horse heart), α-chymotrypsinogen A (bovine pancreas), lactate dehydrogenase (bovine heart), alcohol dehydrogenase (yeast), ovalbumin (egg white), albumin (bovine serum) and urease (jack bean) from Sigma, aldolase (rabbit muscle) from Boehringer and ribonuclease A (bovine pancreas), catalase (bovine liver), ferritin (horse spleen) and blue dextran 2000 from Pharmacia.

Guanidine hydrochloride (GuHCl), sodium dodecyl sulphate (SDS) and sodium iodoacetate of protein chemical grade were obtained from Nacalai Tesque. Lithium dodecyl sulphate (LDS) and *o*-nitrophenol were products of Wako and the latter was recrystallized from ethanol. All other chemicals were of analytical-reagent grade.

Crystalline catechol 2,3-dioxygenase was prepared from *Escherichia coli* W3110 carrying its gene from TOL plasmid according to a previously published method [22], except that Q-Sepharose was used instead of DEAE-cellulose.

Rat pancreatic active phospholipase A_2 and prophospholipase A_2 were purified as reported [23]. To the pancreas homogenate was added 2 m*M* diisopropyl fluorophosphate to prevent activation of the proenzyme.

Small-zone high-performance gel chromatography

Gel chromatography was performed with a Gilson Model 302 liquid delivery module at 25°C. Columns of 600 mm × 7.5 mm I.D. (L-column) and short columns of 75 mm × 7.5 mm I.D. (S-column) of TSK-GEL G3000SW were obtained from Tosoh. The eluents used were (I) 0.05 *M* sodium phosphate containing 0.3 *M* NaCl (pH 7.0), (II) 0.1 *M* sodium pyrophosphate containing 0.3 *M* NaCl (pH 8.3), (III) 6 *M* GuHCl containing 10 m*M* sodium phosphate and 1 m*M* EDTA (pH 6.5) and (IV) 0.1 *M* ammonium acetate (pH 6.8) containing various concentrations of LDS. Protein samples for the experiments with 6 *M* GuHCl and LDS were dissolved in 7 *M* GuHCl (pH 8.5) containing 10 m*M* DL-dithiothreitol and 0.1 *M* Tris–HCl, and 0.1 *M* ammonium acetate containing 0.1% LDS, respectively; otherwise, they were dissolved in the same buffer as for eluent I or II. The effluents were continuously monitored with a Gilson Model 116 detector at 280 and 210/225 nm and/or with a Shodex RI SE-11 differential refractometer.

The distribution coefficients, K_d, of proteins were calculated from the equation $K_d = (V_e - V_0)/(V_t - V_0)$, where V_e, V_0 and V_t are the retention times corresponding to the elution volume of a protein, the void volume and the total available volume, respectively. The V_0 and V_t values were determined from the retention times of blue dextran and water, respectively (see Results and Discussion).

The chromatographic data were analysed by the equation of Ackers [24]:

$$R_e = a_1 + b_1 \mathrm{erf}^{-1}(1 - K_d) \quad (1)$$

where R_e is the Stokes radius, $\mathrm{erf}^{-1}(1 - K_d)$ is the inverse error function of $1 - K_d$ and a_1 and b_1 are empirical constants. The R_e values and molecular weights were taken from the literature [14,25]. The R_e values used for randomly coiled linear polypeptides in 6 *M* GuHCl are the viscosity-based Stokes radii and those for native proteins are the frictional Stokes radii. For native globular proteins the frictional R_e value is virtually identical with the viscosity R_e value over the range of molecular size

used in this study [14,26], and the values for the former are more available [14,18].

In order to relate K_d to the molecular weight (M_r) of a protein, we used eqn. 2 for proteins in 6 *M* GuHCl, eqn. 3 for globular proteins and eqn. 4 for proteins in 0.1% LDS, which were obtained from eqn. 1 based on the relationships between R_e and M_r [25]:

$$M_r^{0.555} = a_2 + b_2\text{erf}^{-1}(1 - K_d) \quad (2)$$

$$M_r^{1/3} = a_3 + b_3\text{erf}^{-1}(1 - K_d) \quad (3)$$

$$M_r^{0.73} = a_4 + b_4\text{erf}^{-1}(1 - K_d) \quad (4)$$

where a_2, b_2, a_3, b_3, a_4 and b_4 are empirical constants.

Hummel–Dreyer gel chromatography

In order to obtain directly the binding curve of *o*-nitrophenol for catechol 2,3-dioxygenase, Hummel–Dreyer gel chromatography [20,21] was performed at 25°C on an S-column with a Tosoh CCPM computer-controlled dual-pump system. The eluents were *o*-nitrophenol solutions of constant concentrations (2.0–96.6 μM), the ligand being dissolved in 0.05 *M* sodium phosphate buffer (pH 7.5). The flow-rate was 0.5 ml/min and the absorbance at 410 nm of the effluents was continuously monitored with a Tosoh UV-8000 detector. The injection volume of the mixtures of catechol 2,3-dioxygenase and *o*-nitrophenol was 25 μl and the total concentration of the enzyme as subunit was kept constant.

Preparation of S-carboxymethylated phospholipase A_2

Purified prophospholipase A_2 or phospholipase A_2 (150 pmol each) was reduced under nitrogen with tributylphosphine (150 nmol) in propanol–0.5 *M* ammonium hydrogencarbonate (pH 8.5) (1:1) for 2 h at room temperature [27], then iodoacetate (900 nmol) in 1 *M* ammonium hydrogencarbonate (pH 8.5) was added. The mixture was incubated for 30 min at 37°C, then the reaction was stopped by the addition of 1 μl of 2-mercaptoethanol. The sample was dried *in vacuo* and then purified as described below.

Analytical procedures

Amino acid analysis was performed as reported previously [28]. The amino acid sequence was determined with an Applied Biosystems Model 470A gas-phase sequenator and a Model 120A PTH analyser.

RESULTS AND DISCUSSION

Confirmation of convenient determination of V_t values by Himmel and Squire

Low-molecular-weight markers, such as dinitrophenylalanine, usually employed for soft gels to determine the V_t value are inadequate for the G3000SW column, as the gels more or less interact with those molecules. Himmel and Squire [11] reported the convenient determination of V_t with water using an absorption detector, so this was first confirmed here.

When water was injected into the column, a derivative pattern, *i.e.*, a positive peak followed by a negative peak, was recorded with an absorption detector, owing to the difference between the refractive indices of the eluents and water. This signal was reproducibly obtained (Table I). We observed that

TABLE I

RANDOM ERRORS IN RELATIVE RETENTION TIMES AND DISTRIBUTION COEFFICIENTS OBTAINED WITH AN S-COLUMN

Eluent I was used and the flow-rate was 0.2 ml/min. The S.D. was calculated from the data for 8–10 determinations for each sample.

Sample	S.D. · 100		
	r_e	K_d[a]	$\text{erf}^{-1}(1 - K_d)$[b]
Blue dextran	0.20	–	–
Thyroglobulin	0.16	0.44	3.8
Ferritin	0.17	0.31	0.59
Serum albumin	0.23	0.42	0.50
α-Chymotrypsinogen A	0.17	0.30	0.29
Water	0.21	–	–

[a,b] These S.D. values were calculated with the equations

$$\text{S.D.}(K_d) = \left[\left(\frac{K_d}{r_i}\right)^2 \text{var}(r_t) + \left(\frac{1}{r_i}\right)^2 \text{var}(r_e) + \left(\frac{K_d}{r_i} - \frac{1}{r_i}\right)^2 \text{var}(r_0)\right]^{\frac{1}{2}}$$

$$\text{S.D.}[\text{erf}^{-1}(1 - K_d)] = 0.5\pi^{\frac{1}{2}}\text{S.D.}(K_d)\exp\{[\text{erf}^{-1}(1 - K_d)]^2\}$$

respectively, where r is the retention time relative to the $\bar{V}_t$ value (*e.g.*, $r_e = V_e/\bar{V}_t$, $r_i = V_i/\bar{V}_t$), V_i is the internal volume and var and S.D. are the variance and the standard deviation, respectively, of the quantity in parentheses.

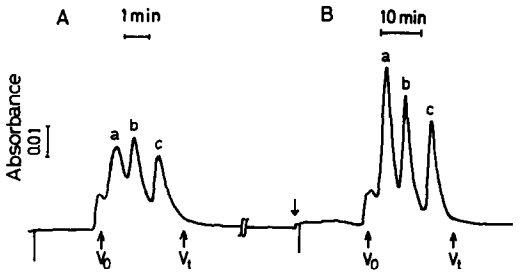

Fig. 1. Elution profiles obtained with the S-column system. Eluent I was used and the flow-rates were (A) 0.5 and (B) 0.08 ml/min. The protein samples used were (a) ferritin, (b) albumin and (c) α-chymotrypsinogen A. The negative deflections indicate the time at which the sample was injected. The flow-rate was changed at the time indicated by the arrow. The chart speeds were (A) 3 cm/min and (B) 10 cm/h.

the retention time of the latter trough obtained with a UV detector coincided with that of water recorded simultaneously with a differential refractometer with a deviation of only 0.2% when the distance between the two detectors connected in series was corrected for the retention times. This deviation was comparable to the random errors and corresponded to an error in K_d of 0.46%. This error was virtually negligible, as discussed below. Therefore, we conveniently determined the V_t value of the column with water, even using an absorption monitor.

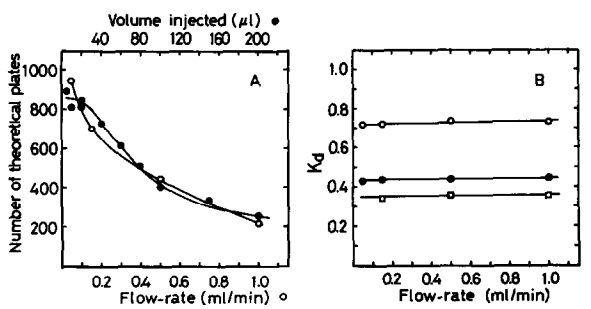

Fig. 2. Characterization of the S-column. Eluent I was used. (A) Effect of the flow-rate and the sample volume on the number of theoretical plates of the S-column with bovine serum albumin. Protein concentration, 3.3 mg/ml. (○) The flow-rate was changed and the sample volume injected remained constant at 10 μl; (●) the sample volume applied was changed at a constant flow-rate of 0.15 ml/min. (B) Dependence of the distribution coefficient, K_d, on flow-rate with the S-column system. The proteins used were (□) aldolase, (●) albumin and (○) α-chymotrypsinogen A. Each sample was injected in a volume of 10 μl.

Performance of the S-column

As Fig. 1 shows, this S-column showed satisfactory resolution for analytical gel chromatography. Its resolution depended on the flow-rate and the sample volume injected (Figs. 1 and 2A), whereas the resolution did not depend on the protein concentration of samples applied up to about 20 mg/ml (data not shown). The resolution increased when the injection volume was decreased, and was no longer improved in a volume of less than about 20 μl at a flow-rate of 0.15 ml/min (Fig. 2A). The S-column had *ca.* 1000 theoretical plates for bovine serum albumin with an injection volume of 10 μl at 0.05 ml/min. The resolving power decreased with an increase in the flow-rate (Fig. 2A). Flow-rates of 0.15–0.5 ml/min suffice for routine analysis, although resolution was sacrificed for a rapid experiment.

The K_d values showed hardly any flow-rate dependence over the range examined (Fig. 2B). These results are consistent with previous observations with an L-column [2]. However, Kakuno *et al.* [8] reported that slightly different V_e values were obtained at 1 and 0.1 ml/min with an L-column (300 mm). The reason for this discrepancy is unknown.

The observed retention times for a given sample were dependent on the injection volume. The time increased with an increase in the volume because it was measured from the moment at which the sample was just injected. This mainly reflected the lag time during which a sample was transferred from the sample loop to the column. Since with a volume below 10 μl the observed values are identical with those extrapolated to zero volume, within experimental error, the sample volume injected remained constant at 10 μl throughout the calibration of the column.

We tested four S-columns and all gave essentially the same K_d values and resolution. The K_d values for given proteins with the S-column were virtually identical with those obtained with an L-column, indicating the homogeneity of the gel matrices [29].

Once the detector baseline had stabilized in our S-column system, baseline drift was not observed even at the range of 0.05 full-scale absorbance when the flow-rate was abruptly changed (Fig. 1). Hence the column was equilibrated at 1.0 ml/min and after about 20 min the flow-rate was changed to the desired value for analysis. In this manner one buffer system was rapidly exchanged with another system.

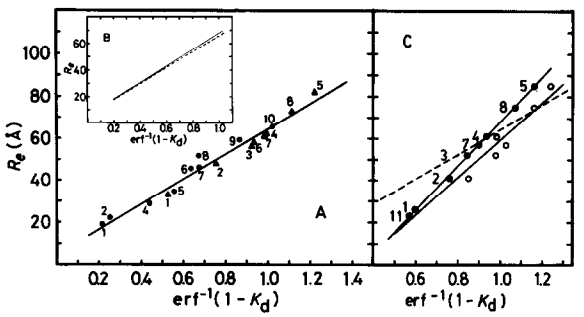

Fig. 3. Correlation between the distribution coefficient, K_d, and Stokes radius, R_e, of proteins with the S-column. Plots of R_e *vs.* $\text{erf}^{-1}(1 - K_d)$ for the data obtained under the following conditions; the pairs of values in parentheses are the a_1 and b_1 values in eqn. 1, respectively: (A) 0.15 ml/min, (●) eluent I and (▲) eluent III (4.89, 59.6); (B) dot-dashed line, 0.15 ml/min, eluent I (5.59, 60.3); solid line, 0.5 ml/min, eluent I (5.48, 62.3); dotted line, 0.15 ml/min, eluent II (pH 8.3) (7.42, 58.0); (C) 0.2 ml/min, (○) eluent IV containing 0.1% LDS (−29.5, 88.5) and (●) eluent IV containing 0.03% LDS (−37.6, 105). The dashed line in (C) is the same as the line in (A). As the calibration line for the L-column in eluent I at 1 ml/min (6.79, 58.0) almost coincided with that for the S-column in eluent II at 0.15 ml/min (dotted line), the line for the L-column has been omitted for clarity. The numbers indicate the following proteins: 1 = ribonuclease A; 2 = α-chymotrypsinogen A; 3 = lactate dehydrogenase; 4 = ovalbumin; 5 = albumin; 6 = alcohol dehydrogenase; 7 = aldolase; 8 = catalase; 9 = ferritin; 10 = urease; 11 = cytochrome *c*.

Calibration of the S-column

Fig. 3 shows plots of R_e *vs.* $\text{erf}^{-1}(1 - K_d)$ for the S-column. In phosphate buffer (eluent I) and 6 *M* GuHCl (eluent III), these plots were satisfactorily linear (correlation coefficients of 0.987 and 0.991, respectively, although non-linear plots have been reported [16,17]). It is likely that the viscosity-based Stokes radii in both media had the same correlation with K_d irrespective of the hydrodynamic shapes of the standards, so we combined these data and calculated the correlation coefficient to be 0.986 (Fig. 3A). These results were consistent with previous results [6,14,30]. Studies on the mechanism of the separation of macromolecules on gel filtration based on these observations have been reported [14,18]. Under various conditions as shown in Fig. 3A and B, the calibration lines approximately coincided with each other.

However, on chromatography of LDS–protein complexes the plots were almost linear, but did not coincide with those obtained from the combined data in eluents I and III (Fig. 3C). We used LDS in place of SDS because LDS is more soluble than SDS at low temperature, *i.e.*, LDS has a lower Krafft point. The a_1 and b_1 values in eqn. 1 were −29.5 and 88.5, respectively, for 0.1% LDS–gel chromatography. The large negative a_1 value compared with that on chromatography in eluents I and III is notable. These values are consistent with previous results with an L-column equilibrated with 0.1% SDS [7]. This indicated that the columns of TSK-GEL G3000SW do not necessarily function in the size-exclusion mode alone in the presence of the dodecyl sulphate salt [7]. On the other hand, Tanford *et al.* [26] reported that a calibration line with native proteins was superimposable on that with proteins reduced and S-alkylated in 0.1% SDS in gel chromatography on a soft gel (Sephadex G-200) column.

In 6 *M* GuHCl solution, all the proteins used are random coils [31], which have the same hydrodynamic shape, so we can reasonably estimate the molecular weights of the polypeptide chain using eqn. 2 [25]. As Fig. 4A shows, the plot was linear (correlation coefficient 0.991), in agreement with our previous results with an L-column [13].

On the other hand, in order to estimate the molecular weights of native proteins by gel chromatography alone, we must assume that both standards and unknowns have the same shape and partial specific volumes [25,29,32]. The calibration line for native proteins with eqn. 3 was almost linear (correlation coefficient 0.973, Fig. 4A), as the standards used were globular proteins. However, one should use this calibration line with caution in calculating M_r values of proteins of unknown

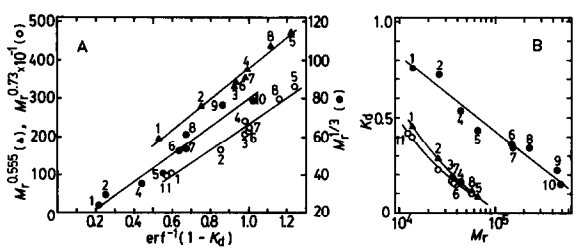

Fig. 4. Correlation between the distribution coefficient, K_d, and the molecular weight, M_r, with the S-column. (A) (▲) Plot of $M_r^{0.555}$ *vs.* $\text{erf}^{-1}(1 - K_d)$ for denatured proteins in 6 *M* GuHCl; (●) plot of $M_r^{1/3}$ *vs.* $\text{erf}^{-1}(1 - K_d)$ for native proteins; (○) plot of $M_r^{0.73}$ *vs.* $\text{erf}^{-1}(1 - K_d)$ for LDS–protein complexes in 0.1% LDS. (B) Plots of K_d *vs.* log M_r. Chromatographic conditions, symbols and numbers as in Fig. 3.

hydrodynamic shape, as has been stressed repeatedly [*e.g.*, 12,14,16–18,25,29,32].

Random errors in gel chromatography on the S-column

We evaluated the random errors in the retention times with several standard samples at a flow-rate of 0.2 ml/min (Table I). The errors were expressed as the standard deviation (S.D.) scaled in terms of a mean V_t value, $\bar{V}_t$. This uncertainty was reported to be mainly due to flow-rate variations [33]. These errors propagate to those in the K_d and $\text{erf}^{-1}(1 - K_d)$ values (Table I).

Although the S.D. values of relative retention time and K_d are almost constant, the error in $\text{erf}^{-1}(1 - K_d)$ showed an unequal distribution on non-linear transformation of K_d (Table I). We therefore performed weighted least-squares analysis with $1/\{\text{var}(R_e) + b_1^2\text{var}[\text{erf}^{-1}(1 - K_d)]\}$ as weights [34], where var is the variance of the quantity in parentheses, assuming that the error in R_e has a constant relative magnitude of 5% [35]. The second term in the denominator in the weights used comes from the random error in gel chromatography. The a_1 and b_1 values obtained without consideration of the second term ($a_1 = 6.32$ and $b_1 = 57.1$) coincided with those obtained with its consideration ($a_1 = 6.31$ and $b_1 = 57.1$). Consequently, the random errors in the raw data for the R_e *vs.* $\text{erf}^{-1}(1 - K_d)$ plot of Ackers do not expand to any significant extent throughout the whole range of K_d values examined.

When the R_e values of proteins with K_d values in the range 0.1–0.75 were estimated from the calibration line in Fig. 3A, the standard deviations of the estimated values were calculated to be at least in the range 0.7–2.0 Å from the following equation:

$$\text{S.D.}(R_e) = \{\text{var}(a_1) + 2\text{erf}^{-1}(1 - K_d)\,\text{cov}(a_1, b_1) + \text{var}(b_1)\text{erf}^{-1}(1 - K_d)^2 + b_1^2\text{var}[\text{erf}^{-1}(1 - K_d)]\}^{\frac{1}{2}} \quad (5)$$

where $\text{cov}(a_1, b_1)$ denotes the covariance between a_1 and b_1. In this instance, we considered the errors in $\text{erf}^{-1}(1 - K_d)$ for a protein in question and in the calibration constants a_1 and b_1, which can approximately be calculated from the weighted least-squares analysis with considering the errors in K_d and R_e for the standard proteins as described above. The uncertainty in $\text{erf}^{-1}(1 - K_d)$ propagates to the R_e value estimated. However, the resulting error in R_e was at most half of the error in R_e obtained by neglecting the random error in K_d. The total error in R_e determined by gel chromatography may depend on the uncertainty in R_e of standard proteins, the error introduced by the use of an empirical relationship between R_e and K_d and/or anomalous elution caused by solute–gel interactions in addition to the minor contribution of the random error in gel chromatography discussed above. This result was consistent with that of a systematic study on the precision of average molecular weights determined by time-based high-performance gel permeation chromatography [33].

Hummel–Dreyer gel chromatography: application to the binding system of catechol 2,3-dioxygenase and o-nitrophenol

Catechol 2,3-dioxygenase catalyses an extra-diol cleavage of catechol to form 2-hydroxymuconate semialdehyde [36]. The enzyme is a homotetramer with a subunit M_r of 35 211 based on the DNA sequence and iron content (one atom per subunit) [22,37]. In order to obtain directly the binding curve of *o*-nitrophenol, a competitive inhibitor [38], we performed Hummel–Dreyer gel chromatography [20,21] because, as described above, the S-column allowed sufficient resolution and rapid separation (Fig. 5).

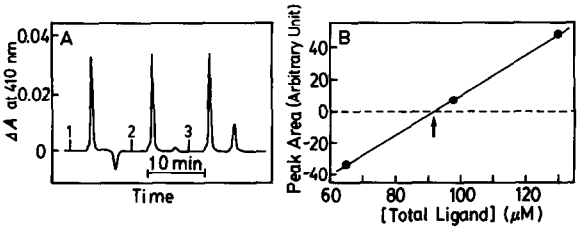

Fig. 5. Hummel–Dreyer gel chromatography for the binding system of catechol 2,3-dioxygenase and *o*-nitrophenol on an S-column at 25°C. The eluent was 14.7 μM *o*-nitrophenol in 50 mM sodium phosphate buffer (pH 7.5). The flow-rate was 0.5 ml/min. (A) Chromatograms. The samples (25 μl) containing various *o*-nitrophenol concentrations (1 = 65.0 μM; 2 = 97.5 μM; 3 = 130 μM) and 141 μM oxygenase per subunit were injected at the times indicated by the positive deflections. (B) Plot of ligand (*o*-nitrophenol) peak area, expressed as positive for the peak or negative for the trough, against the total ligand concentration. The arrow indicates the total *o*-nitrophenol concentration at which the area was zero.

When the free *o*-nitrophenol concentration in the equilibrium mixture of the enzyme and the ligand is equal to that in the eluent, no peak and trough are observed at the elution volume of the ligand on the gel chromatography of the mixture. As the corresponding total ligand concentration can be read from the point at which the area of the plot in Fig. 5B is zero [21], the value of fractional saturation can be easily calculated. To construct a binding curve, several experiments with different ligand concentrations in the eluent are necessary. This method is not time consuming, however, because the times taken for a single run at 0.5 ml/min and for solvent exchange at 1 ml/min were 9 and 20 min, respectively. The binding curve for the enzyme appeared to be nearly hyperbolic, and the dissociation constant was calculated to be 8.1 μM by applying our non-linear least-squares analysis [39].

Determination of the stoichiometry of ligand binding is a prerequisite to investigate the mechanism of protein–ligand interaction. For this, direct methods such as equilibrium dialysis, ultrafiltration [40] and Hummel–Dreyer chromatography [21] are preferable to indirect methods, *e.g.*, spectrometric titration, as the latter cannot detect the presence of the binding sites to which the ligand binding does not induce the spectral change. Further, only the former methods allow us to determine free, not total, ligand concentration directly.

Among the direct methods, the present Hummel–Dreyer method with S-column is more useful and convenient because it is simple, sensitive and not time or sample consuming. In methods using semipermeable membranes, adsorption of samples on the membrane must be considered.

Purification of reduced and S-alkylated proteins for microsequencing: application to phospholipase A_2

With the recent advent of the automated gas-phase sequenator, the amino acid sequences of a protein and its proteolytic digests have been frequently determined to design oligonucleotide probes for cloning of its cDNA. The preparation of reduced and S-alkylated proteins is usually an initial step in protein sequencing strategies. As reduced and S-alkylated proteins are very hydrophobic, special care must be taken to prevent loss due to non-specific adsorption to test-tubes and gels on removing unreacted reagents from a small amount of the polypeptides by column chromatography. Gel chromatography on an S-column in the presence of 0.03% of LDS was found to be suitable for such a purpose.

Takagi *et al.* [7] reported that the resolution of the L-column depended markedly on the salt concentration in the mobile phase containing SDS, and that good separations were obtained over the concentration range 0.05–0.15 M of sodium phosphate; we used 0.1 M ammonium acetate (pH 6.8) or 0.1 M sodium phosphate (pH 7.0), and both solvents were found to be equally effective for the S-column system. We first examined the effect of the LDS concentration on the elution behaviours of standard proteins and phospholipase A_2. With a decrease in the LDS concentration the retention times of S-carboxymethylated proteins increased (Fig. 3C). At LDS concentrations of less than 0.03% (1.1 mM), around the critical micelle concentration [41], the resolution became significantly poorer and the peak of phospholipase A_2 broadened. Therefore, an LDS concentration of at least 0.03% was required for good separation under the conditions employed. As a large excess of LDS may disturb the Edman degradation reaction, we adopted an LDS concentration of 0.03%. We selected a flow-rate of 0.2 ml/min with regard to resolution and speed from the results described above. At this flow-rate an eluted protein could be collected in a volume of 0.2–0.3 ml, and thereby the amount of LDS loaded on a sequenator (less than 0.1 mg) did not affect the Edman reaction.

This method was applied to determine the sequence of the N-terminal extrapeptide of rat pan-

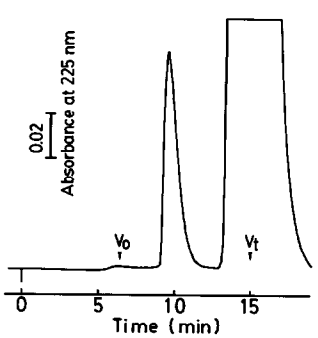

Fig. 6. Elution profile of S-carboxymethylated prophospholipase A_2 on an S-column in 0.03% LDS. The flow-rate was 0.2 ml/min.

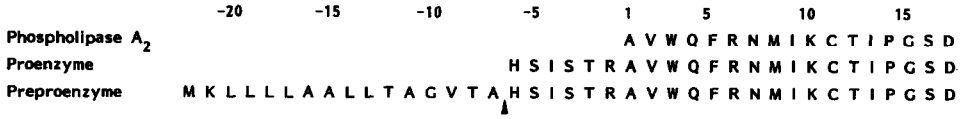

Fig. 7. NH_2-terminal amino acid sequences of rat pancreatic phospholipase A_2 and prophospholipase A_2. The arrowhead indicates a putative signal sequence cleavage site [42].

creatic prophospholipase A_2. The proenzyme was converted into the active enzyme through limited proteolysis of the NH_2-terminal extrapeptide. The phospholipase A_2, prophospholipase A_2 and standard proteins were reduced and S-carboxymethylated as described under Experimental. The dried samples containing a large amount of unreacted reagents were dissolved in 60 μl of 0.1 *M* ammonium acetate in the presence of 0.1% of LDS[a]. The resultant solution was then applied to an S-column as described above. The proenzyme was eluted at the position corresponding to an R_e value of 20.5 Å (Fig. 6). The eluted protein was collected in a 1.5-ml polypropylene tube. The recovery was about 80% as judged by amino acid analysis. The sample was evaporated *in vacuo* over phosphorus pentoxide. As residual ammonium ion may disturb the identification of phenylthiohydantoin-aspartic acid, the sample was further dried after the addition of 100 μl of water. When 0.1 *M* sodium phosphate was used as an elution buffer, this procedure was unnecessary. The dried sample was dissolved in 30 μl of water and then applied to a polybrene-treated glass filter and subjected to automated Edman analysis. The presence of LDS prevented adsorption of the protein on the polypropylene tube.

The NH_2-terminal amino acid sequences of rat pancreatic phospholipase A_2 and prophospholipase A_2 are shown in Fig. 7. The proenzyme has a six-residue long extrapeptide, HSISTR, compared with the active enzyme. The sequence was consistent with processing at a putative signal sequence cleavage site [42] in the amino acid sequence of the preproenzyme deduced from the nucleotide sequence of the phospholipse A_2 cDNA [43].

This method is an alternative to microbore reversed-phase high-performance liquid chromatography (HPLC) [28,44] in microsequencing strategies. Reversed-phase HPLC may be the most powerful technique for the isolation of proteins for microsequencing in many respects, including resolution, sensitivity, speed and the use of volatile organic solvents as mobile phase. Its major drawback, however, is the very low recovery of hydrophobic proteins such as membrane proteins and their low solubility in the organic mobile phase. As the present method utilizes a strong anionic detergent, it may be accommodated to the isolation of membrane proteins of a more hydrophobic nature; the LDS concentration could, if necessary, be adjusted so as to be optimum for the proteins. If proteins are separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis, they can be recovered by electroblotting on to membranes for direct sequencing [45,46]. Although this technique is very useful because of the high resolution of the electrophoresis, it is sometimes hampered by low transfer efficiencies [47]. When a higher resolution than in the present method is desired, S-alkylated proteins separated by 0.03% LDS gel chromatography on an L-column can be similarly handled for microsequencing.

CONCLUSIONS

The short column of TSK-GEL G3000SW is very useul for analytical and micropreparative gel chromatography for the determination of molecular size and weight, the rapid determination of ligand binding parameters and the purification of S-alkylated proteins prepared for microsequencing. Further, it is useful to employ the S-column for analysing the

[a] As the LDS micelles which incorporate low-M_r materials absorbing at 225 nm were eluted just later than the phospholipase A_2 ($M_r \approx 14\,000$), an excess of LDS and low-M_r contaminants may be included in the eluted protein solution when the sample containing an unnecessarily higher concentration of LDS (*e.g.*, 1%) was injected on to the column. A 0.1% LDS solution sufficed for solubilizing S-alkylated phospholipase A_2 and standard proteins used. This method is not suitable for separating proteins of M_r <10 000 because of interference from the LDS micelles eluted and peak broadening of such low-M_r proteins.

protein–protein interactions such as association–dissociation systems as reported previously, *viz.*, a self-associating system of hog kidney D-amino acid oxidase [15].

REFERENCES

1 K. Fukano, K. Komiya, H. Sasaki and T. Hashimoto, *J. Chromatogr.*, 166 (1978) 47.
2 F. E. Regnier and K. M. Gooding, *Anal. Biochem.*, 103 (1980) 1.
3 C. T. Wehr and S. R. Abbott, *J. Chromatogr.*, 185 (1979) 453.
4 N. Ui, *Anal. Biochem.*, 97 (1979) 65.
5 Y. Kato, K. Komiya, H. Sasaki and T. Hashimoto, *J. Chromatogr.*, 190 (1980) 297.
6 T. Imamura, K. Konishi, M. Yokoyama and K. Konishi, *J. Liq. Chromatogr.*, 4 (1981) 613.
7 T. Takagi, K. Takeda and T. Okuno, *J. Chromatogr.*, 208 (1981) 201.
8 T. Kakuno, H. Hiura, O. Ishikawa, M. Umino and Y. Kato, in T. Horio and J. Yamashita (Editors), *Tampakushitsu Koso no Kiso Jikkenho (Basic Experimental Methods for Protein and Enzyme Research)*, Nankodo, Tokyo, 1981, Ch. 2, p. 206.
9 R. A. Jenik and J. W. Porter, *Anal. Biochem.*, 111 (1981) 184.
10 R. C. Montelaro, M. West and C. J. Issel, *Anal. Biochem.*, 114 (1981) 398.
11 M. E. Himmel and P. G. Squire, *Int. J. Pept. Protein Res.*, 17 (1981) 365.
12 K. Horiike, H. Tojo, M. Iwaki, T. Yamano and M. Nozaki, *Biochem. Int.*, 4 (1982) 477.
13 H. Tojo, K. Horiike, K. Shiga, Y. Nishina, R. Miura, H. Watari and T. Yamano, *J. Biochem.*, 92 (1982) 1741.
14 K. Horiike, H. Tojo, T. Yamano and M. Nozaki, *J. Biochem.*, 93 (1983) 99.
15 H. Tojo, K. Horiike, K. Shiga, Y. Nishina, M. Nozaki, H. Watari and T. Yamano, *J. Biochem.*, 95 (1984) 1.
16 M. le Maire, L. P. Aggerbeck, C. Monteilhet, J. P. Andersen and J. V. Møller, *Anal. Biochem.*, 154 (1986) 525.
17 M. le Maire, A. Ghazi, J. V. Møller and L. P. Aggerbeck, *Biochem. J.*, 243 (1987) 399.
18 M. Potschka, *Anal. Biochem.*, 162 (1987) 47.
19 M. le Maire, A. Ghazi, M. Martin and F. Brochard, *J. Biochem.*, 106 (1989) 814.
20 J. P. Hummel and W. J. Dreyer, *Biochim. Biophys. Acta*, 63 (1962) 530.
21 J. M. Andreu, *Methods Enzymol.*, 117 (1985) 346.
22 C. Nakai, K. Hori, H. Kagamiyama, T. Nakazawa and M. Nozaki, *J. Biol. Chem.*, 258 (1983) 2916.
23 T. Ono, H. Tojo, K. Inoue, H. Kagamiyama, T. Yamano and M. Okamoto, *J. Biochem.*, 96 (1984) 785.
24 G. K. Ackers, *J. Biol. Chem.*, 242 (1967) 3237.
25 W. W. Fish, *Methods Membr. Biol.*, 4 (1975) 189.
26 C. Tanford, Y. Nozaki, J. A. Reynolds and S. Makino, *Biochemistry*, 13 (1974) 2369.
27 U. T. Rüegg and J. Rudinger, *Methods Enzymol.*, 47 (1977) 111.
28 H. Tojo, T. Ono, S. Kuramitsu, H. Kagamiyama and M. Okamoto, *J. Biol. Chem.*, 263 (1988) 5724.
29 G. K. Ackers, in H. Neurath and R. L. Hill (Editors), *The Proteins*, Vol. I, Academic Press, New York, 3rd ed., 1975, Ch. 1, p. 1.
30 W. W. Fish, J. A. Reynolds and C. Tanford, *J. Biol. Chem.*, 245 (1970) 5166.
31 S. Lapanje and C. Tanford, *J. Am. Chem. Soc.*, 89 (1967) 5030.
32 K. Kawahara, *Tanpakushitsu Kakusan Koso (Protein, Nucleic Acid and Enzyme)*, 18 (1973) 357.
33 L. Andersson, *J. Chromatogr.*, 216 (1981) 35.
34 W. E. Deming, *Statistical Adjustment of Data*, Wiley, New York, 1943 (Japanese edition, translated by S. Moriguchi, Iwanami Shoten, Tokyo), p. 93.
35 C. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, 1961, Ch. 6, p. 356.
36 M. Nozaki, *Top. Curr. Chem.*, 78 (1979) 145.
37 C. Nakai, H. Kagamiyama, M. Nozaki, T. Nakazawa, S. Inouye, Y. Ebina and A. Nakazawa, *J. Biol. Chem.*, 258 (1983) 2923.
38 K. Hori, T. Hashimoto and M. Nozaki, *J. Biochem.*, 74 (1973) 375.
39 H. Tojo, K. Horiike, K. Shiga, Y. Nishina, H. Watari and T. Yamano, *J. Biol. Chem.*, 260 (1985) 12607.
40 A. J. Sophianopoulos and J. A. Sophianopoulos, *Methods Enzymol.*, 117 (1985) 354.
41 K. Kubo and T. Takagi, *Anal. Biochem.*, 156 (1986) 11.
42 G. von Heijne, *Nucleic Acids Res.*, 14 (1986) 4683.
43 O. Ohara, M. Tamaki, E. Nakamura, Y. Tsuruta, Y. Fujii, M. Shin, H. Teraoka and M. Okamoto, *J. Biochem.*, 99 (1986) 733.
44 E. C. Nice, C. J. Lloyd and A. W. Burgess, *J. Chromatogr.*, 296 (1984) 153.
45 S. Yuen, M. W. Hunkapiller, K. J. Wilson and P. M. Yuan, *Anal. Biochem.*, 168 (1988) 5.
46 R. H. Aebersold, D. B. Teplow, L. E. Hood and S. B. H. Kent, *J. Biol. Chem.*, 261 (1986) 4229.
47 R. G. Feick and J. A. Shiozawa, *Anal. Biochem.*, 187 (1990) 205.