Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1089 (2005) 111–124
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Design of self-interaction chromatography as an analytical tool for predicting protein phase behavior

Tangir Ahamed, Marcel Ottens *, Gijs W.K. van Dedem, Luuk A.M. van der Wielen

Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands

Received 27 January 2005; received in revised form 15 June 2005; accepted 22 June 2005
Available online 11 July 2005

Abstract

Solution conditions under which proteins have a tendency to crystallize correspond to a slightly negative osmotic second virial coefficient (B_{22}). A positive B_{22} value guarantees no crystallization to occur. On the other hand, a B_{22} value within the so called "crystallization slot" thermodynamically supports the crystallization processes but does not guarantee successful crystal growth. It is, however, a prerequisite for protein crystallization that the B_{22} value must be in the slightly negative regime. Self-interaction chromatography (SIC) is designed in this work as an analytical tool for determining B_{22} in a precise and reproducible way. The methodology was demonstrated in detail in terms of its theoretical basis, experimental methodology, troubleshooting and data analysis for different protein samples and solution conditions. The inherent error limit of SIC is found to be comparatively less than other B_{22} measurement techniques. The designed experimental approach was applied for mapping crystallization conditions of a model protein, i.e. lysozyme. Good agreement between the obtained lysozyme B_{22} values and literature values confirms the accuracy of the approach.

Keywords: Self-interaction chromatography; Protein phase behavior; Systematic screening; Predictive crystallization; Lysozyme

1. Introduction

Protein crystallization is one of the critical aspects of structural biology and pharmaceutical biotechnology. In addition, crystallization is one of the demanded techniques in biotechnology and pharmaceutical industries for downstream processing of proteins. Numerous independent variables, including solvent conditions (pH, ionic strength, salt type, temperature, crystallizing agent, etc.) and initial protein concentration, are usually involved in nucleation and growth of protein crystals [1]. Most high-resolution protein structural information is obtained by X-ray diffraction, neutron crystallography or surface plasma resonance of protein crystals. The major obstacle in these processes is often to obtain a diffraction-quality crystal. Because of the involvement of several parameters and the lack of a systematic screening approach, optimum crystallization conditions are traditionally determined by empirical screening. Empirical screening provides neither any insight of crystallization thermodynamics nor any indication of how close the solution conditions were to the ones optimal for growing crystals. Consequently, the approach requires intensive screening of numerous solution conditions blindly and failure is often the case for many proteins, particularly membrane proteins and monoclonal antibodies. It is, therefore, highly desirable to develop high-throughput methods to determine conditions for protein crystallization in a rational manner, reducing the number of crystallization experiments, cost and time.

George and Wilson [2] observed that solution conditions under which proteins have a tendency to crystallize corre-

Abbreviations: B_{22}, osmotic second virial coefficient; B_{HS}, hard sphere/excluded volume contribution to B_{22}; FEC, frontal-exclusion chromatography; kDa, kilo dalton; LALLS, low-angle laser-light scattering; LS, light scattering; MO, membrane osmometry; M_w, molecular weight; NHS, *N*-hydroxysuccinimide; SANS, small-angle neutron scattering; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; SIC, self-interaction chromatography; SLS, static light scattering; *W*, potential of mean force

* Corresponding author. Tel.: +31 152782151; fax: +31 152782355.
E-mail address: m.ottens@tnw.tudelft.nl (M. Ottens).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.06.065

spond to a slightly negative osmotic second virial coefficient (B_{22}), resulting from weak attractive protein self-interactions. They correlated protein crystallization conditions with B_{22} values in the form of so called "crystallization slot" [3]. B_{22} is a thermodynamic parameter that reflects the magnitude and direction of deviations of a non-ideal solution from ideality. At the molecular level, B_{22} characterizes pairwise protein self-interactions including contributions from excluded volume, electrostatic interactions and short-range interactions [4]. According to the McMillan and Mayer [5] solution theory, B_{22} is correlated to the potential of mean force (W), which describes all kinds of possible interactions between two protein molecules in a dilute solution. A negative value of B_{22} indicates protein–protein attraction whereas a positive B_{22} value indicates mutual repulsion. The thermodynamic insight regarding the macromolecular interactions involved in B_{22} and why these interactions are related to protein crystallization were explained [4,6]. There is still doubt whether the crystallization slot guarantees successful crystal growth universally for all kinds of proteins. In this study, we review available literature in order to explore the relationship between B_{22} and phase behavior of different proteins. We also discuss the question whether B_{22} is the only thermodynamic parameter that governs the protein crystallization process.

The link between B_{22} and protein crystallization conditions offers the hope that screening B_{22} values may be useful for the predictive crystallization of proteins that are proven difficult to crystallize. In some cases, solution properties could also be pushed toward the crystallization slot. Nevertheless, there has been little use of B_{22} for predictive crystallization, largely because of the difficulties in B_{22} measurement. A considerable amount of modeling work has been done recently by calculating B_{22} and/or predicting protein phase behavior on a theoretical basis [4,6–10]. However, no model can universally be applied for all kinds of proteins regardless of their molecular mass, size, shape and nature. B_{22} is usually measured experimentally by colloidal characterization techniques, for instance static light [2–4,8,9,11–17], small-angle X-ray [18–20], laser-light [21,22] or neutron [9,23,24] scattering, membrane osmometry [25–27] and sedimentation equilibrium [28]. Unfortunately, all of these methods are too labor-intensive and expensive in terms of both protein and time to allow extensive screening. Moreover, B_{22} measurement by scattering techniques becomes extremely difficult when the solubility of the protein is low ($<5\,\mathrm{mg\,ml^{-1}}$). Another obstacle of the B_{22} aided protein crystallization approach is the inconsistency in B_{22} values measured by different techniques and/or different researchers for exactly the same protein sample and the same solution conditions. In this paper we also discuss possible reasons for this inherent inaccuracy in different techniques. It is, however, essential to have an easy-to-perform B_{22} measurement technique, which provides a precise result using a minimum amount of protein, time and effort.

Self-interaction chromatography (SIC) [29–35] and size-exclusion chromatography (SEC) [36] are two recently developed alternative methods of characterizing weak protein interactions that could potentially meet the requirements of being inexpensive in terms of both time and protein relative to other traditional techniques [37]. SIC measures the interaction of immobilized protein molecules in the stationary phase with free protein molecules in the mobile phase. The average retention of a protein pulse characterizes the protein–protein interaction and hence the value of B_{22}. On the other hand, SEC characterizes the thermodynamic non-ideality of a protein solution as a function of protein concentration. Negative B_{22} values correspond to decreased retention in a SEC column as a function of protein concentration, which consequently reflects the protein–protein attraction and vice versa. The advantage of SIC is that a single injection of dilute protein solution leads to a B_{22} value, whereas several injections of dilute to concentrated protein solution are required for SEC, which obviously costs more protein and time. On the other hand, SIC requires immobilization of protein to the stationary phase, which may sometime cause structural and conformational change of the immobilized protein. Comparatively speaking, the required experimental time and protein consumption in SIC is at least four to five times less than that of SEC. In addition, SIC is better suited to miniaturize the process to the microchip level, which would provide easy and fast screening of crystallization conditions. The SIC principle and methodology can also be used to study interactions among unlike proteins [38].

A systematic crystallization approach was successfully performed to produce diffraction quality crystals of chymotrypsinogen [27] and OmpF porin [12] through B_{22} mapping by static light scattering (SLS) and membrane osmometry (MO), respectively. The SIC approach has been applied so far for predictive crystallization of myoglobin [31] and ribonuclease A [32]. However, no study has been reported for B_{22} mapping and predictive crystallization of proteins that are structurally complex and were previously proven difficult to crystallize by empirical screening. In the present work, SIC methodology was designed with respect to its theoretical basis, experimental methodology, troubleshooting, data analysis and applicability for different kinds of proteins. The accuracy and precision limit of the optimized methodology was compared by reproducing B_{22} trends of the well-known protein, lysozyme.

2. Theory and design of SIC methodology

2.1. Inherent inaccuracy in B_{22} measurement

Lysozyme is the most widely studied protein in the field of B_{22}. We plotted the reported B_{22} values or trends among different literature sources for exactly the same or similar solution conditions (Fig. 1). The possible reason of these variations was suspected to be the source and purity of lysozyme, the employed measurement technique, minor deviations in the solution conditions or experimental errors. Hen

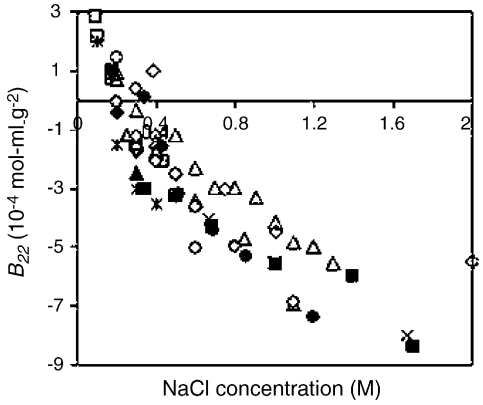

Fig. 1. B_{22} of lysozyme with varying NaCl concentration at pH 4.2–4.7 and temperature 20–25 °C. The corresponding experimental conditions, measurement technique, source of lysozyme and reference are shown in Table 1.

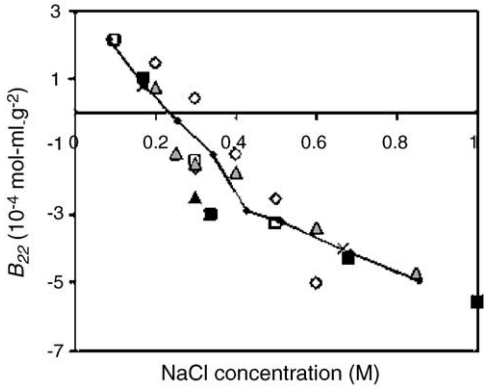

Fig. 2. B_{22} of lysozyme purchased from Sigma with varying NaCl concentration at pH 4.5 and temperature 23–25 °C. The line represents the outcome of this work. The meaning of the other symbols as well as experimental conditions, measurement technique, source of lysozyme and reference are shown in Table 1.

egg lysozyme from different sources is different in terms of purity. For instance lysozyme from Sigma contains more contaminant proteins (ovalbumin and albumin) than those from Seikagaku and Boehringer-Mannheim [15,21,23,36]. We therefore discriminated between different sources of lysozyme and plotted the cases where lysozyme was purchased from Sigma and B_{22} was measured at pH 4.5 and at 23–25 °C (Fig. 2). Fig. 2 still represents a variation of $\pm 1.5 \times 10^{-4}$ mol ml g^{-2} unit. We, therefore, conclude that because of the involvement of only weak protein interactions B_{22} measurement techniques have significant inherent inaccuracy. Virial coefficients are determined from chemical potential as the coefficients of a positive power series of protein concentration. The chemical potential of protein molecules in a given solvent is usually measured by osmotic pressure, light scattering (LS) or sedimentation behavior or retention times in chromatography. Minor inherent inaccuracy in these measuring parameters leads to a huge inaccuracy in the determined B_{22} values, as it is discussed below.

2.1.1. Membrane osmometry

For a dilute protein solution, the B_{22} is defined in terms of the osmotic pressure, Π, by the osmotic virial expansion [39]

$$\Pi = RTc_p \left(\frac{1}{M_w} + B_{22}c_p + \cdots \right) \quad (1)$$

where R is the universal gas constant, T is the temperature in Kelvin, c_p is the concentration of protein and M_w is the molecular weight of the protein. For the description of a dilute protein solution, three-body or higher order interactions are neglected from the virial expansion Eq. (1). Thus, a plot of Π/RTc_p versus c_p is linear for a sufficiently low range of c_p values, with the slope equal to B_{22} and the intercept equal to $1/M_w$. Alternatively, B_{22} can also be measured from the osmotic pressure data at a single point of protein concentration, if the molecular weight and aggregation state of the protein in that solvent condition are known.

Table 1
Solution conditions and references for Figs. 1 and 2

Data symbol	Protein source	pH	Temperature (°C)	Buffer	Measurement technique	Ref.
Black diamond	Sigma	4.5	20	100 mM Na-acetate	SAXS	[20]
Grey Ash	Sigma	4.5	25	100 mM Na-acetate	SAXS	[20]
White diamond	Sigma	4.5	N/A	N/A	LALLS	[21]
Black square	Sigma	4.5	25	N/A	LALLS	[22]
Grey square	Seikagaku	4.7	N/A	50 mM Na-acetate	SLS	[15]
White square	Sigma	4.5	25	Minimal citric acid	SLS	[9]
Black triangle	Sigma	4.5	25	Minimal citric acid	SANS	[9]
Grey triangle	Sigma	4.6	25	40 mM Na-acetate	SLS	[8]
White triangle	Seikagaku	4.7	25	25 mM Na-acetate	SLS	[14]
Black circle	Seikagaku	4.2	25	100 mM Na-acetate	SLS	[11]
Grey circle	Seikagaku	4.6	25	50 mM Na-acetate	SLS	[13]
White circle	Sigma	4.5	23	5 mM Na-Acetate	SIC	[30]
Cross	Sigma	4.5	25	20 mM Na-acetate	SIC	[35]
Star	Seikagaku	4.7	N/A	50 mM Na-acetate	SEC	[36]

N/A, not available; SAXS, small-angle X-ray scattering; LALLS, low-angle laser light scattering; SLS, static light scattering; SANS, small angle neutron scattering; SIC, self-interaction chromatography; SEC, size-exclusion chromatography.

The inherent inaccuracy of B_{22} values measured by MO comes from the inaccuracy in the osmotic pressure measurement and from the partial aggregation of the protein. Membrane osmometers of different manufacturers have an inherent inaccuracy of 0.5–1% in osmotic pressure measurement. According to the information provided by different suppliers, Type 3300 Micro Osmometer of John Morris Scientific and Advanced 3250 osmometer of Advanced Instruments, Inc. have a precision limit of 0.5%. Semi-Micro Osmometer Type Dig. L of KNAUER has a measurement accuracy of <1%. In the work of Haynes et al. [25], the average mean square error of the measured osmotic pressure at different solvent concentration and at different protein concentration was 0.94%. This apparently small inaccuracy in osmotic pressure measurement dramatically affects the B_{22} value. For a simple case of lysozyme (M_w 14,600 g mol^{-1}) at a concentration of 5 mg ml^{-1}, a B_{22} value 0.00×10^{-4} mol ml g^{-2} corresponds to an osmotic pressure of 849 Pa at 25 °C. Only a 1% inaccuracy in osmotic pressure measurement would shift the B_{22} value from 0.00×10^{-4} to $\pm 1.37 \times 10^{-4}$ mol ml g^{-2}.

The measured B_{22} value is more reliable, if the measurement is done from the slope of the plot of Π/RTc_p versus c_p at multipoint protein concentration without any preliminary knowledge of the molecular weight of the protein. However, an additional error may arise from the standard deviation of the slope. We have analyzed the data of Haynes et al. [25] and standard deviations of the slope corresponded to an error of $\pm 0.6 \times 10^{-4}$ mol ml g^{-2} in the calculated B_{22} values. Haynes et al. [25] also found an error margin of $\pm 0.4 \times 10^{-4}$ mol ml g^{-2} only due to the mean square error in osmotic pressure measurement. The overall error then becomes equal to $\pm 1.0 \times 10^{-4}$ mol ml g^{-2}. It is, therefore, concluded that B_{22} values measured by MO have an inherent inaccuracy of $\pm 1.0 \times 10^{-4}$ mol ml g^{-2}.

2.1.2. Light scattering

Macromolecular solutions scatter light due to the thermally induced fluctuation in local concentration. Using this property of protein solutions, light scattering techniques can be used to obtain the so called "static" parameters of a protein such as molecular weight, osmotic second virial coefficient, molecular dimensions and sometimes the radius of gyration. The static light scattering or low-angle laser-light scattering (LALLS) experiment measures the average intensity of light scattered by a protein solution of defined concentration in excess of that scattered by the background solvent. Measurement of B_{22} using this method relies on measuring the intensity of light scattered as a function of the protein concentration. Since protein molecules are usually much smaller than the wavelength of the incident light (<λ/20), their scattering intensity is independent of the scattering angle, in other words, within the Rayleigh limit. In this limit, the Rayleigh ratio, R_θ, is by definition proportional to the scattered light intensity and is related to the M_w and B_{22} by the classic equation [40,41]

$$\frac{Kc_p}{R_\theta} = \frac{1}{M_w} + 2B_{22}c_p \tag{2}$$

where K is an optical or instrumental constant, which can be given by

$$K = \frac{4\pi^2 n_0^2}{N_A \lambda^4}\left(\frac{dn}{dc_p}\right)^2 \tag{3}$$

where n_0 is the refractive index of the solvent, dn/dc_p is the refractive index increment for the protein–solvent pair, N_A is the Avogadro's number and λ is the wavelength of the incident vertically polarized light in vacuum.

Eq. (2) indicates that a plot of Kc_p/R_θ versus c_p allows the determination of M_w and B_{22}. The accuracy of the determined B_{22} values however depends on the measurement accuracy of R_θ, n_0 and dn/dc_p values as well as the linearity of the plot. Since light scattering in the Rayleigh limit is isotropic, the values of R_θ in SLS are usually measured at 90°. In order to determine the absolute R_{90} values for protein solutions, the SLS instrument is calibrated first using toluene or benzene. Toluene and benzene have an established R_{90} value of 1.406×10^{-5} cm^{-1} at 633 nm [8,15,17] and 3.86×10^{-5} cm^{-1} at 488 nm [4,9], respectively. The background scattering of the pure solvent alone is then subtracted first in order to measure the actual increment of scattering due to protein.

Solvents in B_{22} determination system are usually salt/buffer solutions, whose refractive index, n_0, can be determined by a refractometer. The term dn/dc_p is the change of the solution's refractive index with respect to a change in protein concentration, which can be measured using a differential interferometric refractometer. The difference in refractive index between the protein solution and the electrolyte solvent (Δn) is measured at the same wavelength at which light scattering is measured. To obtain the same chemical potential of salt and water in protein solutions as in the pure solvent, a tedious dialysis against the solvent is rigorously required before measuring the Δn, which is often overlooked. The value of dn/dc_p can then be measured by plotting Δn versus c_p as

$$\Delta c = \frac{dn}{dc_p}c_p \tag{4}$$

In order to study the sensitivity of the B_{22} value on the measured R_θ value, we consider a simple case of light scattering intensity of lysozyme (M_w 14,600 g mol^{-1}) at a concentration of 0.002 g ml^{-1} and at a wavelength of 633 nm. We also fix the values of n_0 at 1.33 and dn/dc_p at 1.81 ml g^{-1}, which eventually give an optical constant, K, of 2.37×10^{-7} mol ml cm^{-1} g^{-2}. In this specific case, a B_{22} value 0.00×10^{-4} mol ml g^{-2} corresponds to a R_θ value of 6.91×10^{-6} cm^{-1}. Unfortunately, no information is available about the precision limit of the R_θ value measured by SLS or LALLS from the instrument suppliers. However, incon-

sistency in light scattering data is very common in practice. For this reason, the final data point is usually taken based on the average of at least 50 statistically consistent measurements. Disturbance from dust is a common source of error, which can be minimized by using the statistical dust rejection function and by setting a tight rejection ratio. We, therefore, assumed that the standard deviation of R_θ values would be about 0.05%.

In addition to R_θ, small errors in the optical parameters, i.e. n_0 and dn/dc_p dramatically affect the B_{22} value because of their quadratic dependence. According to the information provided by different suppliers, most interferometric refractometers have a measurement inaccuracy of ∼1%. In practice, a 0.5% inaccuracy is quite normal in the eventual dn/dc_p value. For example, Rosenbaum et al. [13] found the dn/dc_p value of lysozyme to be 1.181 ml g^{-1} with an error bar of ±0.005 ml g^{-1} (0.42%) at 633 nm. We, therefore, approximate the standard deviation of measured n_0 and dn/dc_p values to be 0.5%, which gives an error of approximately 2% in the K value using Eq. (3).

A realistic error margin in B_{22} as measured by light scattering techniques is obtained as follows. If B_{22} is calculated from a single data point of lysozyme concentration considering a 0.05% error in R_θ values and a 2% error in K value, the overall error in B_{22} value is as high as $\pm 2.58 \times 10^{-4}$ mol ml g^{-2}. The error margin decreases with an increasing number of data points at different protein concentrations. For the case of five data points, the standard deviation of the slope corresponds to an error of about $\pm 2.0 \times 10^{-4}$ mol ml g^{-2} in the determined B_{22} value. It is, therefore, quite realistic that a B_{22} value measured by the light scattering technique may have an inherent inaccuracy of $\pm 2.0 \times 10^{-4}$ mol ml g^{-2}. The accuracy limit or error bar is not presented in most published works. However, the extent of the error margin and the experimental reproducibility are obvious from the experimental approach of Curtis et al. [22]. They found an error in the M_w determination on the order of 400 g mol^{-1} for lysozyme and 1200 g mol^{-1} for ovalbumin. If we translate this error in terms of the B_{22} value, the error could be as large as $\pm 4.0 \times 10^{-4}$ mol ml g^{-2}. A similar error is also reported in other work [15]. Rosenbaum et al. [13] also mentioned an estimated uncertainty for the lysozyme B_{22} measurement by their SLS experiment to be ±18 nm^3, which is equivalent to $\pm 0.5 \times 10^{-4}$ mol ml g^{-2}.

2.1.3. Size-exclusion chromatography

Thermodynamic non-ideality of a solute leads to a concentration dependent partition coefficient in a non-interacting column [42]. Using this non-ideal behavior of a protein in a typical SEC column, a theoretical framework was established to determine B_{22} by frontal-exclusion chromatography (FEC) as [36,42]

$$\ln\left(\frac{K_D}{K_0}\right) = 2B_{22}M_w C_i(1 - K_D) \quad (5)$$

where C_i is the plateau value of the protein concentration in the mobile phase, K_0 is the partition coefficient of protein at the limit of infinite dilution and K_D is the local protein distribution coefficient, which is described as

$$K_D \equiv \frac{C_s}{C_m} = \frac{V_r - V_0}{V_t - V_0} \quad (6)$$

where C_s and C_m are the local protein concentrations, respectively, in the pore/intraparticle volume (stationary phase) and interparticle volume (mobile phase), V_r is the average retention volume of the mobile protein, V_0 is the extra-particle or interstitial volume and V_t is the total mobile phase volume. If $\ln(K_D)$ is plotted as a function of $C_i(1 - K_D)$, the slope of the plotted line becomes equal to $2B_{22}M_w$. The B_{22} determination procedure is then to inject different concentrations of protein sample and measure K_D from the retention volumes. Therefore, the accuracy of a determined B_{22} value depends on the run-to-run deviation of retention volume for a fixed protein concentration and the linearity of $\ln(K_D)$ versus $C_i(1 - K_D)$ plot.

Frontal exclusion chromatography was, however, rarely used as a B_{22} measurement technique because it requires a huge amount of protein and long experimental times to reach the plateau region. In contrast, a pulse SEC technique was recently developed [36], where the plateau value of the mobile phase protein concentration, C_i, was replaced by average protein concentration of the mobile phase in the pulse, $\langle C_i \rangle$

$$\ln\left(\frac{K_D}{K_0}\right) = 2B_{22}M_w\langle C_i\rangle(1 - K_D) \quad (7)$$

For the error estimation in SEC, we consider a realistic case with V_0 and V_t values of 6.07 and 10.77 ml, respectively. In such a SEC column, the V_r values vary from 9.85 to 9.88 ml for $\langle C_i \rangle$ values from 0.70 to 2.95 mg ml^{-1}, respectively [36]. The random run-to-run difference in retention volumes in this system is usually within the range of 0.1%. According to error propagation statistics, an error of only ±0.1% in the retention volumes corresponds to an error of ±0.0023 in the K_D values and ±0.00285 in the $\ln(K_D)$ values. Considering the fact of ±0.00285 errors in the $\ln(K_D)$ values, our analysis shows huge error bars in every data point in the direction of Y-axis in the $\langle C_i \rangle(1 - K_D)$ versus $\ln(K_D)$ plot (Fig. 3). The standard error of the slope due to Y-axis error bar was analyzed as described in the ref. [43]. Since every data point shows a huge error bar, the standard error of the slope is as large as 0.0083 ml mg^{-1}, which gives an error of about $\pm 3.0 \times 10^{-4}$ mol ml g^{-2} in the eventual B_{22} value. Therefore, B_{22} values measured by either pulse or frontal elution SEC system would have an inherent inaccuracy of $\pm 3.0 \times 10^{-4}$ mol ml g^{-2}. The error limit could even be higher in a pulse SEC system, because the mobile phase protein concentration in a pulse system changes during transport down the column whose measurement difficulty could be an extra source of error. Since it is quite difficult to measure the accurate average protein concentration of the mobile phase in the pulse, $\langle C_i \rangle$, a reasonably realistic value of $\langle C_i \rangle$ can be determined as the total amount of protein in the pulse divided by

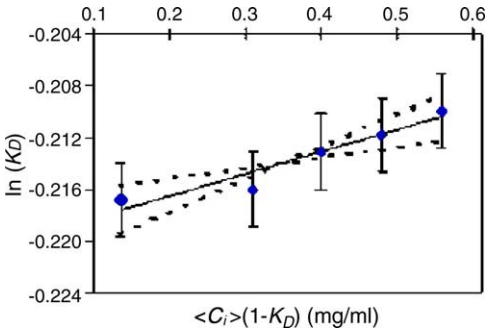

Fig. 3. $\langle C_i \rangle(1 - K_D)$ vs. $\ln(K_D)$ plot considering error margin of ±0.1% in the retention volumes. Solid line represents the trend line according to least-square fitting and dotted lines represent standard error of the trend line according to ref. [43]. This plot was generated using some data of ref. [36].

the volume of the column accessible to the protein. Bloustine et al. [36] showed that the maximum concentration of protein in the eluted mobile phase, C_{max}, can also be used instead of $\langle C_i \rangle$. In such a situation an error of few percentage in the $\langle C_i \rangle$ values is quite optimistic. In practice, C_{max} values differ from respective $\langle C_i \rangle$ values by about 5% [36]. However, this error was neglected in this analysis of error margin of $\pm 3.0 \times 10^{-4}$ mol ml g^{-2}.

2.2. Determination of B_{22} by SIC

2.2.1. Calculation of B_{22} from SIC retention data

Retention of a protein sample in a SEC column is typically characterized by the distribution coefficient, K_{SEC}, which is given by

$$K_{SEC} = \frac{V_r - V_0}{V_i} = \frac{V_r - V_0}{V_t - V_0} \tag{8}$$

where V_r, V_0 and V_t terms are described in Eq. (6) The term $(V_t - V_0)$ was assumed as V_i, which denotes the intra-particle pore volume. The K_{SEC} term in Eq. (8) may vary from 0 to 1 for fully excluded large molecules to fully included small molecules, respectively.

Self-interaction chromatography is essentially a quantitative affinity chromatography system, which estimates weak interactions of mobile protein molecules with immobilized protein molecules. Because of the weak and bi-directional (both attractive and repulsive) nature of protein self-interactions, the SIC system cannot be characterized in terms of association or dissociation constants. In the case of low protein load in the mobile phase, which is essentially the condition for SIC, the slope of the linear region of the adsorption isotherm is however related to the potential of mean force between protein molecules. The distribution coefficient in such a quantitative affinity chromatography system, K_{aff}, can be described as [38,44]

$$K_{aff} = \frac{q}{c} = \frac{\Delta V_{r,aff}}{m} \tag{9}$$

where q is the amount of protein adsorbed per volume of resin, c is the concentration of protein in the mobile phase, m is the amount of resin in terms of volume and the $\Delta V_{r,aff}$ term denotes the changes in retention volume due to protein self-interactions only. The distribution coefficient, K_{aff}, in Eq. (9) is a measure of protein–protein interaction. Positive values of K_{aff} correspond to higher retention of the mobile protein molecules due to attractive interaction with stationary molecules. Similarly, negative values of K_{aff} may also arise from repulsive interaction between stationary and mobile protein molecules. In thermodynamic terms, K_{aff} can also be defined as [30,44]

$$K_{aff} = \frac{\int_{V_i} (e^{-(\Delta G/kT)} - 1)\, dV_i}{V_i} \tag{10}$$

where ΔG is the free energy change of bringing a protein molecule from the interstitial volume into the pore volume so that it interacts with a single immobilized protein molecule, k is the Boltzman constant and T is the absolute temperature. In a typical SIC system, we measure an overall partition coefficient, $K_{overall}$, which is contributed by both size-exclusion and weak-protein interactions. The overall partition coefficient can therefore be represented as

$$K_{overall} = K_{SEC} + K_{aff} \tag{11}$$

Only the K_{aff} term in Eq. (10) is related to protein–protein interaction, and eventually to B_{22}. Since the retention volume in SEC column also varies with the injected protein concentration due to protein–protein interactions [36], the K_{SEC} term as defined in Eq. (8) is also related to protein–protein interaction. However, the injected protein concentration in SIC is usually very dilute and in the linear region of adsorption isotherm. Changes of retention volume in a SEC column as a function of the injected protein concentration is, however, very little and within the range of run-to-run errors of retention volume. Therefore, any relation of K_{SEC} to the protein–protein interactions is neglected. ΔG in Eq. (10) is essentially equivalent to the potential of mean force, W, which is described as the anisotropic interaction free energy required to bring two infinitely spaced solute molecules into a defined separation distance, r, averaged over all possible orientations of the solute molecules. The free energy change, ΔG, is also a function of intermolecular separation distance (r) and possible angular positions/orientations of both immobilized (Ω_1) and mobile (Ω_2) interacting molecules. However, $\Delta G(r, \Omega_1, \Omega_2)$ may not be equal to $\Delta W(r, \Omega_1, \Omega_2)$ in the sense that one mobile protein molecule may simultaneously interact with more than one immobilized molecule or among mobile molecules themselves or with the chromatography resin. In addition, the immobilized protein molecules may lose their rotational freedom and may not be accessible from different angular positions or orientations. Assuming that the experimental SIC system does not encounter these uncertainties, $\Delta G(r, \Omega_1, \Omega_2)$ becomes equal to $\Delta W(r, \Omega_1, \Omega_2)$.

Considering all statistically possible orientations for both immobilized and free protein molecules, the distribution coefficient due to protein–protein interaction (K_{aff}) can be

expressed as discussed by Tessier et al. [30].

$$K_{\text{aff}} = \frac{N \int_{\Omega_1} \int_{\Omega_2} \int_0^{\infty} (e^{-(W(r,\Omega_1,\Omega_2)/kT)} - 1) r^2 dr\, d\Omega_2\, d\Omega_1}{V_i} \quad (12)$$

where N is the total number of immobilized protein molecules accessible for mobile protein molecules. The lower limit of the separation integral in Eq. (12) was taken as zero rather than the center-to-center distance upon intermolecular contact, because the K_{aff} term has to be obtained from chromatography retention data. The nature of protein–protein interactions in the SIC column is similar to that in a real protein solution and the chromatographic retention data represents all sorts of interactions including the excluded volume contribution.

The osmotic second virial coefficient (B_{22}) refers to the interaction between two protein molecules in solution and is rigorously related to the two-body potentials of mean force between protein molecules in solution [5,45]

$$B_{22} = -\int_{\Omega_1} \int_{\Omega_2} \int_0^{\infty} (e^{-(W(r,\Omega_1,\Omega_1)/kT)} - 1)\, r^2 dr\, d\Omega_2 d\Omega_1 \quad (13)$$

Comparing Eqs. (11)–(13) yields

$$B_{22} = \frac{(K_{\text{SEC}} - K_{\text{overall}})V_i}{N} \quad (14)$$

Eq. (14) can be used to calculate B_{22} by the SIC methodology. K_{SEC} can be determined using Eq. (8) for a protein-free column with the same resin material, where no protein–protein interaction can take place and the retention volume of the mobile protein is only determined by size exclusion. In such a protein-free column, the value of K_{aff} is zero. The overall distribution coefficient, K_{overall}, can also be determined using Eq. (8) for an immobilized-protein column, where the retention of the mobile protein is guided by both size exclusion and protein–protein interactions. V_i in Eq. (14) is equal to $(V_t - V_0)$ for the protein-immobilized column. N can also be represented as the total amount of immobilized protein in gram. The unit of B_{22} obtained in this approach is ml g^{-1} which has to be divided by the molecular weight (M_w) of the protein in order to obtain the usual unit of B_{22}, mol ml g^{-2}. Therefore,

$$B_{22} = \frac{(K_{\text{SEC}} - K_{\text{overall}})V_i}{N \cdot M_w} \quad (15)$$

Our approach to the calculation of B_{22} is comparable to that derived in previous studies [30,35] and essentially the same as that of Teske et al. [35] for the case of identically packed immobilized-protein and protein-free columns. However, the immobilized-protein column in our work does not necessarily have to be the same as the protein-free column in terms of column volume and packing integrity. In addition, our approach justifies the use of a protein-free column and does not require the retention volume in the theta condition (when mobile phase proteins have no net interactions with the immobilized proteins), which is quite impractical to determine accurately [34].

An extra hard sphere contribution term (B_{HS}) is disappeared in this derivation because lower limit of the separation integral in Eq. (12) was taken as zero. The B_{HS} contribution to B_{22} is always positive and roughly equal to 6.7 times the molecular volume of the protein [46]. Considering the lysozyme molecule as a hard sphere of 3.11 nm, the excluded volume contribution is 4 times the molecular volume, which is equal to 1.84×10^{-4} mol ml g^{-2}. The B_{HS} term of the larger protein is even smaller in the unit of mol ml g^{-2}.

2.2.2. Selection of stationary phase

For the purpose of SIC, chromatography particles with wide pores are desirable in order to minimize the mass transfer limitation and to ensure that immobilized protein molecules do not block the pore space for the mobile molecule to pass and interact outside the pore surface. In addition, the packed particles should not have any interaction with mobile phase protein. In this work, we used *N*-hyroxysuccinimide (NHS) activated Sepharose FF which consists of a 14-atom (6-aminohexanoic acid) spacer arm between the ligated protein molecules and the surface of the particle. This long spacer arm gives the immobilized protein molecule more flexibility to interact with the mobile phase protein from different angular positions and orientations. The mean particle size, pore diameter and porosity of Sepharose FF are 90 μm (provided by Amersham Biosciences), 50 nm [47] and 0.63 [47], respectively. The NHS groups react with *N*-terminal amino groups of the peptide chains and with the ε-amino groups of lysine residues [48], which provides random orientation of the immobilized protein molecules on the particle surface. In addition, the NHS-activated support shows very fast and complete binding of protein with comparatively minimal protein leakage during storage and chromatography [49]. The only minor disadvantage of NHS-Sepharose is that it may produce some anionic groups on its surface by hydrolysis of NHS, when the coupling reaction is done at high pH and/or high temperature. However, this problem can be minimized and the coupling reaction rate can be controlled in order to obtain the desired coupling concentration by conducting the reaction at pH 6.0 and at a temperature of 4 °C [48].

2.2.3. Optimization of the extent of immobilization

Tessier et al. [30,31] found that the retention of protein depends on the injected concentration at higher surface coverage (∼33%), while the effect of injected concentration on retention volume is negligible at a surface coverage of 17–18% [30,35]. At very high surface coverage, a free protein molecule may have the opportunity to interact simultaneously with multiple immobilized molecules, which results in an injection concentration-dependent retention behavior. In addition, the higher surface coverage may block some par-

ticle pores, which hence become inaccessible to free proteins. We, therefore, controlled the immobilization process to work within a surface coverage of about 15%. The immobilization concentration of 20 mg lysozyme ml^{-1} of packed column corresponds to 15% surface coverage for NHS-activated Sepharose (Appendix A). The incubation time, temperature, pH and protein concentration of the immobilization reaction mixture are the parameters for controlling the immobilization reaction. Only the incubation time was varied in this work and it was found that 12 h of incubation was sufficient to obtain optimum coupling.

2.2.4. Optimization of injection sample

While a l5% surface coverage is used to avoid multi-body interaction, the protein concentration in the mobile phase has also to be low enough to avoid interaction among mobile phase proteins themselves. Since the range of protein self-interactions is very short [4] and the injected protein concentration in the SIC experiment is typically quite dilute, this is not a prominent matter of concern. However, in order to make sure that we are working in the linear region of the adsorption isotherm, the total amount of protein in an injection pulse must be much lower than the amount of immobilized protein in the stationary phase. In that case, the retention of the pulse should not vary with little fluctuations in injection concentration. Patro and Przybycien [29] found no significant difference in the peak position of lysozyme over an injection concentration range of 2–9 mg ml^{-1}. Teske et al. [35] also obtained retention times independent of the mobile phase protein concentration provided that this concentration is less than 0.25 mg ml^{-1}. On the other hand, Tessier et al. [30] observed that the retention time of lysozyme increases when the injection concentration goes below 5 mg ml^{-1}. Considering these variations in previous studies, we have studied the optimum range of injected concentrations and volumes in our system. Using an injection volume of 50 μl and concentrations of 1–5 mg ml^{-1} for a 1.2–1.4 ml columns produced sharp Gaussian peaks with a fairly high detection limit for lysozyme. The peak position was independent of the injected protein concentration in the tested range (1–5 mg ml^{-1}).

2.2.5. Determination of K, K_{SEC} and N

K and K_{SEC} can be determined from the immobilized-protein column and the protein-free column, respectively, using Eq. (8). It is therefore important to inject a pulse of non-interactive fully included and of fully excluded molecules in both columns. Acetone was selected as the fully included molecule because of its small size, delectability in UV at 280 nm and non-interactive nature. The optimum concentration of acetone was found to be 2% (v/v). On the other hand, the interstitial volume was determined by injecting a fully excluded large molecule, i.e. blue dextran ($M_w \sim 2$ mDa). Blue dextran has some interaction with the Sepharose particles and probably with the immobilized protein molecules at low ionic strength. Therefore, the blue dextran pulse was eluted in the presence of 1.0 M NaCl in order to eliminate any interactions with the particle surface and the immobilized proteins.

N is the number of immobilized protein molecules accessible for mobile phase protein. The number of protein molecules immobilized per volume of settled particles can be determined first, then multiplied by the amount of settled particles used to pack the column. Considering the fact that the entire pore spaces in Sepharose FF are not wide enough for the protein molecules [47], it was nevertheless assumed that all the immobilized protein molecules are accessible to mobile molecules. Since only 15% surface coverage is applied, the remaining 85% of the pore surface is still free. Therefore, the immobilized protein molecules entered into the pore spaces are not expected to restrict the entry of mobile molecules. Such a restriction can only be expected in the case of high surface coverage or monolayer coverage. Therefore, the value of N is equal to the total number of protein molecules immobilized onto the stationary phase.

3. Materials and methods

3.1. Materials

Lysozyme from chicken egg white (3× crystallized, dialyzed and lyophilized; product no. L6876) was bought from Sigma-Aldrich Co. *NHS*-activated Sepharose™ 4 Fast Flow (code no. 17-0906-01) was purchased from Amersham Biosciences.

Acetic acid (Baker analyzed, product no. 6052), sodium chloride (Baker analyzed, product no. 0278), hydrochloric acid (36–38%, Baker analyzed, product no. 6081), sodium hydrogen carbonates (ACS grade, Baker analyzed, product no. 0263), potassium hydrogen phosphate anhydrous (ACS grade, Baker analyzed, product no. 0241), potassium dihydrogen phosphate (ACS grade, Baker analyzed, product no. 0240), acetone (Bakers HPLC analyzed, product no. 8142) and sodium hydroxide (Baker analyzed, pellets, product no. 0402) were bought from J.T. Baker. Sodium hydrogen phosphate dihydrate (product no. 6573) and sodium dihydrogen phosphate dodecahydrate (product no. 6345) were bought from Merck. Ethanolamine (redistilled, product no. 41100), tween 80 (product no. P8074), blue dextran (product no. D5717) and magnesium bromide hexahydrate (product no. 518220) were bought from Aldrich. BCA (bicinchoninic acid) protein assay reagents (products 23221 and 23224) were bought from Pierce.

A Tricorn™ 5/50 column (code no. 18-1163-09) and Tricorn™ 5 adapter unit (code no. 18-1153-00) was bought from Amersham Biosciences. Chromatography experiments were done in a Pharmacia FPLC system, which was controlled by Unicorn Version 2.0. Ultracentrifugation experiments were done by a Beckman L-70 ultracentrifuge with a Ti-60 rotor type. Normal centrifugation was done in a Beckman GP centrifuge. All spectrophotometric analyses were done in a Pharmacia spectra UV/visible spectrophotometer.

3.2. Immobilization of protein

An amide linkage is formed between the amino groups of the protein and the NHS activated group of Sepharose in the pH range of 6–9. The coupling reaction is very fast and almost uncontrollable at higher pH and temperature [48]. In addition, NHS groups are hydrolyzed rapidly at higher pH to give free COO^- groups, which makes the particle a weak cation exchanger [49]. In order to avoid this undesirable side reaction, the coupling reaction was done at pH 6.0 and at 4 °C. The protein solution was prepared first at a concentration of 5 mg ml^{-1} in the coupling buffer (0.1 M sodium phosphate, 0.5 M NaCl, pH 6.0). Isopropanol suspended particles were washed five times with ice cold 1 mM HCl by centrifugation. Three milliliters of washed particles were incubated with 10 ml of 1 mM HCl for 15 min at 4 °C for swelling. HCl was removed from the settled particles and immediately replaced by 10 ml of pre-prepared protein solution. The coupling reaction was allowed to proceed at 4 °C with gentle shaking. The desired surface coverage of protein was obtained by manipulating the incubation time. The coupled particles were then washed a few times with ice cold coupling buffer to remove unbound proteins and released NHS. The immobilized particles were incubated again with 10 ml of blocking buffer (1 M ethanolamine, 0.5 M NaCl, 0.1 M Na–phosphate, pH 6.0) for 12 h at 4 °C in order to block any remaining reactive groups. The protein-free particles were prepared in the same way without doing the coupling reaction.

Since NHS, released during the coupling reaction, has a very high UV absorption at 280 nm, it was not possible to determine the amount of immobilized protein from the amount of protein in the wash out solutions. The density of protein immobilized in the particle was therefore determined by a standard BCA technique [50], applied to the solid phase [51].

3.3. Chromatography and data analysis

Tricorn 5/50 (5 mm × 50 mm) columns were packed with both immobilized-protein and protein-free particles at a flow rate of 3 ml min^{-1} for at least 10 min. The flow rate was subsequently reduced to 1 ml min^{-1} for confirming the stability of the bed. The integrity of the packed column was characterized by height equivalent to a theoretical plate analysis, and peak shape and symmetry of a pulse of a small molecule, for instance acetone and/or high salt. The chromatography procedure was accomplished as described by Tessier et al. [30] in an automated Pharmacia FPLC system controlled by Unicorn Version 2.0. The injection sample was prepared at a concentration of 1–2 mg ml^{-1}, unless mentioned otherwise. The column was equilibrated with the appropriate solution until the UV, pH and conductivity base lines became completely straight before every injection. Retention volumes were automatically determined by Unicorn as the peak position. The column was stored at 4 °C in 10 mM sodium phosphate (pH 7.0), when not in use. Each column was used for a period of maximum 4 weeks.

4. Results and discussion

4.1. Inherent inaccuracy of SIC

A theoretical framework for determination of B_{22} by SIC is presented in Section 2.2. According to Eq. (15), the uncertainty in B_{22} may come from errors in the estimation of pulse retention volumes in the columns and in the determination of immobilized protein concentration on the stationary phase. In our experimental set-up, the reproducibility of retention volume was within the limit of ±0.01 ml and the maximum inaccuracy in determining the immobilized protein concentration on the gel particle was ±20%. These amounts of error in V and N in Eq. (15) yield an overall error of maximum $\pm 1.0 \times 10^{-4}$ mol ml g^{-2} in the calculated B_{22} value. A comparative overall error analysis of the different techniques is shown in Fig. 4.

4.2. How efficient is the SIC technology compared to other techniques?

The efficiency of a B_{22} measurement technique is determined by the amount of protein and experimental time required to measure one B_{22} value. Although a fairly good estimation can be made regarding the amount of protein required to determine one B_{22} value, the estimation becomes complicated regarding the time needed to obtain one B_{22}. This is because each method has its own long preparation time, data acquisition difficulties and extraneous complications. In reality, in all of these troubleshooting consume most of the time compared to the real data acquisition experimentation.

In MO, a micro-osmometer may have sample volume as low as 20 μl. Five different concentrations are usually required to compare the Π/RTc_p versus c_p plot. We assume concentration levels of 2, 4, 6, 8 and 10 mg ml^{-1}. Therefore, the minimum amount of protein required to measure

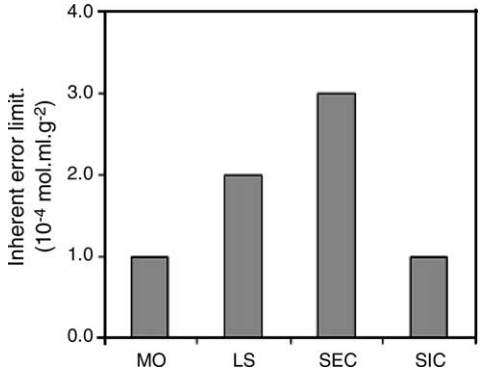

Fig. 4. Inherent inaccuracy limit of different B_{22} measurement techniques.

one B_{22} value is 0.6 mg. Regarding time requirement, once the installation of the osmometry equipment is completed, the duration of each osmotic pressure measurement is about 15 min [25]. MO, however, suffers from some practical problems, for instance fouling and adsorption. Troubleshooting of these difficulties require unusually long time, which are case oriented and difficult to estimate.

In light scattering, 1 ml of sample volume is usually required to place in the sample cell. If we consider five data points with the same protein concentration as estimated for MO, total amount of protein required to measure one B_{22} value is 6 mg. However, light scattering measurement is not usually run-to-run consistent and require a large number of replications in order to validate one data point. Sample preparation and the ability of the LS equipment to measure scattering intensity rapidly over a range of protein concentrations are great challenges. If we assume that all experiments run perfectly and provide acceptable data, 15 min is usually enough to measure the light scattering intensity and the refractive index of a sample.

Five different concentrations of protein are usually injected in a SEC column to obtain a linear relationship in the $\ln(K_D)$ versus $C_i(1-K_D)$ plot. Each protein concentration has to be far apart from the next ones in order to provide better resolution and consequently higher accuracy. For a 20 μl pulse injection with protein concentrations of 10, 20, 30, 40 and 50 mg ml^{-1}, the total protein requirement is 3.0 mg. The elution time of one pulse can be assumed to be 25 min. However, column preparation, characterization and equilibration takes much longer time than the peak elution time.

If the experiment is done in a frontal exclusion system in a 1 ml volume column, the injected sample volume has to be at least 1 ml in order to reach the plateau stage. In a frontal elution system, protein concentrations of 2, 4, 6, 8 and 10 mg ml^{-1} would be enough to provide good resolution. The amount of protein required is then 30 mg in order to obtain one B_{22} value. The elution time will also be at least twice as long as for a pulse system.

Measurement of B_{22} by SIC requires at least 2-pulse injections, one on an immobilized-protein column and another on a protein-free column. Total experimental time for each pulse elution is not more than 25 min for our current set-up. Typically each pulse contains 50 μl of a protein solution at a concentration of 2 mg ml^{-1}. Therefore, two pulses contain a total of 0.2 mg of protein. In addition, optimization of the injection concentration and flow rate requires several pulse at different protein concentration and different flow rate level, which eventually costs more time and protein. A comparatively large amount of protein (about 30 mg) is required to prepare an immobilized-protein column, which can be used for a month for the determination of hundreds of B_{22} values. We assume that an immobilized-protein column is used to determine 160 B_{22} values (20 days × 8 measurements per day) throughout a month. Therefore, an average of minimum 0.45 mg of protein is required to determine a B_{22} value by the SIC technique. However, the SIC method can easily be miniaturized to microchip level, thereby lowering the analysis time and the protein requirement by orders of magnitude.

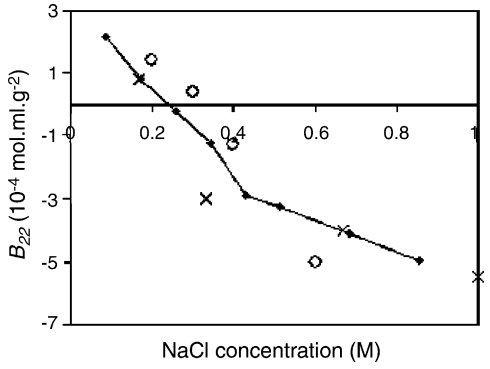

Fig. 5. B_{22} of lysozyme measured by SIC using different gel materials at pH 4.5 and at a temperature of 23–25 °C. The line represents Sepharose from Amersham Biosciences (this work), the circles represent Toyopearl from Tosoh Bioscience [30] and crosses represent cross-linked agarose from Sigma [35].

4.3. Mapping of lysozyme B_{22} profile

Fig. 2 shows that the accuracy of B_{22} data obtained by SIC at pH 4.5 is quite good in comparison to other techniques. However, we have extended our work on lysozyme in order to investigate the reproducibility of the SIC technique in different gel materials, protein immobilization strategies and solution conditions.

The stationary phase and the protein immobilization strategy used in this work were quite different from those in previous studies [30,35]. We have used narrow pore-size particles where protein immobilization took place via a spacer arm. The result obtained in our work shows that the SIC system is able to reproduce B_{22} data irrespective of the type of stationary phase and protein immobilization strategy (Fig. 5). We have also found for our system that a column packed with immobilized lysozyme can be used as long as microorganisms do not degrade it. Microbial degradation deteriorates the packing integrity of the column and the column can no longer produce sharp Gaussian peaks. However, the lifetime of SIC columns can be different depending upon the stability the protein immobilized on it.

The experimental approach was further extended to calculate B_{22} values in conditions available from the literature and in some unknown conditions. It was found earlier that B_{22} trends of lysozyme are quite ideal, decrease smoothly with pH and ionic strength [11], but increase proportionally with temperature [3,20]. The next mapping of B_{22} was done at pH 7.6 in the presence of 10 mM Na–phosphate buffer. There are remarkable variations in published B_{22} values for this condition (Fig. 6). The data obtained in this work fell below SLS data available in the literature. A notable feature of Fig. 6 is that the B_{22} values obtained by SIC at higher NaCl concentrations (>0.5 M) are well below than that of SLS. The reason for this behavior can be explained as simultaneous interaction

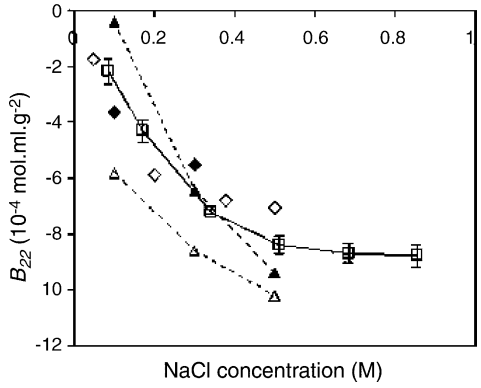

Fig. 6. B_{22} trend of lysozyme at pH 7.6. Black diamond: pH 7.5, 25 °C, SLS [9]; white diamond: pH 7.4, 25 °C, SLS [8]; white rectangle with line: pH 7.6, 10 mM Na–phosphate, SIC [this work]; black triangle with dashed line: pH 7.0, 5 mM bis-tris, SIC [30]; white triangle with dashed line: pH 7.6, 20 mM Na–phosphate, SIC [35].

of a mobile-phase molecule with two or more immobilized molecules [35]. The overall trend of B_{22} as a function of NaCl concentration is obviously due to electrostatic interaction. Since the p*I* of lysozyme is quite high (11.2), electrostatic repulsion is more prominent during self-interaction at low pH and low salt. Lysozyme is almost chargeless at pH 9, where short-range attractions play the vital role.

It is known that B_{22} of lysozyme in $MgBr_2$ shows a minimum at 0.3 M $MgBr_2$ at pH ∼7.6 [11]. This phenomenon was confirmed by the SIC method [30]. The phenomenon was further confirmed in this work and the minimum was found at 0.4 M instead of 0.3 M $MgBr_2$ (Fig. 7). However, the B_{22} of lysozyme does not change significantly with the $MgBr_2$ concentration. Tessier et al. [30] explained this phenomenon as an increase in repulsion at higher $MgBr_2$ concentrations due to binding of the divalent Mg^{2+} to the acidic residues of lysozyme. In order to correctly determine the reason of this behavior, B_{22} was also measured in $MgCl_2$.

The B_{22} trend was found to be steadily decreasing with $MgCl_2$ at pH 4.5 (Fig. 8). However, there is no literature data

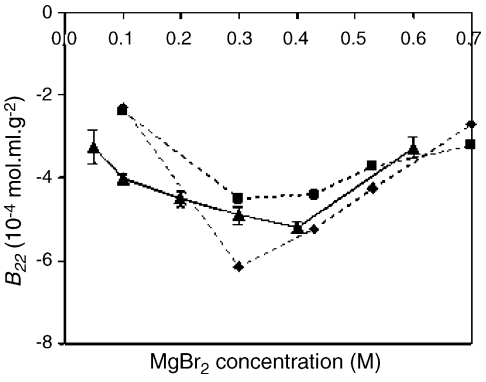

Fig. 7. B_{22} trend of lysozyme as a function of $MgBr_2$ concentration at pH 7.6. Triangle: pH 7.6, 10 mM Na–phosphate, SIC [this work]; diamond: pH 7.8, 5 mM bis-tris, SIC [26]; rectangle: pH 7.8, 20 mM HEPES, 23 °C, SLS [11].

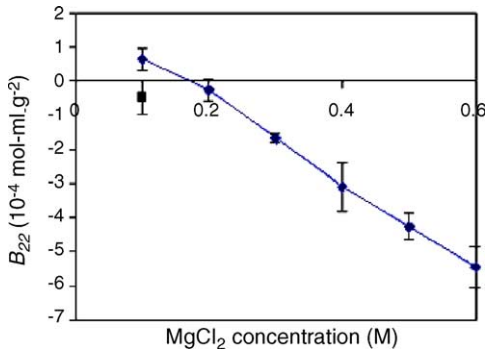

Fig. 8. B_{22} trend of lysozyme as a function of $MgCl_2$ concentration at pH 4.5. Line: pH 4.5, 10 mM Na-acetate, SIC [this work]; rectangle: pH 4.6, 50 mM Na-acetate, 25 °C, SLS [13].

available for this condition except one point. The available data point [13] is lower than that found in this work, probably because of lower ionic strength of the buffer. The B_{22} trend obtained in $MgCl_2$ is comparable with that in NaCl for pH 4.5. The trend line in $MgCl_2$ is slightly lower than that in NaCl. The reason is the presence of more electrolytes at equal molarity in $MgCl_2$ than in NaCl because of the divalency of magnesium. The B_{22} trend of lysozyme was also determined in $MgCl_2$ at pH 7.6 (10 mM Na–phosphate) and we found that the B_{22} does not change much with $MgCl_2$ concentration (Fig. 9). Instead of a minimum, a maximum was found at 0.2–0.3 M $MgCl_2$. The trend was further going down with increasing $MgCl_2$ concentration. Since the effect of $MgCl_2$ and $MgBr_2$ on the B_{22} trend is not very large at pH 7.6, it is hard to determine a minimum or maximum point at a particular ionic strength. It is therefore clear that the B_{22} trend of lysozyme in $MgCl_2$ is similar to that of NaCl but differs between $MgCl_2$ and $MgBr_2$. A likely explanation why trends in $MgCl_2$ are not similar to $MgBr_2$ is the chloride binding affinity of lysozyme. Lysozyme does not exhibit salting in behavior with NaCl due to predominant electrostatic screening of the positively charged protein and/or by adsorption of chloride ions by the protein [52]. Lack of this phenomenon in presence of bromide salt produces a downward peak in the B_{22} trend.

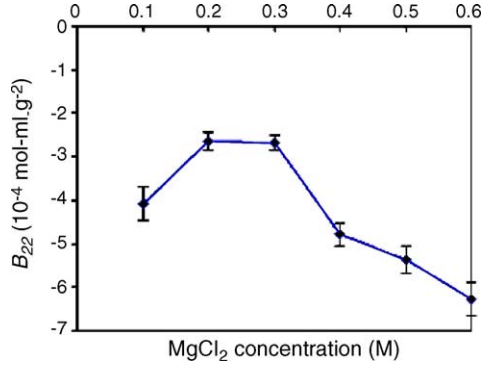

Fig. 9. B_{22} trend of lysozyme as a function of $MgCl_2$ concentration at pH 7.6 (10 mM Na–phosphate).

4.4. Crystallization slot

George and Wilson [2] determined B_{22} values of nine different proteins by SLS at their crystallization conditions and found them to be within a narrow range (between -0.8×10^{-4} and -8.4×10^{-4} mol ml g^{-2}), regardless of the size and nature of the proteins. The pattern of B_{22} for non-crystallization solvent conditions was not studied in detail. It was, however, observed for lysozyme that solution conditions corresponding to positive and highly negative B_{22} values promote no phase separation and amorphous precipitation, respectively. After further investigation of a few other proteins, they defined the so called "crystallization slot" as the range of B_{22} values between -1×10^{-4} and -8×10^{-4} mol ml g^{-2} [3]. This B_{22} based crystallization slot was used thereafter for predictive crystallization of lysozyme [9], chymotrypsinogen [9,27], ribonuclease A [32], myoglobin [31] and OmpF porin [12]. B_{22} clearly has a predictive value for the conditions of protein crystallization. The question one could ask is whether this range of B_{22} values -1×10^{-4} to -8×10^{-4} mol ml g^{-2} applies to all kinds of proteins, regardless of their size, shape, charge, hydrophobicity and surface roughness. In order to explore the versatility and applicability of B_{22} as a predictor of protein phase behavior, crystallization conditions of known proteins were mapped from available literature in terms of B_{22}. B_{22} values of all of these proteins (Fig. 10) at their crystallization conditions fell fairly within the range of -1×10^{-4} to -8×10^{-4} mol ml g^{-2}. Fig. 10 confirms that it is important to have a B_{22} value within the crystallization slot for crystallization of any protein. However, does a B_{22} value within the crystallization slot guarantee successful production of protein crystals? An extended B_{22} mapping was therefore done for proteins, of which the conditions of crystallization, amorphous precipitation and no phase separation were available from the literature (Fig. 11). Fig. 11 supports Fig. 10 in the sense that crystals do not grow in conditions at which the B_{22} value is either positive or largely negative. Interestingly however, crystals were not obtained in a number of cases where the B_{22} values were within the crystallization slot.

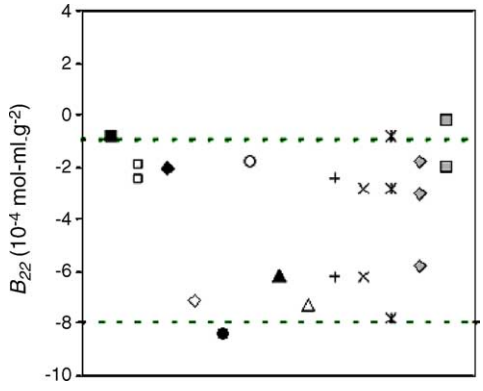

Fig. 10. B_{22} map of different proteins at their crystallization conditions. Solid rectangle: canavalin [2,3]; open rectangles: concanavalin A [2,3]; solid diamonds: bovine serum albumin [2,3]; open diamond: ovostatin [2,3]; solid circle: α-chymotrypsin [2,3]; open circle: satellite tobacco mosaic virus [2,3], solid triangle: ovalbumin [2,3]; open triangle: α-lactalbumin [3]; plus: β-lactoglobulin A [3]; cross: β-lactoglobulin B [3]; star: pepsin [3]; shaded diamond: thaumatin [3]; shaded rectangles: OmpF porin [12].

We have also conducted ultracentrifugal crystallization experiments with lysozyme in six solution conditions, out of which three conditions correspond to crystallization, one corresponds to amorphous precipitation and the remaining two correspond to no phase change (Table 2). According to the previously described models [53,54], for an initial protein concentration of 5 mg ml^{-1} and rotational speed of 45,000 rpm, 8 h of ultracentrifugation was enough to produce crystals in our system (Beckman Ti 60 rotor). After finishing the ultracentrifugation, about 80% of the supernatant was removed gently with a pipette. The remaining solution and pellet were examined visually for the presence of crystals or precipitate. It was unexpectedly found that no phase separation occurred in two samples at pH 7.6 where B_{22} values were -4.3×10^{-4} and -8.8×10^{-4} mol ml g^{-2}. In our experiment pH 7.6 was buffered using K-phosphate, which seems to be unfavorable for growing lysozyme crystals. It was also previously found that phosphate and sulfate ions are comparatively less effective for crystallization of lysozyme [13,55].

It is, therefore, fair to conclude that B_{22} values within the range of 0 to -10^{-3} mol ml g^{-2} are thermodynamically favorable for protein crystallization but do not guarantee successful crystal growth. On the other hand, protein crystallization is difficult or impossible at a condition where the B_{22} value is positive. Successful crystal growth may depend on several other parameters, for instance solubility and the effect of specific ions. Several authors investigated whether any direct relationship exist between protein solubility and B_{22} [3,11,13,16,20,23,56]. Their outcome suggests that a simple correlation may exist, but the relationship is not strong enough to design crystallization experiments. B_{22} may not sufficiently account for all interactions that are reflected in solubility, especially protein–salt interactions [21]. In addition, the crystallization process is significantly affected by the effect of specific ions. B_{22} of lysozyme decreases with increasing chloride ionic strength. However, the presence of phosphate and sulfate as buffering salts is not favorable for lysozyme crystallization even though the B_{22} value is driven into the crystallization slot by extra chloride. Indeed, the solubility of lysozyme is also very high in the presence of phosphate and sulfate ions [56]. Similarly, the solubility of lysozyme is the lowest in buffers containing Na^+ salts compared to other cations at equal ionic strength [57]. Therefore, in addition to B_{22}, the successful design of crystallization experiments may require solubility data and the knowledge of the effect of specific ions on that protein. However, B_{22} is the preliminary guide for systematic screening of protein crystallization conditions in the sense that it must be in the slightly negative regime for crystallization is likely to occur.

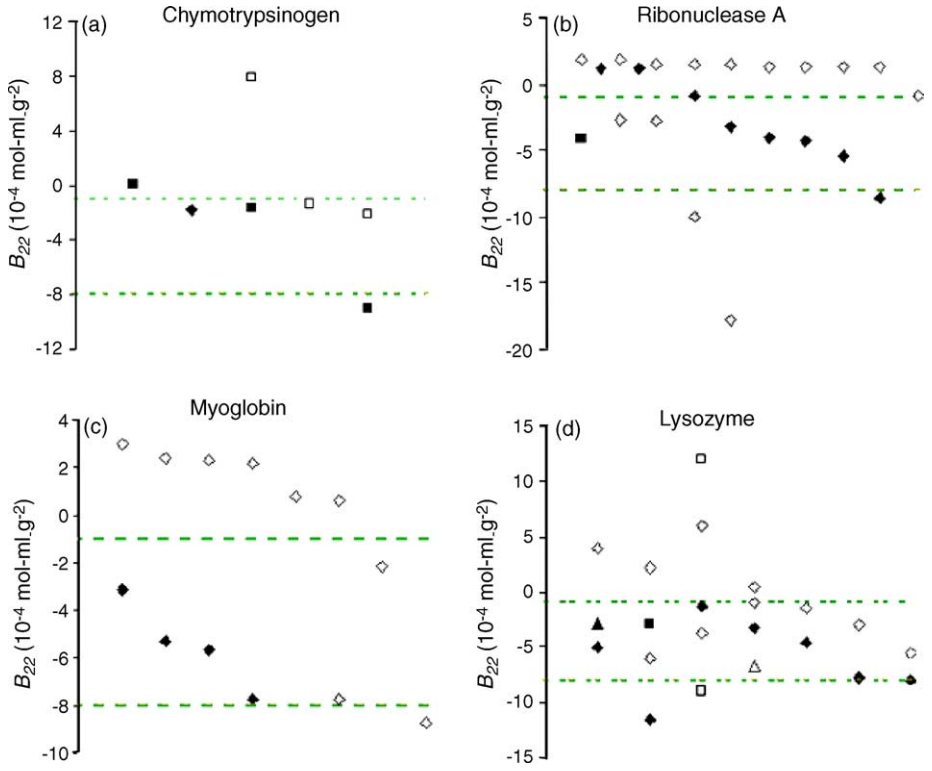

Fig. 11. Phase behavior of proteins as a function of B_{22}. Solid, shaded and white symbols denote crystal, precipitate and no change, respectively. Dashed lines denote the upper and lower boundaries of the crystallization slot. (a) Chymotrypsinogen: rectangular symbols are from ref. [9] and diamond symbol are from ref. [27]. (b) Ribonuclease A: rectangular symbols are from ref. [2,3] and diamond symbols are from ref. [32]. (c) Myoglobin: all symbols are from ref. [31]. (d) Lysozyme: rectangular symbols are from ref. [2], diamond symbols are from ref. [9] and triangular symbols are from ref. [3].

Table 2
Ultracentrifugal crystallization of lysozyme from B_{22} aided prediction

Sample no.	Solution condition	B_{22} (10^{-4}mol ml g^{-2})	After ultracentrifugation
A	pH 4.5 (0.01 M Na-acetate), 20 °C	>30.0	None
B	pH 4.5 (0.01 M Na-acetate), 0.51 M NaCl, 20 °C	∼−2.0	Crystal
C	pH 4.5 (0.01 M Na-acetate), 0.86 M NaCl, 20 °C	∼−4.0	Crystal
D	pH 7.6 (0.01 M K-phosphate), 20 °C	∼0.0	None
E	pH 7.6 (0.01 M K-phosphate), 0.17 M NaCl, 20 °C	∼−4.3	None
F	pH 7.6 (0.01 M K-phosphate), 0.86 M NaCl, 20 °C	∼−8.8	None

A molecular or thermodynamic understanding why a particular range of B_{22} values promotes crystallization was described in the literature [2–4,6,16,58]. Here we recall that negative values of B_{22} indicate that attractive forces between protein molecules are dominant and protein–solvent interactions are less favored than those between protein molecules. A positive value for B_{22} does not completely exclude the possibility of crystallization, but typically requires impractically high concentration of protein in order to bring about any kind of phase separation and the probability of obtaining acceptable crystals is very low. For the negative regime of the B_{22} map, Wilson [59] discriminated between craggs and praggs. Craggs are highly structured microcrystalline aggregates formed at slightly negative B_{22}. On the other hand, praggs are non-specific aggregates formed at highly negative B_{22}, which usually leads to amorphous structures.

5. Conclusion

A theoretical framework was established to correlate self-interaction chromatography retention data with B_{22} value. The approach requires retention data from an immobilized-protein column as well as from a protein-free column for the determination of a B_{22} value. However, the protein-free column does not necessarily have to be the same as the immobilized-protein column in terms of column volume and packing integrity. Details of the chromatography methodology, troubleshooting and data analysis approaches were designed. The reproducibility and accuracy limit of the B_{22} data by the SIC methodology was discussed in comparison to other traditional techniques, and SIC was shown to performs in a superior way. The SIC methodology can even be improved further by miniaturization to microchip level.

The designed methodology was applied for B_{22} mapping of a model protein, i.e. lysozyme. It was also found that a B_{22} value within the crystallization slot is an essential prerequisite of crystallization, but does not guarantee successful crystal growth. In addition to protein–protein interaction, protein solubility and the effect of specific ions also play a vital role for successful crystallization of protein, by mechanism that are not completely understood.

Appendix A. Calculation of surface coverage

The accessible surface area per volume of packed column, also called phase ratio, of a typical chromatography media decreases with increasing mobile phase particle size. The circumradius of the lysozyme molecule is 1.56 nm. An estimate of the phase ratio of Sepharose FF for lysozyme can be obtained from the data of DePhillips and Lenhoff [46]. Although the material used in this paper was neither SP nor CM Sepharose, an approximation can be made for NHS-Sephasore using this data. Interpolating the data in ref [46] the phase ratio for lysozyme is approximately 42.5 $m^2\,ml^{-1}$. In order to obtain 15% surface coverage, the required immobilization concentration is 20 mg of lysozyme ml^{-1} of settled particle for lysozyme.

References

[1] A. McPherson, Methods Enzymol. 114 (1985) 112.
[2] A. George, W.W. Wilson, Acta Cryst. D 50 (1994) 361.
[3] A. George, Y. Chiang, B. Guo, A. Arabshahi, Z. Chi, W.W. Wilson, Methods Enzymol. 276 (1997) 100.
[4] B.L. Neal, D. Asthagiri, O.D. Velev, A.M. Lenhoff, E.W. Kaler, J. Cryst. Growth 196 (1999) 377.
[5] W. McMillan, J. Mayer, J. Chem. Phys. 13 (1945) 276.
[6] B.L. Neal, D. Asthagiri, A.M. Lenhoff, Biophys. J. 75 (1998) 2469.
[7] D. Asthagiri, B.L. Neal, A.M. Lenhoff, Biophys. Chem. 78 (1999) 219.
[8] D.F. Rosenbaum, C.F. Zukoski, J. Cryst. Growth 169 (1996) 752.
[9] O.D. Velev, E.W. Kaler, A.M. Lenhoff, Biophys. J. 75 (1998) 2682.
[10] P. Prinsen, T. Odijk, J. Chem. Phys. 121 (2004) 6525.
[11] B. Guo, S. Kao, H.M. McDonald, A. Asanov, L.L. Combs, W.W. Wilson, J. Cryst. Growth 196 (1999) 424.
[12] C. Hitscherich, J. Kaplan, M. Allaman, J. Wiencek, P.J. Loll, Protein Sci. 9 (2000) 1559.
[13] D.F. Rosenbaum, A. Kulkarni, S. Ramakhishnan, C.F. Zukoski, J. Chem. Phys. 111 (1999) 9882.
[14] R. Piazza, M. Pierno, J. Phys: Condens. Matter 12 (2000) 443.
[15] M. Muschol, F. Rosenberger, J. Chem. Phys. 103 (1995) 10424.
[16] D.F. Rosenbaum, P.C. Zamora, C.F. Zukoski, Phys. Rev. Lett. 76 (1996) 150.
[17] W.W. Wilson, J. Struct. Biol. 142 (2003) 56.
[18] H.V. Porschel, G. Damaschun, Stud. Biophys. 62 (1977) 69.
[19] A.F. Ducruix, J.P. Guilloteau, M.M. Reis-Kautt, A. Tardieu, J. Cryst. Growth 168 (1996) 28.
[20] F. Bonnete, S. Finet, A. Tardieu, J. Cryst. Growth 196 (1999) 403.
[21] R.A. Curtis, J.M. Prausnitz, H.W. Blanch, Biotechnol. Bioeng. 57 (1998) 11.
[22] R.A. Curtis, J. Ulrich, A. Montaser, J.M. Prausnitz, H.W. Blanch, Biotechnol. Bioeng. 79 (2002) 367.
[23] C. Gripon, L. Legrand, I. Rosenman, O. Vidal, M.C. Robert, F. Boue, J. Cryst. Growth 178 (1997) 575.
[24] V. Receveur, D. Durand, M. Desmadril, P. Clamettes, FEBS Lett. 426 (1998) 57.
[25] C.A. Haynes, K. Tamura, H.R. Korfer, H.W. Blanch, J.M. Prausnitz, J. Chem. Phys. 96 (1992) 905.
[26] H.M. Schaink, J.A.M. Smith, Phys. Chem. Chem. Phys. 2 (2000) 1537.
[27] P.E. Pjura, A.M. Lenhoff, S.A. Leonard, A.G. Gittis, J. Mol. Biol. 300 (2000) 235.
[28] J. Behlke, O. Ristau, Biophys. Chem. 79 (1999) 13.
[29] S.Y. Patro, T.M. Przybycien, Biotechnol. Bioeng. 52 (1996) 193.
[30] P.M. Tessier, A.M. Lenhoff, S.I. Sandler, Biophys. J. 82 (2002) 1620.
[31] P.M. Tessier, S.D. Vandrey, B.W. Berger, R. Pazhianur, S.I. Sandler, A.M. Lenhoff, Acta Cryst. D 58 (2002) 1531.
[32] P.M. Tessier, H.R. Johnson, R. Pazhianur, B.W. Berger, J.L. Prentice, B.J. Bahnson, S.I. Sandler, A.M. Lenhoff, Proteins: Struct. Funct. Gen. 50 (2003) 303.
[33] C.D. Garcia, S.C. Holman, C.S. Henry, W.W. Wilson, Biotechnol. Prog. 19 (2003) 575.
[34] C.D. Garcia, D.J. Hadley, W.W. Wilson, C.S. Henry, Biotechnol. Prog. 19 (2003) 1006.
[35] C.A. Teske, H.W. Blanch, J.M. Praustinz, J. Phys. Chem. B 108 (2004) 7437.
[36] J. Bloustine, V. Berejnov, S. Fraden, Biophys. J. 85 (2003) 2619.
[37] P.M. Tessier, A.M. Lenhoff, Curr. Opin. Biotechnol. 14 (2003) 512.
[38] C.A. Teske, H.W. Blanch, J.M. Praustinz, Fluid Phase Equilib. 219 (2004) 139.
[39] W.G. McMillan, J.E. Mayer, J. Chem. Phys. 13 (1945) 276.
[40] B.H. Zimm, J. Chem. Phys. 16 (1948) 1093.
[41] P.J. Wyatt, Anal. Chem. Acta 272 (1993) 1.
[42] L.W. Nichol, R.J. Seizen, D.J. Winzor, Biophys. Chem. 9 (1978) 47.
[43] J.R. Taylor, An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, University Science Books, Mill Valley, California, 1982, p. 159.
[44] J. Stahlberg, B. Jonsson, C. Horvath, Anal. Chem. 63 (1991) 1867.
[45] B.H. Zimm, J. Chem. Phys. 14 (1946) 164.
[46] B.L. Neal, A.M. Lenhoff, AIChE J. 41 (1995) 1010.
[47] P. DePhillips, A.M. Lenhoff, J. Chromatogr. A 883 (2000) 39.
[48] P. Cuatrecasas, I. Parikh, Biochemistry 11 (1972) 2291.
[49] A.P.G. van Sommeren, P.A.G.M. Machielsen, T.C.J. Gribnau, J. Chromatogr. 639 (1993) 23.
[50] P.K. Smith, R.I. Krohn, G.T. Hermason, A.K. Malia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Anal. Biochem. 150 (1985) 76.
[51] A.L. Plant, L. Locascio-Brown, W. Haller, R.A. Durst, Appl. Biochem. Biotechnol. 30 (1991) 83.
[52] P. Retailleau, M. Reis-Kautt, A. Ducruix, Biophys. J. 73 (1997) 2156.
[53] V.V. Barynin, V.R. Melik-Adamyan, Sov. Phys. Crystallogr. 27 (1982) 588.
[54] A.M. Lenhoff, P.E. Pjura, J.G. Dilmore, T.S. Godlewski, J. Cryst. Growth 180 (1997) 113.
[55] M.M. Reis-Keutt, A.F. Ducruix, J. Biol. Chem. 264 (1989) 745.
[56] S. Ruppert, S.I. Sandler, A.M. Lenhoff, Biotechnol. Prog. 17 (2000) 182.
[57] M.L. Broide, T.M. Tominc, M.D. Saxowsky, Phys. Rev. E 53 (1996) 6325.
[58] M. Malfois, F. Bonnete, J. Chem. Phys. 105 (1996) 3290.
[59] W.W. Wilson, Methods: Companion Methods Enzymol, Academic Press, New York, 1990, p. 110.