ISSN 0021-9673

VOL. 705 NO. 1 23 JUNE 1995

Analytical Biotechnology

JOURNAL OF CHROMATOGRAPHY A

INCLUDING ELECTROPHORESIS AND OTHER SEPARATION METHODS

ELSEVIER

JOURNAL OF CHROMATOGRAPHY A

INCLUDING ELECTROPHORESIS AND OTHER SEPARATION METHODS

Scope. The *Journal of Chromatography A* publishes papers on all aspects of **chromatography, electrophoresis** and related methods. Contributions consist mainly of research papers dealing with chromatographic theory, instrumental developments and their applications. In the *Symposium volumes*, which are under separate editorship, proceedings of symposia on chromatography, electrophoresis and related methods are published. *Journal of Chromatography B: Biomedical Applications* —This journal, which is under separate editorship, deals with the following aspects: developments in and applications of chromatographic and electrophoretic techniques related to clinical diagnosis or alterations during medical treatment; screening and profiling of body fluids or tissues related to the analysis of active substances and to metabolic disorders; drug level monitoring and pharmacokinetic studies; clinical toxicology; forensic medicine; veterinary medicine; occupational medicine; results from basic medical research with direct consequences in clinical practice.

Submission of Papers. The preferred medium of submission is on disk with accompanying manuscript (see *Electronic manuscripts* in the Instructions to Authors, which can be obtained from the publisher, Elsevier Science B.V., P.O. Box 330, 1000 AH Amsterdam, Netherlands). Manuscripts (in English; *four* copies are required) should be submitted to: Editorial Office of *Journal of Chromatography A*, P.O. Box 681, 1000 AR Amsterdam, Netherlands, Telefax (+31-20) 485 2304, or to: The Editor of *Journal of Chromatography B: Biomedical Applications*, P.O. Box 681, 1000 AR Amsterdam, Netherlands. Review articles are invited or proposed in writing to the Editors who welcome suggestions for subjects. An outline of the proposed review should first be forwarded to the Editors for preliminary discussion prior to preparation. Submission of an article is understood to imply that the article is original and unpublished and is not being considered for publication elsewhere. For copyright regulations, see below.

Publication information. *Journal of Chromatography A* (ISSN 0021-9673): for 1995 Vols. 683–714 are scheduled for publication. *Journal of Chromatography B: Biomedical Applications* (ISSN 0378-4347): for 1995 Vols. 663–674 are scheduled for publication. Subscription prices for *Journal of Chromatography A*, *Journal of Chromatography B: Biomedical Applications* or a combined subscription are available upon request from the publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries where air delivery via SAL is ensured: Argentina, Australia, Brazil, Canada, China, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request. Claims for missing issues must be made within six months of our publication (mailing) date. Please address all your requests regarding orders and subscription queries to: Elsevier Science B.V., Journal Department, P.O. Box 211, 1000 AE Amsterdam, Netherlands. Tel.: (+31-20) 485 3642; Fax: (+31-20) 485 3598. Customers in the USA and Canada wishing information on this and other Elsevier journals, please contact Journal Information Center, Elsevier Science Inc., 655 Avenue of the Americas, New York, NY 10010, USA, Tel. (+1-212) 633 3750, Telefax (+1-212) 633 3764.

Abstracts/Contents Lists published in Analytical Abstracts, Biochemical Abstracts, Biological Abstracts, Chemical Abstracts, Chemical Titles, Chromatography Abstracts, Current Awareness in Biological Sciences (CABS), Current Contents/Life Sciences, Current Contents/Physical, Chemical & Earth Sciences, Deep-Sea Research/Part B: Oceanographic Literature Review, Excerpta Medica, Index Medicus, Mass Spectrometry Bulletin, PASCAL-CNRS, Referativnyi Zhurnal, Research Alert and Science Citation Index.

US Mailing Notice. *Journal of Chromatography A* (ISSN 0021-9673) is published weekly (total 52 issues) by Elsevier Science B.V., (Sara Burgerhartstraat 25, P.O. Box 211, 1000 AE Amsterdam, Netherlands). Annual subscription price in the USA US$ 5389.00 (US$ price valid in North, Central and South America only) including air speed delivery. Second class postage paid at Jamaica, NY 11431. **USA POSTMASTERS**: Send address changes to *Journal of Chromatography A*, Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003. Airfreight and mailing in the USA by Publications Expediting.

See inside back cover for Publication Schedule, Information for Authors and information on Advertisements.

0021-9673/95/$09.50

Printed in the Netherlands

For Contents see p. VII.

AVAILABLE AT YOUR FINGERTIPS:

ELSEVIER SCIENCE COMPLETE CATALOGUE

NOW AVAILABLE:

THE 1995 ON CD-ROM

These catalogues feature all journals, books and major reference works from Elsevier Science. Furthermore they allow you to access information about the electronic and CD-ROM products now published by Elsevier Science. Demonstration examples of some of these products are included.

Features include:

- All the journals, with complete information about journal editors and editorial boards
- Listings of special issues and volumes
- Listings of recently published papers for many journals
- Complete descriptions and contents lists of book titles
- Clippings of independent reviews of published books
- Book series, dictionaries, reference works
- Electronic and CD-ROM products
- Demonstration versions of electronic products
- Free text search facilities
- Ordering facilities
- Print options
- Hypertext features

ELSEVIER SCIENCE

Ci

Catalogue on **INTERNET**

ELSEVIER SCIENCE

Ccd

Catalogue on **CD-ROM**

Extra features with the Catalogue on Internet

- Alerting facility for new & forthcoming publications
- Updated monthly

ELSEVIER SCIENCE COMPLETE CATALOGUE INTERNET: TRY IT TODAY!

gopher to: gopher.elsevier.nl
WWW: http://www.elsevier.nl/

CD-ROM (published yearly, free of charge)

Please contact:

Customer Service Department
Tel.: +31 (20) 485 3757
Fax: +31 (20) 485 3432
e-mail: nlinfo-f@elsevier.nl

ELSEVIER | PERGAMON | NORTH HOLLAND | EXCERPTA MEDICA

JOURNAL OF CHROMATOGRAPHY A

VOL. 705 (1995)

JOURNAL OF CHROMATOGRAPHY A

INCLUDING ELECTROPHORESIS AND OTHER SEPARATION METHODS

ELSEVIER

Amsterdam – Lausanne – New York – Oxford – Shannon – Tokyo

0021-9673/95/$09.50

Printed in the Netherlands

SPECIAL ISSUE

ANALYTICAL BIOTECHNOLOGY

Guest Editors

G.C. DAVIS
(Indianapolis, IN, USA)

R.M. RIGGIN
(Indianapolis, IN, USA)

VOL. 705, NO. 1 JOURNAL OF CHROMATOGRAPHY A 23 JUNE 1995

CONTENTS

(Abstracts/Contents Lists published in Analytical Abstracts, Biochemical Abstracts, Biological Abstracts, Chemical Abstracts, Chemical Titles, Chromatography Abstracts, Current Awareness in Biological Sciences (CABS), Current Contents/Life Sciences, Current Contents/Physical, Chemical & Earth Sciences, Deep-Sea Research/Part B: Oceanographic Literature Review, Excerpta Medica, Index Medicus, Mass Spectrometry Bulletin, PASCAL-CNRS, Referativnyi Zhurnal, Research Alert and Science Citation Index)

TOPICAL ISSUE ON: ANALYTICAL BIOTECHNOLOGY

ELSEVIER

Journal of Chromatography A, 705 (1995) 1–2

JOURNAL OF CHROMATOGRAPHY A

Preface

The term "analytical biotechnology" has been used to describe the analytical characterization of biosynthetic products, especially proteins. This field has become increasingly important due to the rapid growth and success of the biopharmaceutical industry. Analytical characterization standards for pharmaceutical products are rigorous, and when applied to proteins, meeting these standards requires the use of state of the art technology. Over the past 10–15 years revolutionary advances have been made in analytical technology for protein characterization. HPLC applications for proteins have continued to expand with the advent of rapid separation systems. Capillary electrophoresis has become a useful tool, supplementing HPLC and traditional electrophoresis separation capabilities. Advances have been made in trace level N-terminal sequencing technology, and practical strategies for C-terminal sequencing have become available. Mass spectrometry has become routinely used for protein characterization, greatly increasing our ability to detect small structural changes in proteins (e.g. chemical degradation events during product storage).

This special issue consists of a collection of papers which review some of the more important analytical techniques in this field, with an emphasis on separations, as well as papers describing specific applications. The field of high speed HPLC separations is reviewed by Chen and Horváth. Nguyen et al. review the field of protein mass spectrometry, and include a number of recent, specific examples demonstrating the power of MS in determining protein structure. Bailey reviews the important field of chemical sequencing methods for proteins —both N-terminal and C-terminal strategies. The use of isoelectric focusing (IEF), with particular emphasis on the determination of post-translational processing and chemical modifications of proteins is covered by Gianazza. IEF, while not a quantitative technique, has the advantages of being widely applicable to proteins and offering good resolution of closely related structures. Hu describes a very recent technique, fluorophore-assisted carbohydrate electrophoresis (FACE) which is now widely used to characterize the oligosaccharide structures comprising glycoproteins. Eaton reviews the strategies which have been used to determine the levels of host cell protein (HCP) contaminants in protein pharmaceutical products. The determination of HCPs is challenging since one is faced with the task of detecting a complex mixture of proteins at trace levels in the presence of a large excess of the protein product. Thus both good sensitivity and definable selectivity, for a group of complex analyte proteins are requirements for such assays. The strategies for accurately determining moisture content in proteins, which is a rather more complex task than one might first assume, is reviewed by Towns. Protein stability is generally highly dependent on moisture level, hence moisture content must be accurately measured and controlled. Proteins are hygroscopic and thus readily adsorb or release moisture to the atmosphere, depending on the relative humidity in the surrounding environment.

Harris describes a finding with regard to pro-

SSDI 0021-9673(94)01248-2

tein structures from mammalian cell cultures, the post-translational removal of C-terminal lysine and arginine residues, presumably due to the presence of carboxypeptidase in the process. This finding offers a unique analytical challenge, given limited capabilities in C-terminal sequence analysis techniques. Kalman et al., describe a practical application of capillary electrophoresis to the characterization of protein mutants. Strege and Lagu describe the use of micellar HPLC to the characterization of human growth hormone in fermentation samples. An application of a biospecific affinity sensor for the detection of neu differentiation factor in chromatographic separations is presented by Liu et al. This technique offers the ability to rapidly detect trace levels of a protein with high specificity, based on binding to a selective ligand (e.g. an antibody or receptor).

As one can see from the diversity of analytical methods described in this special issue, the amount of effort required to characterize biosynthetic products is substantial. While it may be hoped that as these methods become more sophisticated, the number of methods needed to describe the purity, identity, and potency of biopharmaceuticals will decrease, the reality may be just the opposite. For the more we learn, the more complex the questions and the further need to push our analytical capabilities.

Indianapolis, IN, USA

G.C. Davis
R.M. Riggin

ELSEVIER

Journal of Chromatography A, 705 (1995) 3–20

JOURNAL OF CHROMATOGRAPHY A

High-speed high-performance liquid chromatography of peptides and proteins

H. Chen, Cs. Horváth*

Department of Chemical Engineering, Yale University, New Haven, CT 06520, USA

Abstract

Over the last thirty years the name HPLC has been synonymous with high-speed liquid chromatography and during the last ten years we have experienced a dramatic increase in the speed of analysis particularly as far as the separation of biological macromolecules, such as proteins, is concerned. With a solid grounding in the chromatographic theories, column technology has been mainly responsible for the advances in this field. Recent development shows that columns packed with micropellicular or gigaporous stationary phases of the bidisperse or the bimodal type facilitate rapid mass transfer between the mobile and stationary phases and thus can deliver high resolution separations in a very short time. This suggests that HPLC has the potential to be the prime analytical technique for on-line monitoring of biotechnological processes in real time. Further enhancement of the speed of separation comes from the use of elevated temperatures. The role of temperature in HPLC has largely been ignored and most commercial instruments are not equipped with appropriate temperature control. Results presented here strongly suggest, however, that elevated column temperature may find increasing use in the HPLC of large molecules. In such analytical applications temperature programming may also play a major role provided columns with low heat capacity, such as packed fused-silica capillaries, gain wider employment in HPLC.

Contents

* Corresponding author.

0021-9673/95/$29.00
SSDI 0021-9673(94)01254-7

1. Introduction

The development of the technique, which we call by the acronym HPLC, commenced about 30 years ago [1]. At that time the high pressure was the dominant new operational and instrumental characteristic and the relatively high speed of the technique was the most important measure of performance. For this reason the name "high-speed liquid chromatography" was also used concurrently with "high-performance liquid chromatography" to denote the new technique [2,3]. Although it has acquired over the years all features associated with high performance, the high speed of separation is still a cardinal feature of HPLC.

High-speed HPLC has been widely used in routine analytical work, method development, process monitoring and quality control, and further progress in these area depends on the development of efficient means for more rapid separations. It is particularly important due to the sequential nature of the separation by column chromatography that allows the separation of only one sample at the time in contrast to planar chromatographic and electrophoretic techniques which facilitates simultaneous separation of several samples. With advances in data acquisition and processing, sample preparation and the chromatographic separation steps are the weakest links in the chain of analytical information gathering and transfer as shown in Fig. 1. Since the preparation of a large number of samples can be carried out parallel, however, the ultimate limiting step remains the chromatographic separation per se.

Recently, process monitoring and control have drawn attention in biotechnology in response to the requirements by federal regulatory agencies [4]. At the present, mostly off-line analytical procedures [5], which may take several hours or even days, are employed for "in-process analysis". The need for on-line monitoring by rapid analytical methods is increasingly recognized to carry out real time monitoring and thus avoid the loss of product and to reduce the total process time and cost [6–8].

Recent advances in HPLC with concomitant

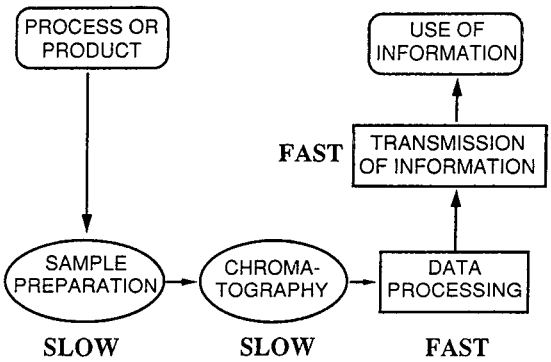


Fig. 1. Schematic illustration of the sample and information flow. Due to the sequential nature of sample handling by column chromatography, HPLC separation is usually the rate-limiting step of the processes. Sample preparation can be carried out in a parallel fashion so that a large number of samples can be simultaneously processed.

enhancement of the separation speed and efficiency strongly suggest that HPLC has the potential to become an important method for on-line monitoring [9]. Although much is expected from the development of various sensors [10] for such purposes, the capability for rapid sequential multicomponent analysis makes a suitable liquid chromatograph to be the ultimate sensor, particularly when the advances anticipated in miniaturization lead to a new generation of instruments. As a separation technique with high resolving power, HPLC is superior in monitoring the concentration of several components over other conventional on-line methods, such as spectroscopic or electrochemical measurements. In the past few years, different approaches have already been developed to monitor by HPLC the fermentation process [11–14] or the effluents from down-stream processing operations [15–17]. By and large HPLC is poised to become a quasi-real time monitoring tool for processes in biotechnology.

Rapid analytical HPLC methods will offer significant advantages also in routine analytical work in the laboratory. Furthermore, analytical HPLC is increasingly used in molecular chromatography to obtain information on the structure of biopolymers, such as proteins, and to study

molecular interactions as well as the chromatographic process proper at the molecular level [18–20]. By increasing the speed of analysis, the analytical productivity and efficiency of both the operator and the instrument can greatly be improved. On the other hand, multidimensional liquid chromatography with coupled columns [21] entails a large number of individual chromatographic runs and is therefore, greatly limited today by the total amount of time required for a complete analysis. Further reduction in the time required for each of the separation steps is necessary to develop the full potential of this highly promising technique and to make multidimensional HPLC, with appropriate automated instrumentation, a powerful analytical method that can be widely used for the determination of the composition of samples containing many components.

In the present report, we shall discuss recent advances and some approaches to achieve rapid analysis of peptides and proteins by HPLC, the peculiarities of the attendant instrumentation and the operational parameters of high-speed HPLC.

2. Theory

The time of chromatographic analysis can be defined as the retention time, t_R, of the most retained sample component. Under isocratic conditions it is given by

$$t_R = \frac{L}{u} \cdot (1 + k') \tag{1}$$

where k' is the retention factor of the last peak, L is the column length and u is the linear flow velocity of the mobile phase. The column length depends on the number of theoretical plates, N, required for the multicomponent separation and on the reduced plate height, h, as

$$L = Nd_p h \tag{2}$$

where d_p is the particle diameter. On the other hand u can be expressed by the reduced flow velocity, ν, which is also termed Peclet number in the chemical engineering literature, as

$$u = \frac{D_m \nu}{d_p} \tag{3}$$

where D_m is the eluite diffusivity in the mobile phase.

According to Darcy's law, the pressure drop across the column is

$$\Delta P = \frac{uL\eta\psi}{d_p^2} \tag{4}$$

where ΔP is the pressure drop and η is the viscosity of mobile phase. The parameter ψ is given by $180(1 + \omega)(1 - \epsilon)^2/\epsilon^2$, where ϵ is the interstitial porosity, 0.4 for random packing, and ω is the volume ratio of the intraparticulate and interstitial void spaces. The latter can be assumed to be unity for totally porous and negligibly small for pellicular sorbents.

For fixed pressure drop and plate number, the shortest analysis time, t_R^*, is achieved at the optimum flow velocity, ν_0 [22] at which the reduced plate height has its minimum value, h_0. Under such conditions t_R^* can be expressed as

$$t_R^* = \frac{N^2 h_0^2 \eta\psi}{\Delta P} \cdot (1 + k') \tag{5}$$

and the particle diameter d_p is obtained as

$$d_p = \sqrt{\frac{h_0 \nu_0 N D_m \eta\psi}{\Delta P}} \tag{6}$$

In the chromatography of small molecules, D_m can be taken as $1 \cdot 10^{-5}$ cm^2/s. If we set N to 5000, ΔP to 300 bar, k' to 3, h_0 to 2, ν_0 to 3 and η to 10^{-2} g cm^{-1} s^{-1}, the separation time calculated from Eq. 5 is approximately 10 s. In this particular case the column length, flow velocity and particle size were evaluated as 1 cm, 0.3 cm/s and 1 μm, respectively, by using Eqs. 2, 3 and 6 with the minimum reduced plate height and optimum reduced velocity.

In the chromatography of large molecules having much smaller diffusivities, however, the requirements may change dramatically. If we use D_m of $5 \cdot 10^{-7}$ cm^2/s instead of $1 \cdot 10^{-5}$ cm^2/s in the case above, we find that for an analysis time

of 10 s, the column length, flow velocity and particle size typically required in macromolecular chromatography would be 0.2 cm, 0.08 cm/s and 0.2 μm, respectively. Whereas such particles are too small for use in packed columns at present, a thick membrane or a stack of thin membranes having 0.2-μm pores, suitable retentive properties, and favorable pore morphology could serve as a chromatographic column for such rapid separations [23].

In practice, the separation of large molecules is generally carried out by using gradient elution. Under linear solvent strength gradient conditions, the analysis time can be estimated by the relationship [24]

$$t_R \approx \frac{L}{u} + \frac{t_G}{1 - \frac{k_f}{k_0}} \tag{7}$$

where t_G is the gradient time and k_0 and k_f are the respective isocratic retention factors of the most retained sample component at the initial and final mobile phase compositions.

Under gradient conditions the plate number of the column can be estimated at high flow velocities by using the expression [25]

$$N \approx \frac{t_G D_m}{1.15 C' d_p^2 \bar{k} \log \frac{k_0}{k_f}} \tag{8}$$

where $\bar{k}$ is the retention factor at the midpoint of the column and the parameter C' depends on the porosity and structure of the packing, the equilibrium constant and the stationary phase configuration, which represents the particular form and structure of the support and the stationary phase proper. Certain novel stationary phase configurations, such as pellicular, gigaporous, and gel-in-a-shell type are of interest in high-speed HPLC of large molecules and will be discussed later.

At fixed plate number and the pressure drop, the advantages of small particle size for a given separation are evident from Eqs. 4 and 8. Both the column length and gradient time decrease with d_p^2 and as a result the reduction in d_p yields fast separations as seen from Eq. 7. However, there is a lower practical limit for the particle diameter and therefore further gains would require the use of stationary phases of special configuration that yields small values of parameter C' in Eq. 8. The resolution in gradient elution does not depend on the column length and the flow velocity [25], thus, flow velocity can be increased, and at the same time the column length reduced, to achieve fast separations at a fixed column inlet pressure.

As seen from Eq. 4, the mobile phase velocity, at the fixed pressure drop, can be increased upon reducing the eluent viscosity by using elevated temperature and thus practicing "superheated" liquid chromatography [26] or even supercritical fluid chromatography with eluents commonly employed in HPLC. Such conditions also lead to an enhancement of the eluite diffusivity in the mobile phase, thus to a reduction of the time of analysis with the same column at a fixed plate number according to Eq. 8 provided the stationary phase is stable at elevated temperatures.

In summary, the employment of small particles, stationary phase configurations with favorable mass transfer properties, and elevated column temperatures combined with the use of steep gradients, high flow velocities, and short columns can be a very effective means to bring about rapid separations of large molecules, such as peptides and proteins.

3. Column design

3.1. Particle size

The effect of the particle size on the analysis time is well understood when using columns packed with conventional sorbents [22,27–30]. Small particles yield enhanced column efficiency by virtue of the relatively small intraparticulate mass transfer resistance due to the short diffusion distances and to a lesser extent due to the small contribution of "eddy diffusion" to the plate height [31]. The enhancement of intraparticular mass transport is particularly important for the rapid separation of large molecules having low diffusivities.

For fixed plate number the retention time can be calculated as a function of particle diameter d_p by combining Eqs. 1–3 as

$$t_R = \frac{(1+k')Nh}{D_m \nu} \cdot d_p^2 \tag{9}$$

As we mentioned in the previous section, high-speed chromatography of large molecules is often carried out at high reduced velocities. Under such conditions the magnitude of intraparticle diffusion resistances in columns packed with porous adsorbents determines the dependence of the reduced plate height on the reduced velocity. In certain practical cases h can be considered to be linearly dependent on ν and the slope of the h vs. ν plot at $k' = 3$ is found to be approximately 0.3 [32]. The results obtained by using Eq. 9 with this assumption and for a given set of practical conditions used in macromolecular separations are shown in Fig. 2. It is seen that columns packed with 1–3-μm monodisperse particles can yield 2000 theoretical plates within a few minutes in the separation of large molecules.

Columns packed with 1.5-μm [33] and 2-μm [34] pellicular as well as 2-μm [35] porous stationary phase particles have been employed in rapid analysis of peptides and proteins. Columns packed with small particles, however, have low permeability and require high column inlet pressure at flow velocities employed for fast separations. The inlet pressure required for columns packed with conventional porous spherical stationary phase particles and operated at high reduced velocities can be determined by

$$\Delta P = \frac{L^2 \eta D_m \psi}{0.3 N d_p^4} \tag{10}$$

The dependence of the column inlet pressure on the particle diameter was calculated by Eq. 10 and the results are also depicted in Fig. 2. Since the pressure is inversely proportional to the fourth power of particle diameter, it increases rapidly with decreasing particle size. As a result the potential of microparticulate stationary phases in HPLC of large molecules is limited by the low column permeability. In such cases, it may be particularly appropriate to reduce the viscosity of the mobile phase by carrying out chromatography at elevated column temperatures with columns having adequate stability for operation at high temperatures. The peculiarities of HPLC at high temperatures will be discussed later in this paper.

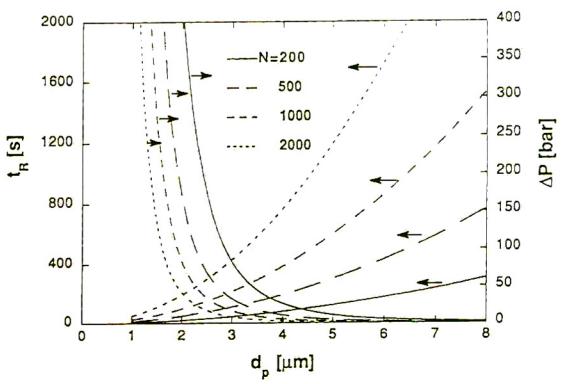


Fig. 2. Dependence of analysis time and pressure drop on particle size at a given plate number for totally porous materials. $k' = 3$, $h = 0.3\nu$, $D_m = 5 \cdot 10^{-7}$ cm^2/s, $L = 3$ cm, $\eta = 0.9$ cP, $\epsilon = 0.4$, and the required N is shown in the figure.

3.2. Stationary phases of novel configuration

In addition to the particle size, the configuration of the stationary phase also plays an important role in determining the speed and efficiency of separation. For rapid HPLC stationary phases of novel configuration, e.g., micropellicular, gigaporous, or gel-in-a-shell type were developed during the past few years and found to have mass transfer properties superior to conventional column packings. Indeed, columns packed with such stationary phases are eminently suitable for the HPLC of biological macromolecules at high efficiency and high speed.

3.2.1. Micropellicular stationary phases

Pellicular stationary phases with an average particle size of 40 μm were first introduced in the mid-1960s [36] after the felicitous experience gained with such stationary phase structure in gas chromatography [37]. Whereas pellicular

stationary phases dominated HPLC in the early years, they were largely replaced in the seventies by totally porous microparticulate bonded phases for the separations of small molecules upon the availability of high performance air classifiers for subsieve particles. Micropellicular stationary phases having small particle diameter (d_p = 1.5–5 μm) shown in Fig. 3 regained interest in the middle eighties for the rapid HPLC of biopolymers [33,34,38–40] and are expected to play an important role in analytical work.

As shown in Fig. 3 the configuration of the actual stationary phase in micropellicular packings is a spherical annulus supported by a fluid impervious, rigid microsphere. The main advantage of such sorbents rests with the rapid mass transfer between the mobile and stationary phases because the diffusion distance in the thin retentive layer of the particles is short. Owing to the solid, fluid-impervious core of the micropellicular packings, columns are stable at high pressures and elevated temperatures unlike those packed with conventional porous materials. Moreover, they also provide improved recoveries in the separation of proteins and peptides. The cavernous interior of conventional sorbents is considered to be responsible for the entrapment of sample. The loading capacity of columns packed with micropellicular stationary phases is relatively low with small molecules, yet with macromolecules it is not much smaller than the loading capacity of totally porous sorbents

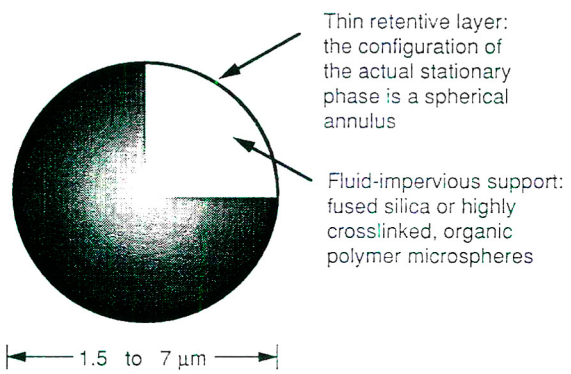


Fig. 3. A schematic illustration of the pellicular stationary phase configuration.

[41]. Because of the small particle diameter (1.5–3 μm) of the micropellicular packings, the permeability of the column is low. This impediment, however, can be mitigated by carrying out the separation at elevated temperature at which the mobile phase viscosity is reduced. The combination of elevated column temperature and the micropellicular stationary phase configuration offers a powerful means to facilitate rapid and efficient separation of large molecules [34,42]. Fast separation of a mixture of four proteins depicted in Fig. 4 was achieved in 6 s at 120°C on a 3-cm HY-TACH column packed with 2-μm pellicular ODS-silica. Fig. 5 shows peptide profiles of tryptic digests of β-lactoglobulin A and methionyl human growth hormone produced by recombinant DNA technology. The chromatograms were obtained on a HY-TACH column similar to that used in Fig. 4. Whereas the quality of the separation does not meet the high standards of the tryptic maps used in protein analy-

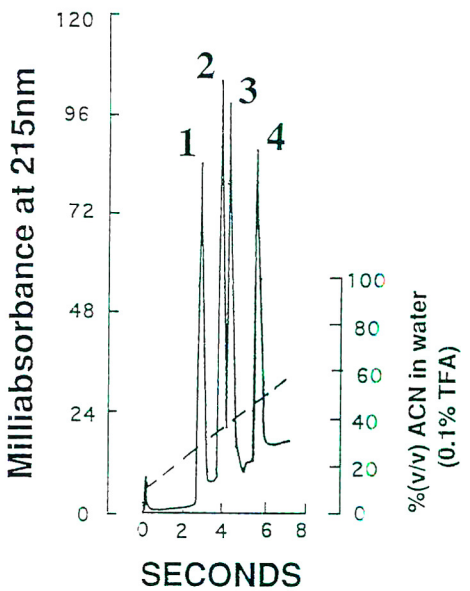


Fig. 4. Rapid separation of standard proteins. Column, 30 × 4.6 mm, packed with 2-μm pellicular ODS-silica; 12 s linear gradient from 10 to 90% (v/v) acetonitrile (ACN) in water containing 0.1% (v/v) trifluoroacetic acid (TFA); temperature, 120°C; flow-rate, 5 ml/min; column inlet pressure, 240 bar. Peaks: 1 = ribonuclease A; 2 = cytochrome *c*; 3 = lysozyme; 4 = β-lactoglobulin B. From Ref. [42].

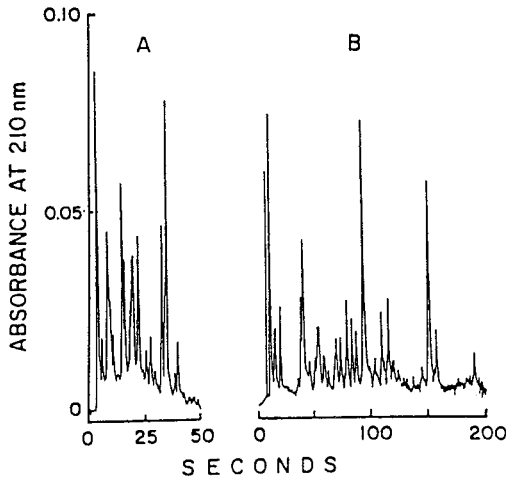


Fig. 5. Chromatographic profiles of tryptic digests of (A) β-lactoglobulin A and (B) methionyl human growth hormone. Column, 30 × 4.6 mm, packed with 2-μm pellicular ODS-silica; linear gradient from 0 to 95% (v/v) acetonitrile in water containing 0.1% (v/v) TFA in (A) 2 min at 5 ml/min and (B) 6 min at 4 ml/min; temperature, 80°C; samples, 5 μg of protein digest each. From Ref. [39].

sis, the time of separation is only a very small fraction of that required by conventional means.

3.2.2. Gigaporous packings

More recently, supports having pore diameters greater than one hundredth of the particle diameter were introduced for fast protein chromatography [43–45]. So far two major types of such gigaporous stationary phases have received great attention. One is prepared by using a bidisperse rigid gigaporous support with an appropriate retentive surface and the column is operated under conditions when mass transfer inside the stationary phase particles involves both intraparticle convection in the gigapores and diffusive transport in the smaller pores. The convective transport in the gigapores is often referred to in the literature as perfusion and such particles are termed perfusive [43]. The other type has a so called "gel-in-a-shell" configuration which is described as a rigid gigaporous support particle filled with a retentive hydrogel [46]. Unlike conventional packed columns, the efficiency of columns packed with the first type of stationary phase does not deteriorate much at high flow velocities even with proteins under conditions of no adsorption. This feature which greatly facilitates high-speed separation of biological macromolecules is attributed to intraparticle convection in the gigapores [43,47]. A schematic illustration of the cross-section of such stationary phase particles is presented in Fig. 6. Columns with gigaporous packings of this type usually have significantly lower loading capacities despite the presence of the small pore regions than those packed with conventional porous column materials.

The favorable mass transfer properties of gigaporous supports manifest in reduced intraparticle mass transfer resistances so that the corresponding reduced plate height increment, h_{intra}, is given by [46]

$$h_{intra} = C \cdot \frac{D'_e}{D_{app}} \cdot \nu \tag{11}$$

As discussed earlier, the magnitude of h_{intra} dominates the reduced plate height at high flow velocities in columns packed with conventional porous materials and the value of h_{intra} is negligible in columns packed with pellicular sorbents. The effects of intraparticle convection is measured appropriately by the relation between the effective diffusivity in the gigapores, D'_e, and the apparent effective diffusivity, D_{app}, which incorporates the effects of both diffusion and convection in the gigapores of a spherical par-

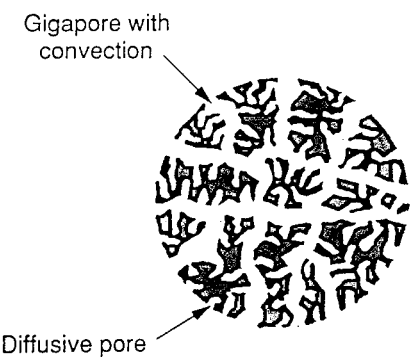


Fig. 6. A schematic illustration of the bidisperse gigaporous stationary phase configuration. After Ref. [43].

ticle [43,47,48]. The latest of these theoretical relationships [47] is given by

$$D_{app} = f\left(1 + \frac{2\nu'}{45}\right)D'_e \qquad (12)$$

where ν' is the reduced velocity in the gigapores and f is the correction factor for diffusive transport in the small pores.

The advantages of gigaporous bidisperse stationary phases are most evident at high flow velocities where the favorable effect of intraparticle convection on the apparent diffusivity is greater than at low flow velocities. The effect depends on the ratio of the diameter of the gigapores to the particle diameter that is assumed to be proportional to the square root of the ratio of the intraparticle flow velocity to the interstitial flow velocity. In Fig. 7 the theoretical dependence of h on ν is illustrated in the absence of intraparticle convection and for three cases with different gigapore to particle diameter ratios as well as for pellicular packings with a chromatographic surface completely exposed to the eluent stream perfusing the column.

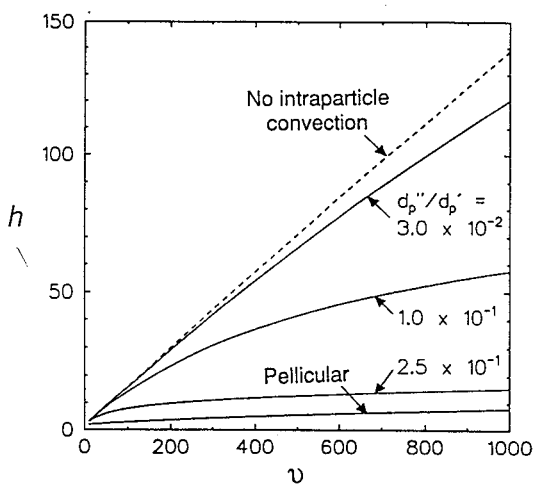


Fig. 7. Illustration of the reduced plate height for a retained eluite ($k' = 2$) as a function of the reduced velocity for various types of column packings. For gigaporous particles, the increasing level of intraparticle convection is expressed by increasing values of the ratio d_p''/d_p', which is equivalent to the ratio of the mean gigapore diameter to the mean interstitial channel diameter of the column packing. From Ref. [47].

The reduced mass transfer resistance in open gigaporous packings allows the use of relatively large particles and high flow velocities thus facilitates rapid analysis of large molecules. This is illustrated in Fig. 8 by the fast separation of a standard protein mixture on a 3 cm long column packed with 20-μm gigaporous adsorbents [16]. It has been shown that the capacity of the column packed with such gigaporous material of bidisperse pore structure does not change strongly with the flow velocity [43], and as a result,

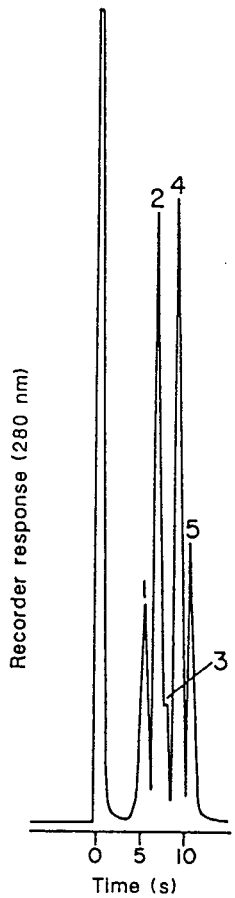


Fig. 8. Reversed-phase separation of proteins. Column, 30 × 2.1 mm, packed with 20-μm highly cross-linked styrene–divinylbenzene having 6000 to 8000 Å gigapores; gradient time, 24 s; flow-rate, 5 ml/min. Peaks: 1 = ribonuclease A; 2 = cytochrome c; 3 = lysozyme; 4 = β-lactoglobulin; 5 = ovabumin. From Ref. [16].

columns packed with such material find use in high-speed separation of peptides and proteins in both analytical and preparative/process-scale applications with modest plate requirements. Nevertheless, in the latter application, the loading capacity of columns packed with open gigaporous stationary phases may be still considered low despite their enhanced mass transfer properties due to intraparticulate convection.

The potential of gigaporous sorbents for rapid chromatography has not yet been fully exploited. For instance, gigaporous silica particles have been available for some time (cf. Table 1), yet, there is a paucity of literature data on their use for fast separation of substances having high molecular mass. Similarly, commercially available gigaporous ceramic hydroxyapatite may also be a useful stationary phase for rapid separation of proteins and nucleic acids.

A recent approach addresses the problem of low capacity by filling the open pores in the rigid support particles with a hydrogel which contains appropriate retentive sites and is permeable at least in part to the biopolymers to be separated. This novel stationary phase configuration shown in Fig. 9 is claimed to combine high loading capacity of the soft gels with the mechanical stability of the rigid particles so that the column can be operated at relatively high flow velocities [46]. Columns packed with this type of stationary phase may offer a felicitous combination of loading capacity, efficiency, speed, resolution, recovery and stability in comparison to the conventional stationary phases [45].

The novel stationary phase configurations of enhanced mass transfer properties described above are expected to find most efficient use in different types of applications where their potential can be fully exploited. Columns packed with pellicular sorbents are the best in obtaining rapid and high resolution analytical separations of biomacromolecules. Their relatively low loading capacity, however, impedes their employment in preparative separations with the exception of micropreparative applications where high speed and efficiency are required [49]. Columns packed with bidisperse gigaporous particles are designed for separations faster than those obtainable with conventional porous materials. They are particularly suited for rapid separations which do not require high efficiency as shown in Fig. 10 [47]. The advantages of such columns are manifest in high-speed analytical separations as well as in preparative-scale purifications. Columns packed with gel-in-a-shell stationary phases may have significantly higher loading capacity than those packed with open bidisperse gigaporous sorbents but their efficiency likely to diminish much faster upon increasing the flow velocity beyond a presently practical value. Thus both types of gigaporous sorbents are expected to find applications in high-speed, large-scale purification of proteins.

3.3. Column dimensions

Several studies have suggested that the resolution of macromolecules is rather insensitive to column length and flow velocity under conditions of gradient elution [25,28,50,51]. Therefore fast separation of peptides and proteins is usually carried out by gradient elution with short columns at high flow velocities often at the maximum permissible column inlet pressure of the system as discussed in the Theory section. Since the number of theoretical plates increases only with the square root of column length under isocratic conditions, also in this case the length of the column should be kept at a minimum for fast separations. The separation of β-lactoglobulin A and B within 15 s on a 1 cm long HYTACH column packed with micropellicular octadecyl-silica under isocratic elution conditions is illustrated in Fig. 11 [34]. This seems to be an extreme case of short columns because in general, 3–5 cm long columns are used for the separations of proteins in contrast with the 10–25 cm long columns employed for analysis of peptide mixtures [51].

The column diameter in the present practice of HPLC ranges mostly from 3 to 5 mm. However, the use of narrow-bore columns having a diameter between 0.5 and 2 mm and micro-bore columns with a diameter smaller than 0.5 mm is gaining significance. Columns with 1 mm inner diameter were used already in the mid-1960s [36]

Table 1
Some commercially available gigaporous silica supports

Trade name	Manufacturer	Mean pore diameter (Å)	Specific surface area (m^2/g)	Specific pore volume (ml/g)	Particle diameter (μm)	Maximum column pressure (p.s.i.)
Zorbax PSM-1000	Rockland Technologies, DE, USA	1000	15	0.75	5	4000
Nucleosil 1000	Macherey–Nagel, Düren, Germany	1000	25	0.8	5, 7, 10	3000
Nucleosil 4000	Macherey–Nagel	4000	10	0.7	5, 7, 10	3000
LiChrospher 1000	Merck, Darmstadt, Germany	1000	30	0.78	10	3000
LiChrospher 4000	Merck	4000	10	0.78	10	3000
SynChropak 1000	SynChrom, IN, USA	1000	25	0.8	7	3000
SynChropak 4000	SynChrom	4000	10	0.7	10	3000

1 p.s.i. = 6894.76 Pa.

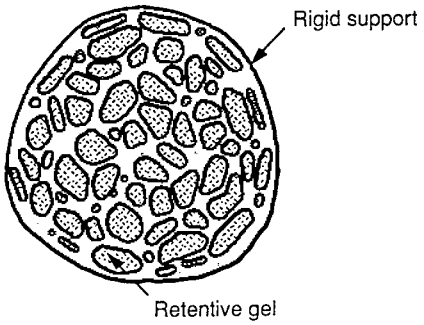


Fig. 9. Artist's rendition of the so called "gel-in-a-shell" stationary phase configuration.

but the instrumental constraints hampered the general acceptance of such narrow-bore columns at the dawn of HPLC. The broad impact of fused-silica capillaries on the practice of gas chromatography and capillary electrophoresis is likely to make capillary liquid chromatography, i.e., micro HPLC with packed quartz capillaries, the method of choice for analytical HPLC. The miniaturization of the column diameter will per se probably not augment separation speed or

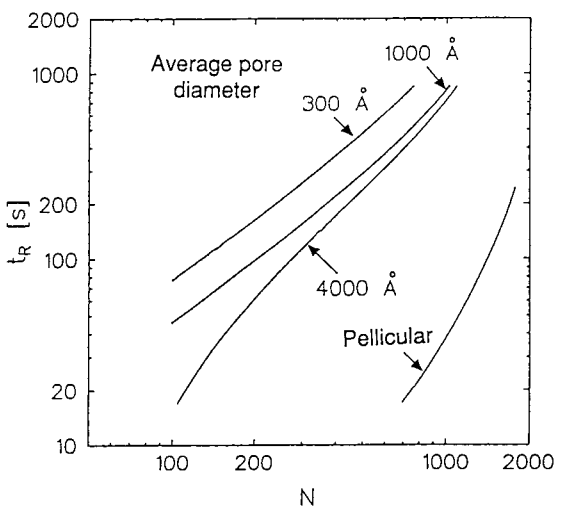


Fig. 10. Separation time versus plate number calculated for 5 cm long columns packed with 8-μm particles of various pore sizes as indicated. Conditions correspond to $D_m = 10^{-6}$ cm^2/s, molecular diameter of the eluite is 40 Å, $k' = 3.3$, and the ratio of intraparticulate and interstitial flow velocities $u'/u = 10^{-2}$. From Ref. [47].

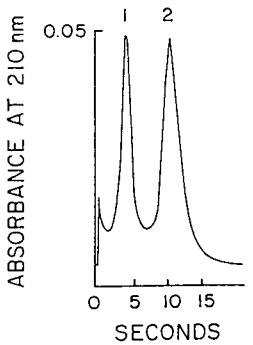


Fig. 11. Separation of β-lactoglobulin B and A by isocratic elution. Column, 10 × 4.6 mm, packed with the micropellicular sorbent; eluent, 36.75% (v/v) acetonitrile in water containing 0.1% (v/v) TFA; flow-rate, 3 ml/min; temperature, 80°C; column inlet pressure, 14.7 MPa; sample size, 10 ng; detection at 210 nm. From Ref. [34].

efficiency in a major way. Yet, there are numerous other considerations, such as environmentally correct low solvent consumption, high mass sensitivity, low heat capacity, and easy conjugation with other analytical techniques, e.g. capillary electrophoresis and mass spectrometry, that may contribute to the progress in the miniaturization of HPLC.

4. Operational parameters

4.1. Temperature

Temperature has been an underrated operational parameter in HPLC and the potential advantages of elevated column temperatures, such as enhanced kinetic and transport properties, have yet to be widely exploited for rapid analysis of biological macromolecules by HPLC. The disinterest in controlling temperature in HPLC is due to the relatively small effect of temperature on selectivity and retention in HPLC of small molecules and to the fact that the magnitude of retention in liquid chromatography, unlike in gas chromatography, can conveniently be modulated by manipulating the mobile phase composition. Thus the relatively high complexity of the instrumentation required

for appropriate temperature control in HPLC at high temperatures and the lack of columns that are stable enough to withstand the harsh conditions at elevated temperatures over an extended period of time have deterred the development of suitable equipment.

Peptides and proteins are delicate molecules which can degrade at elevated temperatures. Therefore they are usually chromatographed at ambient or even sub-ambient temperatures in order to preserve the integrity of the molecular structure. Only recently has the need for rapid analysis of biological macromolecules drawn attention to the use of elevated column temperatures, so that analytical HPLC is routinely carried out at temperatures 20–40°C above ambient temperature [51]. Results obtained by using columns packed with thermally and hydrolytically stable micropellicular stationary phases at temperatures up to 120°C [34,40,42] have demonstrated the advantages of high-temperature HPLC in rapid protein separations.

The effect of temperature on the separation of large molecules has been investigated both theoretically and experimentally [26,42]. At high temperatures the mobile phase viscosity is reduced as shown in Fig. 12, and concomitantly, the diffusivity of the eluite is enhanced. Moreover, the sorption kinetics of the eluite is also accelerated with increasing temperature. Consequently, column efficiency is expected to be higher at elevated column temperatures and this is seen in Fig. 13 from the Van Deemter plots of data obtained at high flow velocities. The effect of temperature on the separation of standard proteins and of the peptides present in the tryptic digest of β-lactoglobulin B obtained on HY-TACH columns is shown by the chromatograms in Figs. 14 and 15, respectively. The two sets of chromatograms illustrate that at super-ambient temperatures where high flow velocities are permitted the separation time was significantly reduced without compromising the resolution as discussed in the Theory section.

In addition to the above described means of reducing particle size and selecting a stationary phase configuration of relatively low mass transfer resistance (pellicular or gigaporous sorbents), the use of elevated column temperature may serve as another useful concomitant for speeding up a separation. Of course, chromatography at high temperatures is a catholic approach to rapid analyses and can be carried out by using any column packed with a sufficiently stable stationary phase.

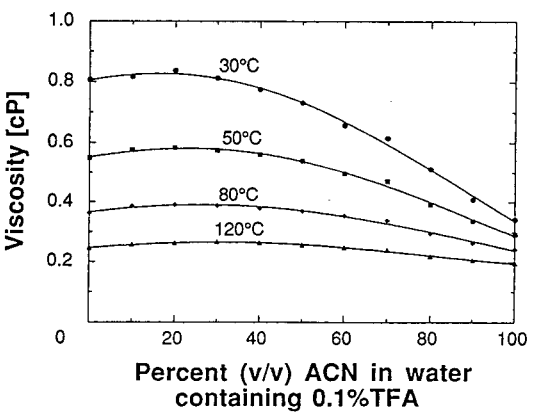


Fig. 12. Viscosity of acetonitrile–water mixtures at different temperatures as a function of the composition. A 30 × 4.6 mm column packed with 2-μm pellicular ODS-silica was used for the measurements and the mixtures contained 0.1% (v/v) TFA. From Ref. [42].

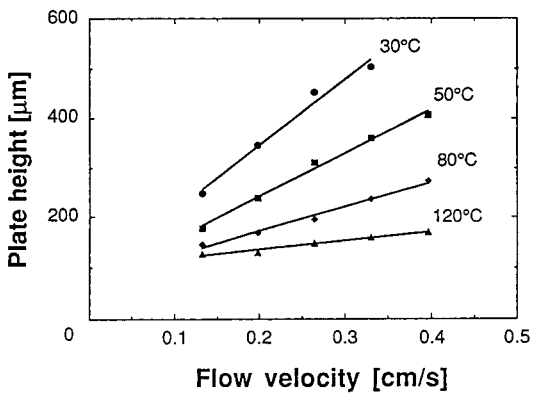


Fig. 13. Plots of plate height versus linear flow velocity measured with ribonuclease A at elevated temperatures in the domain of high reduced flow velocities. Column, 30 × 4.6 mm, packed with 5-μm macroreticular cross-linked polystyrene. Concentration (v/v) of acetonitrile in water containing 0.1% (v/v) TFA: ○ = 25%; □ = 23.5%; ◆ = 22%; ▲ = 18%. From Ref. [42].

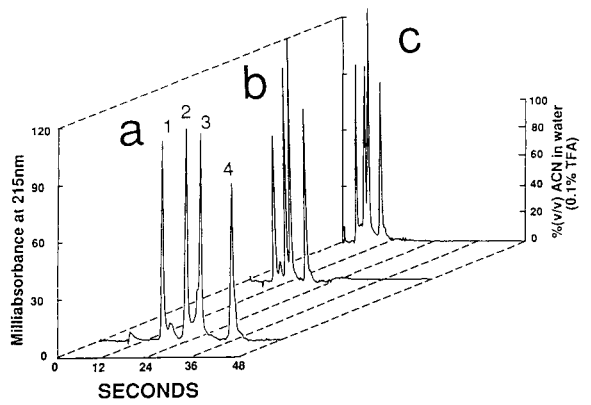


Fig. 14. Chromatograms of standard proteins obtained on a 30 × 4.6 mm I.D. column packed with 2-μm pellicular ODS-silica under three different temperature and flow-rate conditions: (a) 30°C, 2 ml/min; (b) 80°C, 4 ml/min; (c) 120°C, 5 ml/min. Linear gradient of 2.5 ml from 10 to 90% (v/v) acetonitrile in water containing 0.1% (v/v) TFA. Peaks: 1 = ribonuclease A; 2 = cytochrome *c*; 3 = lysozyme; 4 = β-lactoglobulin B. From Ref. [42].

Column stability at the anticipated operating temperatures over an extended period of time is a prerequisite of high-temperature HPLC. Columns packed with pellicular stationary phases have been shown not to change their properties at temperatures up to 120°C for at least 1000 h of operation [42]. Though the thermal stability of

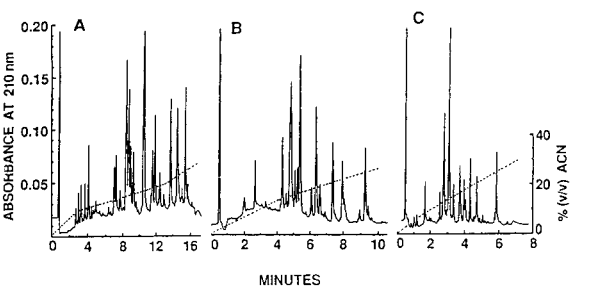


Fig. 15. Effect of temperature and flow-rate on the separation of the tryptic fragments of β-lactoglobulin B. Column, 75 × 4.6 mm, packed with micropellicular C_{18} silica; temperature and flow-rate, (A) 25°C, 0.8 ml/min; (B) 50°C, 1.3 ml/min; (C) 80°C, 2.0 ml/min. The initial column inlet pressure was 34.3 MPa in each case. Sample: 4 μg of reduced and S-carboxymethylated β-lactoglobulin B digest in 20 μl. From Ref. [39].

porous stationary phases especially silica-based porous sorbents in contact with the mobile phase is generally not satisfactory, porous siliceous sterically protected bonded C_{18} stationary phases have been reported to withstand the operating conditions at elevated temperatures up to 90°C [30].

So far the HPLC analysis of proteins at elevated temperatures has been carried out almost exclusively by reversed-phase chromatography. Since this technique employs denaturing conditions even at ambient temperature, further denaturation due to the combined use of the acidic hydro–organic medium, strongly hydrophobic stationary phase, and high temperatures does not cause further complications. Protein analysis at elevated temperatures by using other branches of HPLC may be associated with undesirable conformational changes or further chemical reactions lest the separation is sufficiently fast to avoid such interferences.

HPLC at temperatures above the atmospheric boiling point of the mobile phase may be termed "superheated" liquid chromatography [26] that in many respects, is similar to supercritical fluid chromatography. Both take advantage of enhanced transport properties of the mobile phase to obtain high chromatographic speed and efficiency [26]. The advantages of using superheated or supercritical fluid as the mobile phase over the traditional eluents in HPLC at ambient temperature illustrated in terms of column efficiency by the Van Deemter plots in Fig. 16. Upon the availability of appropriate instruments and columns, superheated liquid chromatography may be superior to supercritical fluid chromatography by offering a greater choice of mobile phases and thus facilitating the modulation of chromatographic retention and selectivity in a wide range [26]. Supercritical fluid chromatography, on the other hand, may be more suitable for the analysis of thermolabile substances and in those instances when the peculiar solvent properties of supercritical carbon dioxide can be utilized for enhanced chromatographic efficiency.

At elevated temperatures undesirable on-column reactions may be accelerated to an extent

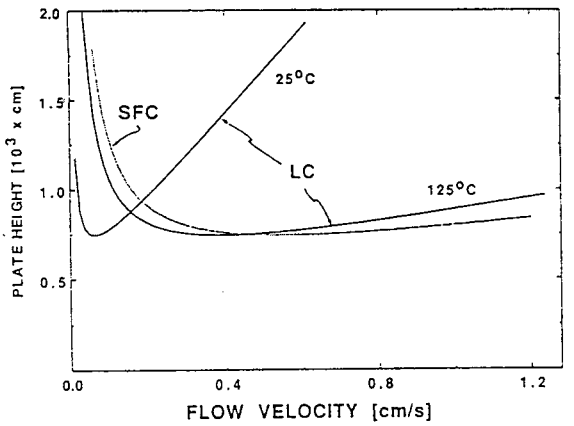


Fig. 16. Van Deemter plots for conditions typical in supercritical fluid chromatography, superheated liquid chromatography, and liquid chromatogaphy at ambient temperatures. Porous 3-μm particles. The diffusivities of the eluite in the three case are 10^{-4}, 10^{-4} and $5 \cdot 10^{-6}$ cm^2 /s, respectively. From Ref. [26].

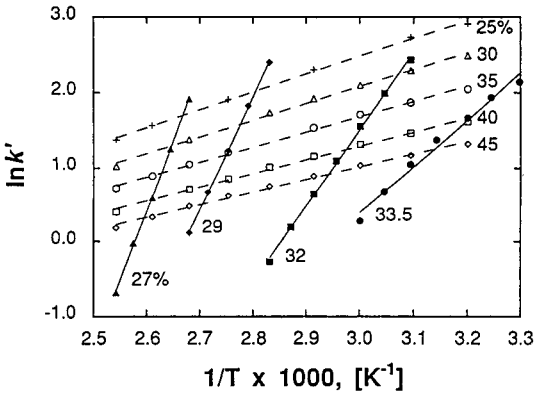


Fig. 17. Plots of the logarithmic retention factor versus the reciprocal absolute temperature in the reversed-phase chromatography of lysozyme (solid lines) and nitrobenzene (dashed lines). The percentage of acetonitrile in the aqueous mobile phase is shown for each set of data. Column, 30 × 4.6 mm, packed with 5-μm macroreticular cross-linked polystyrene; flow-rate: 1 ml/min. In the chromatography of lysozyme the mobile phase contained 0.1% (v/v) trifluoroacetic acid. From Ref. [42].

that may interfere with the chromatographic process. The interplay of the retention process and the on-column reaction is quantitatively measured by the Damköhler number (*Da*), which is defined as the ratio of the residence time in the column to the relaxation time for the on-column reaction [52]. Large *Da* values indicates that the reaction is completed in the column so that only the product(s) of the reaction exit at the column outlet. On the other hand, when *Da* is much smaller than unity, the reaction, such as protein degradation, does not take place to any significant extent during the residence time of the reactive eluite. Although the rate of most on-column reaction increases, the retention time decreases with increasing temperature. Consequently, under conditions of rapid analysis *Da* may remain small and thus significant degradation of the sample may not take place in the column [26].

The retention enthalpies of large molecules are generally greater than those of small ones. This is seen in Fig. 17 from the Van 't Hoff plots of lysozyme and nitrobenzene data obtained in reversed-phase chromatography. For large molecules, therefore, their retention is generally quite sensitive to the changes in temperature that can thus be an effective retention modulator besides affecting the efficiency of the column. This suggests that temperature programming could be a powerful adjunct or even an alternative to mobile phase gradients in the analytical separation of macromolecules by reversed-phase chromatography with columns having low heat capacity.

Temperature programming is eminently suitable for eluting sample components having a wide volatility range and has been widely used in gas chromatography [53]. However, this anisocratic elution mode [21] is rarely used in liquid chromatography although for the separation of large molecules having high retention enthalpies, temperature programming is offers certain advantage. It has been observed, for instance, that the change in selectivity upon changing the temperature is different from that upon changing the mobile phase composition in the separation of peptides [54]. Consequently temperature programming could complement gradient elution. Further, temperature programming might open a narrow elution window for closely related species

and permit a finer tuning of their retention behavior than gradient elution.

Nevertheless, temperature programming can find wide practical applications only when appropriate narrow- or micro-bore columns having low heat capacity and high thermal conductivity are available to facilitate rapid heating and cooling without untoward radial temperature profiles. If packed fused-silica capillaries prevail in the future of HPLC, they may elicit a growing use of temperature programming since it is much easier to program the column temperature than to generate eluent gradient at flow-rates of a few μl/min or nl/min [55].

4.2. Gradient time and flow velocity

As mentioned above, gradient elution is the widely used anisocratic elution mode in the HPLC of peptides and proteins. The magnitude of retention under gradient conditions depends on the flow velocity, the gradient time and volume as well as the starting and final mobile phase compositions. In gradient elution with linear solvent strength gradient, the retention factor at the column midpoint, $\bar{k}$, is given by [28]

$$\bar{k} = \frac{ut_G}{1.15L \log \frac{k_0}{k_f}} \quad (13)$$

Eq. 13 indicates that the magnitude of retention can be easily modulated by changing the gradient time and flow velocity.

With gradient elution, the total analysis time is the sum of the retention time of the last peak and the column regeneration time, i.e., the time it takes to return the column to its initial condition after the gradient run is completed. Columns packed with micropellicular or gigaporous stationary phases require much shorter reequilibration times than those packed with traditional porous stationary phases [43,56]. Thus, such stationary phases of advanced configuration are particularly suitable for use with rapid gradient runs as both the actual separation time and the regeneration time can be significantly reduced.

Whereas gradient elution is used in the separation of sample components having retention factors widely different, in the chromatography of very closely related proteins, isocratic elution with highly efficient columns may be more appropriate as shown in Fig. 11. In addition to the greater resolution, the use of isocratic conditions for closely related sample components has the advantage that the column need not be regenerated and concomitantly the analysis time may be reduced.

5. Instrumentation

Rapid HPLC requires not only stationary phases of low mass transfer and kinetic resistances, but also specially designed instrumentation capable to generate fast eluent gradients and provide temperature control over a wide super-ambient temperature range [8]. Other demanding features of the instrument having low overall system dead volume include a precise sample introduction device with low dispersion as well as a highly sensitive detector with short response time and small flow-cell volume. For separations with high speed and high efficiency by using narrow-bore or micro-bore columns, the demand for very small extra-column dispersion is extremely stringent and consequently the extra-column volume has to be drastically reduced.

As mentioned before it is necessary to control the column temperature in liquid chromatography in order to exploit the potential of high-temperature HPLC and to obtain reproducible results. The HPLC unit must have means for heating the column to the desired temperature without radial temperature gradients that are known to deteriorate its efficiency [57–59]. In Fig. 18 two alternative configurations are shown of an HPLC unit designed for use at elevated temperatures. In configuration A the entire fluid line between pumps and column is located inside the oven, on the other hand, in configuration B the heat exchanger, the mixer, and the injector are outside the oven and the eluent is heated immediately after leaving the injector to the oven temperature by a flash-heater [42]. Selec-

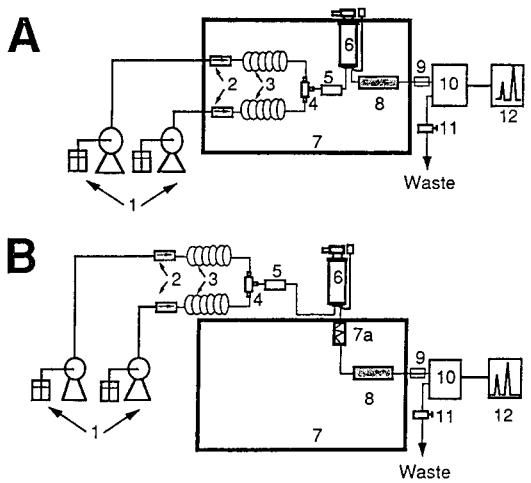


Fig. 18. Flow sheets of the two configurations of the liquid chromatograph for use at temperatures over 100°C with the sample injector inside (A) or outside (B) the oven. 1 = Reservoirs and pumps; 2 = in-line filters; 3 = heat exchangers; 4 = mixer; 5 = inverse in-line filters; 6 = sampling valve; 7a = flash heater; 7 = high-temperature oven; 8 = column; 9 = cooling device; 10 = UV detector; 11 = restrictor; 12 = data processor. From Ref. [42].

tion of the HPLC unit configuration depends on the preferred operating conditions. Configuration A is suitable only for the analysis of samples dissolved in solvents with boiling point higher than the column temperature. In contrast, configuration B places no restriction on the sample solvent and provides the flexibility and convenience needed for quantitative sample introduction over a wide range of super-ambient column temperatures. However, in this case the extra-column dead volume is relatively large due to the flash heater placed between the injector and column.

For high-temperature HPLC the instrumental specifications call for a heat exchanger that can bring the eluent rapidly to column temperature. Specially made thin-walled capillary tubing, called "serpentine tubing", is often employed in such applications [60]. The design of the HPLC system is particularly critical when temperature programming is used, because it requires instantaneous temperature change of both the column and the incoming mobile phase. In general, inadequate heat exchange due to poor construction of the instrument can annihilate the advantages of HPLC at high temperatures [58].

Steep eluent gradients and short columns are often used in the rapid separation of peptides and proteins. The generation of such gradients in a reproducible fashion requires precise and pulseless eluent delivery from the pumps, an efficacious mixing device and low dwell volume. It is recalled that the dwell volume is the volume of the connecting tubing between the inlets of the mixer and the column. Large dwell volume not only delays and distorts the gradient, but also increases the time required for reequilibration of the column. Systems having a dwell volume low enough to obtain dwell times as short as 10 s or less have been used in rapid analysis of peptides and proteins [6,7,56].

One of the most important potential applications of rapid HPLC is in on-line process monitoring, where the design of the sampling system is of particular importance. Especially in-process analysis in biotechnology can benefit from rapid HPLC that can be considered as the ultimate "multi-analyte sensor". In biotechnological applications on-line monitoring of the concentration of certain substances in the fermentation broth usually requires a more elaborate sampling system than that in the effluent from some chromatographic or membrane purification steps. "Dirty" samples from the fermentor, for instance, may require the removal of particulate matter by filtration or dialysis [61,62]. Sometimes it is necessary to dilute the sample before introducing it into the column. Generally, the sampling device should be sterilized in order to avoid contamination of the process stream or the fermentor.

In rapid HPLC separations the band width may be on the time scale of a few seconds and the time constant of the detection system should be less than 100 ms. In the future, separations may be carried out on the millisecond scale and then the time constant of the HPLC unit has be proportionally smaller. In any case, in such applications computers are indispensable for data acquisition and processing. With the aid of high-speed computers, data processing is no

longer the speed-limiting step in the comprehensive analysis scheme. The computer also plays a very important role in the automation of HPLC systems by facilitating the remote control of the instrument and of data collection. Thus the reproducibility and accuracy of the results are enhanced and the analytical results are made available without delay. It should be emphasized again, however, that further progress in high-speed HPLC will inexorably call for concomitant advances in prechromatographic sample treatment.

Acknowledgements

The authors are grateful to one of the referees for the information on siliceous gigaporous supports listed in Table 1. This work was supported by grants No. GM 20933 from the National Institutes of Health, US Department of Health and Human Resources, No. BCS 9014119 from the National Science Foundation, and a grant by the National Foundation of Cancer Research.

References

[1] L.S. Ettre, in Cs. Horváth (Editor), *High-Performance Liquid Chromatography —Advances and Perspectives*, Vol. 1, Academic Press, New York, 1980, pp. 1–74.
[2] J.N. Done, J.H. Knox and J. Loheac, *Applications of High-Speed Liquid Chromatography*, Wiley, New York, 1974.
[3] L.R. Snyder and J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley-Interscience, New York, 1979.
[4] S.K. Paliwal, T.K. Nadler and F.E. Regnier, *TIBTECH*, 11 (1993) 95.
[5] R.C. Dean, *Genetic Eng. News*, 9 (1989) 4.
[6] K.D. Nugent, in J.J. Villafranca (Editor), *Current Research in Protein Chemistry —Techniques, Structure, and Function*, Academic Press, New York, 1990, pp. 233–244.
[7] K. Nugent and K. Olson, *Biochromatography*, 5 (1990) 101.
[8] K.C. Olson and R.L. Gehant, *Biotechnol. Prog.*, 8 (1992) 562.
[9] K. Kalghatgi and Cs. Horváth, in C. Mant and R. Hodges (Editors), *HPLC of Peptides and Proteins —Separation, Analysis, and Conformation*, CRC Press, Boca Raton, FL, 1990, p. 687.
[10] J.V. Twork and A.M. Yacynych (Editors), *Sensors in Bioprocess Control, Bioprocess Technology*, Marcel Dekker, New York, 1990.
[11] S.K. Paliwal, T.K. Nalder, D.I.C. Wang and F.E. Regnier, *Anal. Chem.*, 65 (1993) 3363.
[12] H. Lundstrom, M. Brohjer, B. Osterlof and T. Mokes, *Biotechnol. Bioeng.*, 36 (1990) 1056.
[13] K. Holzhauer-Rieger, W. Zhou and K. Schügerl, *J. Chromatogr.*, 499 (1990) 609.
[14] J. Möller, R. Hidderssen, J. Niehoff and K. Schügerl, *Anal. Chim. Acta*, 190 (1986) 195.
[15] J. Frenz, Ph. van der Schrieck and Cs. Horváth, *J. Chromatogr.*, 330 (1985) 1.
[16] S.P. Fulton, N.B. Afeyan, N.F. Gordon and F.E. Regnier, *J. Chromatogr.*, 547 (1991) 452.
[17] R.E. Cooley and C.E. Stevenson, *Process Control Qual.*, 2 (1992) 43.
[18] J. Frenz, W. Hancock, W. Henzel and Cs. Horváth, in K.M. Gooding and F.E. Regnier (Editors), *HPLC of Biological Macromolecules —Methods and Applications*, Marcel Dekker, New York, 1990, p. 145.
[19] B. Whittman-Liebold (Editor), *Methods in Protein Sequence Analysis*, Springer, New York, 1989, p. 220.
[20] T.E. Hugli (Editor), *Techniques in Protein Chemistry*, Academic Press, San Diego, CA, 1989, p. 339.
[21] Cs. Horváth and W.R. Melander, in E. Heftmann (Editor), *Chromatography —Fundamentals and Applications of Chromatographic and Electrophoretic Methods, Part A: Fundamentals and Techniques*, Elsevier, Amsterdam, New York, 1983, p. 27.
[22] G. Guiochon, in Cs. Horváth (Editor), *High-Performance Liquid Chromatography —Advances and Perspectives*, Vol. 2, Academic Press, New York, 1980, p. 341.
[23] D.D. Frey, R. Van de Water and B. Zhang, *J. Chromatogr.*, 603 (1992) 43.
[24] L.R. Snyder, in Cs. Horváth (Editor), *High-Performance Liquid Chromatography —Advances and Perspectives*, Vol. 1, Academic Press, New York, 1980, p. 207.
[25] L.R. Snyder, M.A. Stadalius and M.A. Quarry, *Anal. Chem.*, 55 (1983) 1412.
[26] F. Antia and Cs. Horváth, *J. Chromatogr.*, 435 (1988) 1.
[27] A.J.P. Martin and R.L.M. Synge, *Biochem. J.*, 35 (1941) 1358.
[28] M.A. Stadalius, M.A. Quarry and L.R. Snyder, *J. Chromatogr.*, 327 (1985) 93.
[29] N.H.C. Cooke, B.G. Archer, K. Olsen and A. Berick, *Anal. Chem.*, 54 (1982) 2277.
[30] J.J. Kirkland and C.H. Dilks, Jr., *LC·GC*, 11 (1993) 292.
[31] J.C. Giddings, *Dynamics of Chromatography*, Marcel Dekker, New York, 1965.
[32] E. Katz, K.L. Ogan and R.P.W. Scott, *J. Chromatogr.*, 270 (1983) 51.
[33] K.K. Unger, G. Jilge, J.N. Kinkel and M.T.W. Hearn, *J. Chromatogr.*, 359 (1986) 61.
[34] K. Kalghatgi and Cs. Horváth, *J. Chromatogr.*, 398 (1987) 335.

[35] N.D. Danielson and J.J. Kirkland, *Anal. Chem.*, 59 (1987) 2501.
[36] Cs. Horváth, B.A. Preiss and S.R. Lipsky, *Anal. Chem.*, 39 (1967) 1422.
[37] Cs. Horváth, *Trennsäulen mit Dünnen Porösen Schichten für die Gaschromatographie*, Johann Wolfgang Goethe University, Frankfurt am Main, 1963.
[38] D.J. Burke, J.K. Duncan, L.C. Dunn, L. Cummings, C.J. Siebert and G.S. Ott, *J. Chromatogr.*, 353 (1986) 425.
[39] K. Kalghatgi and Cs. Horváth, *J. Chromatogr.*, 443 (1988) 343.
[40] Y.-F. Maa and Cs. Horváth, *J. Chromatogr.*, 445 (1988) 71.
[41] L. Várady, K. Kalghatgi and Cs. Horváth, *J. Chromatogr.*, 458 (1988) 207.
[42] H. Chen and Cs. Horváth, *Analytical Methods and Instrumentation*, 1 (1993) 213.
[43] N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang and F.E. Regnier, *J. Chromatogr.*, 519 (1990) 1.
[44] L.L. Lloyd and F.P. Warner, *J. Chromatogr.*, 512 (1990) 365.
[45] J. Horváth, E. Boschetti, L. Guerrier and N. Cooke, *J. Chromatogr.*, 679 (1994) 11.
[46] E. Boschetti, *J. Chromatogr. A*, 658 (1994) 207.
[47] D.D. Frey, E. Schweinheim and Cs. Horváth, *Biotechnol. Prog.*, 9 (1993) 273.
[48] A.E. Rodrigues, L. Zuping and J.M. Loureiro, *AIChE Symp. Ser.*, 84 (1988) 80.
[49] K. Kalghatgi, I. Fellegvári and Cs. Horváth, *J. Chromatogr.*, 604 (1992) 47.
[50] R.M. Moore and R.W. Rodney, *J. Chromatogr.*, 317 (1984) 119.
[51] M.W. Dong, J.R. Gant and B.R. Larsen, *BioChromatography*, 4 (1989) 19.
[52] W.R. Melander, H.-J. Lin, J. Jacobson and Cs. Horváth, *J. Phys. Chem.*, 88 (1984) 4527.
[53] W.E. Harris and H.W. Habgood, *Programmed Temperature Gas Chromatography*, Wiley, New York, 1966.
[54] W.S. Hancock, R.C. Chloupek, J.J. Kirkland and L.R. Snyder, *J. Chromatogr. A*, 686 (1994) 31.
[55] J. Bowermaster and H.M. McNair, *J. Chromatogr. Sci.*, 22 (1984) 165.
[56] K. Kalghatgi and Cs. Horváth, in Cs. Horváth and J.G. Nikelly (Editors), *Analytical Biotechnology: Capillary Electrophoresis and Chromatography*, American Chemical Society, Washington, DC, 1990, p. 162.
[57] R.J. Perchalski and B.J. Wilder, *Anal. Chem.*, 51 (1979) 774.
[58] F.V. Warren and B.A. Bidlingmeyer, *Anal. Chem.*, 60 (1988) 2821.
[59] H.-J. Lin and Cs. Horváth, *Chem. Eng. Sci.*, 36 (1981) 47.
[60] G. Liu, N.M. Djordjevic and F. Erni, *J. Chromatogr.*, 592 (1992) 239.
[61] D. Picque and G. Corrieu, *Biotechnol. Bioeng.*, 40 (1992) 919.
[62] X. Monseur and C.L. Motte, *Anal. Chim. Acta*, 204 (1988) 127.

ELSEVIER

Journal of Chromatography A, 705 (1995) 21–45

JOURNAL OF CHROMATOGRAPHY A

Review

Protein mass spectrometry: applications to analytical biotechnology

D.N. Nguyen*, G.W. Becker, R.M. Riggin

Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

Abstract

The advent of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) in the last 5 years has greatly enhanced the area of protein mass spectrometry. This paper presents an overview of the applications of protein mass spectrometry in the area of analytical biotechnology, particularly as related to biopharmaceutical research and development. These applications include the determination of protein molecular mass, peptide mapping, peptide sequencing, ligand binding, determination of disulfide bonds, active site characterization of enzymes, protein self-association and protein folding/higher order structural characterization.

Contents

* Corresponding author.

0021-9673/95/$29.00
SSDI 0021-9673(94)01256-3

1. Introduction

Various forms of mass spectrometry having the ability to analyze protein structures have been introduced over the past 10–15 years. These techniques have served a central role in the advancement of biotechnology during that time period. The advent of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) in the last 5 years has greatly enhanced the area of protein mass spectrometry. In many areas of research and development (e.g. industrial biotechnology) these techniques are now indispensable tools for examining protein and peptide structures.

Protein mass spectrometry has been the subject of several extensive reviews [1–4], and we therefore do not intend to include a comprehensive review of the field in this paper. Instead, we intend to present an overview of the applications of protein mass spectrometry in the area of analytical biotechnology, particularly as related to biopharmaceutical research and development. Ideally, this examination will prove a useful guide to the general types of protein structural studies which can be addressed by mass spectrometry.

The paper consists of two major sections. In the first section the fundamental aspects of the various modes of protein mass spectrometry are only briefly described since they have been discussed in great detail elsewhere. In the second section various application areas in which mass spectrometry of proteins serves a useful role are discussed. These applications include the determination of protein molecular mass, peptide mapping, peptide sequencing, ligand binding, determination of disulfide bonds, active site characterization of enzymes, protein self-association and protein folding/higher order structural characterization. Each of these applications represents an important research area where many tools can be used, but in which mass spectrometry is becoming increasingly significant.

2. Modes of mass spectrometry

The development and application of mass spectrometry to the analysis of organic molecules began in the late 1950s and early 1960s. During this time period, electron impact (EI) ionization was the only practical ionization technique [5]. The development of chemical ionization (CI) in the late 1960s provided a complementary method for ionization of organic molecules in the gaseous phase [6]. A disadvantage of both methods is that the sample analytes must be in the vapor phase before ionization can occur, often requiring heat for vaporization. Because of their size, involatility and thermal lability, proteins are generally not candidates for study using EI or CI methods.

2.1. Fast atom bombardment mass spectrometry (FAB-MS)

The role of mass spectrometry in protein characterization underwent explosive growth in 1981 when Barber and co-workers [7,8] introduced fast atom bombardment mass spectrometry. The application of FAB-MS allowed scientists to obtain molecular masses of peptides and small proteins routinely and, sometimes, limited structural information was attainable.

In a typical FAB-MS analysis, the sample is dissolved in a non-volatile, viscous matrix such

as glycerol, thioglycerol or a mixture of dithiothreitol and dithioerythritol. The sample matrix mixture, usually about 1 μl, is then applied to a sample probe tip and directly inserted into the mass spectrometer. The sample matrix is bombarded with a high-energy (about 6–8 keV) beam of xenon atoms (Fig. 1).

The mechanism of sample desorption and ionization is not completely understood and is still being investigated. It is likely that a combination of the following mechanisms is responsible for ionization. There are currently two equally favorable mechanisms in describing the ionization process. The first requires that the ions are preformed in the matrix through protonation and complexation with metal ions (Na^+, K^+) [9]. The second theory proposes that the sample is desorbed as neutral molecules which undergo gas-phase ionization in the high-pressure region directly above the matrix–vacuum interface [10]. In practice, the mechanism that predominates the ionization process depends on many factors such as the presence of metal ions in the sample and/or matrix, the concentration of the sample in the matrix and the pK_a of the analyte molecules. In many instances, sample molecules which have extensive hydrogen bonding with the matrix tend to form a homogeneous solution with no concentration of the analyte in a particular locus within the droplet. On the other hand, hydrophobic analytes tend to concentrate on the surface of the droplet. Thus, signal suppression of hydrophilic peptides is often observed in FAB-MS during the analysis of a peptide mixture [11]. This limits one from relying on the FAB mass spectrum of a mixture to determine the concentrations of the analytes present in the sample unless internal standards are carefully applied. Although FAB-MS can be used to analyze a mixture of peptides, practical consideration must be taken. The sample mixture must be free of salt and buffer, hence it should be purified by HPLC or other suitable chromatographic techniques. Peptides should be stored in polypropylene vials to prevent complexation with metal ions. Complete suppression of the analyte signal may be observed if the sample is contaminated with a high concentration of alkaline salts or buffer. When there is a need to concentrate a peptide solution, it must not proceed to dryness since irreversible adsorption to the tube walls is likely. It is also good practice to avoid contact with any glass surface during sample preparation.

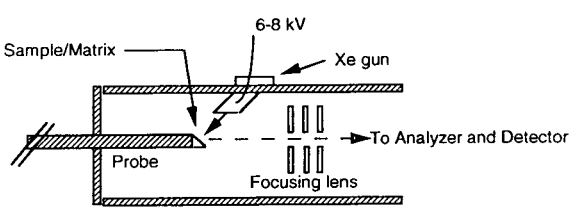


Fig. 1. Ionization source for FAB-MS.

Several drawbacks associated with FAB-MS should also be pointed out. The most obvious drawback is the presence of background peaks in the mass spectrum. The background signals are present throughout the mass spectrum, but it is particularly problematic in the low-molecular mass region (below 300 u). Simply switching to a different matrix allows one to discriminate the background from the analyte signals. A different drawback results directly from the nucleophilicity of the matrix components. For example, dithiothreitol could undergo an S_N2 reaction with samples containing good leaving groups such as halides. An experimental indication that the analyte is reacting with the matrix can be obtained by analyzing the sample using different matrices. If different peak patterns are obtained for different matrices, then it is likely the sample has undergone chemical reactions with the matrix. Other chemical reactions which could occur in the matrix include the reduction of disulfide bonds, the exchange of sulfur by oxygen atoms, ring opening of lactones and hydrolysis [12].

A related technique, continuous-flow FAB (CF-FAB), overcomes some of these problems. In CF-FAB, the sample is introduced continuously by a flow of volatile solvents such as water, methanol and acetonitrile. The use of these volatile solvents eliminates much of the matrix-derived background, thereby lowering the detection limit. CF-FAB also provides mechanical mixing of the sample and ultimately decreases the ion suppressing effect [13]. This technique

allows the analysis of a large number of aqueous samples and can be used to monitor the purity of a synthetic peptide.

Perhaps the greatest drawback of FAB and CF-FAB is the inability to generate good signal intensity with proteins of M_r greater than about 15 000, owing to involatility and thermal instability of proteins, thereby limiting the general applicability of this technique to large proteins. The use of liquid secondary ionization mass spectrometry (LSIMS), where the xenon beam in FAB-MS is replaced with a stream of ions such as Ar^+ or Cs^+, has shown some success with proteins in the M_r range 15 000–20 000. However, these results are far from routine and, in practice, FAB works best in the analysis of proteins with molecular masses of no more than about 5000.

2.2. Plasma desorption mass spectrometry (PD-MS)

The first breakthrough in attempts to solve the problems associated with protein mass spectrometry was the development of field desorption mass spectrometry [14]. In 1974, Friedman and co-workers proposed that rapid heating of a sample could lead to preferential desorption of surface molecules [15]. At the same time, Macfarlane and co-workers discovered that when high-energy fission fragments from ^{252}Cf irradiated thin films of arginine and cysteine, intact molecular ions were observed [16]. Thus, a new ionization technique, plasma desorption, was born. Today, PD-MS is used primarily for molecular mass determination of proteins. This technique provides very little fragmentation, and therefore little structural information can be obtained. The sensitivity of this technique can be impressive with examples of spectra obtained on proteins of M_r about 45 000 (ovalbumin) at the picomole level [17]. Perhaps one of the most useful applications of this technique is its use for peptide mapping by in situ enzymatic digestion [18]. This procedure is normally performed with the same nitrocellulose-bound sample as already used for molecular mass determination. An enzyme solution is applied and after an appropriate time the digestion is terminated by removal of buffer by spin-drying the target. The target containing the digested peptide fragments can then be analyzed by PD-MS.

PD-MS has been a particularly attractive method because of its operational simplicity and high reliability. A schematic diagram of a PD mass spectrometer is shown in Fig. 2. PD-MS is based on spontaneous fission of ^{252}Cf which produces a pair of nuclear fragments such as $^{144}Cs^{20+}$ and $^{106}Tc^{22+}$. These fission fragments, with MeV energy, move in opposite directions. One of these fragments can be oriented to hit the sample foil and ionize the sample molecules. The sample foil containing the adsorbed protein, often referred to as the target, is then mounted on a wheel assembly which can accommodate several samples at one time. The target wheel is positioned on the axis of the cylindrical tube called the flight tube. The ions are accelerated by a grid maintained at ground potential. The main function of the fission fragment detector is to record the time of the fission. The masses of the ions are calculated from the time of a fission event and the time it takes for an ion to reach the detector. The time an ion takes to traverse the flight tube is dependent on its mass/charge ratio. Both detectors produce an electronic pulse when an ion is detected.

In principle, a time-of-flight (TOF) mass spectrometer has no upper mass limit. However, there are several factors which limit the range of samples that can be analyzed by PD-MS. It is clear that the energy density developed by the

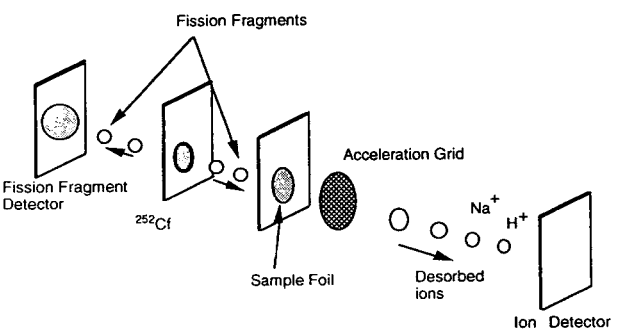


Fig. 2. Schematic diagram of plasma desorption mass spectrometer.

fission of ^{252}Cf is too high and that most of the protein molecules undergo pyrolysis. This problem is very apparent when the molecular mass of the protein approaches 50 000. In this mass range, very few intact molecular ions are detected. Since most of the plasma desorption ion sources are coupled to a TOF analyzer (typical resolution of about 1000), the mass accuracy is far from desirable. This, perhaps, results from the large kinetic energy distribution the ion carries with it [19]. Broadening of the peak to the extent that the isotopic pattern cannot be resolved is a common phenomenon in PD-MS. Therefore, the mass value derived from a PD-MS is often the isotopically averaged value.

2.3. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)

The development of modern laser technology has provided a means of directing a large amount of energy into a sample, leading to desorption of intact molecules rather than thermal decomposition. Matrix-assisted laser desorption ionization was introduced by Tanaka et al. [20] and Karas and Hillenkamp [21]. Two ranges of laser wavelengths, the far-infrared and the far-ultraviolet, are used for desorption of sample molecules. The most commonly used wavelength is between 266 and 366 nm, which is generated from a neodymium/yttrium–aluminum–garnet (Nd:YAG) laser. The reason for choosing this wavelength is that compounds containing π-bonds can be electronically excited. With regard to mass spectrometry, lasers offer two important benefits. First, lasers provide the capability to pulse from a continuous wave down to femtosecond (10^{-15} s). Most laser mass spectrometers use pulses of 100 ns or less to prevent pyrolysis of the proteins. Second, laser beams can be focused to submicrometer diameters, allowing the operator to control the laser beam precisely with an appropriate microscope. The amount of energy deposited in the sample is dependent on laser irradiance or intensity (W cm^{-2}), the pulse width and the absorptivity of the sample. Increasing the laser intensity will lead to an increase in desorption; however, for large proteins too high an intensity will certainly lead to excessive fragmentation.

In MALDI-MS, the sample molecules are mixed with a suitable matrix (1:10 000 ratio). The commonly used matrices are 3-methoxy-4-hydroxycinnamic acid (ferulic acid) or 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) dissolved in a mixture of acetonitrile and 0.1% trifluoroacetic acid (TFA). The final concentration of the matrix is 5–10 g/l. The matrix–analyte mixture is deposited on a silver support and allowed to crystallize by slow evaporation of the matrix solvents. Proper sample preparation is essential in obtaining a good spectrum. Co-crystallization of sample and matrix is critical and will depend on the physical properties of both sample and matrix. The presence of TFA has been found to help achieve the co-crystallization. As mentioned previously, the ability to focus the laser beam precisely on the target allows the accumulation of successive spectra from the same and/or different areas. A simple schematic diagram of a MALDI instrument is shown in Fig. 3. The impact of a laser pulse on the target desorbs a large number of ions, which can cause a saturation effect in the detector. The high yields of low-mass ions saturate the detector, which results in lower efficiency for the high-mass ions because the detector cannot recover fast enough. The solution to this problem is the installation of an ion deflector which deflects the low-mass ions. The detector is also located slightly off-axis from the mass spectrometer. This combination reduces about 80% of the low-mass ions. The low-mass ions dominate the M_r 0–500

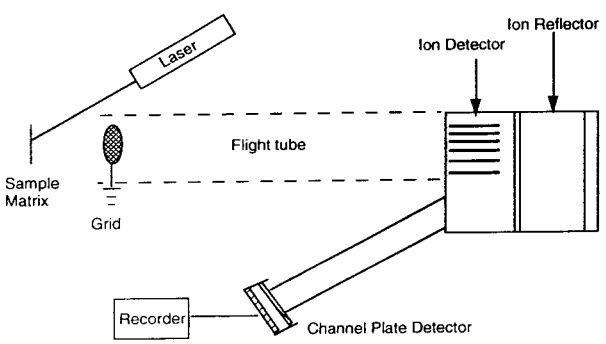


Fig. 3. Schematic diagram of MALDI mass spectrometer.

range and can extend up to 1000. However, small noise background can be seen across the mass spectrum. The low-level noise background is probably the result of fragmentation due to ions colliding with the surfaces in the spectrometer.

The resolution of a MALDI mass spectrometer is much less than desirable, in part due to the use of a TOF mass analyzer. An interesting, but not surprising, fact is that the resolution for low-mass ions is much better than that for high-mass ions. This effect is the result of the spread of the initial kinetic energy of the ions, and also the existence of non-resolved adduct ions [22]. The low resolution of high-mass ions results in uncertainty in mass determination. The uncertainty is about 10–50 u at M_r 10 000 and 100–500 u for proteins of M_r 100 000.

Despite the poor resolution, sensitivity is a major advantage of MALDI-MS. A good signal-to-noise spectrum can be obtained with a few laser pulses from a single spot on the sample target. Since successive spectra can be obtained from a single spot, it can be concluded that this technique is non-destructive. In practice, the actual amount of sample consumed is in the femtomole range. Therefore, for samples available only in limited amounts, the sample can be recovered after the analysis. It should be pointed out that picomole amounts of sample are normally required to form uniform matrix–analyte crystals for the analysis.

A related technique which is still at an early developmental stage is continuous-flow MALDI (CF-MALDI). In CF-MALDI, the sample is introduced to the mass spectrometer via a probe with a liquid matrix (aqueous TFA, ethanol, ethylene glycol, and 3-nitrobenzyl alcohol). The laser beam is set on the opposite side of the probe with the flight tube perpendicular to the probe. In this present configuration, mass spectra of large peptides with molecular masses above 10 000 can be obtained [23–26].

2.4. *Electrospray ionization mass spectrometry (ESI-MS)*

The electrospray ionization process was originally described by Dole et al. [27] in their studies on synthetic and natural polymers of molecular mass in excess of 100 000. However, it was not until 10 years later that electrospray resurfaced as a major technique in mass spectrometry. The application of electrospray to mass spectrometry was largely the result of two different groups almost simultaneously. Yamashita and Fenn [28] coupled atmospheric pressure electrospray to a quadrupole mass spectrometer and Alexandrov et al. [29] coupled it to a magnetic sector mass spectrometer. A few years after the demonstration of the fundamental aspects of electrospray, Fenn and co-workers demonstrated the ability of electrospray to analyze high molecular mass samples such as polyethylene glycol (M_r 17 500) bearing a net charge of up to +23 [30]. It is this unique feature of producing ions with multiple charges that allows electrospray to be used in the analysis of large proteins in a mass analyzer with limited mass range ($m/z < 2500$ u) with an accuracy of better than 0.01% [31,32].

Electrospray is produced by applying a high electrical field to a relatively small flow of liquid from a capillary tube. The electric field causes the liquid surface to be highly charged and a spray of charged liquid droplets forms at the end of the capillary tube. The polarity of the charged droplets can be controlled by the applied polarity on the capillary. The mechanism by which the molecular ions are formed from the charged droplets is not fully understood. Iribane and Thomson [33] proposed the field-assisted ion evaporation model. In this model, the molecular ion formation occurs when the field strength at the surface of the droplet reaches a critical value due to the evaporation of the solvent. Röllgen and co-workers proposed a different mechanism in which the disintegration of the charged droplet occurs via the Rayleigh jet mechanism that leads to very small charged droplets [34]. The bare molecular ions are formed when complete evaporation of the solvent from these micro-droplets occurs. Abbas and Latham [35] showed in a very elegant experiment that the droplets evaporate at a constant rate until the Rayleigh limit (1.1×10^5 elemental charges) is reached. The droplet then abruptly loses 20–25% of its charge and mass. This process is presumed to continue until a bare molecular ion is formed. A

particle of 1 μm requires about sixteen disintegrations to reach the ion evaporation limit. At this point, molecular ion production from the droplet can theoretically occur.

In the development of the electrospray mass spectrometer, transmitting the ions produced at atmospheric pressure into the vacuum chamber of the mass spectrometer presented a tremendous challenge. The major problem is the large amount of condensable vapor produced from the spray nozzle. This problem was overcome by applying a counterflowing curtain gas (20–70°C) to that of the spray and by nebulization. Nitrogen is used for counterflowing curtain gas. The curtain gas is delivered in the opposite direction to the flow of sample. This curtain of dry nitrogen serves to exclude large droplets and particles and to decluster the ions. The nebulization gas is usually air or nitrogen. It is delivered in the same direction of the flow of sample by means of a metal sleeve which wraps around the capillary. Chait and co-workers showed that when the capillary nozzle was operated at around 85°C, counterflowing gas was not necessary [36]. However, thermal destruction of labile peptides and proteins can result from operation at elevated temperature. Fig. 4 shows a schematic diagram of an electrospray source and the formation of molecular ions from charged droplets. These multiply charged ions result from the attachment of protons and/or metal ions (Na^+ or K^+) to the basic and acidic sites on the molecules, respectively. For example, the ε-amino group of lysine can be protonated whereas the side-chain carboxyl group of aspartic acid can carry a sodium ion adduct. Thus, the electrospray mass spectrum shows a distribution of these multiply charged ions. Proteins with a molecular mass of more than 130 000, such as the dimer of bovine serum albumin, have been successfully analyzed by electrospray mass spectrometry [37].

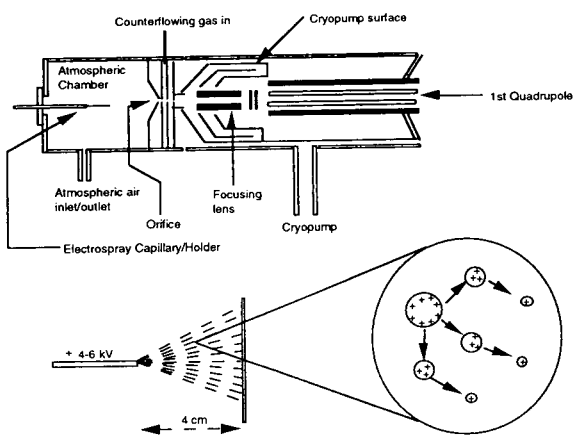


Fig. 4. Electrospray ionization source (above) and formation of molecular ions (below).

The determination of molecular mass from an electrospray mass spectrum is straightforward given the following assumptions. First, the adjacent peaks represent species differing by only one charge. Second, the charge is due to protonation (or some other known ionic species) of the molecular ion. Any two peaks are sufficient to determine the molecular mass. The relationship between the molecular mass (M_r) and the multiply charged ion (M_1) with its charge (Z_1) is described, for the case where the charge is due to protonation, by the equation

$$M_1Z_1 = M_r + 1.0079\,Z_1$$

$$M_2Z_2 = M_r + 1.0079\,Z_2$$

where $M_2 > M_1$. By solving the above equations, the charge of M_1 can be calculated using the following equation ($Z_2 = Z_1 - 1$):

$$Z_1 = (M_2 - 1.0079)/(M_2 - M_1)$$

To date, electrospray has been most widely applied using quadrupole mass spectrometers. Other workers have demonstrated that electrospray can be coupled to other mass analyzers such as magnetic sector [29], time-of-flight [38] and Fourier transform ion cyclotron resonance [39] types. Electrospray MS is gaining popularity among protein chemists because of its simplicity and versatility. The fact that electrospray MS can be coupled directly to HPLC allows the routine use of LC–MS. Applications such as peptide mapping, which could take hours or days using FAB-MS, can now be done with ESI-MS in a few hours.

Unlike FAB-MS, where both positive and negative ions are formed, only positive or negative ions are formed in ESI-MS. The charge of the droplets can be controlled by the polarity of

the applied electrical field. In most instances, peptides and proteins work well in the positive-ion mode. However, for molecules that contain many negative charges, such as oligonucleotides, the sample should be analyzed by negative-ion ESI-MS [40]. The negative ion ESI mass spectrum is characterized by an envelope of multiply charged molecular ions with the form $(M - nH)^{n-}$. Negative-ion ESI-MS generally has lower sensitivity than positive-ion ESI-MS. The decrease in sensitivity is the result of substitution of sodium for hydrogen ions. As the molecular mass of the oligonucleotide increases, the number of phosphate–sodium adducts also increases. This results in a broadening of the peaks and lowering of the measurable current for a given ion. Thus, for samples which are subject to alkali metal attachment, the measured molecular mass will often be overestimated.

3. Protein applications

3.1. Molecular mass determination

The molecular mass of a protein is an important parameter in the biochemical characterization of that protein. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) is a universal technique in protein molecular mass determination. However, the accuracy of SDS-PAGE, ranging from a few per cent for a well behaved globular protein to about 30% for a heavily glycosylated protein, makes the estimation of protein molecular masses extremely tenuous.

With the recent advances in ionization technology described in the previous section, mass spectrometry has become a useful tool for protein molecular mass determination with excellent accuracy. Techniques such as MALDI and ESI allow the determination at low-picomole levels of protein molecular masses exceeding 100 000. Table 1 shows the typical operating characteristics of commercially available instruments.

In the determination of the molecular mass of a protein, it should be remembered that the mass spectrometer will detect all isotopes of every element in the molecule. At high enough resolution, the ionic species consisting of different combinations of isotopes will be resolved, but at lower resolution, a single symmetrical peak will be obtained. In ESI-MS and MALDI-MS, centroiding the peak will afford the average molecular mass of the proteins. For protein analysis, it is not necessary to carry out the analysis at high resolution because the monoisotopic ion becomes undetectable, and even the most abundant peak contains a large combination of isotopes which requires a very high resolution mass spectrometer to achieve accurate analysis. Also, higher resolution is obtained at the expense of sensitivity.

The accuracy of ESI-MS is about 0.01%, which is very good compared with SDS-PAGE. Although ESI-MS can measure molecular masses in excess of 100 000, the analysis can be very complicated. This is because ESI-MS requires

Table 1
Operating characteristics of commercial instruments

Ionization technique	Mass range	Analyzer	Resolution	Total sample required for determination
ESI	150 000[a]	Quadrupole	2000	Picomoles
MALDI	>250 000	Time-of-flight[b]	1000	Picomoles
Plasma desorption	45 000	Time-of-flight	1000	Nanomoles
FAB/LSIMS	5000 15 000 (LSIMS)	Magnetic sector	40 000	Nanomoles

[a] Assuming the proteins can be highly protonated.
[b] Theoretically, the mass range is unlimited.

that the adjacent peaks be resolved from each other, otherwise the charge state cannot be determined. This is probably the main reason why ESI-MS fails in analyzing glycoproteins where heterogeneity from the carbohydrate moiety can give rise to a large number of overlapping peaks. In many cases MALDI-MS can be used successfully to determine the molecular mass of glycoproteins, since fewer charge states are observed compared with ESI-MS.

3.2. Peptide mapping/MS

Peptide mapping is a technique whereby a protein sample is digested either enzymatically or chemically, and the resulting peptides are separated and analyzed. This approach has been used for years for the determination of protein primary sequence. Overlapping sets of peptide fragments were generated using different enzymes or chemical agents and the peptides were separated and sequenced using standard Edman degradation reactions. In 1981, Morris et al. [41] used peptide mapping and FAB-MS to examine or screen protein digests or degradation products. Nowadays, peptide mapping used in conjunction with mass spectrometry (we shall refer to it here as peptide mapping/MS), is used in the following situations: (1) confirmation of a protein sequence, especially of proteins produced through recombinant DNA techniques, (2) detection and identification of post-translational modifications, (3) identification of protein degradation products, (4) identification of protein metabolites, (5) disulfide bond assignments, (6) ligand binding and (7) characterization of enzyme active sites. Each of these applications will be considered individually in the following sections. A general procedure for peptide mapping is shown in Fig. 5. This scheme shows all the options that are available; however, in practice all of these options are seldom used and the ultimate goal of the peptide mapping experiment determines the experimental strategy chosen. Briefly, the protein can be subjected to Edman sequencing and carboxypeptidase digestion to determine the N-terminal sequence and the C-terminal residue, respectively. The failure to

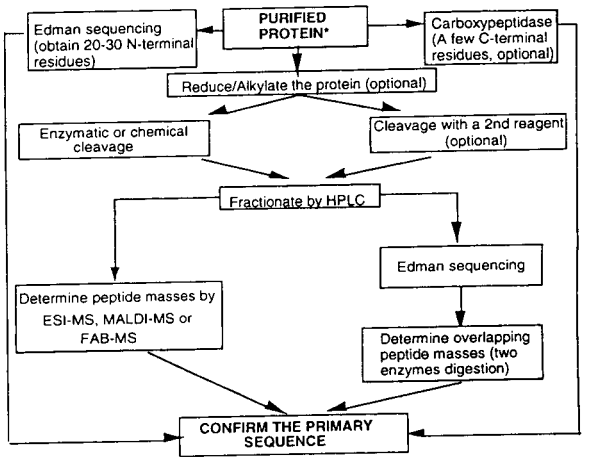


Fig. 5. Procedure for peptide mapping.

generate results from Edman sequencing suggests that the N-terminus of the protein is modified, most commonly by acetylation or formylation. The protein can then be reduced with DTT and alkylated with iodoacetic acid, 4-vinylpyridine or another alkylating agent. The reduced/alkylated protein is then subjected to chemical or enzymatic digestion. In some instances it may be appropriate to omit the reduction/alkylation step (e.g., for proteins containing only a few disulfide bonds or when assigning the disulfide bonds). The protein can be digested with two different enzymes in order to generate overlapping peptide fragments for mass spectral analysis. This option is used when prior information on the protein sequence is not available and it is necessary to align the peptides in the final sequence. The digested fragments can be analyzed directly or be fractionated by HPLC if the mixture is too complex. If desired, the digested fragments can be purified by HPLC and subjected to Edman sequencing. The results from Edman sequencing and mass spectral analysis can be combined to deduce the primary structure of the protein. When the N-terminus of a digested fragment (20 residues or less) is blocked by an acetyl group, the blocking group can be removed by acylaminoacyl-peptide hydrolase [42–44]. Digestion of proteins can be carried

Table 2
Chemical reagents and enzymes for digestion of proteins

Reagents/enzymes	Site of cleavage
1. Trypsin	C-terminus of Arg and Lys
2. *S. aureus* V8, pH 8	C-terminus of Glu and Asp
3. *S. aureus* V8, pH 4	C-terminus of Glu
4. Chymotrypsin	C-terminus of Phe, Tyr, Trp, Leu and Met
5. Endoprotease Lys-C	C-terminus of Lys
6. Endoprotease Asp-N	N-terminus of Asp
7. Clostripain	C-terminus of Arg
8. Pepsin	C-terminus of Phe, Met, Leu and Trp
9. Thermolysin	N-terminus of Leu, Ile, Val, Phe, Met and Ala
10. CNBr	C-terminus of Met
11. Hydroxylamine	Asn–Gly bond
12. Dilute acetic acid	Asp–Pro bond
13. Cysteine cyanylation	N-terminus of Cys
14. NCS	C-terminus of Trp
15. Iodosobenzoic acid	C-terminus of Trp

out chemically or enzymatically. Table 2 lists the common chemical reagents and enzymes that are used in peptide mapping.

As mentioned previously, certain ion suppression effects were observed in FAB-MS with hydrophilic peptides. This effect is virtually nonexistent when peptide mapping is carried out with PD-MS or ESI-MS. The capability of coupling electrospray to HPLC allows the analysis of the digestion fragments without any purification or fractionation. With ESI-MS, full sequencing information is often obtained by performing collision-induced dissociation tandem mass spectrometry (MS–MS). The application of the MS–MS technique in sequencing will be discussed later.

3.2.1. *Confirmation of sequence*

With the proliferation of proteins being produced by recombinant DNA techniques using a variety of expression systems, it is prudent to confirm that the protein sequence obtained is in agreement with that predicted from the DNA sequence. Peptide mapping/MS is an ideal procedure for performing this confirmation of structure in a rapid fashion with minimal consumption of protein sample. The molecular masses of tryptic peptides from the predicted sequence are compared with the molecular masses actually determined from the peptide mapping/MS experiment. This strategy was first suggested in 1984 by Gibson and Biemann [45], who used it to confirm and correct regions from the amino acid sequences of three large proteins, glutaminyl- and glycyl-tRNA synthetase from *Escherichia coli* and methionyl-tRNA synthetase from yeast. Since that time there have been several examples of this same approach applied to other proteins [46–49]. In the past, peptide mapping was a laborious process which included optimization of the HPLC conditions, fractionation of all the digested peptide fragments and analysis of all fractions by either amino acid analysis, peptide sequencing or mass spectrometry or all three. With modern technologies such as ESI-MS and MALDI-MS, peptide mapping/MS of a digested peptide mixture can be performed in a much shorter time.

MALDI-MS has proven useful for peptide mapping/MS and there are several advantages in using these techniques. Since MALDI-MS can tolerate the presence of buffer salts and other impurities, a mixture of peptides obtained from, for example, an enzymatic digest can be analyzed without prior purification. This technique allows a mass spectrum containing ion masses for all peptide fragments to be obtained in only a few minutes after completion of the digest.

Disadvantages of this technique are generally related to the matrix in which the sample is analyzed. The analytes must be uniformly distributed throughout the matrix crystals to ensure that all peptides can be desorbed. The choice of matrix is critical. Each of the common matrices discriminates against mixture components in ways that are sample and matrix specific, thereby necessitating that data be acquired using more than one matrix. Background signals are a problem with any matrix and with some the signals are present up to M_r 1000. Many peptide mixtures resulting from fragmentation of a protein contain peptides with molecular masses well below 1000. These problems are being slowly overcome by the development of new matrices and newer procedures for sample preparation. Stults et al. [50] evaluated the effects of carbohydrate-containing matrices for MALDI-MS analysis of in situ digests of proteins from two-dimensional gels and found fucose–2,5-dihydroxybenzoic acid (DHB) to be the best. Several crystallization methods for MALDI-MS have also been reported, which include vacuum crystallization [51] and stressed matrix crystallization [52,53].

ESI-MS has one major advantage over MALDI-MS, namely the ability to couple an HPLC system directly to the mass spectrometer and perform LC–MS. Application of ESI-MS in peptide mapping/MS offers a rapid means to confirm the primary sequence of proteins. In general, the digestion mixture is separated by RP-HPLC. The output from HPLC is split; a small amount of effluent goes to the mass spectrometer and the rest goes to a fraction collector or waste. Using this set-up the mass of a particular HPLC peak can be assigned readily. The same peak can be collected for Edman sequencing. The results from Edman sequencing determine the N-terminus of the peptide, whereas, the obtained mass determines the C-terminus of the peptide. This procedure can be applied for all HPLC peaks to confirm the primary sequence of proteins. The disadvantage of using ESI-MS in peptide mapping is the time necessary for LC analysis, which could range from 0.5 to 2 h. Fig. 6 shows an example of an LC–MS analysis of the trypsin digest of carboxyamidomethylated hGH. By examining the peaks in the TIC, one can assign the masses to most of the HPLC peaks. However, glycopeptides often show poor response in ESI-MS.

3.2.2. Post-translational modifications

A large number of proteins undergo post-translational modification after their biosynthesis. In some cases, post-translational modifications are critical for the transport and bioactivity of proteins. Post-translational modifications are often not detected in Edman sequencing because the modified residue is either not detected or is destroyed under the harsh chemical environment. Mass spectrometry is probably the best method for the detection of post-translational modifications. Krishna and Wold [54] compiled a useful list of post-translational modifications of proteins. Some of the common post-translational modifications include acylation of the N-terminus, phosphorylation of serine, threonine and tyrosine, oxidation of methionine, glycosylation of serine and asparagine, cleavage of N- and/or C-terminal residues by proteolytic enzymes and formation of C-terminal amides. For simple modifications, such as acetylation or phosphorylation, the analysis is simple and can be carried out using the normal peptide mapping procedure [55]. Caprioli et al. [56] developed a procedure to generate a "sequence-ordered" map from a protease digest by using the overlap information produced from a time-course protease digest of a protein. This technique is based on the fact that the rate of cleavage can differ widely from site to site and that all sites are not equally accessible. Clostripain was used as a limited proteolytic enzyme because it hydrolyzes only at the C-terminal side of arginyl residues. Since not all arginines are accessible to the enzyme, a time-course digestion would produce peptide fragments with large overlapping regions. One other approach is to perform the digestion in oxygen-18-enriched water; only the C-terminal fragment will be unlabeled [57].

Perhaps the most common modification of proteins is glycosylation. Analysis of glycoproteins by mass spectrometry is complicated by the

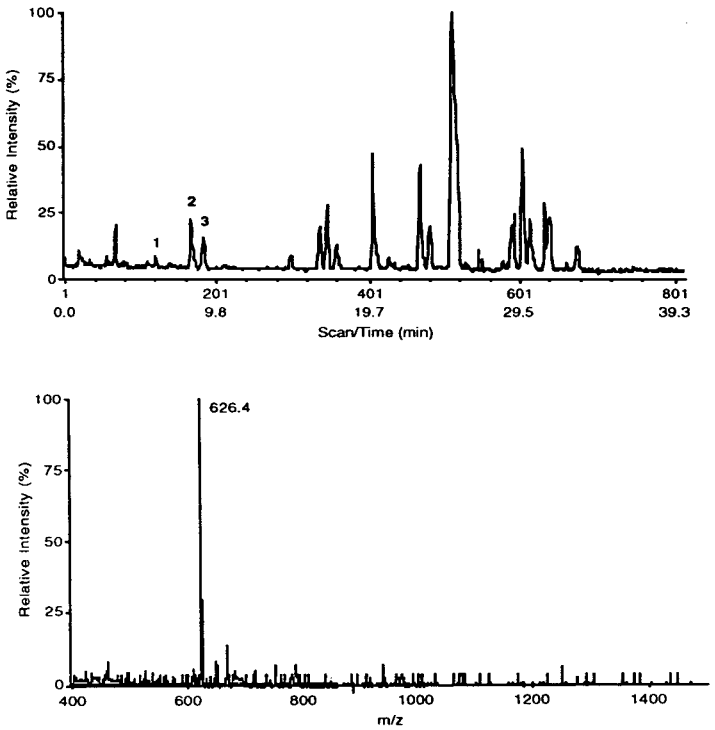


Fig. 6. LC–MS of a trypsin digest of carboxyamidomethylated hGH using ESI-MS. Peak 1 contains an ion at m/z 626.4 which matches the LEDGSPR (T_{14}) fragment of the protein. Every peak in the TIC chromatogram can be analyzed in the same manner to confirm the primary sequence of the protein.

heterogeneity of the carbohydrate moiety. The most reliable method of determining the molecular mass of a heavily glycosylated protein is MALDI-TOF mass spectrometry; the complexity derived from the multiple charge states present in ESI-MS generally renders it impractical for this purpose. However, when the glycoprotein is digested into smaller fragments, FAB, MALDI and ESI mass spectrometry can be successfully applied to the analysis. A classical approach to glycoprotein analysis is to deglycosylate completely using hydrazinolysis, isolate and purify the polysaccharides and then perform structural characterization. A major drawback of this technique is that sites of glycosylation on the protein cannot be determined because of the destruction of the protein. Application of mass spectral techniques to glycoprotein analysis has provided a means of characterizing these glycoproteins efficiently. Sequence analysis of oligosaccharides is outside the scope of this paper, although Sutton et al. [58] have utilized MALDI-MS to identify the site of glycosylation and to obtain the sequence of the carbohydrate moiety. The initial problem with an unknown protein is to determine whether it is glycosylated. One approach has been to apply a highly sensitive lectin screen where a range of lectins must be employed to ensure that all of the common oligosaccharides will be detected [59]. Once a protein is confirmed to be a glycoprotein, it is then fragmented chemically or enzymatically. Hawke et al. [60] have reported that, by using an affinity column, most of the glycopeptides could be separated. The remaining fractions were analyzed by HPLC and screened again for glycopeptides. The isolated glycopeptides were then subjected to mass spectral analysis. The determi-

nation of N-linked oligosaccharides can be performed readily with a "pure" oligopeptide. Digestion of the oligopeptide with peptide-N-glycosidase F (PNGase F) in 50% ^{18}O-labeled water affords a pair of signals (MH^+ and $MH^+ + 2$) due to the partial incorporation of oxygen-18 into the β-carboxyl group of aspartic acid [61]. Unfortunately, there is no known enzyme which can hydrolyze O-linked oligosaccharides. Carr et al. [2] have proposed an elegant procedure using LC–ESI-MS to identify selectively N- and O-linked oligosaccharides in glycoproteins. The glycopeptides are identified using a carbohydrate marker ion. This is done by stepping the "orifice potential" to enhance the low-m/z fragments, in particular m/z 204. This experiment allows one to confirm the presence of a glycoprotein and to localize the glycopeptide fragments in the enzymatic digested chromatogram. By performing a "precursor scan" of m/z 204, glycopeptides can be selectively detected in the presence of other peptides. It should be noted that the spectrum generated using this procedure corresponds to all parent ions that have decomposed to yield m/z 204. O-Linked glycopeptides can be differentiated from N-linked glycopeptides by analyzing a sample that has been treated with PNGase F. Sites of glycosylation can be obtained by MS–MS analysis of the glycopeptides. Edman sequencing can also be used to determine the site of glycosylation (absence of signal where the residue is glycosylated). A schematic diagram of a strategy for glycoprotein analysis is shown in Fig. 7.

One should remember that carbohydrates offer a relatively small number of protonation sites. This point is very important when ESI-MS is used for glycoprotein analysis. For example, if a large oligosaccharide is attached to a small peptide, there might not be enough charge to bring the ions into the mass range of the quadrupole analyzer. In this case, MALDI-MS is a good alternative method.

3.2.3. Protein degradation products

Since almost all degradation products of proteins result in a change in molecular mass compared with the original protein, mass spectrometry can be used to detect such degradation products, in principle. However, mass determination of the intact protein does not yield information as to the location of the chemical modification. Further, for large proteins (M_r greater than approximately 20 000), mass spectrometry generally does not offer sufficient resolution to detect modifications of only 1 u (e.g., as a result of deamidation). Hence the most common use of mass spectrometry with regard to characterization of protein degradation products is the combined use of peptide mapping and mass spectrometry. Examples of this approach include the characterization of deamidated and methionine sulfoxide forms of human growth hormone [63], hydroxylamine cleavage products in insulin-like growth factor I [64] and deamidated forms of tumor necrosis factor [65]. In each of these examples the approach used was to generate peptide fragments (usually by trypsin digestion), isolate those peptide fragments which differed from the fragments arising from the non-degraded protein and characterize the modified peptide fragments by FAB or ESI (usually employing MS–MS to confirm the sequence modification which had occurred). In most cases, N-terminal analysis by Edman degradation is also employed to provide corroborating evidence.

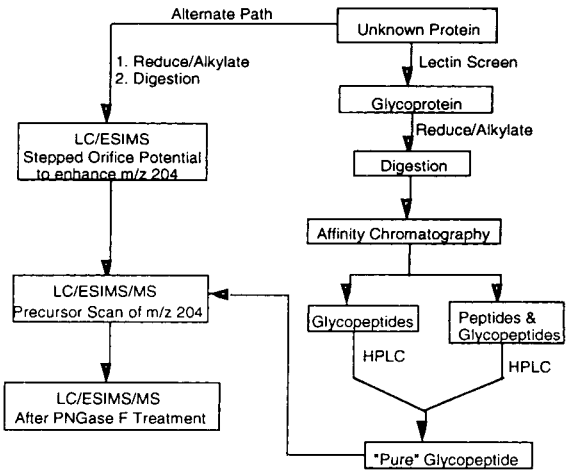


Fig. 7. Strategy for glycoprotein analysis (see Refs. [60] and [62]).

3.2.4. Protein metabolites

Wroblewski and co-workers have used ESI-MS to study the in vitro metabolism of human growth hormone (hGH) [66] and the in vivo metabolism of des(64,65)-human proinsulin [67]. In these examples the protein is proteolyzed, not by the addition of exogenous proteases as in a typical peptide mapping experiment, but by the action of endogenous proteases. An abundance of information concerning the activities involved, the metabolic products and potential metabolic pathways can be obtained from such experiments.

Human growth hormone was incubated with preparations of rat thyroid gland; the products were isolated by reversed-phase HPLC and characterized by a combination of ESI-MS and N-terminal Edman sequencing. The predominant activity that acted on hGH was found to be a chymotrypsin-like serine protease, biochemically similar to rat mast cell protease-I, with cleavages occurring exclusively at Tyr–Phe–Leu–Xaa bonds. The presence of a carboxypeptidase activity was indicated by the detection of metabolites that were truncated by a single amino acid at the C-terminus. The sequence of events leading to the degradation of hGH in this system was found to be initiated by a cleavage between Tyr143–Ser144 to produce a two-chain form of the protein. This was followed by the cleavage of the initial two-chain form at Tyr42–Ser43, liberating the N-terminal peptide Phe1–Phe42. Subsequent events over a 4.5 h incubation led to the degradation of hGH to a set of more than 20 peptides with masses at or below 2300 [66].

The metabolism of des(64,65)-human proinsulin was examined in rats after subcutaneous administration. Circulating insulin-like immunoreactivity in the plasma 25 min after injection was evaluated by anion-exchange and reversed-phase HPLC. Both techniques indicated the presence of a metabolite comprising 5–10% of the circulating immunoreactivity and having the retention characteristics of human insulin. The remainder had retention characteristics of des(64,65)-human proinsulin. The peaks of immunoreactive material were isolated and their structures determined using reversed-phase HPLC and ESI-MS. The major circulating component co-eluted with des(64,65)-human proinsulin and had an identical mass spectrum. Two circulating metabolites were identified. These metabolites co-eluted by reversed-phase HPLC with human insulin and diarginyl(B31,32)-human insulin and had mass spectra identical with the standard compounds. The data indicate that proteolytic processing of des(64,65)-human proinsulin involves an initial tryptic cleavage at the carboxy side of ArgB32, with the formation of human insulin by the subsequent action of a carboxypeptidase to remove the ArgB31–ArgB32 dipeptide from diarginyl(B31,32)-human insulin. The results suggest that some of the pharmacological activity of des(64,65)-human proinsulin may be mediated in part by circulating insulin-like metabolites [67].

3.2.5. Disulfide bond determination

Many proteins, especially secreted or non-cytoplasmic proteins, contain disulfide bonds between pairs of proximal cysteines. The disulfide bonds play a role in the stabilization of the tertiary structure of these proteins. Proteins that are produced through recombinant DNA technology are often refolded from strong denaturants into biologically active forms and it is critical to determine whether the "correct" cysteines are involved in forming the disulfide bonds. Likewise, for naturally occurring proteins, it is important to characterize the disulfide pairing pattern. Some proteins contain free cysteine residues, not involved in disulfide bonds, and it is important to identify these residues. For example, the protease clostripain contains a cysteine at the active site and this free thiol group is essential for the activity [68]. Although analytical methods have been developed to determine the thiol content in a protein [69], the determination of the locations of disulfide bonds is more of a challenge. Disulfide-containing peptides can be identified by their amino acid composition or sequence only if they are purified to homogeneity. For a protein containing many disulfide bonds, this approach cannot be applied. However, total cysteine content can be determined by complete reduction of

the disulfide bonds with a suitable reagent (DTT or β-mercaptoethanol are commonly used) followed by alkylation with iodoacetic acid or iodoacetamide. The total mass increment of the intact protein after this procedure divided by the mass of the alkylating group (e.g., 58 for acetamide) yields the number of cysteines in the protein. If this procedure is repeated, but without prior reduction, only the free cysteines, not involved in disulfide bonds, are alkylated. It should be cautioned that some proteins may have free cysteines buried inside the tertiary structure, preventing them from being alkylated. Therefore, it might be necessary to perform the alkylation in strong denaturing conditions. Again, the mass increment divided by the mass of the alkylating group yields the number of free cysteines in the protein. The difference between the total number of cysteines and the number of free cysteines divided by 2 (two cysteines per disulfide bond) gives the number of disulfide bonds. One of the greatest problems in localizing the disulfide bonds is disulfide exchange. Application of mass spectrometry to solve this problem originates from the observation by Yazdanparast et al. [70] that disulfide bonds are reduced in situ during peptide analysis by FAB-MS.

The determination of the cysteines involved in a disulfide pair is accomplished by a peptide mapping approach. The key to applying this approach successfully is to isolate peptides that contain a single disulfide bond. For large proteins cleavage with cyanogen bromide can be used to generate large peptide fragments that are more susceptible to proteolytic digestion. The protein or large protein fragments are then digested with a suitable protease. Although trypsin can be used, care must be taken that disulfide interchange not occur. If disulfide interchange is a problem, an acid protease such as pepsin should be used. Following proteolytic cleavage, the peptides are isolated and subjected to both Edman sequencing and mass spectral analysis. The peptides that contain multiple sequences are candidates for peptides that contain disulfide bonds. If more than two sequences are found in an isolated peptide peak, this indicates the existence of more than a single disulfide bond and requires subfragmentation with a second protease. Even the existence of only two sequences does not guarantee a single disulfide bond if cleavage has not occurred between the two cysteines of one disulfide bond. Once a set of peptides, each peptide containing a single disulfide bond, is obtained, the disulfide bonds can be assigned unambiguously. If the protein being analyzed has a free sulfhydryl, it will be necessary to alkylate it before starting the fragmentation procedure.

3.2.6. Ligand binding

In the field of drug discovery, a common theme is the identification of low-molecular mass compounds that bind very tightly to a target protein. An example of this is the covalent attachment of protease inhibitors to proteases, often at the active site of the protease. Covalently bound ligands can be studied using the normal peptide mapping procedure. Sall and Kaiser [71] used peptide mapping and ESI-MS to localize the site of binding of methyl 3-(2-methyl-1-oxopropoxy)[1]benzothieno[3,2-*b*]furan-2-carboxylate, a potent and highly selective inhibitor of thrombin, to the active site of thrombin. They were able to show that this compound formed a stable acyl enzyme complex at the active site Ser-205 which, along with His-43 and Asp-99, make up the active site triad of thrombin.

A similar approach was used with the viral enzyme, rhinovirus 3C protease, to show that an inhibitor of this enzyme bound covalently to the active site Cys-146 (Zimmerman and Becker, personal communication). In additional to localizing the site of attachment, information on the nature of the chemical reaction involved can be obtained from the increment in mass that the protein undergoes.

3.2.7. Enzyme active sites

A related application is the characterization of enzyme active sites. The general approach is to utilize a specific affinity label which forms a covalent bond with one of the active site residues. Peptide mapping is then applied to localize the peptide containing the modification.

Several examples of this approach can be found in the recent literature.

The apparent active site of human leukocyte glycoasparaginase, an enzyme involved in the degradation of the N-glycosidic linkage between asparagine and N-acetylglucosamine in various glycopeptides, has been identified by labeling with an inhibitor, 5-diazo-4-oxo-L-norvaline, an asparagine analog [72]. The labeled protein was digested with trypsin and analysis of the peptides by mass spectrometry revealed that the inhibitor was attached through an α-ketone ether linkage to the hydroxyl group of the N-terminal threonine.

Inactivation of histidine ammonia lyase from *Pseudomonas putida* was accomplished by treatment with L-cysteine at pH 10.5 in the presence of oxygen [73]. Inactivation was accompanied by the formation of a new species with a UV absorbance maximum at 340 nm. Following trypsin and staphylococcal V8 protease digestion, a peptide was isolated that was found to contain a modification of mass 184 by ESI-MS. The modification was localized to Ser-143 of this enzyme and the authors conclude that this represents the site for attachment of an electrophilic cofactor required for histidase activity.

3.2.8. Peptide mass maps

An innovative approach to the identification of protein sequences in a large protein sequence database such as the Protein Information Resource (PIR) or the SWISSPROT database has been created by Yates et al. [74]. The approach involves the proteolytic digestion of an unknown protein and then analysis of the peptides by a peptide mapping/MS approach. They have demonstrated that a set of observed masses which is less than 50% of the total number of predicted masses can be used to identify a protein sequence in the database. Mass maps generated by ESI-MS, MALDI-MS and FAB-MS should all work with this approach. When multiple matches are found, tandem mass spectrometry (discussed in the next section) can be used to establish sequence similarity.

3.3. Sequencing

3.3.1. Tandem mass spectrometry

Tandem mass spectrometry (MS–MS) was originally developed for the investigation of ion chemistry. However, its major application currently is in determining the structure of biomolecules. The tandem mass spectrometer could be a multisector magnetic instrument (tandem double focusing) or a triple quadrupole instrument. In both cases, the "first" mass spectrometer (or the first quadrupole) serves as a mass separator. The selected ion (precursor ion) can be induced for fragmentation by colliding with an inert gas. This process is often referred to as collision-induced dissociation (CID). The fragment ions, commonly referred to as product ions, are analyzed in the "second" mass spectrometer (or the third quadrupole). A detailed description of tandem mass spectrometers has been given elsewhere [75]. This section will only describe the use of tandem mass spectrometry in peptide sequencing.

A diagram of tandem mass spectrometry is shown in Fig. 8. A peptide mixture consisting of R, M and N components is ionized. Peptide M is then selected by the first mass analyzer and subsequently induced by colliding with a neutral gas such as argon or helium to give fragment ions A, B and C. These fragment ions can be scanned by the third mass analyzer to give the mass spectrum of the product ions. The use of a collision gas is essential to produce a significant number of intense product ions.

In peptides, the fragmentation generally occurs at the peptide backbone. However, cleavage at peripheral bonds can occur, particularly when using magnetic sector mass analyzers. For peptides with mass greater than 2500, fragmentation seldom occurs because the vibrational energy

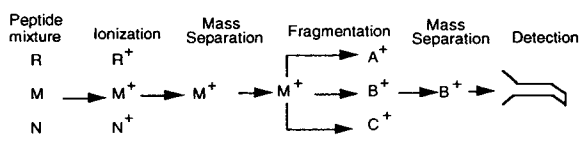


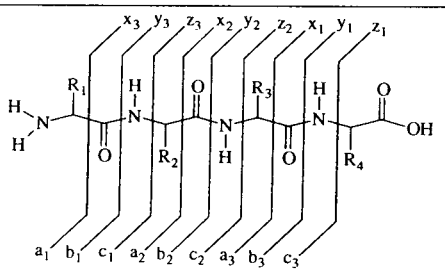
Fig. 8. MS–MS of R, M and N peptide mixture.

induced by collision is distributed through a larger number of bonds (except for post-source decay MALDI-MS where fragmentation of peptide with molecular mass up to 3000 Da can occur). In order to understand peptide fragmentation, a nomenclature system was proposed by Roepstorff and Fohlman [76]. This system was soon modified to use lower case rather than capital letters to avoid confusion, e.g. C_n could be mistaken for a cysteine residue at position *n* [77]. Table 3 shows the common types of fragment ions.

The mechanism for fragmentation is not fully understood at the present time. The formation of a_n, b_n and y_n is generally accepted to involve initial protonation of the amide nitrogen atom. d_n Ions retain the positive charge at the N-terminus and are formed as a result of the loss of an alkene from the side-chain of the other terminus of a_n ions. d_n Ions provide a means to differentiate the isomeric amino acids leucine and isoleucine. If leucine were the amino acid at the terminus, an isopropyl radical would be eliminated. On the other hand, isoleucine would eliminate a methyl or an ethyl radical [78,79]. Johnson et al. [80] proposed mechanisms for the formation of the fragment ions. It is important to know that these mechanisms are chemically plausible, but there are few experimental data to support them except for the observed ions. In the most common form of ESI-MS, in which quadruple mass analyzers are used, b and y fragment ions predominate.

Tandem mass spectrometry has been used to study a large number of peptides and proteins. Some of the notable works include the studies of bradykinin [81], substance P [82], somatostatin [83] and thioredoxin [84]. So far, most of the proteins studied are well characterized and have known primary sequences. For an unknown protein, the tandem mass spectrum can be very complicated because some proteins exhibit both N- and C-terminal fragments. There are also fragments that do not fit in any of the proposed fragment ions. Despite these complications, mass spectrometry offers protein chemists a means to obtain sequencing information much faster than any other analytical method. There are also

Table 3
Common types of fragment ions



Ion	Type of cleavage	Terminus	Structure
a_n	Backbone	N	$^{+}NH{=}CHR$
b_n	Backbone	N	$-NHCHRCO^{+}$
c_n	Backbone	N	$-[NHCHRCONH_2]H^{+}$
x_n	Backbone	C	$[OCNHCHRCO_2H]H^{+}$
y_n	Backbone	C	$H_3^{+}NCHRCO-$
z_n	Backbone	C	$[CHRCONHCHRCO_2H]H^{+}$
d_n	Side-chain	N	$[-NHCH{=}CHR']H^{+}$
v_n	Side-chain	C	$[HN{=}CHCONHCHRCO_2H]H^{+}$
w_n	Side-chain	C	$[R'CH{=}CHCONHCHRCO_2H]H^{+}$

computer programs which can be used to assist in the analysis of spectral data of peptides. Some of the commonly used programs include MacProMass [85], Mac Mass [86] and COMPOST [87].

3.3.2. Post-source decay MALDI-MS

In addition to tandem mass spectrometry, post-source decay (PSD) is a new technique for peptide sequencing. It has been observed that large peptide and protein ions are unstable on their way through a time-of-flight instrument. The metastable decay of these ions has been detected using two-stage reflectron MALDI-MS. The term post-source decay was introduced by Spengler and co-workers [88–91] to describe the "delayed" fragmentation of the desorbed analyte ions. The fragmentation was due to the multiple collisions of analyte ions with matrix ions during the early plume expansion and ion acceleration followed by collisions with residual or admitted gas molecules in the field-free drift region of a MALDI instrument. Unimolecular decomposition of the precursor ion gives rise to fragmentation patterns which can be used to determine the sequence of the peptide. The fragment ions travel with nearly the same velocities as their precursors, but with lower kinetic energies. Thus, by applying a retarding field at the reflector, one can separate the product ions from the precursor ions. Because of the lower kinetic energies of the product ions, they do not penetrate into the retarding field as deeply as their precursors; therefore, they leave the reflector earlier and arrive sooner at the detector. PSD-MALDI offers many advantages over conventional tandem mass spectrometry. The instrumental sensitivity for product ions is at least two orders of magnitude better than tandem mass spectrometry owing to the high ion transmission and high ion yield (20–80% of precursor ion). Also, MALDI-MS offers a much longer dissociation time range than a conventional collision cell which allows the activation and decay of peptides of M_r up to 3000. The fragmentation patterns of PSD-MALDI are very similar to that of the low-energy CID [92]. Typically, a_n, b_n, C_n, $a_n - 17$, $b_n - 17$, x_n, y_n and z_n ions are observed in a mass spectrum. The loss of 17 u is due to loss of ammonia from arginine residues.

There are drawbacks associating with PSD-MALDI which should be mentioned. First, mass calibration can be difficult because one would have to perform tedious calibration with known precursor ion and product ion masses. Therefore, calibration is normally done using computer software. This software takes into account the instrument geometry, electrical field parameters and the flight times of precursor and product ions to calculate the ion masses. Second, for large peptides (M_r above 1500), it is often necessary to increase the laser irradiance to have sufficient desorption. However, increasing the laser irradiance causes a decrease in resolution which ultimately increases the uncertainty in mass assignment of these peptides. Third, because of the complexity of the fragmentation patterns and the lack of algorithms for cleavage interpretation, analysis of PSD mass spectra requires much more time than it takes to acquire the mass spectrum. Despite these difficulties, PSD-MALDI has gained acceptance in peptide sequencing. For example, Zambias et al. [93] utilized PSD-MALDI for the analysis of a covalently bound peptoid (modified peptide) to a polymeric bead. Hoyes et al. [94] utilized PSD-MALDI for the analysis of ACTH peptide (M_r 2466). Spengler and Kaufmann [95] utilized PSD-MALDI to localize the site of a post-translational modification of a peptide. The analysis of a lipopeptide (M_r 1863) located the fatty acid on the N-terminus of the peptide. One interesting, perhaps very important, application of PSD-MALDI is the possibility of using this technique for the characterization of unknown proteins. Yu et al. [96] performed PSD-MALDI on an unknown tryptic fragment. The resulting primary sequence of this peptide was compared with existing sequences in a database to suggest that this protein resembles CHO MCP-1 protein.

3.3.3. Carboxypeptidases

Carboxypeptidases, exoproteases which cleave only the C-terminal residues in proteins, have been used in protein sequencing with variable success. Four carboxypeptidases, carboxypepti-

dase P (CPP), A (CPA), B (CPB) or Y (CPY), or a combination of these have been most commonly used in sequencing applications. In general, the rates of cleavage by these enzymes are dependent on the polarity of the C-terminal amino acid side-chain, making sequencing progress extremely variable. Because of this variability, carboxypeptidases have traditionally been used to determine only the C-terminal residue of a protein by identifying the amino acid released by its chromatographic retention time [97].

A different approach is to monitor the truncated protein instead of the released amino acid and the techniques of ESI-MS, PD-MS and MALDI-MS now make this approach feasible. A digestion reaction is carried out using carboxypeptidase(s) of choice. At selected intervals, aliquots are removed and the reaction is quenched and stored for later analysis. Using PD-MS, the digestion can be carried out directly on the nitrocellulose support [98,99]. With ESI-MS, the digestion mixture can be monitored by continuous infusion. Smith and Duffin [100] used carboxypeptidase P to sequence interleukin 3 (IL-3) and superoxide dismutase (SOD). IL-3 has a C-terminal sequence of ...TTLSLAIF. The sequencing was performed in 15 m*M* ammonium acetate buffer (pH 4.0). Under these conditions, six amino acid residues were digested by CPP. Additional reaction time did not release any further residues. Superoxide dismutase has a C-terminal sequence of ...CGVIGIAK. Digestion of this protein yielded only one residue. However, after reducing and alkylating the protein with iodoacetic acid, three residues were released by the enzyme. The sequencing did not proceed past the glycine residue even with extended reaction time. Schär et al. [101] reported the use of a mixture of CPA and CPB and also CPY for the digestion of synthetic parathyroid hormone using MALDI-MS. After about 2 h of digestion, 21 amino acid residues were removed from the protein. It was found that CPY digests valine rapidly and very slowly at histidine residues, hence it was not possible to localize all the truncated proteins. Also, truncated peptides of M_r below 1200 were difficult to detect because of interference from the matrix. Rosnack and Stroh [102] described the use of a low-flow reactor to monitor the digestion of glucagon and apomyoglobin using ESI-MS. The protein and enzyme (CPP) were mixed and infused into a fused-silica reactor with continuous monitoring by ESI-MS. Glucagon gave sequence ions for the first 19 amino acids; however, apomyoglobin gave only 50% of the first 30 amino acids. Attempts were made to sequence cytochrome *c* and carbonic anhydrase. Unfortunately, no sequence information was obtained from these digestions.

The application of carboxypeptidases to protein sequencing has been investigated in our laboratory. We decided to use ESI-MS in combination with CPP and CPY for protein sequencing. Glucagon (HSQGTFTSDYSKYLDSRRAQDFVQWLMNT) was used as a model compound with a mixture of CPP and CPY as reagents. The detection of the truncated forms of glucagon was carried out by ESI-MS. The combination of CPP and CPY was found to be necessary for successful digestion. For example, digestion with CPY alone was retarded at the tryptophan residue and virtually stopped at the aspartic acid residue. Digestion with CPP alone proceeded past these residues but was retarded at the leucine residue. A combination of CPP and CPY resulted in digestion of the entire glucagon molecule in about 70 min. (Fig. 9). Peaks associated with all of the predicted glucagon fragments could be identified in the mass spectrum taken at 70 min although there was a high background. Obviously, the sequence of glucagon is known and this greatly facilitates the analysis of the mass spectrum. An unknown protein would present a much greater challenge and at present the technique is probably best applied to the confirmation of a known or predicted sequence. The ability to obtain sequence information readily from the C-terminus of a protein is, however, an important advance in protein characterization.

3.4. Non-covalent interactions

Non-covalently bound ligands pose a problem in the analysis of these complexes by most ionization techniques. Henion and co-workers

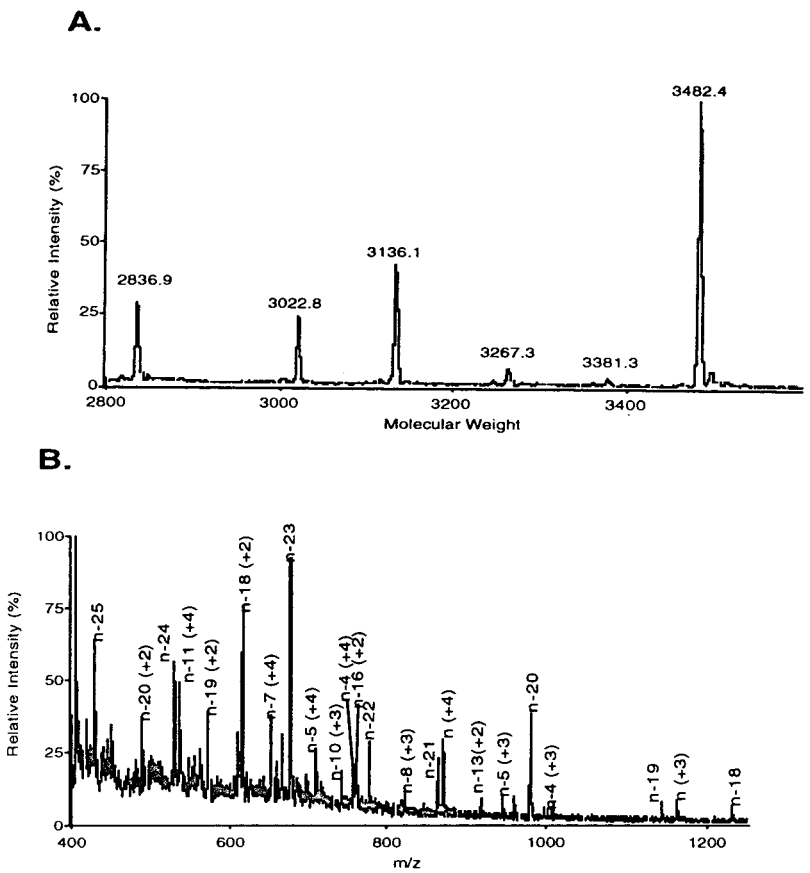


Fig. 9. Sequencing of glucagon using a mixture of CPP and CPY. (A) Reconstructed mass spectrum of digestion using CPP and CPY after 15 min. Data were acquired in the chromatographic mode. (B) The actual ion m/z of digestion after 70 min; n refers to the glucagon and $n-18$ refers to glucagon minus 18 amino acid residues from the C-terminus. The charges of the ions are shown in parentheses. The charge is not specified when the ion is in a +1 charge state. The +1 charge states were assigned by examination of the isotopic pattern of the peak. A total of 21 nmol of glucagon was consumed. See to Table 1 for peak assignments.

suggested that non-covalent molecular association complexes might be detectable under the "soft ionization" conditions offered by ESI and that the reaction might be monitored in a real-time mode [103]. The first successful application of ESI-MS to the detection of non-covalent receptor–ligand complexes was reported by the same laboratory [104]. The macrolide FK506, an immunosuppressive agent, inhibits T-cell activation when in a complex with the cytoplasmic receptor FKBP. FKBP was mixed with a slight excess of FK506 at pH 7.5, and a new signal appeared corresponding to the complex. Two important controls were performed to ensure that covalent adducts did not form. First, when FK506 was combined with denatured FKBP, the signal for the complex was not observed. Second, when FK506 was replaced with a more potent reagent (higher affinity toward the receptor), the peak corresponding to the new complex was much stronger in intensity.

ESI-MS has also been used to study the hydrolysis of a hexasaccharide of N-acetylglucosamine (NAG_6) by lysozyme. A mass spectrum immediately after mixing the enzyme and substrate reveals a new protonated peak which corresponds to the lysozyme–NAG_6 complex. During the time-course reaction, the spectrum

shows the disappearance of lysozyme–NAG_6 and appearance of lysozyme–NAG_4 and lysozyme–NAG_3 [105].

Kata and Chait [106] demonstrated the utility of ESI-MS in analyzing protein–ligand complexes by studying the heme–globin interaction. The oxygen-carrying protein myoglobin contains a non-covalently bound heme group in the hydrophobic pocket of the native globin chain. At low pH values, unfolding of the protein results in complete protonation of buried histidine residues, resulting in a higher charge-state and a loss of the heme group. A mass spectrum obtained from an aqueous myoglobin solution at pH 3.35 (completely denatured protein) shows a single distribution of peaks and the total absence of ions correspond to the heme–globin complex. However, at pH 3.9, where both native and denatured forms of myoglobin coexist, the mass spectrum exhibits two distinct distributions of peaks. One set of peaks corresponds to the denatured myoglobin (higher charge state) and the other set corresponds to the intact heme–globin complex. These results suggest that native, non-covalent complexes of proteins and cofactors in solution can be preserved in the gas phase and observed by mass spectrometry.

Many proteins self-associate to form non-covalent dimers or higher oligomers. For example, human growth hormone (hGH), under certain conditions, will form a tightly complexed, non-covalent dimer [107]. This dimer is stable under aqueous conditions and can be resolved from the monomer by high-resolution size-exclusion chromatography. Re-chromatography of the isolated dimer on the same system reveals that this fraction is still dimeric with the retention time expected for the dimer and has not reverted to monomer. However, addition of 30% acetonitrile to the sample before re-chromatography caused the conversion of the dimer into monomer. Likewise, treatment with SDS sample buffer followed by analysis by SDS-PAGE shows that the dimer has been converted into monomer. This hGH dimer has been studied by ESI-MS (Fig. 10). Under the conditions normally used to obtained protein spectra, 0.1–1.0% acetic acid–50% acetonitrile, only the spectrum

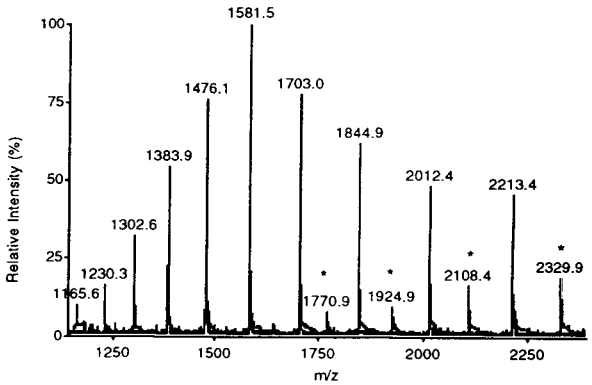


Fig. 10. Spectrum of a non-covalent dimer of human growth hormone. The spectrum was obtained on a Sciex API-III instrument by infusing at 2.5 μl/min a sample of hGH dimer at a concentration of 1 mg/ml in 0.1% acetic acid. The m/z values corresponding exclusively to the dimer are marked with asterisks.

of monomeric hGH was observed. This is not surprising, since 30% acetonitrile had been shown to dissociate the dimer. However, by omitting the acetonitrile and infusing a solution at a concentration of 1 mg/ml in 0.1% acetic acid, evidence for the dimer was obtained. The spectrum appears to be an overlay of the spectrum of the dimer and the spectrum of the monomer. To rule out the possibility that the observed dimer spectrum was the result of dimer formation in the mass spectrometer, a solution of monomeric hGH was studied under identical conditions and no evidence of dimer formation was observed.

3.5. Protein folding/higher order structural characterization

3.5.1. Deuterium exchange

While mass spectrometry is most commonly used to assess the primary structure of proteins or peptides, there are various experimental approaches available which can yield information concerning protein folding and higher order (secondary, tertiary or quaternary) structure. One such approach is the determination of hydrogen–deuterium exchange rates under controlled solution conditions. In one application

the exchange rates of the various amide hydrogens in cytochrome *c* were determined using HPLC–FAB-MS [108]. Cytochrome *c* was incubated in ^2H_2O at various temperatures. After incubation for preset time intervals, the protein was transferred into a solution in which deuterium exchange is very slow (pH 2–3, 23°C) and digested with pepsin. The number of deuterium atoms incorporated into each of the proteolytic peptides was determined by HPLC–FAB-MS. This approach can provide information regarding the secondary or tertiary structure of the protein, since amides which exchange readily have a greater solvent accessibility than do amides which exchange slowly. In cases where a biosynthetic form of a protein is being structurally characterized, this approach can be especially useful if a native (natural source) form of the protein is available as a comparator.

3.5.2. *Proteolysis*

Another useful strategy for protein structure evaluation is limited proteolytic digestion. This approach involves digestion of the protein with a proteolytic enzyme, under conditions such that the rate of proteolysis is sufficiently slow to allow determination of the kinetics of formation of the various peptide fragments. This approach has been used to compare the structures of wild-type and recombinant yeast (mutant) calmodulins, using ESI-MS [109]. In this study, the calmodulins were digested with trypsin using varying protein/enzyme ratios. After digestion for selected time intervals, the reaction was terminated by the addition of soybean trypsin inhibitor. The peptide solution was desalted and the characteristic fragments were determined by ESI-MS (without chromatographic separation). The use of ESI-MS directly, without prior HPLC separation, makes this a particularly rapid assay method. However, the lack of an HPLC separation makes the technique more susceptible to matrix interference (e.g., salts) and the quantitative capability of such a method is therefore relatively limited. Nonetheless, this approach provides significant information regarding the structure of the recombinant protein.

3.5.3. *Charge state*

The charge-state distribution observed for a protein in ESI-MS can provide significant information regarding protein folding. For example, ESI-MS has been used to study the mechanism of refolding of acid-denatured myoglobin [110]. In this study, the ESI-MS analysis was conducted at neutral pH, since the heme group was demonstrated to remain attached to the protein under these conditions. Myoglobin was initially denatured in 10% acetic acid and then refolded by adjusting to various pH levels (5–8) by addition of ammonia solution. The molecular mass and charge-state distribution of the partially, or fully, refolded protein was then determined by ESI-MS analysis. Based on the pH dependence of the observed molecular mass (affected by the presence or absence of the heme group) and charge state distribution, the ESI-MS data indicated that myoglobin refolds in two major steps. The unfolded polypeptide chain first refolds to form a "native-like" structure without the heme group, and the binding cavity of this structure then binds the heme group by noncovalent interaction. This example clearly demonstrates that ESI-MS can provide useful information regarding protein folding pathways, and this information can be used in conjunction with other more traditional techniques (e.g., circular dichroism) to obtain a more complete understanding of these complex phenomena.

3.5.4. *Cross-linking reagents*

Addition of cross-linking reagent to proteins, followed by subsequent structural characterization of the linked domains, has been a traditional approach used to probe protein secondary and tertiary structure. Mass spectrometry clearly can provide an additional analytical tool to be used in this approach. The interfacing domains in recombinant human erythropoietin (EPO) have been characterized in this manner [111]. In this case, amino groups were selectively cross-linked by specific cross-linkers such as disuccinimidyl suberate or dithiobis(succinimidyl propionate). The linked regions were then characterized by trypsin digestion followed by HPLC separation. The isolated tryptic peptides were characterized

by sequencing (Edman degradation) and mass spectrometry. These data were used to determine which lysine groups in EPO were adjacent in the three-dimensional structure. The study also demonstrated that non-glycosylated and glycosylated forms of EPO have a high degree of similarity with regard to the protein conformation. A similar approach has been described for studying the subunit association of proteins [112]. In this case the proteins are cross-linked with glutaraldehyde and the molecular mass of the adduct is determined by MALDI-MS. The validity of this approach was confirmed by analyzing a variety of proteins having various known states of association.

References

[1] A.L. Burlingame, T.A. Baillie and D.H. Russell, *Anal. Chem.*, 64 (1992) 467R.
[2] S.A. Carr, *Adv. Drug. Deliv. Rev.*, 4 (1990) 113.
[3] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
[4] J.A. McCloskey (Editor), *Methods Enzymol.*, 193 (1990) 960.
[5] K. Biemann, *Mass Spectrometry, Applications to Organic Chemistry*, McGraw-Hill, New York, 1962.
[6] M.S.B. Munson and F.H. Field, *J. Am. Chem. Soc.*, 88 (1966) 2621.
[7] M. Barber, R.S. Bordoli, R.D. Sedgwick and A.N. Taylor, *Nature*, 293 (1981) 270.
[8] M. Barber, R.S. Bordoli, R.D. Sedgwick, G.J. Elliot and A.N. Taylor, *Anal. Chem.*, 54 (1982) 645A.
[9] T. Keough, *Anal. Chem.*, 58 (1985) 2027.
[10] J.A. Sunner, R. Kulatunga and P. Kebarle, *Anal. Chem.*, 58 (1986) 1312.
[11] S. Naylor, A.F. Findeis, B.W. Gibson and D.H. Williams, *J. Am. Chem. Soc.*, 108 (1986) 6359.
[12] L.D. Detter, O.W. Hand, R.G. Cooks and R.A. Walton, *Mass Spectrom. Rev.*, 7 (1988) 465.
[13] R.M. Caprioli, W.T. Moore and T. Fan, *Rapid Commun. Mass Spectrom.*, 1 (1987) 15.
[14] H.D. Beckey, *Principles of Field Ionization and Field Desorption Mass Spectrometry (International Series in Analytical Chemistry*, Vol. 61), Pergamon Press, Oxford, 1977.
[15] R.J. Bueler, E. Flanigan, L.J. Greene and L. Friedman, *J. Am. Chem. Soc.*, 96 (1974) 3990.
[16] D.F. Torgerson, R.P. Skowronski and R.D. Macfarlane, *Biophys. Res. Commun.*, 60 (1974) 616.
[17] G. Jonsson, A. Hedin, P. Hakansson, B.U.M. Sunqvist, H. Bennich and P. Roepstorff, *Rapid Commun. Mass Spectrom.*, 3 (1989) 190.
[18] A. Tsabopoulos, G.W. Becker, J.L. Occolowitz and I. Jardine, *Anal. Chem.*, 171 (1988) 113.
[19] N. Furstenau, W. Knippelberg, F.R. Krueger, G. Weisz and K. Wien, *Z. Naturforsch., Teil A*, 32 (1977) 711.
[20] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida and T. Yoshida, *Rapid Commun. Mass Spectrom.*, 2 (1988) 151.
[21] M. Karas and F. Hillenkamp, *Anal. Chem.*, 60 (1988) 2299.
[22] R.C. Beavis and B.T. Chait, *Rapid Commun. Mass Spectrom.*, 3 (1989) 233.
[23] D.S. Nagra and L. Li, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 671a.
[24] R.M. Whittal, D.S. Nagra and L. Li, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 672a.
[25] D.S. Nagra and L. Li, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 987.
[26] R.M. Whittal, D.S. Nagra and L. Li, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 985.
[27] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson and M.B. Alice, *J. Chem. Phys.*, 49 (1968) 2240.
[28] M. Yamashita and J.B. Fenn, *J. Phys. Chem.*, 88 (1984) 4451 and 4671.
[29] M.L. Alexandrov, L.N. Gall, N.V. Krasnov, V.I. Nikolaev, V.A. Pavlenko and V.A. Shkurov, *J. Anal. Chem. USSR*, 40 (1985; transl. 1986) 1227.
[30] S.F. Wong, C.K. Meng and J.B. Fenn, *J. Phys. Chem.*, 92 (1988) 546.
[31] T.R. Covey, R.F. Bonner, B.I. Shushan and J. Henion, *Rapid Commun. Mass Spectrom.*, 2 (1988) 249.
[32] M. Mann, C.K. Meng and J.B. Fenn, *Anal. Chem.*, 61 (1989) 1702.
[33] J.V. Iribane and B.A. Thomson, *J. Chem. Phys.*, 64 (1976) 2287; 71 (1979) 4451.
[34] C. Schmelzeisen-Redeker, L. Bueftering, and F.W. Röllgen, *Int. J. Mass Spectrom. Ion. Processes*, 30 (1967) 663.
[35] M.A. Abbas and J. Latham, *J. Fluid Mech.*, 30 (1967) 663.
[36] S.K. Chowdhury, V. Katta and B.T. Chait, *Rapid Commun. Mass Spectrom.*, 4 (1990) 81.
[37] J.A. Loo, H.R. Udseth and R.D. Smith, *Anal. Biochem.*, 179 (1989) 404.
[38] O.A. Mirgorodskaya, A.A. Shevchenko, I.V. Chernushevich, A.F. Dodonov and A.I. Miroshnikov, *Anal. Chem.*, 66 (1994) 99.
[39] K.D. Henry, E.R. Williams, B.H. Wang, F.W. MacLafferty, J. Shabanowitz and D.F. Hunt, *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 9075.
[40] T.R. Covey, R.F. Bonner, B.I. Shushan and J. Henion, *Rapid Commun. Mass Spectrom.*, 2 (1988) 249.

[41] H.R. Morris, M. Panico, M. Barber, R.S. Bordoli, R.D. Sedgwick and A.N. Taylor, *Biochem. Biophys. Res. Commun.*, 101 (1981) 623.
[42] K. Kobayashi and J.A. Smith, *J. Biol. Chem.*, 262 (1987) 11435.
[43] G. Rhadhakrishna and F. Wold, *J. Protein Chem.*, 9 (1990) 309.
[44] S. Tsunasawa, H. Takakura and F. Sakiyama, *J. Protein Chem.*, 9 (1990) 265.
[45] B.W. Gibson and K. Biemann, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 1956.
[46] K. Biemann, *Biomed. Environ. Mass Spectrom.*, 16 (1988) 99.
[47] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
[48] V. Ling, A.W. Guzzetta, E. Canova-Davis, J.T. Stults, W.S. Hancock, T.R. Covey and B.I. Shushan, *Anal. Chem.*, 63 (1991) 2909.
[49] K. Biemann, *Annu. Rev. Biochem.*, 61 (1992) 977.
[50] J.T. Stults, T.M. Billeci, W.J. Henzel and S.C. Wong, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 421a.
[51] S.R. Weinberger, K.O. Boernsen, J.W. Finchy, V. Robertson and B.D. Musselman, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 775a.
[52] G.B. Hurst, T.J. Czartoski and M.V. Buchanan, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 975.
[53] F. Xiang and R.C. Beavis, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, 1994, p. 974.
[54] K.G. Krishna and F. Wold, in K. Imahori (Editor), *Methods in Protein Sequence Analysis*, Plenum Press, New York, 1993, pp. 167–172.
[55] T. Covey, B. Shushan, R. Bonner and W. Schröder, in H. Jörnvall, J. Höög and A. Gustavson (Editors), *Methods in Protein Sequence Analysis*, 1991, p. 249–256.
[56] R.M. Caprioli, B. Whaley, K.K. Mock and J.S. Cottrell, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry II*, Academic Press, San Diego, 1991, pp. 497–510.
[57] K. Rose, M.G. Simona, R.E. Offord, C.P. Prior, B. Otto and D.R. Thatcher, *Biochem. J.*, 215 (1983) 273.
[58] C.W. Sutton, J.A. O'Neill and J.S. Cottrell, *Anal. Biochem.*, 218 (1994) 34.
[59] K.L. Hsi, L. Chen, D.H. Hawke, L.R. Zieske and P.M. Yuan, *Anal. Chem.*, 198 (1991) 238.
[60] D.H. Hawke, K.L. Hsi, L.R. Zieske, L. Chen and P.M. Yuan, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry III*, Academic Press, San Diego, 1992, pp. 315–326.
[61] J. Gonzales, T. Takao, H. Hori, V. Besada, R. Rodriguez, G. Padron and Y. Shimonishi, *Anal. Biochem.*, 205 (1992) 151.
[62] S.A. Carr, M.J. Huddleston and M.F. Bean, *Protein Sci.*, 2 (1993) 183.
[63] G.W. Becker, P.M. Tackitt, W.W. Bromer, D.S. Lefeber and R.M. Riggin, *Biotechnol. Appl. Biochem.*, 10 (1988) 326.
[64] E. Canova-Davis, M. Eng, V. Mukku, D.H. Reifsnyder, C.V. Olson and V.T. Ling, *Biochem. J.*, 285 (1992) 207.
[65] S.A. Carr, M.F. Bean, M.E. Hemling and G.D. Roberts, in A.L. Burlingame and J.A. McCloskey (Editors), *Biological Mass Spectrometry*, Elsevier, Amsterdam, 1990, pp. 621–652.
[66] V.J. Wroblewski, R.E. Kaiser and G.W. Becker, *Pharm. Res.*, 10 (1993) 1106.
[67] V.J. Wroblewski, M. Masnyk and R.E. Kaiser, *Diabetes*, 42 (1993) 1407.
[68] A.M. Gilles, A. Lecroisey and B. Keil, *Eur. J. Biochem.*, 145 (1984) 469.
[69] T.W. Thannhauser, Y. Yonishi and H.A. Scheraga, *Methods Enzymol.*, 143 (1987) 115.
[70] Y. Yazdanparast, P. Andrews, D.L. Smith and J.E. Dixon, *Anal. Biochem.*, 153 (1986) 348.
[71] D.J. Sall and R.E. Kaiser, Jr., *J. Med. Chem.*, 36 (1993) 2350.
[72] V. Kaartinen, J.C. Williams, J. Tomich, J.R. Yates, III, L.E. Hood and I. Mononen, *J. Biol. Chem.*, 266 (1991) 5860.
[73] D. Hernandez, J.G. Stroh and A.T. Phillips, *Arch. Biochem. Biophys.*, 307 (1993) 126.
[74] J.R. Yates, III, S. Speicher, P.R. Griffin and T. Hunkapiller, *Anal. Biochem.*, 214 (1993) 397.
[75] R.A. Yost and R.K. Boyd, *Methods Enzymol.*, 193 (1990) 154.
[76] P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.*, 11 (1984) 601.
[77] A.A. Tuinman and G.R. Pettit, *Int. J. Pept. Protein. Res.*, 36 (1990) 331.
[78] R.S. Johnson, S.A. Martin and K. Biemann, *Int. J. Mass Spectrom. Ion Processes*, 86 (1988) 137.
[79] J.T. Stults and J.T. Watson, *Biomed. Environ. Mass Spectrom.*, 14 (1987) 583.
[80] R.S. Johnson, S.A. Martin, K. Biemann, J.T. Stults and J.T. Watson, *Anal. Chem.*, 59 (1987) 2621.
[81] L.C.E. Taylor and L. Poulter, in *Proceedings of the American Society for Mass Spectrometry*, American Society for Mass Spectrometry, San Francisco, 1988.
[82] H.A. Scoble, S.A. Martin and K. Biemann, *Biomed. J.*, 245 (1987) 621.
[83] S.A. Martin and K. Biemann, *Int. J. Mass Spectrom. Ion Processes*, 78 (1987) 213.
[84] R.S. Johnson and K. Biemann, *Biochemistry*, 26 (1987) 1209.
[85] T.D. Lee, *MacProMass*, Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA.

[86] J.-C. Gesquiere, *Mac Mass*, Faculté de Pharmacie, URA 1309 CNRS, Institut Pasteur, Lille.
[87] K. Biemann, *COMPOST*, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA.
[88] B. Spengler, D. Kirsch and R. Kaufmann, *J. Phys. Chem.*, 96 (1992) 9678.
[89] R. Kaufmann, B. Spengler and F. Lützenkirchen, *Rapid Commun. Mass Spectrom.*, 7 (1993) 902.
[90] B. Spengler, D. Kirsch, R. Kaufmann and E. Jaeger, *Rapid Commun. Mass Spectrom.*, 6 (1992) 105.
[91] B. Spengler, D. Kirsch and R. Kaufmann, *Rapid Commun. Mass Spectrom.*, 5 (1991) 198.
[92] J.C. Rouse, W. Yu and S.A. Martin, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 676.
[93] R.B. Zambias, D.A. Boulton and P.R. Griffin, *Tetrahedron Lett.*, 35 (1994) 4283.
[94] J.B. Hoyes, S.G. Curbishley, P. Doorbar, P. Tatterton, R.H. Bateman, B. Beer and J. Lockett, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 682.
[95] B. Spengler and R. Kaufmann, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993*, p. 774a.
[96] W. Yu, M. Flemming, I. Mazsarroff, H. Scoble and S. Martin, in *Proceedings of the 42nd ASMS Conference on Mass Spectrometry and Allied Topics, Chicago, Illinois, May 29–June 3, 1994*, p. 442.
[97] K. Breddam, *Carlsberg Res. Commun.*, 51 (1986) 83.
[98] K. Klarskov, K. Breddam and P. Roepstorff, *Anal. Biochem.*, 180 (1989) 38.
[99] R. Wang, R.J. Cotter, J.F. Meschia and S.S. Sisodia, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry III*, Academic Press, San Diego, 1992, pp. 505–513.
[100] C.E. Smith and K.L. Duffin, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry IV*, Academic Press, San Diego, 1993, pp. 463–470.
[101] M. Schär, K.O. Börnsen and E. Gassmann, *Rapid Commun. Mass Spectrom.*, 5 (1991) 319.
[102] K.J. Rosnack and J.G. Stroh, *Rapid Commun. Mass Spectrom.*, 6 (1992) 637.
[103] E.D. Lee, W. Mück, T.R. Covey and J.D. Henion, *J. Am. Chem. Soc.*, 111 (1989) 4600.
[104] B. Ganem, Y.T. Li and J.D. Henion, *J. Am. Chem. Soc.*, 113 (1991) 6294.
[105] B. Ganem, Y.T. Li and J.D. Henion, *J. Am. Chem. Soc.*, 113 (1991) 7818.
[106] V. Kata and B.T. Chait, *J. Am. Chem. Soc.*, 113 (1991) 8534.
[107] G.W. Becker, R.R. Bowsher, W.C. MacKellar, M.L. Poor, P.M. Tackitt and R.M. Riggin. *Biotechnol. Appl. Biochem.*, 9 (1987) 478.
[108] Z. Zhang and D.L. Smith, *Protein Sci.*, 2 (1993) 522.
[109] S.E. Brockerhoff, C.G. Edmonds and T.N. Davis, *Protein Sci.*, 1 (1992) 504.
[110] R. Feng and Y. Konishi, *J. Am. Soc. Mass Spectrom.*, 4 (1993) 638.
[111] M. Haniu, L.O. Narhi, T. Arakawa, S. Elliott and M.F. Rohde, *Protein Sci.*, 2 (1993) 1441.
[112] T.B. Farmer and R.M. Caprioli, *Biol. Mass Spectrom.*, 20 (1991) 796.

ELSEVIER

Journal of Chromatography A, 705 (1995) 47–65

JOURNAL OF CHROMATOGRAPHY A

Review

Chemical methods of protein sequence analysis

Jerome M. Bailey

Beckman Research Institute of the City of Hope, Division of Immunology, 1450 E. Duarte Road, Duarte, CA 91010, USA

Abstract

Chemical methods of protein sequence determination are reviewed with particular emphasis on methodology for increasing the sensitivity of amino-terminal sequence analysis and on progress toward the development of an automated procedure for sequential degradation from the carboxy-terminus.

Contents

1. Introduction

Since it is the amino acid sequence which determines the manner in which the polypeptide chain can fold to form the secondary and tertiary structures necessary for biological function, determination of the protein primary structure or sequence is a necessary first step in understanding how a particular function of a protein relates to its structure.

Currently, protein sequence determination is accomplished either by sequencing the protein on an automated sequencer using chemical methods for successive degradation from the amino-terminus or by sequencing the gene for that protein using established DNA sequencing meth-

0021-9673/95/$29.00
SSDI 0021-9673(94)01250-4

odology. Although protein sequencing can be considered to be more difficult and slower than DNA sequencing, it often provides information not obtainable by the latter method. Such information includes: (1) identification of post-translational modifications which are not predictable from the gene sequence. This has become particularly important in view of the ever increasing numbers of recognized post-translational modifications [1], (2) determination of partial protein sequence information which can be used for the design of oligonucleotide probes complementary to predicted gene sequences. In many cases, these oligonucleotide probes, often obtained from low abundance proteins, have been the only route to the cloning of a particular gene, (3) characterization of recombinant proteins in order to confirm that the predicted structure conforms to the expressed product.

These applications have created an increased need for improved methods of protein sequencing from the amino-terminus which can be performed faster and with smaller quantities of sample and for a procedure capable of a sequential degradation from the carboxy-terminus.

This manuscript reviews progress in chemical methodology relating to the development of more sensitive and faster methods of amino-terminal sequence analysis and recent methodology which shows promise toward the development of a routine method for a sequential degradation from the carboxy-terminus.

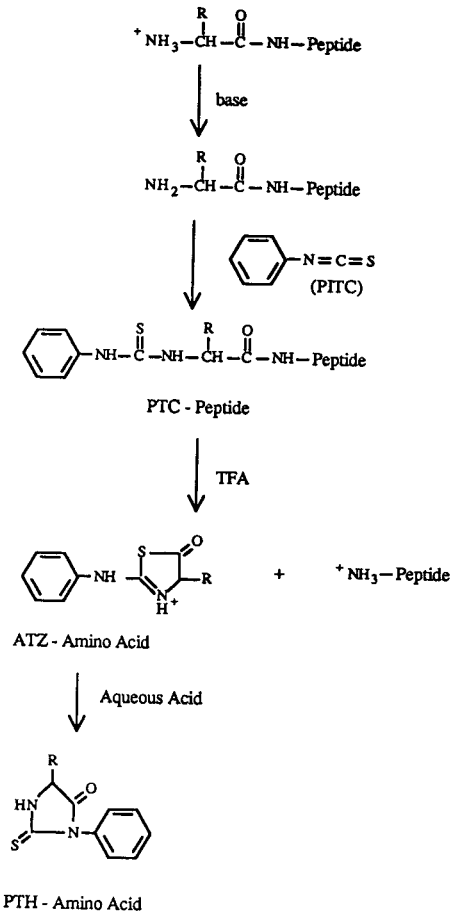


Fig. 1. Edman chemistry for amino-terminal sequence analysis.

2. Amino-terminal sequence analysis

The most widely used method of protein sequence analysis is based on the automated Edman chemistry [2]. In this method (See Fig. 1), the free amino terminus of a polypeptide is reacted with phenylisothiocyanate (PITC) in the presence of base to form a phenylthiocarbamyl peptide (PTC peptide) (this step is commonly referred to as the "coupling step"). Subsequent treatment of the modified polypeptide with acid cleaves the modified N-terminal amino acid as an anilinothiazolinone (ATZ) (this step is referred to as the "cleavage step") and leaves the shortened peptide ready for continued sequence analysis. The thiazolinone is then treated with aqueous acid in order to convert it the more thermodynamically stable phenylthiohydantoin (PTH) derivative (the "conversion step") which is then identified and quantitated by reversed-phase HPLC.

The amount of protein required for sequence analysis with current methodologies (10 to 50 pmol of sample for 10 to 20 cycles of sequence information), while steadily decreasing over the years, has not kept pace with the recent advances in protein purification techniques which

now permit isolation of 50 to 100 femtomoles of sample [3]. The major obstacles to increasing the sensitivity of protein sequencing down to the femtomolar level involve: difficulty in handling such small amounts of protein without losses; the intrinsic detectability of the released PTH amino acids which have relatively low extinction coefficients; the intrinsic background noise associated with absorbance measurements at the wavelength used for PTH amino acid detection (269 nm); the chemical background, caused by the reagents used in sequencing, visible on the HPLC chromatograms in which the PTH amino acids are detected; and the background due to PTH amino acids formed from nonspecific cleavage along the polypeptide chain. The need for a unified approach to solve these problems by redesign of the sequencing chemistry, sample supports, detection methods, and instrumentation has been discussed [4].

2.1. Instrumentation

The chemistry as originally described by Edman in 1950 is still practiced today with only minor variations in all of the commercially available protein sequenators. The performance criteria routinely expected today has been achieved primarily through advances in instrumentation. The original spinning-cup sequenator [5] was capable of sequencing 15 residues in 24 h using approximately 250 nmoles of protein. The combination of reversed-phase HPLC for detection of the PTH amino acids [6], the use of polybrene as a carrier for proteins and peptides [7,8], and modifications to the spinning cup sequenator [9–11] allowed the sensitivity to be reduced to approximately 1 nmole. The introduction of the "gas-phase" sequenator [12,13] and on-line, narrow-bore (2 mm I.D. columns) HPLC detection of the PTH amino acids, reduced the sample requirements to 10–100 picomoles [14].

Since the introduction of the gas-phase sequenator, only minor advances have been made in instrument development. The introduction of the continuous flow reactor (CFR) [15] simplified and miniaturized the reaction cartridge resulting in lower backgrounds and facilitated application of the sequencer to polyvinylidene difluoride (PVDF) membranes (commonly used for electrotransfer from protein gels). A recently described protocol for optimization of the reaction times, miniaturization of the reaction cartridge, and HPLC separation of PTH amino acids was found to permit the usual 40–50 min cycle time to be reduced to 25 min [16]. A commercial reaction cartridge (Blott cartridge, Applied Biosystems) designed specifically for samples on polymeric supports (PVDF) has been introduced and evaluated [17]. Other efforts to reduce chemical background in automated sequencers has involved the evaluation of alternative coupling bases, such as diisopropylethylamine, reduced concentrations of phenylisothiocyanate, and diode-array detection of the PTH amino acids [18]. Some more recent studies showing sequence data obtained at the 1–10 picomole level have advocated routine sequencer maintenance [19] and the use of reduced levels of polybrene and extensive precycling [20]. The successful adaptation of a microbore HPLC (Michrom BioResources) to a commercial sequencer (Applied Biosystems) showed an approximately 2–3 fold enhancement of the limit of detectability of the released PTH amino acids [21]. A compact gas-phase instrument with minimized liquid flow systems [22] was demonstrated to produce less sequencer background peaks at the 5–10 picomole level.

Recently published methods for the separation of the PTH amino acids by capillary HPLC [23] and capillary electrophoresis [24] have shown femtomolar levels of detection of the PTH amino acids, however these techniques require the use of low microliter and subnanoliter injection volumes, respectively, and are thus not of current practical value, since current automated sequencer technologies dissolve the PTH amino acids in a 50–100 μl volume for injection. Use of only a small fraction of this volume would negate any value in the increased sensitivity of detection with these methodologies until a sequencer generating the PTH amino acid in suitably low volumes is developed.

2.2. Solid supports for sequence analysis

Modern gas-phase sequenators when first described relied on the use of glass fiber filters coated with polybrene for retaining the non-covalently applied peptide or protein to be sequenced [12]. Although the proteins separated by one and two dimensional gel electrophoresis could be electroblotted onto these supports [25,26], the introduction of PVDF membranes [27] has largely superseded the use of glass fiber filters in most laboratories. The PVDF membranes were found to give superior yields for electroblotting and for sequence analysis and found to be easier to stain for the presence of protein samples [27–29]. More recent studies comparing the efficiency of sequence analysis and blotting from a series of different commercially available membranes, including glass fiber, PVDF, and polypropylene, have been described [3,30]. An alternative support for non-covalent protein or peptide immobilization consisting of a biphasic column, containing hydrophobic (reversed phase) and hydrophilic (ion-exchange) portions, has recently been introduced by Hewlett-Packard [31]. The biphasic column offers several advantages in that sample clean-up prior to sequencing is greatly facilitated and the lack of necessity for the use of polybrene to assist in sample retention greatly reduces the chemical background during detection of the PTH amino acids. Other preliminary work shows promising results with the use of Zitex (porous PTFE) supports for protein sequencing. Originally used for automated C-terminal sequencing [32], these supports have shown improved repetitive yields for N-terminal protein sequencing as compared to PVDF supports and do not require the use of polybrene for sample retention, resulting in reduced chemical backgrounds.

The non-covalent methods of sample application, while convenient, involve compromises in the choice of solvents and reagents which can be utilized. The types of solvents that may be used are limited to those that are relatively non-polar and that do not cause elution of the protein or peptide. The base and acid reagents used in the coupling and cleavage step are delivered in the gas phase in order to prevent sample elution from the support. A method, termed "solid-phase sequencing", where the peptide or protein to be sequenced was covalently coupled to a solid support prior to sequence analysis was introduced by Laursen in 1966 [33]. The potential advantages to this approach stem from the fact that proteins which are covalently linked to an insoluble matrix can be easily separated from reagent and reaction by-products. The choice of solvents for sequencing can be optimized for ability to remove excess reagents without the need to worry about sample elution. This can lead to faster instrument cycle times and reduction in the amount of UV absorbing background.

Since the introduction of the solid-phase approach to N-terminal protein sequencing, several different types of functionalized supports have been described for the covalent immobilization of polypeptide samples. These include polystyrene resins, polyacrylamide resins, and glass beads substituted with aminoalkyl or aminophenyl groups [34]. Typically these amino functionalized supports are activated for protein coupling with bifunctional reagents such as phenylene diisothiocyanate (DITC). The DITC group is capable of forming a stable thiourea linkage to the support and the peptide N-terminal amino group or ε-amino group of lysine side chains. Recently glass beads derivatized with isothiocyanate, aminophenyl and aminoethylaminopropyl groups [35], glass fiber sheets functionalized with aminophenyl groups [36], and PVDF membranes derivatized with aryl amines and DITC [37] have been used for the covalent immobilization of polypeptides for N-terminal sequencing. The polypeptides are either immobilized by coupling between the epsilon-amino groups of the lysine and the isothiocyanate groups on the solid support using the established DITC chemistry or by the coupling of the carbodiimide activated C-terminal carboxyl groups of the polypeptides and the amino groups on the matrix.

Despite the potential advantages of the solid-phase sequencing approach, the method has never gained routine acceptance. This stems in

part from the difficulties in obtaining high yield, routine methods for covalent coupling, partial blockage of the N-terminal amino group and low yields of Lys when using amino groups for covalent attachment, and low yields of Asp and Glu when using carboxyl groups for covalent attachment. The commercial availability of DITC and arylamine functionalized PVDF membranes for covalent immobilization of proteins and peptides has sparked some renewed interest in the solid-phase approach [38,39]. The solid-phase methodology has been shown to be most useful for solving specific problems not accessible to gas phase sequencers, such as the determination of phosphorylation sites [40–42] and glycosylation sites [43,44].

2.3. Chemistry

Numerous attempts have been made to increase the sensitivity of Edman degradation through the use of radiolabeled, chromophoric, or fluorescent reagents. The vast majority of reagents that have been utilized for this purpose are used in place of the standard PITC reagent and have relied on the isothiocyanate group as the electrophilic group used to mediate the coupling reaction. Isothiocyanates form a stable thiourea group with the N-terminal amino acid. The sulfur atom of a thiourea is thus perfectly placed so that upon acidification a kinetically favored five-membered ring could form which specifically cleaves only the N-terminal amino acid. The use of isothiocyanates also has the advantage in that the isothiocyanate group does not cause any complications in the sequencing reaction due to reaction with amino acid side chains.

4-(N,N'-Dimethylamino)azobenzene-4'-isothiocyanate (DABITC), a highly chromophoric reagent first described by Chang et al. [45] has the potential for improving the sensitivity of detection of the amino acid derivatives, but suffers from a lack of complete derivatization of the N-terminal amino group. It thus requires a follow-up coupling with PITC to prevent serious overlap problems. This reagent has primarily been used as a manual sequencing reagent with a DABITC–PITC double coupling procedure [46], although it has been used in automated solid-phase sequencing [47]. More recently, Aebersold and co-workers [48,49] reported a DABITC–PITC solid-phase sequencing method in which proteins were immobilized on DITC-derivatized aminopropyl glass-fiber sheets. Sequence analysis was performed at the 20–50 picomole level, a substantial improvement over previous methods, but still less sensitive than current gas-phase sequence analysis. Fluorescent reagents, such as fluorescein isothiocyanate [50,51] and dansyl-containing isothiocyanates [52–56] have also been evaluated as sensitivity enhancing reagents. Although synthetic amino acid derivatives prepared using these reagents show subpicomole sensitivity by HPLC analysis, they have not surpassed the sensitivity of gas-phase Edman degradation during automated sequence analysis. In general, it has been found that the use of large bulky chromophores on the isothiocyanate reagent interferes with the efficiency of the derivatization and cleavage reactions of the Edman degradation. An examination of several isothiocyanates [57] showed that, when an isothiocyanate reagent contains an electron withdrawing group the coupling reaction is favored and the cleavage reaction is slowed, and when the isothiocyanate contains an electron donating group the coupling reaction is slowed and the cleavage reaction is favored. It was thus concluded that PITC was the optimal choice of the reagents tested for the Edman chemistry because it offered the best balance between the rates of the coupling and cleavage reactions. Most chromophoric and fluorescent compounds are relatively large when compared to a phenyl ring and when such a compound that contains a reactive isothiocyanate group is substituted for PITC, the coupling and/or cleavage reaction is kinetically disfavored by a combination of steric and electronic effects.

The use of radiolabeled reagents is not a successful approach, since radiolabeled reagents undergo autoradiodegradation which results in decreasing product yields and increasing amounts of labeled by-products. Modified phenyl isothiocyanates such as 4-(*tert.*-butoxycar-

bonylaminomethyl)-PITC, which are designed to react with post-column fluorescent reagents, have also been investigated [58] but have been found to undergo side reactions during the cleavage reaction resulting in loss of the amino group [54].

An alternative to the use of modified Edman reagents is the reaction of the anilinothiazolinone (ATZ)-amino acid intermediate with sensitivity-enhancing nucleophilic reagents (Fig. 2). The use of radiolabeled amines produced amino acid derivatives which could be detected at the femtomole level [59,60], but the handling of radioactive materials was inconvenient. Horn et al. [61] have extended earlier studies on the use

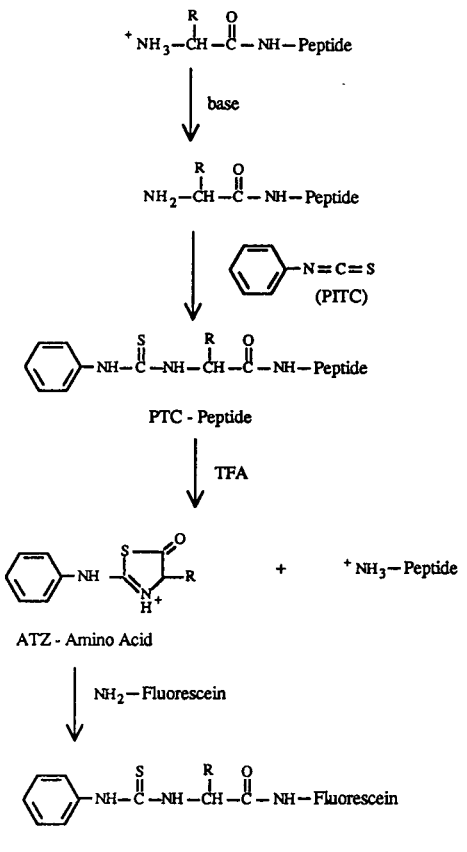


Fig. 2. Reaction of the anilinothiazolinone-amino acid intermediate with sensitivity enhancing nucleophilic reagents.

of MeOH–HCl as a conversion reagent [62] to include chromophoric or fluorophoric alcohols, resulting in the formation of phenylthiocarbamyl amino acid esters. Tsugita et al. [63] have recently reported a modification of the Edman degradation scheme, in which ATZ amino acids are reacted with 4-aminofluorescein resulting in highly fluorescent, phenylthiocarbamyl amino acid aminofluorescein amides (PTCAF-amino acids). PTCAF-amino acids were separated by reversed-phase HPLC and were detectable at the 0.1–1 femtomole level. Several known and unknown protein samples were reported to be sequenced at the 100 femtomole to 10 picomole level using an Applied Biosystems 477A sequencer. Work performed in our laboratory with this method demonstrated low yields with the hydrophilic amino acids, in particular threonine, histidine, glutamate, lysine, and glutamine, and the total lack of yield with aspartate. Recent studies concerning the aminolysis of the ATZ-amino acids by Pavlik et al. [64], showed that many of the ATZ-amino acids, in particular the hydrophilic amino acids, can rearrange so rapidly to the more thermodynamically stable PTH amino acids that by the time the ATZ-amino acid is brought over to the conversion flask of an automated instrument anywhere from 5–70% of the amino acid has already been converted. Once an ATZ-amino acid has converted to a PTH it would not be capable of reacting with aminofluorescein. A recent report describes chemistry which converts the PTH amino acids to the phenylthiocarbonyl derivatives and then back to an ATZ derivative, which in turn, is reacted with aminofluorescein [65]. This procedure was reported to solve the difficulties of derivatization with all of the amino acids, except aspartate which was found not to derivatize.

2.4. Use of mass spectrometry for detection

Recent advances in mass spectrometry have made the use of a mass spectrometer as a detector in place of chromatographic methods a viable option with potential advantages that include, speed, sensitivity, and the ability to analyze post-translationally modified amino acids

since detection is based on mass and not retention time. The use of mass spectrometry for protein sequencing has followed two general approaches.

The first approach involves the substitution of the phenylisothiocyanate reagent of the Edman chemistry with an alternative reagent to generate an amino acid with enhanced detectability by mass spectrometry since PTH amino acids are not detected with high sensitivity. A number of reagents for this purpose have been described in recent years, but to date none have been reduced to practice. The use of 4-nitrophenyl isothiocyanate as a modified Edman reagent to generate amino acid analogues detectable in the femtomolar range by negative ion chemical ionization mass spectrometry has been described [66]. Replacement of the phenylisothiocyanate reagent in the coupling reaction with 3-[4′(ethylene - N,N,N - trimethylamino)phenyl] - 2 - isothiocyanate was shown to produce thiohydantoin amino acid derivatives with quaternary amine groups with femtomole levels of detectability [67]. A disadvantage of this reagent is that covalent attachment of the sample to a solid support is required in order to prevent elution of the sample due to the polar nature of the coupling reagent. The use of dimethylaminopropyl isothiocyanate as a sequencing reagent was shown to proceed with kinetics equivalent to phenyl isothiocyanate and was found to generate amino acid derivatives with the easily ionized dimethylaminopropyl group [68]. A potential advantage of this reagent over reagents with permanently charged quaternary amine groups is that the dimethylaminopropyl group will not become charged until the cleavage step, thereby eliminating the need for covalent attachment of the polypeptide sample. Thiohydantoin amino acids containing the dimethylaminopropyl and trimethylaminopropyl groups were both synthesized and found to be detectable by electrospray mass spectrometry with equal sensitivity in the femtomolar range [68]. The synthesis of 5-acetoxy-2-phenylthiazole derivatives for all of the naturally occurring amino acids was recently described [69]. These amino acids are the products formed during the thiobenzoylation method of protein sequencing and were found to be detectable by gas chromatography–mass spectrometry at the femtomole level.

The second approach involving the use of mass spectrometry involves the generation of a defined series of peptide fragments (called "ladder sequencing") by performing the coupling step of the Edman chemistry in the presence of a small amount of phenylisocyanate as a chain termination reagent [70]. The ladder of peptides, each differing from the next by one amino acid, are then analyzed by matrix-assisted laser-desorption ionization, time of flight mass spectrometry (MALDI-TOF-MS). The difference in mass between each peptide corresponds to the amino acid removed, thus the sequence of the peptide is read by the mass difference between successive peaks. The chief advantages of the method are its speed and ability to analyze post-translationally modified amino acids. The disadvantages are that the method still requires low picomole amounts of sample, and thus is currently not any more sensitive than automated methods of sequence analysis based on chromatographic detection. The technique is limited to peptides with masses less than 3000–4000 daltons due to the increased mass errors of MALDI-TOF-MS above these masses, and the method can not distinguish the isobaric amino acids Leu and Ile.

Despite current limitations of the above described methodologies, continued work combined with predicted future advances in the sensitivity of detection by mass spectrometry, suggest that the combination of chemical methodology for sequencing with mass spectrometric methods of detection show much promise.

3. Carboxy-terminal sequence analysis

The last few years have seen a renewed interest in the development of a chemical method for the sequential C-terminal sequence analysis of proteins and peptides. Such a method would be analogous and complimentary to the Edman degradation commonly used for N-terminal sequence analysis [2]. It would also be invaluable for the sequence analysis of proteins

with naturally occurring N-terminal blocking groups, for the detection of post-translational processing at the carboxy-terminus of expressed gene products, and for assistance in the design of oligonucleotide probes for gene cloning. Although a number of methods have been described, the method known as the "thiocyanate method", first described in 1926 [71], has been the most widely studied and appears to offer the most promise due to its similarity to current methods of N-terminal sequence analysis. The field of C-terminal sequencing was extensively reviewed in 1991 [72], so this review will concentrate mainly on more recent developments in the field and in particular, the thiocyanate method.

3.1. Chemistry

The thiocyanate method involves the reaction of a protein or peptide with an isothiocyanate reagent, in the presence of a carboxylic acid activating reagent (such as acetic anhydride), to form a peptidylthiohydantoin (Fig. 3). The derivatized amino acid is then hydrolyzed to yield a shortened peptide or protein and a thiohydantoin amino acid. As in the Edman degradation the thiohydantoin amino acid is then identified

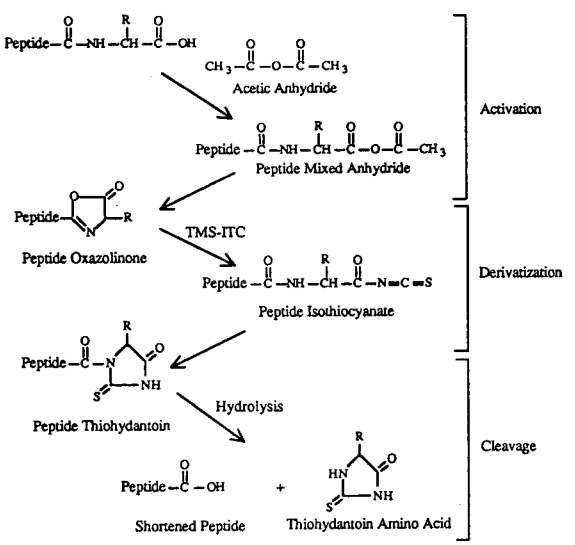


Fig. 3. Reaction scheme for the thiocyanate method of carboxy-terminal sequence analysis.

and quantitated by reversed-phase HPLC. Since the thiohydantoin amino acids produced with this methodology have UV absorption spectra and extinction coefficients similar to those of the phenylthiohydantoin amino acids formed during the Edman degradation, the sensitivity of the thiocyanate method is anticipated to be similar to that of current N-terminal methods (10–200 pmol of applied sample). Traditionally, the sequence of steps necessary for one complete cycle of the thiocyanate chemistry have been referred to as the activation, derivatization, and cleavage steps. The activation step refers to the treatment of the C-terminal carboxylic acid with a reagent that activates it for reaction with a nucleophilic reagent by converting it to an electrophilic group, such as a mixed anhydride or oxazolinone (Fig. 3). The derivatization step refers to the reaction of the activated C-terminal carboxylic acid with an isothiocyanate reagent to derivatize it to a thiohydantoin and the cleavage step refers to the specific hydrolysis of the peptidylthiohydantoin to form a shortened peptide or protein and a thiohydantoin amino acid.

A number of problems have prevented this methodology from becoming a routine method in the protein chemistry laboratory. These include the severity of the conditions used for the derivatization reaction, blockage of the shortened peptide to continued sequencing by the reagents used for the cleavage reaction, the inability to sequence through some of the amino acids commonly found in proteins (especially proline), long reaction times, large amounts of required sample, and the need to covalently couple the sample to be sequenced to a solid support.

3.1.1. The derivatization reaction

Although the derivatization of amino acids to their corresponding thiohydantoins has been studied since this reaction was first reported in 1911 [73], the mechanism of peptidylthiohydantoin formation by acetic anhydride and thiocyanate ions is still not well understood. Early experiments confirmed the assumption that thiocyanic acid is first formed when acetic anhydride and acetic acid interact with ammonium thiocyanate and it is the thiocyanic acid which

actually reacts to form the thiohydantoin [74]. The differing ability of the various salts of thiocyanic acid to form a thiohydantoin was reasoned to result from their ability to form thiocyanic acid on interaction with acetic anhydride and acetic acid [75]. When more convenient methods for the preparation of thiocyanic acid became available, thiocyanic acid, in the presence of acetic anhydride, was found to be more reactive for the formation of 2-thiohydantoins than were the thiocyanate salts. As a result, thiocyanic acid has been used by Kubo et al. [76] and more recently by Inglis et al. [77] for the sequential degradation of peptides from the C-terminus. However, one of the principal drawbacks of thiocyanic acid is that it tends to be self reactive, even at ambient temperature, and quickly loses its ability to derivatize the peptide. Furthermore, these polymeric thiocyanic acid products are intensely UV absorbing at the wavelengths used for thiohydantoin detection and subsequently interfere with the HPLC identification of the released thiohydantoin amino acid. The instability of the free thiocyanic acid presents difficulties when the chemistry is automated, since the reagents needs to be stable to storage in a reagent bottle at room temperature. As discussed by Inglis et al. [78], one way to help stabilize the thiocyanic acid is to refrigerate it while in the automated instrument. A less costly and more convenient solution to the problems caused by the use of thiocyanic acid was presented with the introduction of trimethylsilylisothiocyanate (TMS-ITC) (Fig. 3) for derivatization of the C-terminal amino acid to a thiohydantoin [79]. The trimethylsilyl group offered two advantages, (1) it stabilized the thiocyanate sufficiently so that self reaction was no longer a problem, and (2) it did not compromise the ability of the thiocyanate to form thiohydantoins. This is consistent with the observation that silylated amines have often been found to be better nucleophiles than the corresponding unsubstituted amines [80].

The use of guanidine thiocyanate [81], benzoyl isothiocyanate [82] and tributyltin isothiocyanate [83] have also been described for the formation of peptidylthiohydantoins. Both guanidine thiocyanate and tributyltin isothiocyanate require the use of a chloride containing activation reagent, such as acetyl chloride or FMOC-chloride. The chloride ion is required for release of the thiocyanate ion from the guanidine or tributyltin molecule. Once released the thiocyanate ion is then free to react with the peptide oxazolinone to form a peptidylthiohydantoin. Benzoyl isothiocyanate (BITC) offers the potential advantage in that it does not require a separate activating reagent. Reaction of BITC with a C-terminal carboxylate was found to form a benzoic acid mixed anhydride which then required the addition of pyridine (or other catalyst such as triazine, imidazole or tetrazole) for formation of an oxazolinone. Once formed the oxazolinone can then react with the isothiocyanate ion, liberated when BITC reacted with the C-terminal carboxylate, to form a peptidylthiohydantoin. Although, all three of these reagents can effectively form peptidylthiohydantoins, they have not been found to offer any significant advantages over the acetic anhydride/TMS-ITC chemistry in terms of reaction times, yields, and ability to derivatize all twenty common occurring amino acids.

The intermediate involved in thiohydantoin formation has been a subject of study for many years. An oxazolinone was postulated to be a necessary intermediate during the synthesis of amino acid thiohydantoins with acetic anhydride and ammonium thiocyanate when this reaction was first studied [74]. The racemization of the C-terminal amino acids observed on reaction with acetic anhydride and TMS-ITC [84] suggests that reaction of peptides with acetic anhydride forms a peptide oxazolinone (Fig. 3). This is consistent with the above postulated mechanism. The formation of oxazolinones is known to cause racemization of amino acids [85–87]. Further evidence of an oxazolinone intermediate during the formation of amino acid thiohydantoins was obtained by Csonka and Nicolet [85]. Additional studies described by Cornforth [88] actually demonstrated the formation of an oxazolinone intermediate in the formation of thiohydantoins by the combined use of absorption spectra and polarimetry to follow the rate of oxazolinone

formation. In fact, once the oxazolinone was formed, the reaction with isothiocyanic acid was found to be facile enough to occur readily at 0°C in the case of 2-phenyl-4-benzyl-5-oxazolinone [88].

A number of reagents for activation of the carboxylic acid other than acetic anhydride have been described. The use of acetic anhydride was shown to cause problems with the sequence analysis of certain amino acids [84] and found to contribute to difficulties in the sequence analysis of protein samples [89]. The use of 9-fluorenylmethyl chloroformate [89], 2-ethyl-5′-phenylisoxazoliumsulfonate [90], 1,1,3,3-tetramethyl-chlorouronium chloride [91], and 2-halo-1-methylpyridinium salts [92] have all been described.

An alternative approach for the formation of peptidylthiohydantoins involved the use of diphenyl phosphoroisothiocyanatidate (DPP-ITC) and pyridine [93]. This chemistry combined the activation and derivatization steps, eliminating the need for a separate activation step. Although the use of DPP-ITC for C-terminal sequencing was first described by Kenner et al. in 1953 [94], it was never actively pursued, most likely since it took several days for the reaction to go to completion. These authors proposed a reaction mechanism whereby isothiocyanate ion was liberated by exchange with the C-terminal carboxylate forming an acyl phosphate. The acyl phosphate was in turn postulated to be attacked by the isothiocyanate ion to form the desired acyl isothiocyanate and diphenyl phosphate ion. Work in our laboratory suggested an alternative mechanism involving first attack by the C-terminal carboxylate (formed by base treatment of the polypeptide) on the phosphate of DPP-ITC to form a pentacovalent acylphosphorylisothiocyanate. Introduction of pyridine, imidazole, triazine, or tetrazole was then found to promote the rapid rearrangement of the pentacovalent acylphosphorylisothiocyanate to the acylisothiocyanate, with concomitant release of diphenylphosphate, which then rapidly cyclized to a thiohydantoin. The entire reaction was found to be complete in less than 10 min at 50°C [93]. Pyridine has recently been shown to promote a similar type of reaction by removing carbon dioxide from the mixed anhydride formed on reaction of an amino acid carboxylate with ethyl chloroformate to form an amino acid ester [95]. The success of this reaction was dependent on pyridine being present in a molar excess with respect to the ethyl chloroformate. A similar mechanism, as that proposed by us for the reaction of DPP-ITC with carboxylates to form an acylisothiocyanate, was proposed for the reaction of diphenylphosphoryl azide with peptide carboxylates to form an acyl azide [96]. These authors postulated that the diphenylphosphoryl azide reacts with the C-terminal carboxylate to form a pentacovalent phosphorus compound (which was noted to be resistant to racemization) and rearranged via an intramolecular mechanism to form an acyl azide. The use of the DPP-ITC/pyridine reaction, by elimination of the oxazolinone intermediate, was found to result in faster cycle times and improved the ability to sequence previously difficult amino acids, in particular, aspartate, glutamate, serine, threonine, and proline (see below).

3.1.2. The cleavage reaction

The cleavage reaction has been extensively studied since the thiocyanate chemistry for C-terminal degradation was first proposed by Schlack and Kumpf in 1926 [71]. In their procedure, strong base (1 *M* sodium hydroxide) was used to liberate the amino acid thiohydantoin and generate a new carboxyl-terminal amino acid. Since this work was published, numerous groups have tried to reduce the severity of the conditions required in order to apply this chemistry to the sequential degradation of proteins from the carboxyl terminal end. Lesser concentrations of sodium hydroxide [97,98] than originally used by Schlack and Kumpf [71] and of barium hydroxide [99] were found to effectively cleave peptidylthiohydantoins. Other groups [100,101] used acidic conditions based on the original procedure used by Johnson and Nicolet [73] for the de-acetylation of amino acid thiohydantoins. These authors added concentrated hydrochloric acid to the coupling solution to cause cleavage of the peptidylthiohydantoin

bond. Unlike hydroxide which was shown to cause breakdown of the thiohydantoin amino acids [102], hydrochloric acid was shown not to destroy the amino acid thiohydantoins [103,104]. Cromwell and Stark [105] showed that concentrated hydrochloric acid could be used to cleave the thiohydantoin amino acid at room temperature. The major drawback with this procedure was that when applied to proteins no more than two or three cycles could be performed. Yamashita [106] found that cleavage of petidylthiohydantoins could be done in a repetitive manner with a protonated cation-exchange resin. Stark [104] reported that certain organic bases such as morpholine or piperidine could be substituted for sodium hydroxide, and along the same lines, Kubo et al. [76] reported that aqueous triethylamine (0.5 *M*) could be used to effectively cleave peptidylthiohydantoins. Stark [104] appeared to have solved the cleavage problem by introducing acetohydroxamic acid in aqueous pyridine at pH 8.2 as a cleavage reagent. This reagent was shown to rapidly and specifically cleave peptidylthiohydantoins at room temperature and at mild pH. However, a more recent study found that, although acetohydroxamate is an excellent cleavage reagent for the first amino acid, it forms a stable peptidyl hydroxamate ester which is difficult to hydrolyze, and which can partially or completely block (depending on the conditions employed) the shortened peptide from further sequencing [84]. Primary amines, such as N-butylamine in trifluoroethanol, were also found to be excellent cleavage reagents by Inglis et al. [77], but were recently shown by Hawke and Boyd [107] and Inglis et al. [78] to permit only one cycle of sequencing since they form a stable amide at the C-terminus of the shortened peptide, effectively blocking the shortened peptide to further sequencing.

In another study by Inglis et al. [78], an aqueous solution of potassium hydroxide containing 33% methanol and dithioerythritol (DTE) or dithiothreitol (DTT) was employed for the cleavage reaction. Although the presence of DTE or DTT has been shown to protect the released thiohydantoin amino acid from degradation under these basic conditions [77,108], the mercapto group of these molecules has also been shown to form an adduct at the C-terminus of the shortened peptide resulting in the formation of a percentage of shortened peptide blocked to further degradation at each cycle [108]. This may partially explain the decrease in the repetitive yield with each cycle observed by Inglis et al. [77].

A dilute solution of aqueous triethylamine was demonstrated to cause rapid and quantitative cleavage of a wide variety of peptidylthiohydantoins in the solution phase [84], but when applied to automated sequencing on the solid phase, several problems became apparent which resulted in poor repetitive yields and increased background [109,110].

A new cleavage reagent, sodium trimethylsilanolate, was introduced and demonstrated to cause rapid and quantitative cleavage of peptidylthiohydantoins [110]. This reagent was found not to suffer from the limitations associated with previous reagents, such as: the generation of UV absorbing background peaks which interfere with identification of the released thiohydantoin amino acids, partial or complete blockage of the shortened peptide to continued sequencing, and destruction of the released thiohydantoin amino acid. Use of sodium trimethylsilanolate in alcoholic solvents and under an inert atmosphere, such as argon, has permitted the extended sequencing of peptides and proteins, covalently coupled to polyethylene supports and non-covalently applied to Zitex strips, respectively [32,93].

Another approach described involves the treatment of the peptidylthiohydantoin with an alkylating agent in the presence of base, followed by cleavage with trimethylsilyl isothiocyanate in the presence of trifluoroacetic acid [91]. There are two potential advantages of this approach. The first is that the need to reform the C-terminal carboxylate on the shortened peptide is obviated since the shortened peptide is derivatized to a peptidylthiohydantoin simultaneous with cleavage. The second is that the alkylation step could be used to introduce a highly chromophoric or fluorescent group that

would permit a more sensitive detection of the thiohydantoin amino acids. However, in practice, this approach still suffers from several disadvantages. The initial derivatization to a peptidylthiohydantoin in cycle 1 still requires the use of a separate activation and isothiocyanate reagent limiting the number of amino acids which can be successfully derivatized to a thiohydantoin; several of the amino acid side chains have been found to interfere with the alkylation reaction thereby permitting successful sequencing of only 13 of the 20 naturally occurring amino acids with this approach [91]; the method requires covalent attachment of protein samples to a solid support in order to prevent loss of the sample; and the alkylated thiohydantoin amino acids are unstable and must be analyzed immediately after formation. Despite these limitations, the method still shows promise, and with continued work and incorporation of advances from other laboratories, it is likely that the limitations associated with this approach can be solved.

3.2. Solid supports for C-terminal sequence analysis

Most of the studies involving C-terminal sequence analysis have employed samples which are covalently coupled to a solid support. The use of glass beads for the covalent immobilization of peptide samples [77,79,89,111–113] suffers from chemical instability especially during the basic conditions used for the cleavage step in the thiocyanate chemistry. The use of carboxylic acid modified PVDF [110], DITC-activated amino PVDF [78,108], a disuccinimidoyl carbonate polyamide resin [107] and aminomethylpolystyrene beads [91] have also been considered as alternatives to silica supports. In our own work, both the glass and PVDF-based supports were found to be unstable in the presence of the cleavage reagent, sodium trimethylsilanolate [83]. A more suitable support was found to be a carboxylic acid modified polyethylene film [83].

As discussed above for N-terminal sequence analysis, there are several problems with the use of covalent methods for C-terminal sequence analysis. These include the lack of a routine method for high yield covalent coupling. This stems from the fact that each protein or peptide sample often has quite unique properties and it is difficult to develop a procedure which works well for all of the samples. Additionally, low yields or no yields of lysine often result when covalently coupling to peptide amino groups [93]. Recent work demonstrating the C-terminal sequence analysis of proteins non-covalently applied to Zitex supports has demonstrated that the use these supports can permit the routine C-terminal sequence analysis of polypeptides without the need for covalent coupling, simplifying the procedure and eliminating the loss of sample all too often associated with covalent coupling procedures [32,93].

3.3. Proline

The derivatization of C-terminal proline to a thiohydantoin has been a major impediment to the development of a routine method of C-terminal sequence analysis of proteins and peptides. Since the method was first described in 1926 [71], the derivatization of C-terminal proline has been problematic. While over the years a few investigators have reported the derivatization of proline, either with the free amino acid or on a peptide, to a thiohydantoin [76,77,114], others have been unable to obtain any experimental evidence for the formation of a thiohydantoin derivative of proline [84,99,103,104,115]. As pointed out by Stark [104], cyclization to form a peptidylthiohydantoin with proline would require the quaternization of the imino nitrogen, thereby potentially resulting in the simultaneous cyclization and cleavage of C-terminal proline. This would lead to a gap at proline, since the method would continue on to the next residue. Recently, Inglis and De Luca [116] described the successful synthesis of thiohydantoin proline from N-acetylproline. This was done by the one-pot reaction of acetic anhydride, acetic acid, trifluoroacetic acid, and ammonium thiocyanate

with N-acetyl proline. Experiments performed in our laboratory reproduced this work and developed a scaled up procedure which permitted several hundred milligrams of thiohydantoin proline to be produced [117]. However, application of this procedure to a tripeptide containing a C-terminal proline by Inglis and De Luca [116] and similar attempts in our laboratory showed that this method did not work well for peptide samples. Application of the procedure to the tripeptide, N-acetyl-Ala-Phe-Pro, in our laboratory, found that thiohydantoin proline was formed in low yield (approx. 1–2% of theoretical). Recovery of the peptide products after the reaction revealed that approximately half of the starting peptide was unchanged and the remaining half had been decarboxylated at the C-terminus, thereby blocking it to C-terminal sequence analysis. This was most likely caused by the high concentration of trifluoroacetic acid, the excess of acetic anhydride present, and the high temperature (80°C) at which the the reaction was performed.

The poor reaction with C-terminal proline most likely stems from the fact that proline cannot form the necessary oxazolinone for efficient reaction with the isothiocyanate. Work in our laboratory has solved this problem by the use of diphenyl phosphoroisothiocyanatidate and pyridine. Reaction of this reagent with C-terminal proline directly forms the acylisothiocyanate without the need for oxazolinone formation. Once the acylisothiocyanate is formed, the addition of either liquid or gas phase acid followed by water releases the proline as a thiohydantoin amino acid derivative [117]. Unlike thiohydantoin formation with the other 19 naturally occurring amino acids, C-terminal proline thiohydantoin requires the addition of acid to provide a hydrogen ion for protonation of the thiohydantoin ring nitrogen. This step is necessary for stabilization of the proline thiohydantoin ring. The resulting quaternary amine containing thiohydantoin can then be readily hydrolyzed to a shortened peptide and thiohydantoin proline by introduction of water vapor or by the addition of sodium trimethylsilanolate (the reagent normally used for cleavage of peptidylthiohydantoins). The automation of this chemistry has allowed proline to be analyzed in a sequential fashion without affecting the chemical degradation of the other amino acids [117].

3.4. Instrumentation

The instrumentation used for automated C-terminal sequencing has involved only minor modifications to existing commercial protein sequenators. Instruments which have been utilized for automated C-terminal sequence analysis include: modules of the Knauer sequencer [81], the commercially available Applied Biosystems Model 477A [91], and the commercially available Hewlett-Packard Protein Sequencing system [117]. A recent manuscript describes the construction of a compact sequencer designed specifically for automated C-terminal sequence analysis [32]. In all of these instruments the "conversion flask" is used as a holding vessel for the thiohydantoin amino acid. After the cleavage reaction the thiohydantoin amino acid is transferred to the conversion flask, where it is then dried and then re-dissolved in a solvent suitable for on-line HPLC injection.

3.5. Examples of automated C-terminal sequence analysis

The chemistry and instrumentation developed in our laboratory now permits the derivatization and identification of all of the twenty naturally occurring amino acids. Derivatization of the C-terminal amino acid to a thiohydantoin is accomplished with diphenyl phophoroisothiocyanatidate (liquid phase) and pyridine (gas phase). The peptide is then extensively washed with ethylacetate and acetonitrile to remove reaction by-products. The peptide is then treated briefly with gas-phase trifluoroacetic acid, followed by water vapor in case the C-terminal residue is a proline (this treatment has no effect on residues which are not proline). The derivatized amino acid is then specifically cleaved with sodium tri-

methylsilanolate to generate a shortened peptide or protein which is ready for continued sequencing. The thiohydantoin amino acid derivative is then quantitated and identified by reversed-phase HPLC. Peptides are covalently attached to carboxylic acid-modified polyethylene film prior to sequencing and proteins are non-covalently applied to Zitex strips. Automated sequencing was performed on a compact sequencer constructed in our laboratory [32].

Fig. 4 shows the reversed-phase separation of the thiohydantoin amino acid standards (400 pmol) [117]. The extinction coefficients of the thiohydantoins [approx. 17 500 [105]) are similar to the phenylthiohydantoins and therefore would be expected to have similar limits of detectability. The major difficulty with the HPLC separation of the thiohydantoin amino acids has been the greater hydrophilicity of the thiohydantoins as compared to the phenylthiohydantoins. Most commercially available C_{18} reversed-phase HPLC columns are not ideal for the separation of such polar molecules. This is reflected most with the thiohydantoin derivatives of the polar amino acids such as Asp, Asn, Gly, and Gln. These derivatives are poorly retained, even in the absence of organic solvents, and consequently these residues are poorly resolved and often elute as broad peaks which are difficult to identify and quantitate, especially under automated sequencing conditions. As with PTH-Ile, the TH-Ile analogue elutes as a doublet due to the formation of the allo-isomer during synthesis. This results in the formation of two diastereomeric derivatives of TH-Ile which are resolved during the separation. Although Thr and Ser sequence in good yield, the nature of the derivatives formed is not well understood. The threonine derivative obtained does not co-elute with a dehydrothreonine standard (labelled as T in Fig. 4), but rather elutes 2 min earlier. It has equivalent UV absorbance at 265 and 319 nm, suggesting that it may not be a simple thiohydantoin threonine. The same UV absorbance properties were observed with the serine analogue. The elution position of the Thr and Ser analogues is shown with arrows in Fig. 4. C-Terminal cysteine is also readily sequenced with this methodology, forming a derivative indistinguishable from that formed with serine. The lysine standard shown in Fig. 4 is acetylated at the ε amino group during the synthesis of the analogue. However, during actual automated sequencing, the epsilon acetyl analogue is not formed unless the protein sample is treated with acetic anhydride prior to sequencing. In the absence of any pretreatment lysine forms a thiohydantoin derivative which co-elutes with TH-Phe (see Fig. 5, cycle 3).

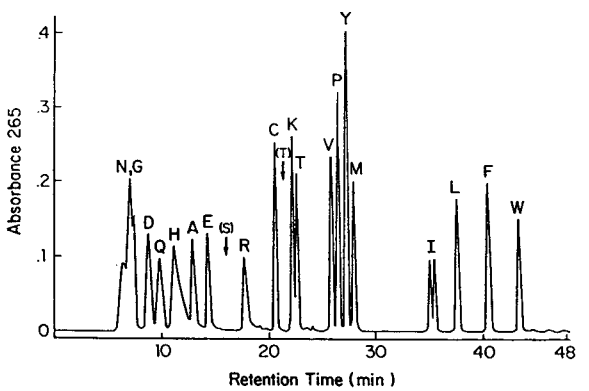


Fig. 4. High-performance liquid chromatographic separation of the amino acid thiohydantoin standards. The amino acid thiohydantoins standards (400 pmol) were separated on a C_{18} Reliasil column [117]. Absorbance was monitored at 265 nm at 0.4 AUFS. The thiohydantoin derivative of cysteine is S-methylated and the side ε-amino group of the lysine derivative is acetylated. The elution positions of the Thr and Ser analogues formed during sequencing are indicated with arrows.

Figs. 5 and 6 show three cycles of automated sequencing of hemoglobin α-chain (4.1 nmol) and bovine serum albumin (700 pmol), respectively. Both proteins were non-covalently applied to Zitex strips. Fig. 7 shows three cycles of sequence analysis of the tripeptide, LAP (15 nmol), covalently coupled to carboxylic acid-modified polyethylene. The low yields of leucine in cycle 3 are due to covalent attachment of this amino acid to the solid support [93]. This methodology currently is capable of providing 2–3 cycles of C-terminal sequence information with

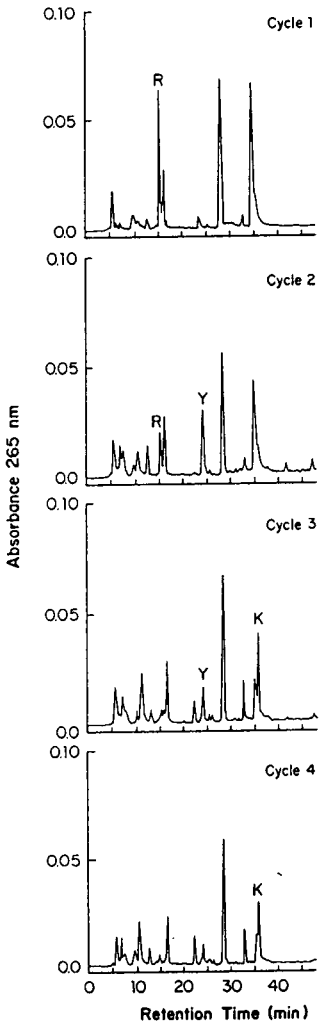


Fig. 5. Automated C-terminal sequencing of hemoglobin α-chain (4.1 nmol) non-covalently applied to Zitex. The sequence of the C-terminus is –Lys-Tyr-Arg.

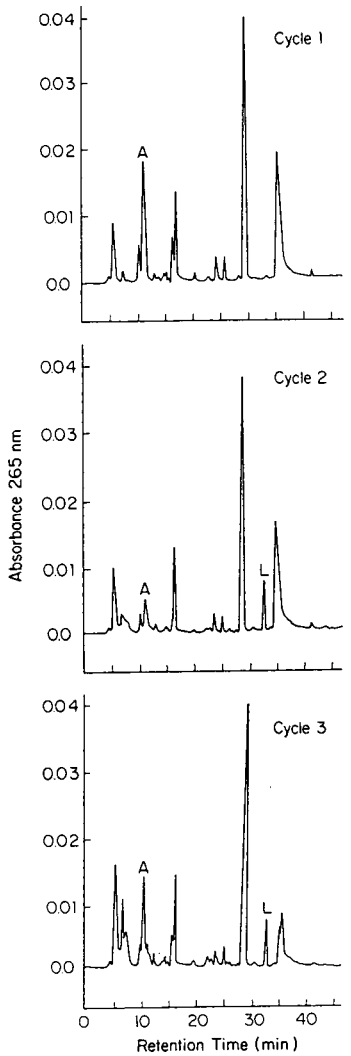


Fig. 6. Automated C-terminal sequencing of bovine serum albumin (700 pmol) non-covalently applied to Zitex. The sequence of the C-terminus is –Ala-Leu-Ala.

noncovalently applied proteins and 5–6 cycles on most covalently attached peptides.

4. Conclusions

There are a number of promising chemical approaches to improving the sensitivity of sequence analysis from the amino-terminus. The recent advances in the use of fluorescent detection and the use of mass spectrometry for detection of the released amino acid derivatives both show promise and suggest that a more sensitive and faster method for amino-terminal sequencing will be possible in the not too distant future.

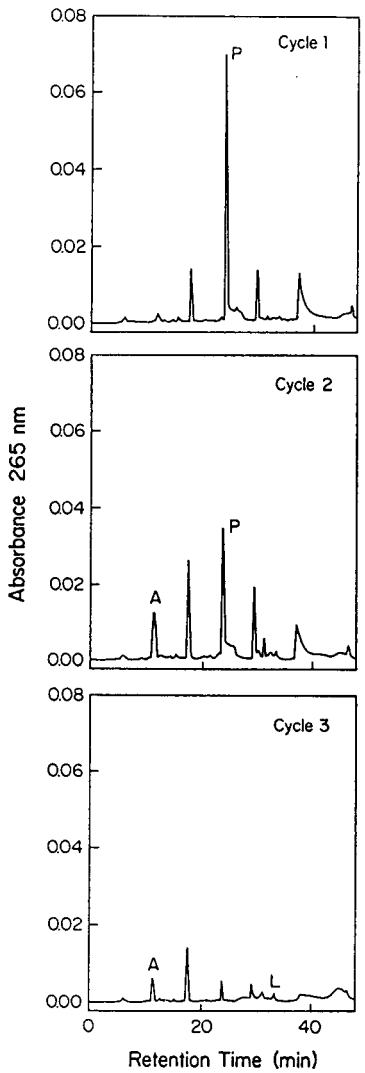


Fig. 7. Automated C-terminal sequencing of the tripeptide, LAP (15 nmol) covalently coupled to carboxylic acid modified polyethylene.

Advances in the development of chemistry for automated C-terminal sequencing now permit for the first time the ability to sequentially degrade all twenty of the common amino acids, an essential first step in the development of a routine automated procedure. The ability to sequence proteins which can be non-covalently applied to solid supports and the adaptability of the chemistry to a wide range of commercially available protein sequenators is anticipated to speed the development of this long awaited procedure.

Acknowledgments

Appreciation is expressed to Dr. John Shively for helpful discussions and to Oanh Tu and Gilbert Issai for technical assistance. This work was supported by NIH grant GM46022.

References

[1] R.G. Krishna, and F. Wold, in K. Imahori and F. Sakiyama (Editors), *Methods in Protein Sequence Analysis*, Plenum, 1993, pp. 167–172.
[2] P. Edman, *Acta Chem. Scand.*, 4 (1950) 283–293.
[3] R.J. Simpson, R.L. Moritz, G.S. Begg, M.R. Rubira and E.C. Nice, *Anal. Biochem.*, 177 (1989) 221–236.
[4] S. Kent, R. Hood, R. Abersold, D. Teplow, L. Smith, V. Farnsworth, P. Cartier, W. Hines, P. Hughes and C. Dodd, *Biotechniques*, 5 (1987) 314–321.
[5] P. Edman and G. Begg, *Eur. J. Biochem.*, 1 (1967) 80–91.
[6] C.L. Zimmermann, E. Appella and J.J. Pisano, *Anal. Biochem.* 77 (1977) 569–573.
[7] G.E. Tarr, J.F. Beecher, M. Bell and D.J. McKean, *Anal. Biochem.*, 84 (1978) 622–627.
[8] D.G. Klapper, C.E. Wilde and J.D. Capra, *Anal. Biochem.*, 85 (1978) 126–131.
[9] B. Wittmann-Liebold, *Hoppe-Seyler's J. Physiol. Chem.*, 354 (1973) 1415–1431.
[10] M.W. Hunkapiller and L.E. Hood, *Biochemistry*, 17 (1978) 2124–2133.
[11] J.E. Shively, *Methods Enzymol.*, 79 (1981) 31–48.
[12] R.M. Hewick, M.W. Hunkapiller, L.E. Hood and W.J. Dreyer, *J. Biol. Chem.*, 256 (1981) 7990–7997.
[13] D.H. Hawke, D.C. Harris and J.E. Shively, *Anal. Biochem.*, 147 (1985) 315–330.
[14] R.J. Simpson, J.A. Smith, R.L. Moritz, M.J. O'Hare, P.S. Rudland, J.R. Morrison, C.J. Lloyd, B. Grego, A.W. Burgess and E.C. Nice, *Eur. J. Biochem.*, 153 (1985) 629–637.
[15] J.E. Shively, P. Miller and M. Ronk, *Anal. Biochem.*, 163 (1987) 517–529.
[16] N.F. Totty, M.D. Waterfield and J.J. Hsuan, *Protein Science*, 1 (1992) 1215–1224.
[17] D.F. Reim, P. Hembach and D.W. Speicher, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: III*, Academic Press, 1992, pp. 53–60.
[18] V. Farnsworth, W. Carson and H. Krutzsch, *Peptide Research*, 4 (1991) 245–251.

[19] D. Atherton, J. Fernandez, M. DeMott, L. Andrews and S.M. Mische, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: IV*, Academic Press, 1993, pp. 409–418.
[20] H. Erdjument-Bromage, S. Geromanos, A. Chodera and P. Tempst, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: IV*, Academic Press, 1993, pp. 419–426.
[21] R.W. Blacher and J.H., Wieser, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: IV*, Academic Press, 1993, pp. 427–433.
[22] J. Calaycay, M. Rusnak and J.E. Shively, *Anal. Biochem.*, 192 (1991) 23–31.
[23] R.L. Moritz and R.J. Simpson, in K. Imahori and F. Sakiyama (Editors), *Methods in Protein Sequence Analysis*, Plenum, 1993, pp. 3–10.
[24] K.C. Waldron and N.J. Dovichi, *Anal. Chem.*, 64 (1992) 1396–1399.
[25] R.H. Aebersold, D.B. Teplow, L.E. Hood and S.B.H. Kent, *J. Biol. Chem.*, 261 (1986) 4229–4238.
[26] J. Vandekerckhove, G. Bauw, M. Puype, J. Van Damme and M. Van Montagu, *Eur. J. Biochem.*, 152 (1985) 9–19.
[27] P. Matsudaira, *J. Biol. Chem.*, 261 (1987) 10035–10038.
[28] Q. Xu and J.E. Shively, *Anal. Biochem.*, 170 (1988) 19–30.
[29] M. Moos, N.Y. Nguyen and T. Liu, *J. Biol. Chem.*, 263 (1988) 6005–6008.
[30] C. Eckerskorn and F. Lottspeich, *Electrophoresis*, 14 (1993) 831–838.
[31] K. Granlund-Moyer, *Am. Biotech. Lab.*, (1992) 34.
[32] J.M. Bailey, M. Rusnak and J.E. Shively, *Analytical Biochemistry*, 212 (1993) 366–374.
[33] R.A. Laursen, *J. Am. Chem. Soc.*, 88 (1966) 5344–5346.
[34] R.A. Laursen and W. Machleidt, *Methods Biochem. Anal.*, 26 (1980) 201–284.
[35] Liang Song-Ping and R.A. Laursen, *Anal. Biochem.*, 188 (1990) 366–373.
[36] R.H. Aebersold, D.G. Pipes, R.E.H. Wettenhall, H. Nika and L.E. Hood, *Anal. Biochem.*, 187 (1990) 56–65.
[37] D.J.C. Pappin, J.M. Coull and H. Koester, in J.J. Villafranca (Editor), *Current Research in Protein Chemistry*, Academic Press, 1990, pp. 191–202.
[38] J.M. Coull, D.J.C. Pappin, J. Mark, R. Aebersold and H. Koster, *Anal. Biochem.*, 194 (1991) 110–120.
[39] A. Admon and D.S. King, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: III*, Academic Press, 1992, pp. 45–51.
[40] R. Aebersold, J.D. Watts, H.D. Morrison and E.J. Bures, *Anal. Biochem.*, 199 (1991) 51–60.
[41] R.E.H. Wettenhall, R.H. Aebersold and L.E. Hood, *Method Enzymol.*, 201 (1991) 186–199.
[42] S. Sullivan and T.W. Wong, *Anal. Biochem.*, 197 (1991) 65–68.
[43] J.L. Abernethy, Y. Wang, A.E. Eckhardt and R.L. Hill, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: III*, Academic Press, 1992, pp. 277–286.
[44] P.J. Neane and F.P. Barry, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry: IV*, Academic Press, 1993, pp. 153–159.
[45] J.Y. Chang, E.H. Creaser and K.W. Bentley, *Biochem. J.*, 153 (1976) 607–611.
[46] J.Y. Chang, D. Brauer and B. Wittmann-Liebold, *FEBS Lett.*, 93 (1978) 205–214.
[47] J. Salnikow, A. Lehmann and B. Wittmann-Liebold, in M. Elzinga (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1982, pp. 181–188.
[48] R.H. Aebersold, G.D. Pipes, H. Nika, L.E. Hood and S.B. Kent, *Biochemistry*, 27 (1988) 6860–6867.
[49] R.H. Aebersold, H. Nika, G.D. Pipes, R.E.H. Wettenhall, S.M. Clark, L.E. Hood and S.B.H. Kent, in B. Wittmann-Liebold (Editor), *Methods in Protein Sequence Analysis*, Springer-Verlag, Berlin, 1989, pp. 79–97.
[50] H. Maeda and H. Kawauchi, *Biochem. Biophys. Res. Commun.*, 31 (1968) 188–192.
[51] K. Muramoto, H. Kamiya and H. Kawauchi, *Anal. Biochem.*, 141 (1984) 446–450.
[52] H. Hirano and B. Wittmann-Liebold, *Biol. Chem. Hoppe-Seyler*, 367 (1986) 1259–1265.
[53] H. Hirano and B. Wittmann-Liebold, in B. Wittmann-Liebold (Editor), *Methods in Protein Sequence Analysis*, Springer-Verlag, Berlin, 1989, pp. 42–51.
[54] S.W. Jin, G.X. Chen, Z. Palacz and B. Wittmann-Liebold, *FEBS Lett.*, 198 (1986) 150–154.
[55] S.W. Jin, S.Z. Xu, S.L. Zhang and T.B. Tang, in: B. Wittmann-Liebold (Editor), *Methods in Protein Sequence Analysis*, Springer-Verlag, Berling, 1989, pp. 34–41.
[56] J. Salnikow, Z. Palacz and B. Wittmann-Liebold, in K.A. Walsh (Editor), *Methods in Protein Sequence Analysis – 1986*, Humana Press, Clifton, NJ, 1987, pp. 247–260.
[57] G.E. Tarr, in: J.E. Shively (Editor), *Methods of Protein Microcharacterization*, Humana Press, Clifton, NJ, 1986, pp. 155–194.
[58] J.J. L'Italien and S.B.H. Kent, *J. Chromatogr.*, 283 (1984) 149–156.
[59] J.K. Inman and E. Appella, *Methods Enzymol.*, 47 (1977) 374–385.
[60] A. Tsugita, I. Arai, M. Kamo and C.S. Jone, *J. Biochem.*, 103 (1988) 399–401.
[61] M.J. Horn, S.L. Early and S.G. Magil, in T.E. Hugli (Editor), *Techniques in Protein Chemistry*, Academic Press, San Diego, CA, 1989, 51–58.
[62] M.N. Margolies, A. Brauer, C. Oman, D.G. Klapper and M.J. Horn, in M. Elzinga (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1989, pp. 189–203.
[63] A. Tsugita, M. Kamo, C.S. Jone and N. Shikama, *J. Biochem.*, 106 (1989) 60–65.

[64] M. Pavlik, Z. Voburka, I. Kluh and V. Kostka, *Anal. Biochem.*, 201 (1992) 9–16.
[65] V. Farnsworth and K. Steinberg, *Anal. Biochem.*, 215 (1993) 190–199.
[66] S. Waidyanatha and R.J. Anderegg, *39th ASMS Conference Proceedings*, (1991) 1400–1401.
[67] R. Aebersold, E.J. Bures, M. Namchuk, M.H. Goghari, B. Shushan and T.C. Covey, *Protein Science*, 1 (1992) 494–503.
[68] J.M. Bailey, O. Tu, C. Basic, G. Issai and J.E. Shively, in J.W. Crabb (Editor), *Techniques in Protein Chemistry V*, Academic Press, 1994, pp. 169–178.
[69] M.L. Stolowitz, C.-S. Kim, S.R. Marsh and L. Hood, in K. Imahori and F. Sakiyama (Editors), *Methods in Protein Sequence Analysis*, Plenum, 1993, pp. 37–44.
[70] B. Chait, R. Wang, R.C. Beavis and S.B.H. Kent, *Science*, 262 (1993) 89–92.
[71] P. Schlack and W. Kumpf, *Z. Physiol. Chem.*, 154 (1926) 125–170.
[72] A.S. Inglis, *Anal. Biochem.*, 195 (1991) 183–196.
[73] T.B. Johnson and B.H. Nicolet, *J. Am. Chem. Soc.*, 33 (1911) 1973–1978.
[74] T.B. Johnson and W.M. Scott, *J. Am. Chem. Soc.*, 35 (1913) 1136–1143.
[75] T.B. Johnson, A.J. Hill and B.H. Bailey, *J. Am. Chem. Soc.*, 37 (1915) 2406–2416.
[76] H. Kubo, T. Nakajima and Z. Tamura, *Chem. Pharm. Bull.*, 19 (1971) 210–211.
[77] A.S. Inglis, J.F.K. Wilshire, F. Casagranda and R.L. Laslett, in B. Wittmann-Liebold (Editor), *Methods in Protein Sequence Analysis*, Springer, 1989, pp. 137–144.
[78] A.S. Inglis, R.L. Moritz, G.S. Begg, G.E. Reid, R.J. Simpson, H. Graffunder, L. Matschull and B. Wittmann-Liebold, in H. Jornvall, J.O. Hoog and A.-M. Gustavsson (Editors), *Met. Protein Sequence Analysis*, Birkhauser-Verlag, Basel, 1991, pp. 23–34.
[79] D.H. Hawke, H.-W. Lahm, J.E. Shively and C.W. Todd, *Anal. Biochem.*, 166 (1987) 298–307.
[80] I. Fleming, in D.N. Jones (Editor), *Comprehensive Organic Chemistry*, Vol. 3, Pergamon, 1979, pp. 541–686.
[81] B. Wittmann-Liebold, L. Matschull, U. Pilling, H.-A. Bradaczek and H. Graffunder, in H. Jornvall, J.O. Hoog and A.-M. Gustavsson (Editors), *Met. Protein Sequence Analysis*, Birkhauser-Verlag, Basel, 1991, pp. 9–22.
[82] D.H. Hawke and V.L. Boyd, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry: II*, Academic Press, 1991, pp. 107–114.
[83] N.S. Shenoy, J.M. Bailey and J.E. Shively, *Protein Science*, 1 (1992) 58–67.
[84] J.M. Bailey and J.E. Shively, *Biochemistry*, 29 (1990) 3145–3156.
[85] F.A. Csonka and B.H. Nicolet, *J. Biol. Chem.*, 99 (1933) 213–216.
[86] H.E. Carter and C.M. Stevens, *J. Biol. Chem.*, 133 (1940) 117–128.
[87] M. Goodman and L. Levine, *J. Am. Chem. Soc.*, 86 (1964) 2918–2922.
[88] J.W. Cornforth, *The Chemistry of Penicillin*, Princeton University Press, 1949, pp. 688–848.
[89] H.-A. Bradaczek and B. Wittmann-Liebold, *J. Prot. Chem.*, 11 (1992) 365.
[90] V.L. Boyd, D.H. Hawke and T.G. Gieser, *Tetrahedron Lett.*, 31 (1990) 3849–3852.
[91] V.L. Boyd, M. Bozzini, G. Zon, R.L. Noble and R.J. Mattaliano, *Anal. Biochem.*, 206 (1992) 344–352.
[92] N.R. Shenoy, J.E. Shively and J.M. Bailey, *J. Protein Chemistry*, 12 (1993) 195–205.
[93] J.M. Bailey, F. Nikfarjam, N.S. Shenoy and J.E. Shively, *Protein Science*, 1 (1992) 1622–1633.
[94] G.W. Kenner, H.G. Khorana and R.J. Stedman, *J. Chem. Soc. (London)*, (1953) 673–678.
[95] P. Husek, *J. Chromatogr.*, 552 (1991) 289–299.
[96] T. Shioiri and S.-I. Yamada, *Chem. Pharm. Bull.*, 22 (1974) 855–858.
[97] S.G. Waley and J.J. Watson, *J. Chem. Soc.*, (1951) 2394–2397.
[98] A. Kjaer and P. Eriksen, *Acta Chem. Scand.*, 6 (1952) 448–450.
[99] R.A. Turner and G. Schmerzler, *Biochim. Biophys. Acta.*, 13 (1954) 553–559.
[100] J. Tibbs, *Nature*, 168 (1951) 910.
[101] V.H. Baptist and H.B. Bull, *J. Am. Chem. Soc.*, 75 (1953) 1727–1729.
[102] E. Scoffone and A. Turco, *Ric. Sci.*, 26 (1956) 865–871.
[103] S.W. Fox, T.L. Hurst, J.F. Griffith and O. Underwood, *J. Am. Chem. Soc.*, 77 (1955) 3119–3122.
[104] G.R. Stark, *Biochem.*, 7 (1968) 1796–1807.
[105] L.D. Cromwell and G.R. Stark, *Biochem.*, 8 (1969) 4735–4740.
[106] S. Yamashita, *Biochem. Biophys. Acta.*, 229 (1971) 301–309.
[107] D. Hawke and V.L. Boyd, in H. Jornvall, J.O. Hoog and A.-M. Gustavsson (Editors), *Met. Protein Sequence Analysis*, Birkhauser-Verlag, Basel, 1991, pp. 35–45.
[108] C.G. Miller, C.-T. Kong and J.E. Shively, in T.E. Hugli (Editor), *Techniques in Protein Chemistry*, Academic Press, 1989, pp. 67–78.
[109] J.M. Bailey, N.S. Shenoy, M. Ronk and J.E. Shively, *Protein Science*, 1 (1992) 68–80.
[110] J.M. Bailey and J.E. Shively, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry: II*, Academic Press, 1991, 115–129.
[111] M.J. Williams and B. Kassell, *FEBS Lett.*, 54 (1975) 353–357.
[112] M. Rangarajan and A. Darbre, *Biochem. J.*, 157 (1976) 307–316.
[113] J.L. Meuth, D.E. Harris, F.E. Dwulet, M.L. Crowl-Powers and F.R.N. Gurd, *Biochem.*, 21 (1982) 3750–3757.
[114] S. Yamashita and N. Ishikawa, *Proc. Hoshi. Pharm.*, 13 (1971) 136–138.

[115] T. Suzuki, K.-D. Song, Y. Itagaki and K. Tuzimura, *Org. Mass Spectrometry*, 11 (1976) 557–568.
[116] A.S. Inglis and C. De Luca, in K. Imahori and F. Sakiyama (Editors), *Methods in Protein Sequence Analysis*, Plenum, 1993, pp. 71–78.
[117] J.M. Bailey and J.E. Shively, in J.E. Shively (Editor), *Micromethods for Protein Structural Analysis: A Companion to Methods in Enzymology*, Academic Press, 1994, in press.

ELSEVIER

Journal of Chromatography A, 705 (1995) 67–87

JOURNAL OF CHROMATOGRAPHY A

Review

Isoelectric focusing as a tool for the investigation of post-translational processing and chemical modifications of proteins

Elisabetta Gianazza

Istituto di Scienze Farmacologiche, via Balzaretti 9, I-20133 Milan, Italy

Abstract

It has been demonstrated that good agreement may be observed between computed and experimental isoelectric point (p*I*) values when proteins of known sequence are focused under denaturing conditions on immobilized pH gradient IPG slabs, at least in the pH range 4–7.5. Hence, discrepancies between expected and found in this experimental set-up may be reliably ascribed to some kind of post-transcriptional processing, or chemical modification, having taken place in the sample. This evaluation is made easier when the comparison is set between the p*I* of a parent molecule and that (or those) of one to several of its derivatives as resolved in a single experiment (for instance, as a spot row in two-dimensional maps); no previous knowledge is required in these cases about the amino acid composition of the primary structure. The effects on protein surface charge are discussed in this review mainly for two biologically relevant processes, glycosylation and phosphorylation. Then, the p*I* shifts are analysed for some protein modifications that may occur naturally but can also be artefactually elicited, such as NH_2 terminus blocking, deamidation and thiol redox reactions. Finally, carboxymethylation and carbamylation are used to exemplify chemical treatments often applied in connection with electrophoretic techniques and involving charged residues. Procedures to be applied in order to verify whether a given modification has occurred, and often relying on the focusing of a treated specimen, are detailed in each section. Numerical examples on model proteins are also discussed. As an important field of application of the above concepts may be genetic engineering, an exhaustive bibliographic list dealing with p*I* evaluation and structural assessment on recombinant proteins is included.

Contents

0021-9673/95/$29.00
SSDI 0021-9673(94)01251-2

1. Introduction

According to an optimistic conception, proteins synthesized in heterologous 'model' systems (in vitro translation of mRNA, monoclonal antibody secretion by hybridomas, protein overproduction in transfected cells) are expected to be both size and charge homogeneous. A less confident picture still assumes that in experimental settings some steps of the physiological processing may not take place, and the final product should then correspond to simplified 'proforms' as opposed to the mature, complex structures one would purify from natural sources. On the contrary, the proteins produced are often heterogeneous (or microheterogeneous), and in a few instances wrong and unexpected processing does occur.

Worries about the outcome of uncontrolled metabolic pathways have been clearly asserted by National Drug Administrations. Proteins obtained by biotechnological procedures, before they are registered for therapeutic use, need to be shown to be identical with their natural counterparts according to the criteria of both size and surface charge, the latter being assessed by isoelectric focusing [1].

Surprisingly, this point has received only mediocre attention in the scientific literature. In a search for 1984–93 in the EBSCO-MedLine bibliography, 138 entries were extracted by the key words Post-translational 'and' Isoelectric Point, and 62 by Protein Heterogeneity 'and' Biotechnology 'or' Recombinant. In less than 20 out of these 200 items were expected and found properties of a protein compared, and in many instances just the size of the molecule was assessed.

The aim of this paper is to overview the general issue of enzymatic and non-enzymatic modifications of proteins, with reference to their effect on surface charge. Basic concepts and general references will be mostly discussed (in the reference list, *Annual Reviews of Biochemistry and of Cell Biology*, *Advances in Protein Chemistry*, *Methods in Enzymology* and IRL's *Practical Approach* series are most often cited); in keeping with the theme of this issue, specific examples dealing with biotechnological applications are covered in a separate section.

Post-translational modifications may be grouped into two kinds. (a) The first covers changes in a protein primary structure, with proteolysis at peptide bonds [2–4]. It includes the removal of intervening sequences or of leader peptides [5] as for intracellular processing, as well as the action of proteolytic cascades or of 'converting enzymes' in the metabolism of secreted molecules. Most of these events are relevant to biotechnologically engineered proteins [for examples, see [6–10]). Subunits containing unlike peptides usually result in vivo from proteolysis of larger preforms; in vitro, on the other hand, they must be assembled with low yields from individual components. Hydrophobic stretches prompting secretion into the culture medium as much as polyhistidine runs exploited for one-step purification by IMAC are often cloned 5′ to protein coding sequences [11–13]. Formylmethionine is the primary N-terminus of all bacterially synthesized proteins. Although in the latter example, and also in some single amino acid clipping events from the NH_2 [14] or the COOH terminus [15], molecular mass reduction is negligible, these modifications are typically identified by size fractionation techniques, including SDS-PAGE. (b) Attention will be mostly devoted to the second mode of processing, in which the size is slightly affected and no peptide bond is cleaved. This includes both enzymatic [16–18] and non-enzymatic [17,19] covalent modifications in vivo, and in vitro

chemical reactions [20–23]. The list of topics covered is detailed in the Contents. A short bibliographic survey at the end covers the modifications not dealt with in detail.

As a rule, post-translational processing confers upon proteins specific properties that make the holo forms substantially differ from the corresponding apo forms (as discussed in detail for glycoproteins in Ref. [24]). Parameters such as distribution or half-life may be influenced, but most often the very function of the protein is affected. Fine tuning of protein interactions may thus be driven by differential processing along the cell cycle (e.g., [25,26]), at various ontogenetic steps [27–30], in different tissues [31,32] or under varying physiological [33,34] or pathological [35–41] conditions. This evidence, and the possibility of immunological reaction to variant protein forms, explains the concern against improperly processed recombinant proteins for therapeutic use.

With this observation, we are back to isoelectric focusing as an analytical tool to detect protein variants resulting from differential processing. While the discriminating power of the technique will be discussed at length in a later section, it is stressed here that, to be recognized as isoforms derived from a single primary structure, various protein components require to share (and to show) common properties. (a) The simplest parameter is merely quantitative: the only, or the major, components of purified, or partially purified, samples are readily recognized. As trivial as this notation may be, a further common property is to link such a set of peptides, i.e., their purification 'history'. (b) The most complex identifiers, in turn, rely on biospecificity: antigens may be detected by immune reactions, enzymes by zymograms and binding proteins by affinity blotting. In both these extreme cases, 1-D separations are adequate for assessing relationships among protein bands. (c) In general, however, the only similarity one can immediately appreciate rests on similar molecular masses. This corresponds visually to the grouping of protein spots into horizontal, or slanted, rows in 2-D maps. In some cases, the parent molecule and its derivatives are present in similar amounts, in others the apo-protein prevails, and in still other instances one of the modified forms is the most represented. The baricentre of a row may thus correspond to the first (= most acidic, e.g., HSP 27 in HUVEC 2-D [42]), to the middle (= average p*I*, e.g., in serum maps, transferrin or α_1-antitrypsin [43]) or to the last component (= most alkaline, e.g., apoA-I [43]). How to assess the kind of relationships among protein isoforms will be detailed in the main sections of this review. A strategy to screen for coordinate pairs of polypeptides by the computer-assisted analysis of 2-D patterns has been devised by Lemkin et al. [44]. It is based on the assumption that, if there is a post-translational modification in a protein in the transition between two functional states of a biological system, in many cases the sum of the protein concentration of a precursor–product pair in one is equivalent to that in the other. In addition to the identification of candidate pairs with a structural relationship, cues to the nature of the post-translational modification that might relate them could also be derived from comparison of the isoelectric point and apparent molecular mass.

As with all electrophoretic techniques, IEF and 2-D (= IEF + SDS-PAGE), followed by any in situ staining protocol, allow for protein analysis on minute amounts of material, without the need for prior purification. In contrast, 2-D maps may evaluate qualitatively and quantitatively all protein components in a complex sample.

Emphasis is given here to electrophoretic techniques, because with them a wealth of accurate chemico-physical and, in many instances, functional information may be gathered within a short time, with the use of simple and commonplace equipment and of inexpensive reagents. Ready-made IPG slabs make it possible even for inexperienced workers to combine ease with high-quality standards for reproducibility and resolution in p*I* assessment.

NMR may be successfully applied to investigate protein structure [45,46] and even to assess the homogeneity of protein preparations [47]. Very good resolution of macromolecules on the basis of their charge is now being obtained by CE [48]. An application to the identification of

chemical modification sites on metalloproteins can be found in Ref. [49]. Unsurpassed accuracy in mass evaluation is provided by MS [50–52]. Mass variations brought about by the post-translational processing exceed in most instances the confidence limits, as good as 0.001–0.02% with ESI, for a mass range >100 000 [53]. [Δ_{mass} ($100\,000_{NH_2}$) → ($100\,000_{CH_3CONH-}$) = 0.057%; (100 000) → ($100\,000_{PO_4}$) = 0.095%]. Identification of post-translationally modified amino acids is discussed in Ref. [54]; an example of the determination of the glysosylation patterns, disulfide linkages and protein heterogeneities of baculovirus-expressed proteins by MS is given in Ref. [55]. CE and MS more and more often are being coupled off- and on-line in a set-up that duplicates the sequence of 2-D electrophoresis. While it is possible that in the near future these procedures will become standard, and the dedicated equipment will be available to most research institutions, these techniques are just leaving their prototype stage, and their costs still exceed by far the investment required by standard electrophoretic equipment.

Thus, in the meantime, why not exploit available techniques with all of their possibilities?

2. Protein titration curves and p*K* of the side-chain dissociating groups

A titration curve describes the relationship between pH and surface charge of a molecule [56,57]. For a protein, such a curve results from the contribution of the side-chains of different charged amino acids, with various p*K*s and varying relative abundance. When no influence from neighbouring segments of the molecule is exerted on the dissociation of any group, and each charged residue may be treated independently, then the surface charge of a protein results from the sum of individual contributions. This situation is experimentally approached for proteins in their unfolded state when short- and long-range interactions between different parts of the molecule are prevented by high ionic strength and/or chaotropic agents (guanidinium hydrochloride, urea). Under the assumptions above, the course of a protein titration curve may be modelled from its primary structure. The p*K* values for amino acid side-chains and for NH_2 and COOH termini to be used in such a computation are those experimentally derived from a panel of model compounds [57].

Conversely, the experimental assessment of a protein titration curve allows the evaluation of the number and p*K* of its dissociating groups. Two procedures, electrometric [58] and electrophoretic [59–61] titration, may be applied. The former requires several milligrams of a pure protein and a pH meter; the analysis may be carried out over a wide range, including pH extremes. The electrophoretic titration, corresponding to migration of a protein across a pH gradient, is limited to the ca. 4–10 range, where arginine groups are fully protonated throughout. As an advantage, the protein to be analysed needs neither to be homogeneous nor to amount to more than tens of micrograms. What is actually obtained with this technique is a 'pH–mobility' curve. However, for a given protein, the latter is proportional to a true titration, as long as size and shape of the protein are constant across the experimental range, and the viscous parameter dictating u is not affected by pH. In contrast, no direct comparison between the absolute surface charge on different proteins is usually possible, as the above parameter does change from one molecule to another. In fact, this electrophoretic technique has mostly been used to study the pH dependence of binding phenomena (protein–protein [62–64] and protein–ligand interactions, either in solution [65,66] or matrix-bound, as in affinity electrophoresis [67,68]), although inference about structure could be derived by pairwise comparisons between protein isoforms [69,70].

When titration curves are evaluated under native conditions, the discrepancies between computed and found describe the interactions of individual charged groups with their environment, and the connected p*K* shifts. A different approach for the experimental assessment of individual p*K*s into native proteins is through the analysis of the pH dependence of EPR signals

[71]. In a few cases, when the tertiary structure of a protein is known, the p*K* of specific groups can be computed while the shielding effect of neighbouring residues is taken into account [72–77].

As with any chemico-physical datum, p*K*s vary with temperature, slightly for acidic and more extensively for basic residues. Reference should then always be made to unambigously specified experimental conditions (see below).

3. Isoelectric point

The p*I* is a singularity point in a titration curve, corresponding to the pH at which the surface charge of a protein equals zero, hence its electrophoretic mobility is also zero [78–80]. If referred to a random coil, the p*I* only depends on a protein amino acid composition, and can be computed from analytical or sequence data. In contrast, when interactions are allowed to occur either between amino acid stretches within a protein in its native structure or between ampholytes and appropriate additives in the analysis medium, the experimental p*I* may be subtly altered. As a rule, when moved to a hydrophobic environment, a group shifts its dissociation so as to favour the uncharged form (i.e., the p*K* of an acid increases and that of a base decreases) [57]. Whether the result of a structural change (e.g., several well characterized variants of haemoglobin [81,82] or α_1-antitrypsin [83,84]) or brought about by the experimental conditions (D,L-ANS-amino acids and cyclodextrins [85], neutral–neutral protein mutants in urea–detergent [86,87]), these p*I* changes can be exploited in favourable cases for the resolution of 'neutral' isoforms.

The hypothesis of a random coil for a protein spatial arrangement implies that no restrictions exist on the relative movement between its different segments. This rule would be violated either by the presence of intra-chain covalent bonds, such as with –S–S– bridges, or by structural constraints, such as with turns forced between adjacent residues by proline. Cystein and cystine, both neutral amino acids in acidic to slightly basic buffers, give a differential contribution to a protein surface charge at high pH. A knowledge of the oxidation form of sulfur-containing amino acids is thus required for a reliable assessment of the p*I* of alkaline proteins.

3.1. Experimental assessment of pI by IEF

Proteins are separated purely on the basis of their p*I* values by the electrophoretic technique of isoelectric focusing [88], in which amphoteric compounds migrating along a pH gradient stop moving when they reach a region in the separation medium whose pH matches their isoelectric point. Any molecule diffusing out of the isoelectric region becomes charged, to be moved back by electrophoresis to its equilibrium zone, where it concentrates or 'focuses'.

A stable pH gradient within an anticonvective matrix may be established essentially in two ways. The former (CA-IEF) [89,90] makes use of a large number of amphoteric buffers, whose p*I*s, each slightly different from the next one, evenly fill the pH span to be covered in the experiment. The separation medium, within the boundaries set by the electrodes, or by two conductivity barriers in contact with them (concentrated solutions of an acid and a base, respectively), is buffered at varying pH by the sequence of the buffers (carrier ampholytes, CA) at their p*I*s. The pH spans covered with this technique vary between 2 and ca. 6 pH units; the gradient is intrinsically unstable with time and may be upset by the presence in the system of non-amphoteric ionic compounds (salts).

Any gradient decay is prevented when the buffering groups dictating the pH are covalently bound within the network of the separation matrix, as with immobilized pH gradients (IPG) [91,92]. These are obtained by the copolymerization of acidic and basic acrylamido derivatives together with acrylamide and bisacrylamide monomers. The gels are allowed to set after pouring a gradient from two limiting solutions whose composition as for the buffering compounds is adjusted to result in a linear pH course between the stated extremes. Gradients prepared in this way may be cast as wide (7.5 pH

units) or as narrow (0.1 pH unit) as required, in order to optimize the resolution of any protein mixture. For specific purposes, non-linear gradients may also be devised and precisely cast [93].

The p*I* of an unknown protein may be experimentally measured after an IEF run by interpolating the focusing position of the relevant band over the actual pH gradient course. For IPGs the latter may be assumed to correspond to the figures computed by modelling, provided that (i) gradient pouring is linear, (ii) the incorporation efficiency is the same for all individual monomers [94], and (iii) the p*K* of the dissociating groups is accurately known under the experimental conditions (effect of temperature [95]; effect of additives [96]). The last point has recently been reassessed by Bjellqvist et al. [97]. They established a relevant pH scale for IPG runs in 8 *M* urea at 20°C, while validating a correspondence between the focusing position on such a gradient and the protein p*I* as computed from its known sequence (see below). Conversely, an experimental evaluation of the pH course in an IPG is hardly possible, e.g., with a surface electrode. Readings may be performed on aliquots of eluted isoelectric buffers when IPGs are run after reswelling in carrier ampholytes (mixed-bed, or hybrid, isoelectric focusing) [98]. The latter approach is standard after CA-IEF, but the accuracy of such measurements is usually low, not only because of the vagaries in the cutting and elution steps but mostly for the systematic effects, which are difficult to compensate for, of differences in temperature (on CA [99,100]; on proteins [101]) and solvent (on CA [99,102,103]; urea on proteins [104,105]) between separation and pH testing.

3.2. *Theoretical calculations of pI*

At least in the pH range 4–7.5 [106], the accordance between the experimental p*I* of a protein and the figure computed for a random-coiled structure from type and number of dissociating side-chains is so close (within one third to half of a charge unit) as to allow the assignment of a peptide spot in a 2-D map to a known sequence (the second positional parameter, M_r, is also immediately derived from amino acid composition). Any discrepancy from the expected focusing position, and exceeding the experimental error, may then be assumed to imply some structural modification in the species under investigation. By this approach, Bjellqvist and co-workers could identify either four [106] or 18–20 [107] cases of blocked NH_2 terminus; for other proteins, they raised doubts about the peptide being glycosylated, or correctly sequenced [107]. Otherwise, the largest discrepancy was observed for a protein not containing His residues and whose p*I* then depended markedly on the p*K* of the actual N-terminus [107]. Another interesting example of comparison between computed and found p*I* values for mouse MHC class I antigens is discussed in Ref. [108]. As stressed several times, the computation of p*I* is in principle restricted to fully denatured structures. However, for many proteins neither the surface charge distribution nor hydrophilic/hydrophobic balance noticeably influences the intrinsic p*K*s of the charged amino acids, hence the p*I* of the protein and the assumptions above may be extended to functionally folded structures.

For native proteins, 'structural modifications' might imply non-covalent binding as for the holo forms of carrier proteins and prosthetic enzymes, and the covalent (–S–S– bridges) and non-covalent interactions between different subunits within oligomeric structures. For denatured proteins, only covalent modifications are detected after electrophoresis in urea media (except for the 'hydrophobic affinity' effects brought about by addition of detergents to the separation medium; see above).

To be discriminated by IEF, the parent and modified forms of a protein need to differ at pH = p*I* by an integral or fractional number of units in their surface charge. In turn, this charge difference will be proportional to the p*K*–p*I* difference for the charged group involved in the structural change. The extent of p*I* shift for a given charge difference then depends on the buffering power (at p*I*) of the protein, treated as a polyprotic buffer. Less precisely, it depends

inversely on the total surface charge of the protein (at pI) and on its size [107].

The discrepancies between theoretical and experimental pI values in Bjellqvist et al.'s survey [107] were of the order of a few hundredths of a pH unit, i.e., about one order of magnitude above the current resolving power with narrow-range IPGs (in a side-to-side comparison of different isoforms, a few thousandths of a pH unit are a sufficient ΔpI for unambiguous resolution). These data imply that some of the shifts connected with a covalent modification might not exceed the confidence limit for pI estimate; caution is then required for discriminating between the parent molecule and its processed form when dealing with a single protein spot (i.e., when assessing, for instance, whether a COOH terminus is free or blocked).

The finding of peptide spots in a row on a 2-D map (i.e., with similar or identical size, in a slanted or a horizontal row) is diagnostic of the presence of different isoforms deriving from a parent molecule, although relatedness should be positively confirmed by functional (e.g., immunological) or biochemical (e.g., V8 peptide mapping) tests. By the same approach, in order to identify the nature of the bound residue, the number and relative pI of the resolved spots are to be compared with number and pK of the putative sites for a given post-translational modification. The correspondence of the actual protein pattern with expectation is, however, only a prerequisite to the experimental validation of the hypothesis through chemical and biochemical approaches (specific removal, or analytical identification, of the modifying groups).

4. Post-translational modifications

In surveys on protein processing, tens of amino acid modifications have been listed [16,109]. Many of them are highly specific, whether to a single protein or to a class thereof (just one example in Ref. [110]). Other are commonplace, and among them mainly two processes will be stressed, namely glycosylation and phosphorylation, whose biological role is largely understood. Then some protein modifications that may occur naturally but can also be artefactually elicited will be discussed. Finally, some chemical treatments in common use that are relevant to electrophoretic techniques are analysed.

4.1. Glycoproteins

Sugar moieties are essential in protein targeting to various cellular and extracellular compartments, and take part in signalling and recognition [24,111–115]. Different glycoforms may be tissue-specific [24]. As an extreme case, no Thy-1 molecules are common between rat brain and thymus, despite the amino acid sequence being identical [116]. A marked difference in the glycosylation patterns between two physiological states (baseline vs. acute phase reaction) has been described for a number of serum components (including α_1-acid glycoprotein [117] and α_2-macroglobulin [118]), and also for basolateral membrane protein CE 9 in rat hepatocytes upon administration of the peroxisome proliferation inducer ciprofibrate [119]. Two interesting examples of cytoplasmic non-glycosylated proteins having a secretory glycosylated counterpart are those of ferritin [120] and PAI-1 [121]. Most uncommon, in contrast, is a remodelling in the oligosaccharide chains once the glycoproteins have reached their final compartment; sialic acid removal from serum proteins amounts to a signal for uptake and catabolism [122].

As a rule, the product of protein glycosylation is highly heterogeneous. In animal cells, the last sugar in every saccharide chain is always charged. Sialic acid is not to be found in plants whereas, as a rule, bacteria synthesize no glycoprotein at all [123]. Thus, animal glycoproteins resolve in a 2-D map as a slanted row of spindle-shaped spots. The pK of sialic acid is around 3, hence a charge difference between the various isoforms is to be maintained throughout a wide pH range. α_1-Acid glycoprotein is possibly the only example of a protein the resolution of whose isoforms at pH near pI depends on the varying extent of sialic acid titration [124]. Removal of sialic acid by neuraminidase treatment

reduces (but seldom abolishes) protein microheterogeneity (some residues are extremely resistant to enzyme digestion). The kinetics of this treatment resolve intermediate bands corresponding to the stepwise removal of sialic acid; their number matches the number of such residues. Sialidase activity is often found in body fluids: careful handling conditions, without repeated cycles of freezing and thawing, are then required for proper sample storage. In order to emphasize the charge shift contributed by amino acid mutation while disregarding any metabolic effect on the glycosylation process and to simplify the electrophoretic pattern, for some glycoproteins, including α_1-acid glycoprotein [125], genetic analysis is currently performed on the desialylated apoproteins. Complete removal of the sugar chains may be obtained with endoglycosidases (examples in Refs. [126] and [127]). These usually require the protein to be in a completely unfolded state, such as after treatment with SDS followed by saturation with a non-ionic detergent. Changes in M_r always ensue, whereas for sialylated molecules the increase in p*I* just duplicates what is obtained with neuraminidase alone. For non-sialylated glycoproteins, such as those of vegetable origin, no charge shift is observed [for instance, on 7 S globulin from *Glycine max.* (β-conglycinin); unpublished data].

Enzyme treatment is one of the procedures for assessing the presence of glycyl moieties, to be applied to fully processed molecules. Another experimental approach may be affinity detection after blotting [128] using lectins specific for a given sugar type [129]. Even upon aspecific staining the behaviour of glycoproteins is sometimes peculiar: they may give metachromatic shadows with Coomassie Blue staining or (when heavily sialylated) bind silver very poorly. Sugar moieties are often part of an immunogen epitope, so many neuraminidase- and endoglycosidase-treated proteins lose some or even most of their immunological reactivity. On purified proteins, chemical analysis can qualitatively and quantitatively specify all bound glycyl moieties ([130]; just one example in Ref. [131]).

Synthesis within a controlled experimental system allows one to monitor p*I* and M_r shifts as the newly assembled protein travels from ER and Golgi to its final compartment [132–135]; to determine the incorporation of tritiated precursors [136,137]; or to prevent glycosylation altogether in the presence of inhibitors [138,139].

Among electrophoretic techniques, CE is becoming a practicable alternative to IEF when 'fingerprinting' oligosaccharides (e.g., [140]).

Non-enzymatic glucosylation will be dealt with later.

4.2. Phosphoproteins

Phosphorylation [141–145] is most often a reversible process, whereupon one (or few) residues are added or released at precise locations, by highly specific enzymes, in response to proper stimuli [146–155]. The holo and apo forms of the affected proteins differ in their biological role, one being the active and the other the inactive form of an enzyme [146], or else one acts as a stimulatory and the other as an inhibitory factor in a regulation process [156]. The in ↔ out flux of phosphate groups thus parallels the on ↔ off switch of a protein function. The current pattern of phosphorylation may thus be highly variable depending on the physiological state of the extracted tissue; it actually defines such a state. Care must be taken to avoid any interference from the extraction conditions, since some phosphate groups are labile; phosphatase inhibitors are available for this purpose [145]. In some cases many sites in a protein may become phosphorylated, but the number of physilogically relevant residues, whose occupancy entails a functional change, is restricted. Addition and removal of phosphate groups might result in an M_r change for proteins resolved by SDS-PAGE [157–160]. On the other hand the one-charge (below pH 6) or two-charge (above pH 7) shift connected with each acidic group always involves a Δp*I*, to be resolved by IEF.

The hypothesis that a group of protein bands corresponds to a phosphorylation train may be tested by phosphatase treatment; as a rule, the 'out' step may be performed stoichiometrically

with broad-specificity enzymes [161,162]. In contrast, the 'in' process is usually brought about only by narrow-specificity phosphorylases [163–165]. As a result, whereas reduction of the IEF pattern to a single most basic band is often possible [166,167], the converse, i.e., the 'saturation' of putative phosphorylation sites so as to give a single most acidic protein derivative, is not a biologically sound proposition.

The action of protein kinases, however, may be duplicated in vitro [167]. The experimental set-up in which phosphorylation is tested typically involves the use of ^{32}P [166,168]. Radioactive tagging allows one to appreciate a phenomenon involving low-abundance proteins, such as most of the enzymes and regulatory factors alluded to above. Moreover, only the newly synthesized phosphoproteins enter a qualitative and quantitative evaluation. For the former point, while the physiologically relevant form (or forms) takes up the label, the baseline heterogeneity of the protein is immaterial. For the latter, the incorporated radioactivity is evaluated with reference to a zero background; even when possible, immunological detection or protein staining would require quantification of each species within a sometimes complex train before and after the relevant stimulus, and comparison between the 'baricenter' of the two band rows. Monitoring the 'out' step would be much more difficult with this approach.

Antisera may be developed specific for the phosphorylated versus the apo form of a protein [169]; monoclonal antibodies directed against phosphotyrosine are commercially available. Other experimental approaches to the analysis of phosphoproteins are discussed in Ref. [170]; special interest is devoted to ^{31}P NMR studies [171]. Phosphopeptide analysis is reviewed in Ref. [172].

4.3. Reactions at NH_2 groups

A blocked NH_2 terminus is common finding when Edman degradation is attempted on blotted proteins. A number of residues may react, by chemical equilibrium or enzymatic action, with the free NH_2 group and, while the evidence of a block is easily obtained, no clue is directly available as to the chemical nature of the blocking residue. A stepwise treatment of electroblotted proteins has been proposed in order to expose, whenever possible, the NH_2 terminus [173]. Proteins containing acetylserine or acetylthreonine can be deblocked on-membrane by exposure to TFA vapours and N-formylated proteins by treatment with HCl solution. Pyrrolidone carboxylic acid residues may be removed, with varying efficiency, by pyroglutamate aminopeptidase. N-acetylated proteins are first digested on-membrane with trypsin in order to generate the N-terminal peptide fragment, that is now available to the deblocking action of acylamino acid-releasing enzyme. The percentage of NH_2 terminus-blocked proteins found by different workers in different tissue extracts is highly variable, and artefacts from the sample preparation procedure have been implicated. These might include oxidation, reaction with aldehydes (e.g., HCHO as a contaminant in Tris buffers) and cyclization of glutamic acid to pyroglutamate. Whether a physiologically important modification or an artefact, the blocking is a major nuisance for protein identification through partial sequencing, as it forces the experimenter to resort to protein digestion and peptide purification prior to internal sequencing.

Amino-terminal [174] and side-chain [175] acetylation may be treated together, and also with fatty acylations [176–178], as for their effects on acid–base balance: the removal of an amino group becomes evident only when pI is below pK_{NH_2}. For acidic to neutral proteins, however, the pI shift is easily seen [106]. In contrast to reactions with C_1 or C_2 groups, lipid modifications may be assessed, in newly synthesized proteins, by monitoring the incorporation of ^3H-labelled fatty acids (myristate [179], palmitate [180,181]).

Core histones can be reversibly acetylated at distinct lysine residues within the N-terminal protein domains [182]; this process is assumed to be involved in changes in chromatin structure and function during different nuclear processes. A dynamic equilibrium between apo and modified forms is maintained by two enzyme ac-

tivities, histone acetyltransferase and histone deacetylase [183]. Owing to their high p*I*, the histones cannot be run to equilibrium in CA-IEF and bind to the matrix on IPGs [184] (Righetti et al., however, subsequently presented data about the feasibility of alkaline IPG not interfering with histone migration). Their acetylated or otherwise modified derivatives are then resolved either by acid–urea–Triton electrophoresis [185] or by CE [186].

A peculiar example of reaction with the NH_2 terminus, and of a diagnostic use of IEF, involves haemoglobin and glucose. The adduct formed rearranges to give a chemically stable product, which makes the overall process irreversible [187]. The equilibrium between free and bound glucose depends only on its concentration: the determination of the glucosylated haemoglobin thus evaluates the average glycaemia over the lifespan of the protein, which amounts to several weeks. This makes glucosylated haemoglobin (HbA_{1c}) a valuable index in the follow-up of diabetic patients [188,189]. However, an important physiological property of the reacted protein, namely the O_2 dissociation curve, is also modified. In diabetes, the effects of generalized protein glucosylation are manifold [190]. One of the most serious involves the alteration of the filtering properties towards either anionic or cationic molecules by the kidneys, which behave in fact more as an ion exchanger than as a sieve. By detailing this example and that of histones above, it is intended to stress that the issue of p*I* shift after chemical modification treated so far only from a chemico-physical standpoint may also have major physio-pathological impact. For proteins other than Hb, non-enzymatic protein glucosylation [191–193] involves lysyl residues, instead of NH_2 termini (albumin [194], histones [195]).

4.4. Reactions at COOH groups

Deamidation is a most general phenomenon connected with protein ageing: for molecules with long half-lives, its quantification involves evaluating the protein turnover [16]. Δ (charge) for the transition $CONH_2 \rightarrow COOH$ varies across the pH range 3.5–5.5, corresponding to the titration of the carboxyl group. As a result, the p*I*s of parent and modified molecules may be difficult, or impossible, to sort out for very acidic molecules. Moreover, the electrophoretic migration of the two forms, differing at high pH, become more and more similar at low pH: at pK_{COOH}, the distance between the two protein bands is reduced to about half in comparison with $pH = pK + 1$, and becomes negligible at $pH = pK - 1$. This observation may in principle be exploited to verify whether two protein isoforms actually differ by a deamidation step: the electrophoretic titration [57–59] of the proteins to be compared is run across an acidic pH range, and the shape of the two curves is analysed. The same feature would be observed also for neutral → acidic amino acid mutations. As a rule, however, the deamidated side-product is a minor percentage in comparison with the amidated parent molecule, while two allelic variants in a protein polymorphism are expressed with the same abundance. Another indirect test for deamidation of Gln and Asn residues may rely on the differential specificity of V8 protease for Glu-C and Asp-C in different buffers, in comparison with the intensitivity to digestion of the amidated counterparts [196,197]. Another analytical approach, were both isoforms purified in substantial amounts, might involve the reaction with [^{14}C]glycine methyl ester in presence of soluble carbodiimide (resulting in the amidation of free COOH groups) [198]. Extensive chemical deamidation at alkaline pH, or blocking of free COOH, would result either in two proteins with lower but identical p*I*, or in polycations with identical mobility. These reactions could be performed on the unresolved isoforms, but neither of them is quantitative so as to permit an unequivocal interpretation of the results. Sequencing of the relevant peptide is then a complex yet reliable approach to assess the nature of the structural difference between the proteins under investigation. Extraction and treatment under alkaline conditions might result in artefactual protein deamidation; low temperature and as short as possible exposure times are thus essential to minimize deterioration.

The occurrence, detection and biosynthesis of carboxy-terminal amides and isoprenylation and

methylation at carboxy-terminal cysteine residues are discussed in Refs. [199] and [200], respectively, and exemplified in Ref. [201].

4.5. Reactions at SH groups

Many heteropolymeric proteins with interchain –S–S– bridges derive from continuous sequences processed to their quaternary structure through peptide bond hydrolysis, with or without the removal of an amino acid stretch [2–4]. In fewer cases, independently synthesized subunits are covalently linked to one another after their assembly [202]. In a single-chain protein, the position of the disulfide bridges is dictated by the folding process, which in turn depends on the primary sequence; in vitro, air oxidation is adequate to form disulfides between the appropriate Cys pairs [203,204]. Under physiological conditions this step, which requires hours when testing a fully denatured and reduced protein, is completed alongside the translational process [205,206]. Cycling between reduced and oxidized (thiol SH $\rightleftharpoons$ disulfide –S–S–) forms is thought to take part in metabolic regulation through the redox control of enzyme activities [207]. However, heterogeneity such as for the oxidation state of a purified protein is often the result of an artefact, most likely oxidation by atmospheric oxygen at alkaline pH. For a purified protein, if a single SH residue is available per molecule, then only dimers may be formed, whereas a more complex situation might result in multiple banding [208]. While the heteropolymers are expected to have p*I*s intermediate between those of the parent components, the homodimers are indistinguishable from the monomeric form as long as the dissociation of SH does not contribute to the protein surface charge. Thus, acidic and neutral proteins during an isoelectric focusing separation are not exposed to such a pH as to elicit an artefactual oxidation, nor can this process, had it taken place beforehand, be detected by electrophoresis [209], although a functional test, such as zymogramming an enzyme preparation, may reveal a reduced specific activity. In contrast, for alkaline proteins, both the conditions of migration favour the oxidation process and the resulting p*I* shift can be identified by IEF [210,211]. While buffer addition with low concentrations of a reducing agent, for instance DTT, is compatible with zonal electrophoresis, their weak acid nature makes thiolic reagents hardly compatible with isoelectric focusing. In fact, when incorporated into an IPG gel, the additive tends to migrate at pH < p*K*, and hence it does not contribute to the reducing potential, whereas the non-depleted portion interferes with buffering power and pH course in the alkaline part of the gradient [209,212]. Most recent prescriptions to avoid protein oxidation on IEF include reduction of the IPG matrix with ascorbic acid [213], running under paraffin oil to exclude access of O_2 [210,214], and a certain migration of thiolic reagent from the cathodic strip during the last phase of the run [214]. With CA-IEF, even the low concentrations of thiolic reagents contributed by sample buffers distort the pH gradient above pH $\approx$ 7.5, with loss of alkaline proteins, when applied, as customary, near the cathode [209,212].

One alternative to the above problems could be to dispose altogether of reducing agents both in the sample and in the gel by alkylating any free SH on proteins. This approach has been proposed for SDS-PAGE samples in order to avoid Cys reoxidation and erratic banding patterns [215]. Moreover, treatment with iodoacetamide has been suggested as the final equilibration step between the first and second dimensions of a 2-D separation, in order to reduce background staining with many silvering procedures [216]. Carboxymethylation for IEF, or 2-D mapping, has been discussed a few times, but no consensus protocol could be arrived at. When applied to complex mixtures of proteins, such as whole tissue extracts, reduction and alkylation with iodoacetamide result for most of the components in neat, single spots. However, in comparison with a run of standard sample preparations, some spots disappear, or change their focusing position, or grossly increase their staining intensity. No simple trend can be observed for these anomalies, and none of the possible explanations for these findings (incomplete selectivity of the reaction towards SH versus NH_2 groups; partial deamidation of the

akylating reagent; differential silver deposition on the alternative nucleophilic centres, SH and –S–C–) is easily proved for any specific case [209]. These perplexities notwithstanding, carboxymethylation has been proposed as an analytical tool. Treatment with different ratios of iodoacetamide and iodoacetic acid results in the formation of a train of bands differing by one charge unit, and whose number exceeds by one that of Cys residues [23,217]. Intrinsically restricted in its application to charge-homogeneous proteins, and largely outdated by current sequencing techniques, this suggestion still deserves mention because, as is usually the case with techniques relying on electrophoretic separations, it may be applied even to incompletely purified proteins, as long as the relevant bands may be unambiguously identified, while allowing one to count integral numbers of amino acid residues per protein molecule, in contrast to standard analytical protocols following HCl hydrolysis. The above procedure introduces the concept of chemically derived charge trains, that is most often realized by stepwise carbamylation. SH and –S–C– forms of a protein migrate to the same position in IEF when their p*I* is acidic to slightly basic, but differ in their surface charge at alkaline pH. Monomeric and polymeric forms of a protein migrate with different mobility in a sieving medium under native conditions: only the fastest migrating band will be detected after addition of just SDS if the inter-subunit interactions are non-covalent in nature, whereas a reducing agent will be required if there exist cystine bonds. For a peptide with intrachain –S–S– bridges, the apparent M_r in SDS-PAGE is lower than expected from sequence data in the absence of a reducing agent. After reduction with β-mercaptoethanol, the estimated value fits the logarithmic relationship with R_F, and seems to grow larger upon carboxymethylation.

4.6. Carbamylation

The carbamylation protocol involves boiling homogeneous proteins, of neutral to alkaline p*I*, in a concentrated urea solution for increasing lengths of time [218]. From the decomposition of urea, carbamic acid is formed, which reacts with free NH_2. A series of protein derivatives of decreasing p*I* are obtained, evenly spaced along the p*I* axis at first, then focusing closer and closer as the influence of the residual buffering groups (His and Glu/Asp) becomes more important at pH = p*I*. Individual spots are conventionally labelled as 0 (parent molecule), $-1, \ldots, -n$ (loss of 1 to n positive charges). These mixtures of modified proteins have been used for years as landmarks for calibrating the pH axis in 2-D maps [218,219]. Although extremely efficient [220], the classical protocols for 2-Ds with CA-IEF in capillary tubes could not achieve absolute positional reproducibility in the spot pattern, for a number of reasons, e.g., deformability of the IEF gel rod, pH drift with time and gradient distortion from salts or thiolic reagents in the sample buffer. At the same time, for the natural protein spots that could be taken as landmarks, being common to many if not all the tissue extracts, the p*I* in urea was not known with accuracy [218]. The inclusion of the carbamylated proteins was to set a reference grid, not intended to specify a chemico-physical parameter such as p*I* but merely a relative position. Virtually all of these problems have now been solved with IPG technology. A relevant pH scale has been detailed [97], and accurate pH courses for many formulations covering narrow to wide pH gradients have been computed and published [93,221–223]. Moreover, the reproducibility of gradient pouring has been tested, and the positional reproducibility of a 2-D map, when IPGs supported on GelBond plastic backing are used in the first dimension, has been found to be better than 0.5 mm [224]. This allows one to read directly the p*I* of a protein spot, and its M_r, from its absolute position in the x–y plane. Moreover, the recent work by Bjellqvist et al. [107] established the correct p*I*, under denaturing conditions, for a number of ubiquitous proteins. In another context, carbamylation by treatment with sodium cyanate had been proposed for reducing the average p*I* of IgGs to be used in electroimmunodiffusion (Laurell's rockets) for the determination of alkaline proteins, whose electrophoretic mobility would be very

low at pH 8.2 (i.e., at the average p*I* of native IgGs) according to standard protocols [225].

Maleylation and its effects on p*I* and SDS-PAGE mobility have been discussed in Ref. [226].

4.7. Miscellaneous

In the bibliographic search mentioned in the Introduction, although most references dealt with glyco- and phosphoproteins, in three cases the authors dismissed any covalent modification to the protein under investigation and in 17 they had evidence of a post-translational processing but could not specify its nature.

An amide-linked ethanolamine–phosphoglycerol is reported in Ref. [227]; an example of methylation [228,229] is discussed in Ref. [230]. ADP-ribosylation [231,232] and ubiquitination [233,234] were analysed in histones ([235] and [236], respectively). Of the possible effects of covalent coenzyme binding to proteins [237,238], only one old example could be found [239].

5. Numerical examples

Although this review focuses on theory, while urging to experiment, it was felt that at least one figure with a few numerical examples were needed. Three proteins were considered as model structures, A, B and C (Fig. 1), differing in their amino acid composition (top panels, and left to right) and buffering power at p*I* (see below). For each of them, Δp*I* was computed [240] for a number of hypothetical post-translational modifications. Rounded figures were used for all p*K* values (see upper right panel; standard one-letter abbreviations are used for amino acids, COOH and NH_2 represent protein termini and H_2PO_4 and HPO_4^- indicate pK_1 and pK_2 of a phosphate group). In order to reduce the number of variables to be taken into account in the calculation [240], D and E counts were combined, and no Y was included in the hypothetical sequences. The postulated modifications are marked in Fig. 1 as follows: (1) for glycoproteins, presence of 1 sialic acid; (2) for glycoproteins, presence of 3 sialic acids; (3) for glycoproteins, presence of 5 sialic acids; (4) for phosphoproteins, presence of 1 phosphate group; (5) blocked NH_2 terminus; (6) deamidation of 1 Q/N group; (7) monomer with thiol redox as –S–S– plus SH; (8) dimer assembled via 3 –S–S– bridges; (9) for carbamylated proteins, reaction at 1 K; (10) for carbamylated proteins, reaction at 2 K; (11) for carbamylated proteins, reaction at 4 K; (12) for carbamylated proteins, reaction at 6 K.

The associated Δp*I* are plotted clockwise for each model protein (lower panels, from left to right); since these values vary by over three orders of magnitude, a logarithmic scale had to be used for their representation (bottom right).

In protein B, p*I* = 5.165, $\beta_{pI} = 5.68$ mequiv l^{-1} pH^{-1}, at 1 m*M* concentration, the p*I* shifts observed for glycosylation, phosphorylation, NH_2 blocking and carbamylation differ from one another by a maximum of 0.006 pH unit, since they actually amount to the addition, or to the removal, of one unit charge, either negative or positive. Owing to the buffering effect of D and E at and below p*I*, in stepwise processes the shifts become lower and lower, to approach a constant value. One hypothetical deamidation, in contrast, results in a fractional surface charge variation, hence the substantially lower p*I* decrement for the $CONH_2 \rightarrow COOH$ transition. No influence is observed from the redox state of C residues, since pK_{SH} is much higher than p*I*.

The same trend is observed for protein A, p*I* = 4.500, $\beta_{pI} = 14.3$ mequiv l^{-1} pH^{-1}, except that the Δp*I*s are much lower both in absolute terms and as differences among various one-unit changes. Moreover, the p*I* shift associated with a fractional charge variation (i.e., for the deamidated protein) amounts to a much lower percentage than in the instance above. This is, of course, the effect of the higher buffering power at p*I* allowed by the much larger number of D and E residues in this structure.

Conversely, the buffering power of protein C is very low, $\beta_{pI} = 5.02$ mequiv l^{-1} pH^{-1}, at and below the isoelectric point, p*I* = 9.014, and the shifts between subsequent steps in glycosylation and carbamylation become larger and larger. At

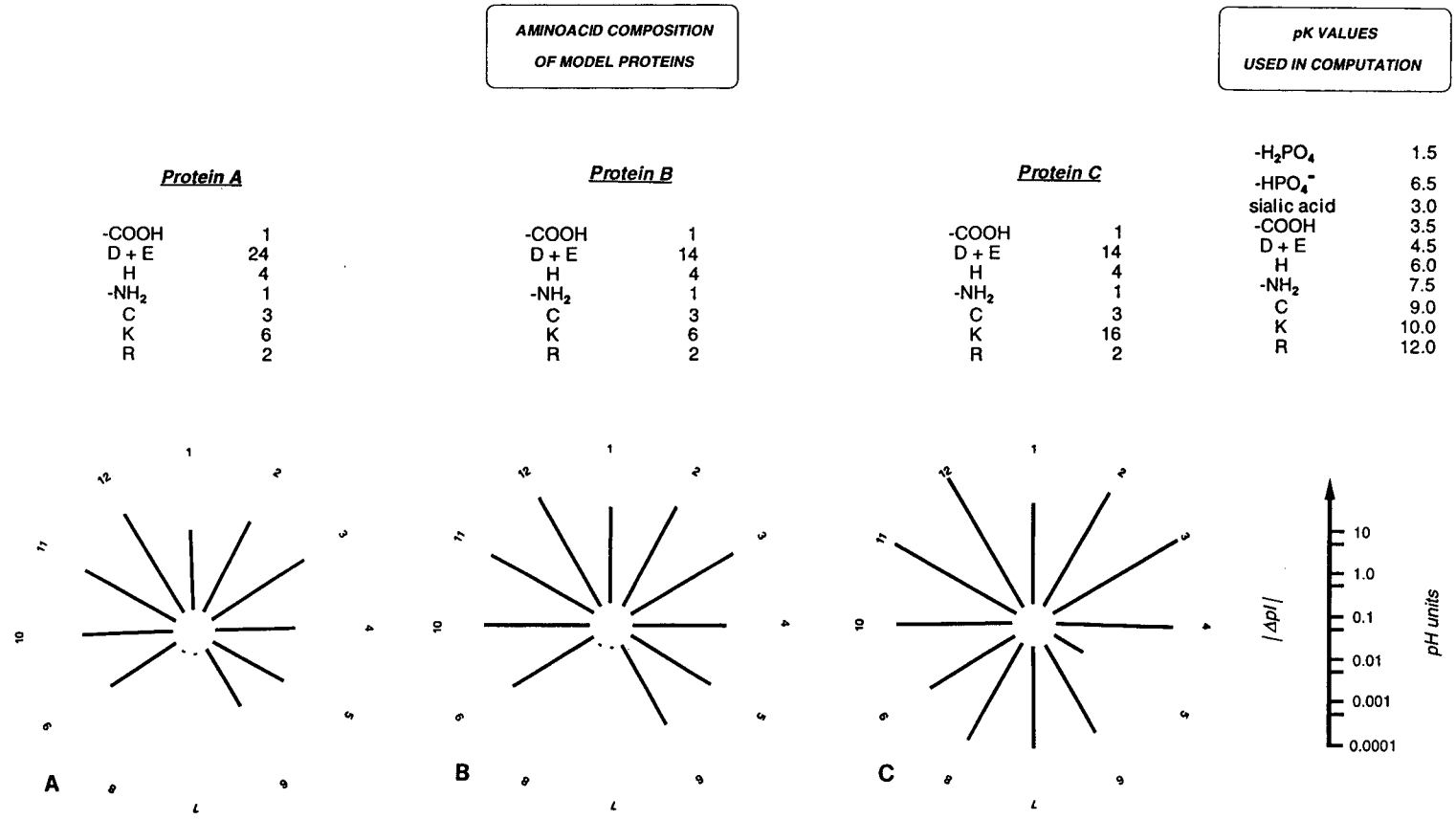


Fig. 1. Expected Δp*I* for various post-translational modifications in different hypothetical protein structures. The isoelectric point for the three model proteins A, B and C, whose amino acid compositions are given in the top panels, was computed by the program described in Ref. [233]. The p*K* values used in the calculations are tabulated on the right. New p*I*s were evaluated after the following changes, corresponding to the simulation of some kind of post-translational modification: case (1) addition of 1, case (2) 3 or case (3) 5 sialic acid residues; case (4) addition of 1 phosphate group; case (5) removal by blocking of NH_2 terminus; case (6) deamidation, i.e., addition of 1 D/E residue; case (7) removal by oxidation of 2 C residues; case (8) removal by oxidation of all C residues resulting in protein dimerization; case (9) removal by carbamylation of 1, case (10) 2, case (11) 4 or, case (12) 6 K residues. The resulting Δp*I*s are represented clockwise according to a logarithmic scale (bottom right) in the diagrams of the lower row. See text for more details.

pH 9, K is not 100% protonated and carbamylation thus involves fractional charge decrements. NH_2 terminus blocking results in a p*I* shift of only 0.006 pH unit, as p*I* is much higher than pK_{NH_2}. Phosphorylation brings about a two unit charge variation, whereas the removal (by oxidation) of each SH group corresponds to one-half charge difference (pK_{SH} equals p*I*).

6. Recombinant proteins

The reports in Table 1 [241–260] are the only ones from the bibliographic search where the issue of charge heterogeneity of recombinant proteins was addressed. In most cases, however, the properties of the biotechnological product were just assessed instead of being compared with the natural product. The chemico-physical parameters of the latter were possibly hardly known in some instances, and gene cloning for heterologous expression was actually a short-cut in comparison with a lengthy and cumbersome purification strategy.

Most of the problems connected with gene expression technology, including co- and post-translational processing, modification and secretion in different heterologous systems, are dis-

Table 1
Comparison between natural and recombinant proteins

Protein	Gene source	Expression system	Correct properties	No glycosylation	Heterogeneous glycosylation	Incorrect properties	Ref.
Myosin light chain	*Drosophila*	In vitro	×				[241]
Acid phosphatase	Man	In vitro	×				[242]
Peptidyl prolyl isomerase	Man	*E. coli*	×				[243]
M protein	Parainfluenza virus	*E. coli*				×	[244]
α-Interferon	Man	*E. coli*		×			[245]
Structural proteins	Semliki Forest virus	Yeast				×	[246]
G protein βγ-subunits	Man	Insect cells	×				[247]
Interferon γ-receptor	Mouse	Insect cells			×		[248]
Manganese peroxidase	*Phanerochaete chrysosporium*	Insect cells				×	[249]
HLA-B27 + E3/19K	Man adenovirus	Insect cells			×	(×)	[250]
Interleukin-9	Man	T-cell lines			×		[251]
T1	Mouse	Fibroblasts			×		[252]
gp130	Man	Mouse melanoma	×				[253]
Fib2 of fibronectin	Man	Mouse L cells			×		[254]
TGF-βIII receptor	Rat	COS cells			×		[255]
Tissue factor	Man	CHO cells			×		[256]
Proenkephalin	Rat	CHO cells			×	(×)	[257]

cussed in Refs. [258] and [259]. From the list, the number of cases for which correct properties are observed equals the entries for which incorrect properties are reported. It is no surprise to find that no glycosylation occurs in *Escherichia coli* [243]. Complete fidelity to the model, as might be expected, is obtained when a human protein is expressed by the same histotype in a different animal species [253]. The glycosylation pattern of virus proteins is known to be host-specific [24]. However, especially interesting is the observation that when expressed in *Saccharomyces cerevisiae* the E1 envelope protein of Semliki forest virus receives yeast-characteristic outer-chain glycans [246]. Phosphorylation is reported in Ref. [257].

Two examples of fusion, or chimeric, proteins found heterogeneous with respect to p*I* and/or to M_r are discussed in Refs. [259] and [260].

Finally, the case of a shift in isoelectric point between cellular and secreted forms of a monoclonal antibody against gp-41 of HIV-1 virus is reported in Ref. [261].

Acknowledgements

I thank (in alphabetical order) Paolo Arosio, Bengt Bjellqvist, Franco Bonomi, Fabrizio Celentano, Pietro Ciceri, Paola Coari, Monica Di Luca, Monica Galliano, Luciana Mussoni, Giuseppe Pollio, Franca Rossini, Angelo Viotti and Mara Zocchi. I am most grateful to the Laboratory Director, Professor Cesare R. Sirtori, and to my colleagues Maria Rosa Lovati and Cristina Manzoni for continuous support.

Abbreviations

CA = carrier ampholyte; CE = capillary electrophoresis; 1- and 2-D = one- and two-dimensional electrophoresis; EPR = electron spin resonance; ER = endoplasmic reticulum; ESI = electrospay ionization; HUVEC = human umbilical cord endothelial cells; IEF = isoelectric point; IMAC = immobilized metal chelate chromatography; IPG = immobilized pH gradient; MHC = major histocompatibility complex; MS = mass spectrometry; NMR = nuclear magnetic resonance; PAI-2 = plasminogen activator inhibitor-2; p*I* = isoelectric point; SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis; TFA = trifluoroacetic acid; Thy-1 = cell-surface glycoprotein, member of the immunoglobulin superfamily; u = electrophoretic mobility.

References

[1] A.F. Bristow, in E.L.V. Harris and S. Angdal (Editors), *Protein Purification Applications: A Practical Approach*, IRL Press, Oxford, 1990, Ch. 2, p. 29.
[2] A. Hershko and M. Fry, *Annu. Rev. Biochem.*, 44 (1975) 775.
[3] W. Wickner, *Annu. Rev. Biochem.*, 48 (1979) 23.
[4] H. Holzer and P.C. Heinrich, *Annu. Rev. Biochem.*, 49 (1980) 63.
[5] M.S. Briggs and L.M. Gierasch, *Adv. Protein Chem.*, 38 (1986) 109.
[6] D. Gerlach, R. Kraft and D. Behnke, *Zentralbl. Bakteriol. Mikrobiol. Hyg.*, 269 (1988) 314.
[7] T. Sato, P.W. Oeller and A. Theologis, *J. Biol. Chem.*, 266 (1991) 3752.
[8] E. Canova-Davis, M. Eng, V. Mukku, D.H. Reifsnyder, C.V. Olson and V.T. Ling, *Biochem. J.*, 285 (1992) 207.
[9] D. Chattopadhyay, D.B. Evans, M.R. Deibel, Jr., A.F. Vosters, F.M. Eckenrode, H.M. Einsphahr, J.O. Hui, A.G. Tomasselli, H.A. Zucher-Neely, R.L. Heinrikson ans S.K. Sharman, *J. Biol. Chem.*, 267 (1992) 14227.
[10] S. Jauris-Heipke, R. Fuchs, M. Motz, V. Preac-Mursic, E. Schwab, E. Soutschek, G. Will and B. Wilske, *Med. Microbiol. Immunol. (Berl.)*, 182 (1993) 37.
[11] S.J. Brewer and H.M. Sassenfeld, in E.L.V. Harris and S. Angdal (Editors), *Protein Purification Applications: A Practical Approach*, IRL Press, Oxford, 1990, p. 91.
[12] A.J. Dorner and R.J. Kaufman, *Methods Enzymol.*, 185 (1990) 577.
[13] A.R. Rees, M.J.E. Sternberg and R. Wetzel (Editors), *Protein Engineering: A Practical Approach*, IRL Press, Oxford, 1992.
[14] A.F. Strachan, W.F. Brandt, P. Woo, D.R. van der Westhuyzen, G.A. Coetzee, M.C. deBeer, E.G. Shephard and F.C. deBeer, *J. Biol. Chem.*, 264 (1989) 18368.
[15] J.J. Billadello, D.R. Roman, A.M. Grace, B.E. Sobel and A.W. Strauss, *J. Biol. Chem.* 260 (1985) 14988.
[16] F. Wold, *Annu. Rev. Biochem.*, 50 (1981) 783.
[17] F. Wold and K. Moldave (Editors), *Methods in Enzymology*, Vol. 106, Academic Press, New York, 1984.

[18] F. Wold and K. Moldave (Editors), *Methods in Enzymology*, Vol. 107, Academic Press, New York, 1984.
[19] J.J. Harding, *Adv. Protein Chem.*, 37 (1985) 247.
[20] T. Imoto and H. Yamada, in T.E. Creighton (Editor), *Protein Function: A Practical Approach*, IRL Press, Oxford, 1989, p. 247.
[21] E.T. Kaiser, D.S. Lawrence and S.E. Rokita, *Annu. Rev. Biochem.*, 54 (1985) 565.
[22] G.K. Ackers and F.R. Smith, *Annu. Rev. Biochem.*, 54 (1985) 597.
[23] M. Hollecker, in T.E. Creighton (Editor), *Protein Structure: A Practical Approach*, IRL Press, Oxford, 1989, p. 145.
[24] T.W. Rademacher, R.B. Parekh and R.A. Dwek, *Annu. Rev. Biochem.*, 57 (1988) 785.
[25] K.W. Adolph and M.K. Song, *Biochemistry*, 24 (1985) 345.
[26] S.K. Howlett, *Cell*, 45 (1986) 387.
[27] D.A. Knecht, E.D. Green, W.F. Loomis and R.L. Dimond, *Dev. Biol.*, 107 (1985) 490.
[28] G.P. Livi, N.A. Woychik and R.L. Dimond, *Differentiation*, 30 (1985) 83.
[29] J.M. Edelberg, J.J. Enghild, S.V. Pizzo and M. Gonzalez-Gronow, *J. Clin. Invest.*, 86 (1990) 107.
[30] M. Takahashi, H. Homma and M. Matsui, *Biochem. J.*, 293 (1993) 795.
[31] K. Reske, W. Ballhausen, W. Henkes, B. Opel, U. Machein, A. Reske-Kunz, F.J. Schneider and P. Steinlein, *Behring Inst. Mitt.*, 81 (1987) 46.
[32] D.R. Joseph, W. Lawrence and B.J. Danzo, *Mol. Endocrinol.*, 6 (1992) 1127.
[33] G.V. Avvakumov and O.A. Strel'chyonok, *Biochim. Biophys. Acta*, 925 (1987) 11.
[34] R.H. Lustig, D.W. Pfaff and C.V. Mobbs, *Endocrinology*, 124 (1989) 1863.
[35] C.M. Cheney, K.G. Miller, T.J. Lang and A. Shearn, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 6422.
[36] C.H. Park, K. Artzt and D. Bennett, *Dev. Genet.*, 10 (1989) 53.
[37] W.G. McLean, C. Pekiner, N.A. Cullum and I.F. Casson, *Mol. Neurobiol.*, 6 (1992) 225.
[38] D. Gautreau, S.R. Zetlan, G.D. Mazur and M.R. Goldsmith, *Dev. Biol.*, 157 (1993) 60.
[39] J.E. Zahner and C.M. Cheney, *Mol. Cell. Biol.*, 13 (1993) 217.
[40] K.M. Verbanac and E.C. Heath, *J. Biol. Chem.*, 261 (1986) 9979.
[41] D.T. Curiel, C. Vogelmeier, R.C. Hubbard, L.E. Stier and R.G. Crystal, *Mol. Cell. Biol.*, 10 (1990) 47.
[42] G.R. Guy, R. Philip and Y.H. Tan, *Electrophoresis*, 15 (1994) 417–440.
[43] D.F. Hochstrasser and J.-D. Tissot, *Adv. Electr.*, 6 (1993) 267.
[44] P.F. Lemkin, P. Sonderegger and L.E. Lipkin, *Clin. Chem.*, 30 (1984) 1965.
[45] G. Wagner, *Prog. Nucl. Magn. Reson. Spectrosc.*, 22 (1990) 101.
[46] G.C.K. Roberts (Editor), *NMR of Macromolecules: A Practical Approach*, IRL Press, Oxford, 1993.
[47] G.O. Daumy, D. Delgarno, A.S. McColl, J.M. Merenda and G.R. Schulte, *Biochim. Biophys. Acta*, 967 (1988) 326.
[48] W.D. Pickering, *LC · GC*, 7 (1989) 752.
[49] I.-J. Chang, H.B. Gray and M. Albin, *Anal. Biochem.*, 212 (1993) 24.
[50] J.A. McCloskey (Editor), *Methods in Enzymology*, Vol. 193, Academic Press, New York, 1990.
[51] S.A. Carr, M.E. Hemling, M.F. Bean and G.D. Roberts, *Anal. Chem.*, 63 (1991) 2802.
[52] B.T. Chait and S.B.H. Kent, *Science*, 257 (1992) 1885.
[53] P. Roepstorff, *Trends Anal. Chem.*, 12 (1993) 413.
[54] S.A. Carr and K. Biemann, *Methods Enzymol.*, 106 (1984) 29.
[55] T.P. Knepper, U. Arbogast, J. Schreurs and M.L. Deinzer, *Biochemistry*, 31 (1992) 11651.
[57] C. Tanford, *Adv. Protein Chem.*, 17 (1962) 69.
[58] Y. Nozaki and C. Tanford, *Methods Enzymol.*, 11 (1967) 715.
[59] A. Rosengren, B. Bjellqvist and V. Gasparic, in B.J. Radola and D. Graesslin (Editors), *Electrofocusing and Isotachophoresis*, Walter Gruyter, Berlin, 1977, p. 165.
[60] P.G. Righetti, R. Krishnamoorthy, C. Lapoumeroulie and D. Labie, *J. Chromatogr.*, 177 (1979) 219.
[61] L. Valentini, E. Gianazza and P.G. Righetti, *J. Biochem. Biophys. Methods*, 3 (1980) 323.
[62] P.G. Righetti, G. Gacon, E. Gianazza, D. Lostanlen and J.C. Kaplan, *Biochem. Biophys. Res. Commun.*, 85 (1978) 1575.
[63] D. Lonstanlen, G. Gacon and J.C. Kaplan, *Eur. J. Biochem.*, 112 (1980) 179.
[64] E. Gianazza and P. Arnaud, *Electrophoresis*, 2 (1981) 247.
[65] R. Krishnamoorthy, A. Bianchi-Bosisio, D. Labie and P.G. Righetti, *FEBS Lett.*, 94 (1978) 319.
[66] J. Constans, M. Viau, C. Gouaillard, C. Bouissou and A. Clerc, in B.J. Radola (Editor), *Electrophoresis '79*, Walter de Gruyter, Berlin, 1980, p. 701.
[67] K. Ek and P.G. Righetti, *Electrophoresis*, 1 (1980) 137.
[68] K. Ek, E. Gianazza and P.G. Righetti, *Biochim. Biophys. Acta*, 626 (1980) 356.
[69] P.G. Righetti, R. Krishnamoorthy, E. Gianazza and D. Labie, *J. Chromatogr.*, 166 (1978) 455.
[70] E. Gianazza and P. Arosio, *Biochim. Biophys. Acta*, 625 (1980) 310.
[71] L.E. Erickson and R.A. Alberty, *Methods Enzymol.*, 6 (1963) 895.
[72] A. Warshel and J. Åqvist, *Annu. Rev. Biophys. Biophys. Chem.*, 20 (1991) 267.
[73] S.T. Russell and A. Warshel, *J. Mol. Biol.*, 185 (1985) 389.
[74] M.K. Gilson and B.H. Honig, *Nature*, 330 (1987) 84.
[75] M.J.E. Sternberg, F.R.F. Hayes, A.J. Russell, P.G. Thomas and A.R. Fersht, *Nature*, 330 (1987) 86.

[76] A. Warshel, F. Sussman and J.-K. Hwang, *J. Mol. Biol.*, 201 (1988) 139.
[77] T. Takahashi, H. Nakamura and A. Wada, *Biopolymers*, 32 (1992) 897.
[78] S.P.L. So/rensen, K. Linderstro/m-Lang and E. Lund, *C. R. Trav. Lab. Carlsberg, Sér. Chim.*, 16 (1926) 5.
[79] R.K. Cannan, *Chem. Rev.*, 30 (1942) 295.
[80] R.A. Alberty, in A.H. Neurath and K. Bailey (Editors), *The Proteins: Chemistry, Biological Activity and Methods*, Vol. I, Academic Press, New York, 1953, p. 461.
[81] J.B. Whitney, III, R.R. Cobb, R.A. Popp and T.W. O'Rourke, *Proc. Natl. Acad. Sci. U.S.A.*, 82 (1985) 7646.
[82] P.G. Righetti and G. Cossu, *Trends Anal. Chem.*, 5 (1986) 147.
[83] U. McKusik, *Mendelian Inheritance in Man*, Johns Hopkins University Press, Baltimore, 1992, p. 92.
[84] A. Görg, W. Postel, J. Weser, W. Patutschnick and H. Cleve, *Am. J. Hum. Genet.*, 37 (1985) 922.
[85] P.G. Righetti, C. Ettori, P. Chafey and J.P. Wahrmann, *Electrophoresis*, 11 (1990) 1.
[86] G. Saglio, G. Ricco, U. Mazza, C. Camaschella, P.G. Pich, A.M. Giani, E. Gianazza, P.G. Righetti, B. Giglioni, P. Comi, M. Gusmeroli and S. Ottolenghi, *Proc. Natl. Acad. Sci. U.S.A.*, 76 (1979) 3420.
[87] K. Altland, P. Becher and A. Banzhoff, *Electrophoresis*, 8 (1987) 293.
[88] H. Ribe, *Ann. N.Y. Acad. Sci.*, 209 (1973) 11.
[89] H. Svenson, *Acta Chem. Scand.*, 15 (1961) 325.
[90] P.G. Righetti, *Isoelectric Focusing: Theory, Methodology and Applications*, Elsevier, Amsterdam, 1983.
[91] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Görg, W. Postel and R. Westermeier, *J. Biochem. Biophys. Methods*, 6 (1982) 317.
[92] P.G. Righetti, *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam, 1990.
[93] E. Gianazza, P. Giacon, B. Sahlin and P.G. Righetti, *Electrophoresis*, 6 (1985) 53.
[94] P.G. Righetti, K. Ek and B. Bjellqvist, *J. Chromatogr.*, 291 (1984) 31.
[95] P.G. Righetti, E. Gianazza and B. Bjellqvist, *J. Biochem. Biophys. Methods*, 8 (1983) 89.
[96] E. Gianazza, G. Artoni and P.G. Righetti, *Electrophoresis*, 4 (1983) 321.
[97] B. Bjellqvist, C. Pasquali, F. Ravier, J.C. Sanchez and D. Hochstrasser, *Electrophoresis*, 14 (1993) 1357.
[98] P.G. Righetti, A. Morelli, C. Gelfi and R. Westermeier, *J. Biochem. Biophys. Methods*, 13 (1986) 151.
[99] W.J. Gelsema and C.L. deLigny, *J. Chromatogr.*, 130 (1977) 41.
[100] S. Fredriksson, in B.J. Radola and D. Graesslin (Editors), *Isoelectric Focusing and Isotachophoresis*, Walter de Gruyter, Berlin, 1977, p. 71.
[101] A. Görg, W. Postel, C. Friedrich, R. Kuick, J.R. Strahler and S.M. Hanash, *Electrophoresis*, 12 (1991) 653.
[102] W.J. Gelsema, C.L. deLigny and N.G. van der Veen, *J. Chromatogr.*, 149 (1977) 149.
[103] W.J. Gelsema, C.L. deLigny and N.G. van der Veen, *J. Chromatogr.*, 171 (1979) 171.
[104] N. Ui, *Ann. N.Y. Acad. Sci.*, 209 (1973) 198.
[105] D.P. Goldenberg and T.E. Creighton, *Anal. Biochem.*, 138 (1984) 1.
[106] B. Bjellqvist, G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.-C. Sanchez, S. Frutiger and D. Hochstrasser, *Electrophoresis*, 14 (1993) 1023.
[107] B. Bjellqvist, B. Sasse, E. Olsen and J.E. Celis, *Electrophoresis*, 15 (1994) 529.
[108] H.G. Ljunggren, M. Oudshoorn-Snoek, M.G. Masucci and H.L. Ploegh, *Immunogenetics*, 32 (1990) 440.
[109] B.C. Johnson (Editor), *Posttranslational Covalent Modifications of Proteins*, Academic Press, New York, 1983.
[110] A.A. vanLandeghem, J.B. Soons, R.A. Wever, M.W. Mul-Steinbusch and T. Antonissen-Zijda, *Clin. Chim. Acta*, 153 (19850 217.
[111] M. Fukuda and A. Kobata (Editors), *Glycobiology: A Practical Approach*, IRL Press, Oxford, 1993.
[112] R.S. Spiro, *Adv. Protein Chem.*, 27 (1973) 349.
[113] R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 45 (1976) 217.
[114] C.B. Hirshberg and M.D. Snider, *Annu. Rev. Biochem.*, 56 (1987) 63.
[115] G.W. Hart, R.S. Haltiwanger, G.D. Holt and W.G. Kelly, *Annu. Rev. Biochem.*, 58 (1989) 841.
[116] R.B. Parekh, A.G.D. Tse, R.A. Dwek, A.F. Williams and T.W. Rademacher, *EMBO J.*, 6 (1987) 1233.
[117] K. Altland, T. Roeder, H.M. Jakin, H.-G. Zimmer and V. Neuhoff, *Clin. Chem.*, 28 (1982) 1000.
[118] B. Silvestrini, A. Guglielmotti, L. Saso and C.Y. Cheng, *Clin. Chem.*, 35 (1989) 2207.
[119] J.R. Bartles, S. Khuon, X.H. Lin, L.Q. Zhang, J.K. Reddy, M.S. Rao, S.T. Isoye, C.L. Nehme and B.E. Fayos, *Cancer Res.*, 50 (1990) 669.
[120] P. Aisen and I. Listowsky, *Annu. Rev. Biochem.*, 49 (1980) 357.
[121] D. Belin, *Thromb. Haemost.*, 70 (1993) 144.
[122] G. Ashwell and J. Harford, *Annu. Rev. Biochem.*, 51 (1982) 531.
[123] J. Lechner and F. Wieland, *Annu. Rev. Biochem.*, 58 (1989) 173.
[124] K. Schmid and J.P. Binette, *Nature*, 190 (1961) 630.
[125] S. Weidinger, T. Müller, F. Schwarzfischer and H. Cleve, *Hum. Genet.*, 77 (1987) 286.
[126] H. Asao, T. Takeshita, M. Nakamura, K. Nagata and K. Sugamura, *Int. Immunol.*, 2 (1990) 469.
[127] W.R. Hudgins, B. Hampton, W.H. Burgess and J.F. Perdue, *J. Biol. Chem.*, 267 (1992) 8153.
[128] H. Lis and N. Sharon, *Annu. Rev. Biochem.*, 55 (1986) 35.
[129] R. Hawkes, *Anal. Biochem.*, 123 (1982) 143.
[130] R.A. Dwek, C.J. Edge, D.J. Harvey and M.R. Wormald, *Annu. Rev. Biochem.*, 62 (1993) 65.

[131] H. Wang, S.J. Segal and S.S. Koide, *Endocrinology*, 123 (1988) 795.
[132] K. Ikuta, K. Nakajima, S. Ueda, S. Kato and K. Irai, *J. Gen. Virol.*, 66 (1985) 1131.
[133] K.J. Colley and J.U. Baezinger, *J. Biol. Chem.*, 262 (1987) 3415.
[134] M.M. Hussain and V.I. Zannis, *Biochemistry*, 29 (1990) 209.
[135] S. van Weely, J.M. Aerts, M.B. van Leeuwen, J.C. Heikoop, W.E. Barranger, J.M. Tager and A.W. Schram, *Eur. J. Biochem.*, 191 (1990) 669.
[136] H. Rouhandeh and R. Cohrs, *J. Virol.*, 52 (1984) 1005.
[137] Y. Bourbonnais, S. Fortin and P. Crine, *Biochem. Cell Biol.*, 64 (1986) 1262.
[138] A.D. Elbein, *Annu. Rev. Biochem.*, 56 (1987) 497.
[139] N.I. Perrone-Bizzozero and L.I. Benowitz, *J. Neurochem.*, 48 (1987) 644.
[140] M. Taverna, A. Baillet, D. Biou, M. Schlüter, R. Werner and D. Ferrier, *Electrophoresis*, 13 (1992) 359.
[141] G. Taborsky, *Adv. Protein Chem.*, 28 (1974) 1.
[142] C.S. Rubin and O.M. Rosen, *Annu. Rev. Biochem.*, 44 (1975) 831.
[143] T. Hunter and B.M. Sefton (editors), *Methods in Enzymology*, Vol. 200, Academic Press, New York, 1991.
[144] T. Hunter and B.M. Sefton (editors), *Methods in Enzymology*, Vol. 201, Academic Press, New York, 1991.
[145] D.G. Hardie (editor), *Protein Phosphorylation: A Practical Approach*, IRL Press, Oxford, 1993.
[146] E.G. Krebs and J.A. Beavo, *Annu. Rev. Biochem.*, 48 (1979) 923.
[147] B.M. Sefton and M.-A. Campbell, *Annu. Rev. Cell Biol.*, 7 (1991) 257.
[148] E.J. Nestler and P. Greengard, *Protein Phosphorylation in the Nervous System*, Wiley, New York, 1984.
[149] A. Curtis, V. Lyons and G. Fink, *J. Endocrinol.*, 105 (1985) 163.
[150] M.M. Black, J.M. Aletta and L.A. Greene, *J. Cell Biol.*, 103 (1986) 545.
[151] P. Chretien and J. Landry, *J. Cell. Physiol.*, 137 (1988) 157.
[152] R. Balhorn and R. Chalkley, *Methods Enzymol.*, 40 (1989) 138.
[153] W.F. Patton, M.R. Dhanak and B.S. Jacobson, *Electrophoresis*, 11 (1990) 79.
[154] K.J. Koller, M.T. Lipari and D.V. Goeddel, *J. Biol. Chem.*, 268 (1993) 5997.
[155] J.F. Santaren, J.C. Ramirez and J.M. Almendral, *J. Virol.*, 67 (1993) 5126.
[156] T. Hunter and M. Karin, *Cell*, 70 (1992) 375.
[157] H.J. Xu, S.X. Hu, T. Hashimoto, R. Takahashi and W.F. Benedict, *Oncogene*, 4 (1989) 807.
[158] S.C. Francesconi and S. Eisenberg, *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 4089.
[159] B. Suter and R. Steward, *Cell*, 67 (1991) 917.
[160] I. Sadowski, D. Niedbala, K. Wood and M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 10510.
[161] S. Shenolikar and T.S. Ingebritsen, *Methods Enzymol.*, 107 (1984) 102.
[162] P. Cohen, *Annu. Rev. Biochem.*, 58 (1989) 453.
[163] P.J. Roach, *Methods Enzymol.*, 107 (1984) 81.
[164] T. Hunter and J.A. Cooper, *Annu. Rev. Biochem.*, 54 (1985) 897.
[165] A.M. Edelman, D.K. Blumenthal and E.G. Krebs, *Annu. Rev. Biochem.*, 56 (1987) 567.
[166] L.A. Pon, J.A. Hartigan and N.R. Orme-Johnson, *J. Biol. Chem.*, 261 (1986) 13309.
[167] V. Skibeli, O. Anderson and K.M. Gautvik, *Gen. Comp. Endocrinol.*, 80 (1990) 333.
[168] P.M. Grob, A.H. Ross, H. Koprowski and M. Bothwell, *J. Biol. Chem.*, 260 (1985) 8044.
[169] G.S. Bennett and C. DiLullo, *J. Cell Biol.*, 100 (1985) 1799.
[170] T.M. Martensen, *Methods Enzymol.*, 107 (1984) 3.
[171] M. Brauer and B.D. Sykes, *Methods Enzymol.*, 107 (1984) 36.
[172] P. van der Geer and T. Hunter, *Electrophoresis*, 15 (1994) 544.
[173] H. Hirano, S. Komatsu, H. Kajiwara, Y. Takagi and S. Tsunasawa, *Electrophoresis*, 14 (1993) 839.
[174] S. Tsunasawa and F. Sakiyama, *Methods Enzymol.*, 106 (1984) 165.
[175] V.G. Allfrey, E.A. DiPaola and R. Sterner, *Methods Enzymol.*, 107 (1984) 224.
[176] A.M. Schultz, L.E. Henderson and S. Oroszlan, *Annu. Rev. Cell Biol.*, 4 (1988) 611.
[177] D.A. Towler, J.I. Gordon, S.P. Adams and L. Glaser, *Annu. Rev. Biochem.*, 57 (1988) 69.
[178] N.M. Hooper and A.J. Turner (Editors), *Lipid Modification of Proteins: A Practical Approach*, IRL Press, Oxford, 1992.
[179] I. Marie, J. Svab, N. Robert, J. Galabru and A.G. Hovanessian, *J. Biol. Chem.*, 265 (1990) 18601.
[180] M. Vai, L. Popolo, R. Grandori, E. Lacana and L. Alberghina, *Biochim. Biophys. Acta*, 1038 (1990) 277.
[181] D.K. Vorbroker, C. Dey, T.E. Weaver and J.A. Whitsett, *Biochim. Biophys. Acta*, 1105 (1992) 161.
[182] I. Isenberg, *Annu. Rev. Biochem.*, 48 (1979) 159.
[183] G. Brosch, E.I. Georgieva, G. López-Rodas, H. Lindner and P. Loidl, *J. Biol. Chem.*, 267 (1992) 20561.
[184] P.G. Righetti, M. Delpech, F. Moisand, J. Kruh and D. Labie, *Electrophoresis*, 4 (1983) 393.
[185] R.W. Lennox and L.H. Cohen, *Methods Enzymol.*, 170 (1988) 533.
[186] H. Lindner, W. Helliger, A. Dirschlmayer, M. Jaquemar and B. Puschendorf, *Biochem. J.*, 283 (1992) 467.
[187] R.J. Koenig, S.H. Blobstein and A. Cerami, *J. Biol. Chem.*, 252 (1977) 2992.
[188] S. Rahbar, *Clin. Chim. Acta*, 22 (1968) 296.
[189] K.H. Gabbay, *N. Engl. J. Med.*, 295 (1976) 443.
[190] M. Brownlee and A. Cerami, *Annu. Rev. Biochem.*, 50 (1981) 385.

[191] J.O. Jeppson, B. Franzen and V.O. Nilsson, *Sci. Tools*, 25 (1978) 69.
[192] L. Beccaria, G. Chiumello, E. Gianazza, B. Luppis and P.G. Righetti, *Am. J. Haematol.*, 4 (1978) 367.
[193] R. Flückiger and P.M. Gallop, *Methods Enzymol.*, 106 (1984) 77.
[194] G. Candiano, G.M. Ghiggeri, G. Delfino, C. Queirolo, E. Gianazza and P.G. Righetti, *Electrophoresis*, 5 (1984) 217.
[195] H.M. Liebich, E. Gesele, C. Wirth, J. Wöll, K. Jobst and A. Lakatos, *Biol. Mass Spectrom*, 22 (1993) 121.
[196] J. Houmard and G.R. Drapeau, *Proc. Natl. Acad. Sci. U.S.A.*, 69 (1972) 306.
[197] R.J. Beynon and J.S. Bond (Editors), *Proteolytic Enzymes: A Practical Approach*, IRL Press, Oxford, 1989.
[198] K.L. Carraway and D.E. Koshland, *Methods Enzymol.*, 25 (1972) 616.
[199] G. Kreil, *Methods Enzymol.*, 106 (1984) 218.
[200] S. Clarke, *Annu. Rev. Biochem.*, 61 (1992) 355.
[201] L. Gutierrez, A.I. Magee, C.J. Marshall and J.F. Hancock, *EMBO J.*, 8 (1989) 1093.
[202] S.M. Hurtley and A. Helenius, *Annu. Rev. Cell Biol.*, 5 (1989) 277.
[203] D.B. Wetlaufer, *Methods Enzymol.*, 107 (1984) 301.
[204] C. Tanford, *Adv. Protein Chem.*, 23 (1968) 121.
[205] T.E. Chreighton, *Methods Enzymol.*, 107 (1984) 305.
[206] D.A. Hillson, N. Lambert and R.E. Freedman, *Methods Enzymol.*, 107 (1984) 281.
[207] H.F. Gilbert, *Methods Enzymol.*, 107 (1984) 330.
[208] P.G. Righetti, E. Gianazza, A.M. Gianni, P. Comi, B. Giglioni, S. Ottolenghi, C. Secchi and L. Rossi-Bernardi, *J. Biochem. Biophys. Methods*, 1 (1979) 47.
[209] E. Gianazza and P. DePonti, *Electrophoresis*, 14 (1993) 1259.
[210] K. Altland and U. Rossman, *Electrophoresis*, 6 (1985) 314.
[211] P.G. Righetti, B. Barzaghi, E. Sarubbi, A. Soffientini and G. Cassani, *J. Chromatogr.*, 470 (1989) 337.
[212] P.G. Righetti, G. Tudor and E. Gianazza, *J. Biochem. Biophys. Methods*, 6 (1982) 219.
[213] M. Chiari, C. Chiesa, P.G. Righetti, M. Corti, J. Tikam and R. Shorr, *J. Chromatogr.*, 499 (1990) 699.
[214] K. Altland, P. Becher, U. Rossman and B. Bjellqvist, *Electrophoresis*, 9 (1988) 474.
[215] L.C. Lane, *Anal. Biochem.*, 86 (1978) 655.
[216] A. Görg, W. Postel, J. Weser, S. Günther, J.R. Strahler, S.M. Hanash and L. Somerlot, *Electrophoresis*, 8 (1987) 122.
[217] T.E. Chreighton, *Nature*, 284 (1980) 487.
[218] N.L. Anderson and B.J. Hickman, *Anal. Biochem.*, 93 (1979) 312.
[219] S.L. Tollaksen, J.J. Edwards and N.G. Anderson, *Electrophoresis*, 2 (1981) 155.
[220] J. Taylor, N.L. Anderson and N.G. Anderson, *Electrophoresis*, 4 (1983) 338.
[221] E. Gianazza, F. Celentano, G. Dossi, B. Bjellqvist and P.G. Righetti, *Electrophoresis*, 5 (1984) 88.
[222] E. Gianazza, S. Astrua-Testori and P.G. Righetti, *Electrophoresis*, 6 (1985) 113.
[223] E. Gianazza, F. Celentano, S. Magenes, C. Ettori and P.G. Righetti, *Electrophoresis*, 10 (1989) 806.
[224] E. Gianazza, S. Astrua-Testori, P. Caccia, P. Giacon, L. Quaglia and P.G. Righetti, *Electrophoresis*, 7 (1986) 76.
[225] N.H. Axelsen, J. Kro/ll and B. Weeke (Editors), *A Manual of Quantitative Immunoelectrophoresis: Methods and Applications*, Universitetsforlaget, Oslo, 1973.
[226] D. Bobb, *Ann. N.Y. Acad. Sci.*, 209 (1973) 225.
[227] S.W. Whitehart, P. Shenbagamurthi, L. Chen, R.J. Cotter and G.W. Hart, *J. Biol. Chem.*, 264 (1989) 14334.
[228] W.K. Paik, *Methods Enzymol.*, 106 (1984) 265.
[229] W.K. Paik and P. DiMaria, *Methods Enzymol.*, 106 (1984) 274.
[230] W.K. Paik, Y.B. Cho, B. Frost and S. Kim, *Biochem. Cell Biol.*, 67 (1989) 602.
[231] K. Ueda and O. Hayaishi, *Annu. Rev. Biochem.*, 54 (1985) 73.
[232] K. Wielckens, R. Bredehorst and H. Hilz, *Methods Enzymol.*, 106 (1984) 472.
[233] D. Finley and V. Chau, *Annu. Rev. Cell Biol.*, 7 (1991) 25.
[234] H. Busch, *Methods Enzymol.*, 106 (1984) 238.
[235] H. Lindner, J. Wesierska-Gadek, W. Helliger, B. Puschendorf and G. Saubermann, *J. Chromatogr.*, 472 (1989) 243.
[236] K. Shimogawara and S. Muto, *Arch. Biochem. Biophys.*, 294 (1992) 193.
[237] Y. Morino and F. Nagashima, *Methods Enzymol.*, 106 (1984) 116.
[238] T.P. Singer and W.S. McIntire, *Methods Enzymol.*, 106 (1984) 369.
[239] P.G. Righetti and J.W. Drysdale, *Ann. N.Y. Acad. Sci.*, 209 (1973) 163.
[240] F. Celentano, E. Gianazza, G. Dossi and P.G. Righetti, *Chemometrics*, 1 (1987) 349.
[241] J. Toffenetti, D. Mischke and M.L. Pardue, *J. Cell Biol.*, 104 (1987) 19.
[242] G. Paradis, J.Y. Dube, P. Chapdelaine and R.R. Tremblay, *Biochem. Cell Biol.*, 65 (1987) 921.
[243] T.F. Holzman, D.A. Egan, R. Edalji, R.L. Simmer, R. Helfrich, A. Taylor and N.S. Burres, *J. Biol. Chem.*, 266 (1991) 2474.
[244] M. Kawano, H. Bando, S. Ohgimoto, K. Okamoto, K. Kondo, M. Tsurudome, M. Nishio and Y. Ito, *Virology*, 179 (1990) 857.
[245] G.R. Adolf, I. Kalsner, H. Ahorn, I. Maurer-Fogy and K. Cantell, *Biochem. J.*, 276 (1991) 511.
[246] S. Keranen, *Gene*, 48 (1986) 267.
[247] J.D. Robishaw, V.K. Kalman and K.L. Proulx, *Biochem. J.*, 286 (1992) 677.
[248] M. Fountoulakis, E.J. Schlaeger, R. Gentz, J.F. Juranville, M. Manneberg, L. Ozmen and G. Garotta, *Eur. J. Biochem.*, 198 (1991) 441.

[249] E.A. Pease, S.D. Aust and M. Tien, *Biochem. Biophys. Res. Commun.*, 179 (1991) 897.
[250] F. Levy and S. Kvist, *Int. Immunol.*, 2 (1990) 995.
[251] Y.C. Yang, S. Ricciardi, A. Ciarletta, J. Calvetti, K. Kelleher and S.C. Clark, *Blood*, 74 (1989) 1880.
[252] A.K. Werenskiold, *Eur. J. Biochem.*, 204 (1992) 1041.
[253] A.P. Albino, L.H. Graf, Jr., R.R. Kantor, S. Silagi and L.J. Old, *Mol. Cell. Biol.*, 5 (1985) 692.
[254] K. Ichihara-Tanaka, K. Titani and K. Sekiguchi, *J. Biol. Chem.*, 265 (1990) 401.
[255] X.F. Wang, H.Y. Lin, E. Ng-Eaton, J. Downward, H.F. Lodish and R.A. Weinberg, *Cell*, 67 (1991) 797.
[256] A. Rehemtulla, M. Pepe and T.S. Edgington, *Thromb. Haemost.*, 65 (1991) 521.
[257] I. Lindenberg, E. Shaw, J. Finley, D. Leone and P. Deininger, *Endocrinology*, 128 (1991) 1849.
[258] D.V. Goeddel (Editor), *Methods in Enzymology*, Vol. 185, Academic Press, New York, 1990.
[259] M.W. Wathen, R.J. Brideau, D.R. Thomsen and B.R. Murphy, *J. Gen. Virol.*, 70 (1989) 2625.
[260] S. Mullner, B. Karbe-Thonges and D. Tripier, *Anal. Biochem.*, 210 (1993) 366.
[261] E. Wenish, S. Reiter, S. Hinger, F. Steindl, C. Tauere, A. Jungbauer, H. Katinger and P.G. Righetti, *Electrophoresis*, 11 (1990) 966.

ELSEVIER

Journal of Chromatography A, 705 (1995) 89–103

JOURNAL OF CHROMATOGRAPHY A

Review

Fluorophore-assisted carbohydrate electrophoresis Technology and applications

Guo-Fu Hu

Center for Biochemical and Biophysical Sciences and Medicine, Harvard Medical School, S.G. Mudd Building, 250 Longwood Avenue, Boston, MA 02115, USA

Abstract

Carbohydrates, in particular the complex carbohydrates conjugated to proteins and lipids, have important functions in a variety of biological systems. Their isolation and structural determination–prerequisites for elucidation of their biological functions–have been technical challenges for many decades. Almost all available chromatographic and electrophoretic methods as well as NMR and MS have been applied to carbohydrate analysis but none has proved satisfactory in terms of simplicity, sensitivity, reproducibility, cost and requirement for materials. Recently, a technique called fluorophore-assisted carbohydrate electrophoresis was developed which is very promising. It separates fluorescently-labeled carbohydrates on polyacrylamide gels and uses a charge-coupled device camera to detect and quantitate the products. This review describes the principles of the method and its applications to several aspects of research on carbohydrate-containing biological biomolecules.

Contents

1. Introduction

Carbohydrates are ubiquitous in nature and found in all living species. They are central to energy generation and storage, on the one hand, and provide mechanical support for cells, on the other. Thus, the oxidation of carbohydrates, $nO_2 + (CH_2O)_n \rightarrow nCO_2 + nH_2O$, is the primary metabolic energy-generating process throughout the biosphere, and the cellulose of

0021-9673/95/$29.00
SSDI 0021-9673(94)001252-0

woody plants, the cell walls of bacteria, and the exoskeletons of insects and crustacea are all carbohydrates. Indeed, cellulose and chitin are the two most abundant biopolymers on earth. Complex carbohydrates have been shown to be bioactive and to participate in cellular and biochemical interactions in a wide variety of biological systems [1]. For example, the carbohydrate moieties of glycoproteins affect both their physicochemical and biological properties including folding and three-dimensional structure, solubility and stability, circulatory life time [2,3], susceptibility to proteases [4], biological activity and its modulation [5], molecular and cellular recognition [6], immunogenicity [7], cell adhesion [8], egg fertilization [9], lymphocyte homing [10] and bacterial and viral interactions with host tissues [11,12]. Furthermore, carbohydrates in recombinant proteins confer important activities on, as well as influence the stability and bioavailability of, genetically engineered glycoprotein drugs [13–17]. More than a dozen glycoproteins of therapeutic interest have been produced by recombinant DNA techniques [18]. Impressive examples are glycocerebrosidase for the treatment of Gaucher's disease [19,20], erythropoietin for the treatment of anaemia in hemodialysis patients [21] and tissue plasminogen activator as a thrombolytic agent [22]. More recently, the recruitment of leukocytes to injured tissue was found to occur via interaction of its cell surface carbohydrate Sialyl-Lewis [Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc] with endothelial cell selectins which are expressed in response to cytokines released during the inflammatory response [23–25]. This discovery has created intense interest in the development of sugar-based anti-inflammatory and anti-tumor drugs [26,27]. In addition, changes in protein glycosylation have been shown to be useful molecular markers in the diagnosis of several human diseases [28–34]. With the increased awareness of the biological significance of protein glycosylation, there has been a growing demand for rapid, convenient and cost-effective analytical methods to characterize both the carbohydrate content and chemical structure of the carbohydrate moieties of glycoproteins.

2. Current methodology

The complete characterization of a glycoprotein requires the analysis of the primary structure and conformation of both the carbohydrate side chains and the protein to which they are attached; and identification of the glycosylation site(s), the anomeric specificity of the linkage and the pattern of carbohydrate heterogeneity at each glycosylation site. The characterization of the protein part is relatively easy owing to the availability of well-established methods, techniques and equipment, whereas full elucidation of carbohydrate structures —historically tedious and laborious— continues to be a challenge. Numerous sugar chain variations can be formed from a small number of monosaccharide units. For example, two amino acids can form only two dipeptides, but two monosaccharides can form as many as 32 disaccharides since the linkage can occur at any of the four hydroxyl groups per monosaccharide, exist in either of two anomeric forms and involve either furanose or pyranose rings. The number of isomers increases geometrically as the number of constituents increases since branching becomes a possibility. In addition, because of the lack of template in their biosynthesis, sugar chains are microheterogeneous even at a single glycosylation site, and very often glycoproteins have more than one glycosylation site. In addition to single glucose and N-acetyl glucosamine moieties which are linked to Lys and Ser/Thr residues, respectively, complex carbohydrates can be attached to a protein in either of two linkages. In N-linked oligosaccharides, N-acetylglucosamine (GlcNAc) is the reducing terminal monosaccharide linked to the amide group of an asparagine side chain in an Asn–X–Ser(Thr) sequence. In O-linked oligosaccharides, N-acetylgalactosamine usually is the reducing terminal sugar linked to the hydroxyl group of a Ser or Thr residue in the polypeptide backbone. O-linked sugar chains are usually short and formed by the stepwise transfer of a monosaccharide to the Ser and Thr residue from its nucleotide derivative. In contrast, N-linked sugar chains are synthesized by a series of additions and removals that include lipid-linked

intermediates [35]. They are usually larger than O-linked sugar chains and can be divided into three groups: complex-type sugar chains, high mannose-type sugar chains and hybrid-type sugar chains [5]. All three types contain a common pentasaccharide core structure, Manα1-6(Man-α1-3)Manβ1-4GlcNAcβ1-4GlcNAc.

The first step in the structural characterization of a carbohydrate chain is the release of the oligosaccharides from the glycoprotein to generate an intact pool of oligosaccharides that are subsequently separated from one another. These are difficult tasks. Both enzymatic and chemical methods have been used to cleave oligosaccharides from glycoproteins. Two classes of enzymes have been identified that can release asparagine-linked oligosaccharide chains, and both are now commercially available. *endo*-β-N-Acetylglucosaminidases (EC 3.2.1.96) hydrolyse the glycosidic bond between two N-acetylglucosamine residues [36,37], thus leaving one GlcNAc residue on the protein. Peptide-N-(N-acetyl-β-glucosaminyl)-asparagine amidase (EC 3.2.2.18) hydrolyses GlcNAc–Asn linkages. The oligosaccharides cleaved by this enzyme are thus intact and possess a reducing terminal. It has proved to be the most efficient enzyme for releasing N-linked oligosaccharides from glycoproteins [38]. Enzymatic release of O-linked carbohydrate chains from glycoproteins by endo-α-N-acetylgalactosaminidase (EC 3.2.1.97) is not used widely since its substrate specificity is quite strict [39]. It only recognizes the disaccharide Galβ1-3GalNAc.

Chemical release of oligosaccharides has been applied successfully to both N-linked and O-linked sugars [40]. For the latter, β-elimination in the presence of an alkaline solution of $NaBH_4$ is often used [41]. Hydrazine and trifluoromethanesulfonic acid (TFMS) can be used to cleave both N-linked and O-linked sugar chains [42,43]. Hydrazinolysis has increasingly received attention since it can discriminate between N-linked and O-linked sugars [44]. It also has the advantage that it generates oligosaccharides with reducing termini which allow subsequent labelling and derivatization for easy downstream separation and detection. Hydrazine unavoidably degrades the protein part of the glycoprotein. If the protein part must be kept intact, enzymatic cleavage or hydrolysis with TFMS should be used.

Separation of heterogenous glycoforms from one another to yield homogenous oligosaccharides is yet another very difficult step in carbohydrate analysis, since the monosaccharide constituents of complex carbohydrates usually have very similar chemical structures. Often the compositions of different oligosaccharides differ by only one or two monosaccharides, and in some cases the oligosaccharides have the same composition but differ only in one or two anomeric linkages. Chromatographic techniques such as GLC, TLC, low-pressure LC and HPLC have all been used for carbohydrate separation and purification [44–55]. Methods for separation have been developed based on both size and charge. Anionic oligosaccharides containing sialic acid, uronic acid, phosphate or sulfate groups can be separated by ion-exchange chromatography. For uncharged oligosaccharides, gel permeation chromatography on Bio-Gel P-4 usually is a good choice. Neutral sugars can also be separated by ion-exchange chromatography in borate buffer if they have *cis*-hydroxyl groups which allow formation of a charged borate complex. In addition, affinity chromatography on immobilized lectin columns is used widely to separate oligosaccharides and differentiate different sugar structures. Since the specificity of this method is very high, it not only allows the separation of structurally similar oligosaccharides, but also provides information about their terminal structures [56]. Recently, NMR and MS have gained prominence and are now the most important methods for detailed elucidation of carbohydrate structure [57–59]. By a combination of the currently available techniques, several thousand complex carbohydrates have now been characterized and reported. However, these methods are still not ideal: they are time-consuming, involve expensive equipment and require expert interpretation of the data. Also, a relatively large amount of starting material is required which sometimes is very difficult to obtain from biological samples. Clearly, the

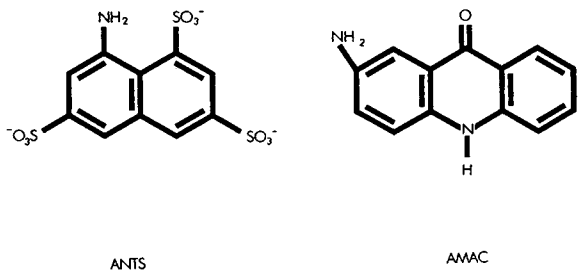
existence (or introduction) of a rapid, simple, sensitive, cost-effective method would fill a critical niche in carbohydrate analysis.

For this reason, a method based on electrophoretic separation of carbohydrates fluorescently-labeled at the aldehydic reducing termini shows promise for the structural characterization of oligosaccharides, and may be particularly useful for multiple sample analysis and comparison [60–67]. The method has been termed fluorophore-assisted carbohydrate electrophoresis (FACE) or the less euphonic polyacrylamide gel electrophoresis of fluorophore-labeled saccharides (PAGEFS). This review describes the principles of FACE and its application to several aspects of carbohydrate analysis.

3. Fluorophore-assisted carbohydrate electrophoresis

FACE technology was first described by Jackson in 1990 [63]. It combines the high resolution and simplicity of polyacrylamide gel electrophoresis (PAGE) with the sensitivity and visibility of fluorescence. In principle, carbohydrates that have a reducing terminus (aldehydic carbon) are reacted with a fluorophore that has a primary amino group. The resultant Schiff base is stabilized by reductive amination with sodium cyanoborohydride to yield the final stable fluorescently labeled derivatives which are separated by PAGE under appropriate conditions and followed by detection and quantification of the fluorescence of the bands under UV illumination.

The fluorophores that have been used most often are 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS) [62–65] and aminoacridone (AMAC) [60,61] whose chemical structures are shown in Fig. 1. ANTS has three sulfonic acid groups, an excitation maximum at 365 nm and an emission that peaks at 515 nm. ANTS labeling not only confers fluorescence on the carbohydrate but also adds three negative charges that enable the previously neutral molecule to migrate in an electric field. Both acidic and neutral



ANTS AMAC

Fig. 1. Chemical structures of ANTS and AMAC.

saccharides can be labeled with ANTS. Separation of ANTS-labeled carbohydrate derivatives is based both on charge and size. The mobility of neutral saccharide derivatives depends mostly on their size as all possess the same charge contributed by ANTS. However, the conformation of the saccharides and the configuration of their hydroxyl groups may also contribute to their migration, presumably by perturbing their effective size and interaction with the gel matrix. Thus, glucose, galactose and mannose are resolved readily, as are maltose (Glcα1-4Glc), isomaltose (Glcα1-6Glc) and cellobiose (Glcβ1-4Glc) [63].

Several electrophoretic buffer systems have been tested for the separation of fluorescently labeled saccharides [61,63,64]. For ANTS-labeled saccharides, the Tris–glycine system of Laemmli [68], commonly used for sodium dodecyl sulfate (SDS)-PAGE, gives satisfactory results. SDS is not required for separation to occur and is therefore omitted throughout the procedure. The concentration of the gel can be varied and optimized for specific purposes. Generally, 30% acrylamide is adequate for most of the N-linked oligosaccharides. A gradient gel may be necessary to separate saccharides that have a wide range of size and charge. A ladder of glucose polymers ranging from glucose to maltooligosaccharides of approximately 30 glucose units have been well separated in a 20 to 40% gradient gel [63].

The other fluorophore, AMAC, has a 425 nm excitation wavelength and an emission maximum at 520 nm. It is uncharged at neutral and alkaline conditions and is thus useful for the analysis of

small acidic saccharides. AMAC derivatives of neutral saccharides do not migrate in the Tris–glycine buffer system. They can migrate into the gel and be separated in a borate-containing buffer owing to the formation of a charged borate complex with the *cis*-hydroxyl groups of the saccharides. Therefore, AMAC serves to differentiate neutral and acidic saccharides by electrophoresis in borate and non-borate buffer systems. Another feature of AMAC is that it undergoes a significant reaction with sialic acid by an unknown mechanism [61]. All commonly occurring monosaccharides (Glc, Gal, Man, Fuc, SA, GlcNAc and GalNAc) can be labeled with AMAC and separated subsequently from one another by PAGE in a borate buffer. AMAC is the reagent most often used for monosaccharide analysis.

The chemical reaction of ANTS with a reducing sugar is shown in Fig. 2. The derivatization of carbohydrates with either ANTS or AMAC should have 1:1 stoichiometry and complete modification has been obtained with up to 100 nmol glucose, lactose, maltopentaose [63] and Gal-6-SO_3^- under the conditions described in the legends to Figs. 3 and 4 for ANTS and AMAC labeling, respectively. The emission wavelength of the fluorophores for both ANTS and AMAC derivatives of saccharides produces yellow bands under UV illumination after appropriate filtration. A 10 pmol band is easily visualized. One simple way to view and record FACE gels is with a standard long wavelength UV light box and the type of arrangements commonly employed for photographing ethidium bromide-stained DNA gels. Greater sensitivity can be achieved by recording a digitized image of the FACE gel with a cooled, charge-coupled device (CCD) camera. With both ANTS and AMAC-labeled saccharides, as little as 0.1 pmol per band can be detected with the Glycoscan system of Millipore (Bedford, MA, USA) or the FACE imaging system of Glyko (Novato, CA, USA). Since the derivatization is stoichiometric and quantitative, the intrinsic fluorescence of the polyacrylamide gel and the buffers is low, and the linear range of the CCD camera is wide, quantitative carbohydrate analysis can be achieved easily by direct comparison of the fluorescence intensity of a given band with that of a known amount of internal standard.

4. Oligosaccharide profiling by FACE

In studies involving glycoproteins, it is often necessary to compare the degree of glycosylation of two or more glycoproteins. For example, the glycosylation of a recombinant glycoprotein must be compared to that of its native counterpart. This comparison is usually performed either "indirectly" by monosaccharide composition analysis or "directly" by oligosaccharide profile analysis. Experimentally, composition analysis is easier, and for this reason has commonly been preferred. However, like peptide mapping and fingerprint analysis of a protein, profiling analysis of the oligosaccharides released from glycoproteins is more informative and gives the properties of the oligosaccharides such as charge, size and linkage.

N-Linked oligosaccharide analysis by FACE starts with enzymatic release of glycans from the glycoproteins. Peptide N-glycosidase F (PNGF) has proved to be an ideal glycosidase for N-linked glycan release. It is, in fact, an amidase

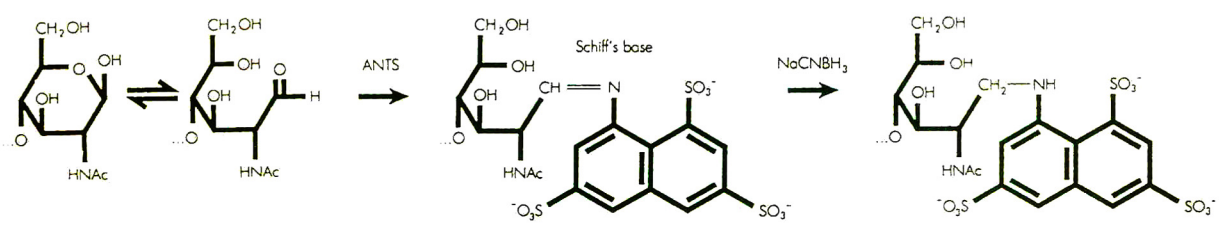


Fig. 2. Derivatization of a reducing sugar by ANTS.

rather than a glycosidase since it cleaves the GlcNAc–Asn linkage. All three types of N-linked glycans can be released by PNGF under appropriate conditions. The liberated oligosaccharides have a reducing terminus which is ideal for subsequent derivatization with ANTS. The amount of glycoprotein required to obtain a clear and informative profile of N-linked oligosaccharide varies with both the carbohydrate content of the glycoprotein and the microheterogeneity of the sugar chains. In most cases, 50 μg of glycoprotein are adequate. Conditions for the release of N-linked oligosaccharides by PNGF and the subsequent derivatization of the glycans with ANTS from various amounts of glycoproteins have been defined [65]. The N-linked oligosaccharide profiles of six glycoproteins are shown in Fig. 3. Most of the N-linked saccharides migrate somewhere between G_5 and G_{12}.

N-Linked oligosaccharide profiling is very useful for comparison of the glycosylation patterns of glycoproteins. Multiple samples can be analyzed side-by-side on the same gel and, hence, the method is ideal for examining of the oligosaccharide chains of a glycoprotein isolated from different species or at different stages during development or cellular differentiation. It can also be used to identify changes in glycosylation that might accompany certain physiological or pathological conditions. For recombinant glycoproteins, FACE oligosaccharide profiling promises to be a simple, convenient means for quality control to monitor batch-to-batch consistency. This is particularly important when the carbohydrate moiety participates in the biological function of the glycoprotein.

Profiling of O-linked oligosaccharides of glycorpoteins is essentially the same as for N-linked oligosaccharides except for two changes. Instead of PNGF, hydrazinolysis is used to release the oligosaccharide from the proteins, and a higher percentage of acrylamide gel must be used because O-linked oligosaccharides are usually small and contain neuraminic acid. As an example, the O-linked oligosaccharides obtained from the glycoprotein fetuin are shown in Fig. 4.

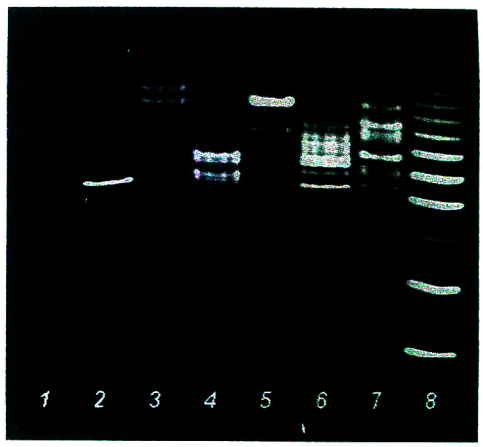


Fig. 3. N-Linked oligosaccharide profiles of glycoproteins. A 50-μl aliquot of a 1 mg/ml solution of a glycoprotein in water was combined with 50 μl of 40 m*M* sodium phosphate buffer, pH 8.6. SDS and β-mercaptoethanol were added to final concentrations of 0.1% and 50 m*M*, respectively, and the mixture was boiled for 5 min to denature the protein. It was cooled on ice and Nondient P40 (NP-40) was added to a final concentration of 0.5%. (If necessary, 0.2% SDS can be used to denature the protein, but the concentration of NP-40 must then be increased to give a final concentration that is five times higher than SDS so that PNGF will be fully active.) PNGF, 5 mU, was added to cleave the oligosaccharides from the glycoprotein and the mixture was incubated at 37°C. After 2 h, 3 volumes of ice-cold ethanol were added to precipitate the protein. The supernatant, containing released oligosaccharides, was dried in a centrifugal vacuum evaporator at a temperature ≤45°C to avoid oligosaccharide degradation. The residue was resuspended in 5 μl of 0.15 *M* ANTS in 15% (v/v) acetic acid. Then 5 μl of 1.0 *M* sodium cyanoborohydride in dimethyl sulfoxide (DMSO) were added, and the suspension was mixed, briefly centrifuged to bring the reactants to the tip of the tube, and incubated at 45°C for 3 h. After drying in a centrifugal vacuum evaporator the fluorescent-labeled oligosaccharides were redissolved in 20 μl of 20% glycerol and 4 μl of the sample was used for electrophoresis. Each lane corresponds to the oligosaccharides released from 10 μg glycoprotein. A 30% acrylamide gel was used with the Laemmli [68] buffer system but with SDS omitted. Electrophoresis was carried out at 5°C, 15 mA with an upper limit setting of 800 V and 60 W. A fluorescent image of the gel was recorded with a CCD camera under UV illumination. Lanes: 1 = control, no glycoprotein; 2 = transferrin; 3 = carboxypeptidase Y; 4 = fetuin; 5 = asialofetuin; 6 = α_1-acid glycoprotein; 7 = α_1-antitrypsin; 8 = a standard ladder of glucose polymers in which the lowest band is G_2.

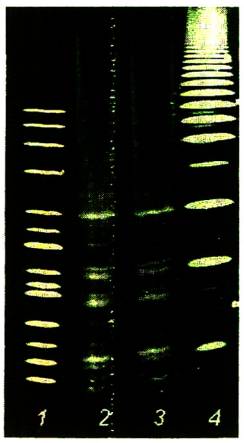


Fig. 4. O-Linked oligosaccharides of fetuin. A 50-μl volume of 1 mg/ml fetuin in a conical reaction tube was dried in a centrifugal vacuum evaporator and placed in a vacuum dessicator over P_2O_5 to remove all traces of water. The sample was dried overnight, 50 μl of anhydrous hydrazine were added and the tube was vortexed, overlayed with dry nitrogen and capped tightly. After 3 h incubation in a sand bath at 60°C, the sample was again dried in the centrifugal vacuum evaporator and resuspended in 30 μl of 0.2 *M* $NaHCO_3$, pH 10.7. Acetic anhydride, 3 μl, was added to re-N-acetylate the primary amino groups generated during hydrazinolysis. The sample was incubated on ice for 15 min, dried in the centrifugal vacuum evaporator and labeled with ANTS as described in Fig. 3. The labeled oligosaccharide mixture was electrophoresed in a 40% acrylamide gel and the products were visualized with the FACE system. Lanes: 1 = the migration pattern of some known oligosaccharides: maltoheptaose, maltohexaose, maltopentaose, maltotetraose, maltotriose, cellotriose, galactobiose, maltose, lactose, galactosylgalactose, N-acetylgalactosamine, galactose, glucose and 6-deoxyglucose, from top to bottom, respectively; 2, 3 = the O-linked oligosaccharides from 10 μg fetuin; 4 = a standard ladder of glucose polymers in which the lowest band is glucose itself.

The conditions for hydrazinolysis of O-linked oligosaccharides have been described by Patel et al. [44].

Both N-linked and O-linked oligosaccharide profiling can be carried out either analytically or preparatively. Individual oligosaccharide bands can be cut from a preparative gel, eluted into water and subjected to subsequent analysis for monosaccharide composition and oligosaccharide sequence.

5. Monosaccharide composition analysis by FACE

Most glycoproteins contain up to seven different kinds of common monosaccharides: mannose (Man), fucose (Fuc), galactose (Gal), glucose (Glc), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc) and N-acetylneuraminic acid (Neu5Ac). Such monosaccharides can be divided into three groups: neutral sugars comprising Glc, Gal, Man and Fuc; amino sugars including GlcNAc and GalNAc; and acidic sugars containing Neu5Ac. Amino sugars and neutral sugars react with both ANTS and AMAC. Neu5Ac lacks an aldehydic reducing group and thus does not react with ANTS. It can, however, form a fluorescent derivative with AMAC by an as yet unknown mechanism. Monosaccharide composition analysis is therefore usually carried out with AMAC. As shown in Fig. 5, AMAC derivatives of the seven monosaccharides can be separated by PAGE in a buffer that contains borate. Monosaccharide composition analysis can be done with both an intact glycoprotein and purified oligosaccharides such as those eluted from a preparative oligosaccharide profiling gel. Since the thermal and acidic stabilities among the three groups of monosaccharides differ, three different hydrolysis conditions have been developed to achieve optimal recovery of neutral, amino and acidic sugars, respectively. Table 1 lists conditions for hydrolysis of the three types of sugars from both intact glycoproteins and isolated oligosaccharides. For neutral sugars in an intact glycoprotein, more than 90% of the sugars can be recovered by hydrolysis in 2 *M* TFA at 100°C for 5 h. For amino sugars, 4 *M* HCl at 100°C for 3 h is usually adequate for the purpose. After hydrolysis, re-N-acetylation of the amino sugars must be performed to reverse whatever deacetylation has occurred during this process. This can be accomplished by incubating the samples on ice with 1 *M* acetic anhydride in 0.2 *M* $NaHCO_3$, pH 10.7, for 15 min. For neutral and amino sugars from a purified oligosaccharide, the hydrolysis time can be reduced to 1 h

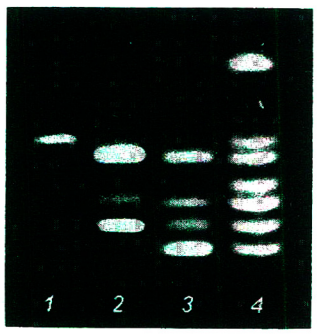


Fig. 5. Monosaccharide composition analysis of human transferrin. Three tubes containing 50 μg human transferrin each were dried in a centrifugal vacuum evaporator. A 100-μl volume of 0.1 *M* trifluoroacetic acid (TFA) was added to tube 1 and it was incubated at 80°C for 30 min to hydrolyze neuraminic acid. A 100-μl volume of 2 *M* TFA was added to tube 2 and it was incubated at 100°C for 5 h to hydrolyze neutral sugars. 100 μl of 4 *M* HCl was added to tube 3 and it was incubated at 100°C for 3 h to hydrolyze amine sugars. After incubation, the samples were again dried in a centrifugal vacuum evaporator. Tube 3 was subjected to re-N-acetylation as described in the legend to Fig. 4. For AMAC labeling, the samples were resuspended in 2.5 μl of 15% (v/v) acetic acid, mixed with 2.5 μl of 0.2 *M* AMAC in DMSO, followed by 5 μl of sodium cyanoborohydride, mixed and incubated at 45°C for 3 h. AMAC-labeled monosaccharides were dried, redissolved in 5 μl of DMSO, and diluted with 20% glycerol to a final volume of 20 μl. A 4-μl aliquot of this solution was then applied to a monosaccharide composition gel. Concentrations of acrylamide in the separating and stacking gels are 20 and 4%, respectively, and the electrophoresis buffer is 0.1 *M* Tris base–borate/boric acid, pH 8.3. Electrophoresis was carried out at 25 mA per gel for 2.5 h. Lanes: 1 = neuraminic acid; 2 = neutral sugars; 3 = amine sugars; 4 = monosaccharide standards: GalNAc, Neu5Ac, Man, Fuc, Gal and GlcNAc, from top to bottom, respectively.

and 30 min, respectively. Neu5Ac, which is very labile, requires special attention. Appropriate conditions for Neu5Ac analysis from glycoproteins and purified oligosaccharides are 0.1 *M* TFA or 0.1 *M* HCl at 80°C for 30min, respectively. Moreover, the labeling efficiency of Neu5Ac by AMAC is much lower than that of neutral and amino sugars, presumably because of a different reaction mechanism. Therefore, a Neu5Ac labeling control should be used every time. N-Acetyllactosamine is here recommended as a monocomposition control to monitor the efficiency of the hydrolysis, re-N-acetylation and fluorophore labeling reactions.

It should be noted that determination of monosaccharide composition by FACE has been thought not to compare favorably with other methods available for monosaccharide analysis. Thus, ion-exchange chromatography at high pH with electrochemical detection [69] or reversed-phase chromatography of precolumn derivatized sugars [70,71] have been shown to provide an accurate determination of monosaccharide composition. However, when combined with other applications of FACE such as oligosaccharide profiling and sequencing, monosaccharide analysis by the same method becomes a convenient extension of this technology.

6. Oligosaccharide sequencing by FACE

Oligosaccharide sequence analysis is the most difficult step in the complete structural determination of a glycoprotein. Unlike protein and DNA sequencing, it requires identification of (i) the particular carbon atom involved in each linkage, (ii) the anomeric specificity of the linkage and (iii) the identity and order of each monomer. Several physical, chemical and enzymatic methods have been developed and refined for oligosaccharide sequencing [72–75].

Table 1
Conditions for hydrolysis of intact glycoproteins and isolated oligosaccharides

Analysis	Intact glycoproteins	Isolated oligosaccharides
Neuraminic acids	0.1 *M* TFA, 80°C, 30 min	0.1 *M* HCl, 80°C, 30 min
Neutral sugars	2 *M* TFA, 100°C, 5 h	2 *M* TFA, 100°C, 1 h
Amino sugars	4 *M* HCl, 100°C, 3 h	4 *M* HCl, 100°C, 30 min

Sequential digestion with exoglycosidases of known specificities is the one most commonly used for this purpose [75].

Oligosaccharide sequencing is a third major potential application of FACE technology. Oligosaccharide profiling and monosaccharide composition analyses are prerequisites for sequence determination. Even more importantly, the oligosaccharide must be available in a homogeneous form. This can be achieved by oligosaccharide preparative electrophoresis. The oligosaccharide isolated in this way would have been fluorescently labeled already with ANTS which remains at the reducing end throughout the procedure. The digestion products can thus be analyzed for a shift in mobility without additional treatment. Further, the composition analysis of the isolated oligosaccharide provides preliminary information about the oligosaccharide such as content of neuraminic acid and mannose. This information is very useful in determining the appropriate exoglycosidases to apply. Moreover, the selectivity of the releasing enzymes is also helpful in oligosaccharide sequencing. PNGF releases all three types of N-linked oligosaccharides from glycoproteins, whereas endoglycosidase H only releases high-mannose and hybrid oligosaccharides. Therefore, if an oligosaccharide isolated from a PNGF-released profile does not appear on the endoglycosidase H-released profile, it must be a complex-type oligosaccharide. Based on the Neu5Ac content from monosaccharide composition analysis, it is easy to decide whether neuraminidase or β-galactosidase should be used to cleave the outermost monosaccharide. Subsequent to the first exoglycosidase digestion, an aliquot of the product mixture can be taken for gel analysis to determine the mobilities of the remaining oligosaccharides. The next exoglycosidase digestion can be carried out with either the digestion mixture or the gel-purified product from the previous digestion. Repetitive treatment of the same digestion mixture with different exoglycosidases will generate the final core structure, Manβ1-4GlcNAcβ1-4GlcNAc or Manβ1-4GlcNAcβ1-4(Fucα1-6) GlcNAc. Hence, the oligosaccharide sequence can be deduced from knowledge of the general structural features of N-linked oligosaccharides, the degradation patterns resulting from digestion with exoglycosidases of known specificity, and the pre-determined migration patterns of known sequence. Table 2 lists the mobility of ANTS-labeled N-linked oligosaccharide standards as defined by Glyko. The migration values are expressed as a degree of polymerization, DP, which describes the mobility of an oligosaccharide relative to a mixture of glucose polymers. For example, a band that migrates in a position corresponding to maltotetraose (G_4) has a DP of 4, and to maltopentaose (G_5) a DP of 5. An oligosaccharide that migrates halfway between G_4 and G_5 has a DP of 4.5. Table 3 shows the shift in mobility of an ANTS-labeled oligosaccharide that would occur on removal of a single monosaccharide. These values can serve to correlate mobility shifts measured in the sequencing gel with the number of monosaccharides released sequentially from an oligosaccharide. It should be kept in mind that photobleaching of a fluorophore-labeled oligosaccharide will inevitably occur during UV-assisted excision from a preparative gel, and this precludes accurate quantitation of isolated oligosaccharides. Quantitation of exoglycosidase-released monosaccharides by monosaccharide composition gel analysis is likely to be less than satisfactory since both the exoglycosidase and the buffer may interfere with the efficiency of derivatization.

The position and anomeric character of a glycosidic linkage can be determined by the specificity of the exoglycosidase employed. Three neuraminidases with distinct specificities are commercially available. Neuraminidase I releases only α2-3 linked neuraminic acid, neuraminidase II cleaves both α2-3 and α2-6 linkages, whereas neuraminidase III is non-specific and removes all neuraminic acids (α2-3,-6,-8 linkages) from glycoproteins. This allows the linkage of a neuraminic acid in a glycoprotein to be determined readily by appropriate use of these enzymes.

Recently, a new strategy for sequencing oligosaccharides has been developed [76]. This so-called reagent-array method of analysis involves

Table 2
Mobility of ANTS-labeled N-linked oligosaccharides

Oligosaccharides	DP
Oligomannose 9	8.8
Oligomannose 8	8.1
Oligomannose 7	7.4
Oligomannose 6	6.5
Oligomannose 5	5.7
Disialylated (2α2-6), galactosylated bi-antennary	5.4
Disialylated (2α2-6), galactosylated bi-antennary, core-fucosylated	5.7
Disialylated (2α2-3), galactosylated bi-antennary	6.4
Disialylated (2α2-3), galactosylated bi-antennary, core-fucosylated	6.7
Trisialylated (2α2-6, 1α2-3), galactosylated bi-antennary	6.2
Trisialylated (1α2-6, 2α2-3), galactosylated bi-antennary	6.7
Asialo-, galactosylated bi-antennary	7.8
Asialo-, galactosylated bi-antennary, core-fucosylated	8.5
Asialo-, galactosylated tri-antennary	9.6
Asialo-, galactosylated tri-antennary, core-fucosylated	10.5
Asialo-, galactosylated tetra-antennary	11.5
Asialo-, galactosylated tetra-antennary, core-fucosylated	12.1
Asialo-, degalacto, bi-antennary	5.6
Asialo-, degalacto, bi-antennary, core-fucosylated	6.2
Asialo-, degalacto, tri-antennary	6.4
Asialo-, degalacto, tri-antennary, core-fucosylated	7.1
Asialo-, degalacto, tetra-antennary	7.3
Asialo-, degalacto, tetra-antennary, core-fucosylated	7.9
Conserved tri-mannosyl core	4.2
Conserved tri-mannosyl core, core-fucosylated	4.8
Manα1-6Manβ1-4GlcNAcβ1-4GlcNAc	3.3
Manα1-6Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc	4.0
Manβ1-4GlcNAcβ1-4GlcNAc	2.5
Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc	3.2

incubating each of nine aliquots of an oligosaccharide to be sequenced with a set of defined, preprepared mixtures of exoglycosidases, i.e. the reagent array. Each sample glycan will be digested from the non-reducing end until only a limit glycan fragment remains. Then, the products from each of the nine digestion mixtures are combined and subjected to gel permeation chromatography on Bio-Gel P-4. This generates a signature of the hydrodynamic volume/relative molar ratio of the stop-point fragments. By comparing this signature with the theoretical signatures of oligosaccharides with known sequence, the structure of the glycan can be deduced by a best-fit computer program. The reagent-array method has been modified by Oxford Glycosystems and a carbohydrate sequencer (RAAM 1000 GlycoSequencer; Oxford Glycosystems, Abingdon, UK) based on this technology has been developed and is now commercially available.

Table 3
Monosaccharide contribution to oligosaccharide DP[a]

Monosaccharide	Mobility shift, DP units
Neu5Ac	−1
Gal	+1
GlcNAc	+0.75
Man	+0.75
Fuc	+0.6

[a] Defined by Glyko, Novato, CA, USA.

FACE oligosaccharide sequencing uses the

same principle [75]. In this case, the digestion products need not be combined, but can be analyzed side-by-side in different lanes of the same gel. Fig. 6 shows an example of this method applied to an oligosaccharide isolated from human serum glycoproteins. Typically a FACE sequencing experiment would involve setting up five separate enzyme digests. About 50–100 pmol of ANTS-labeled oligosaccharide would be used in each of the five tubes. The first tube does not receive any enzyme and is used as a size marker for the location of the starting material on the gel. The second sample is digested with neuraminidase III, which releases all the neuraminic acids. The third tube contains neuraminidase III and β-galactosidase which will release both neuraminic acids and galactose. The fourth tube contains neuraminidase III, β-galactosidase and β-N-acetylhexosaminidase which will release neuraminic acids, galactose and N-acetylglucosamine. The fifth tube contains all of the above enzymes plus α-mannosidase which will degrade the residual oligosaccharide to its core structure.

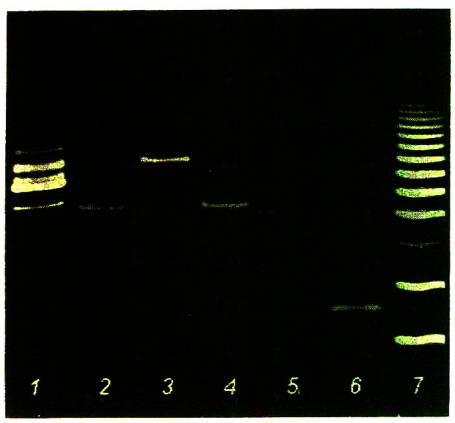


Fig. 6. Sequence analysis of an oligosaccharide from human serum glycoproteins. N-Linked oligosaccharides were released from 50 μl of normal human serum and labeled with ANTS under the conditions described in the legend to Fig. 3. After preparative N-linked oligosaccharide gel electrophoresis the oligosaccharide with a DP value of 5.3 was isolated by aqueous extraction of the gel slice. Lane 1 shows the migration pattern of all the ANTS-labeled oligosaccharides released from serum glycoproteins. The isolated DP 5.3 oligosaccharide was divided into five equal fractions, each of which was incubated with the indicated combinations of glycosidases in 50 m*M* sodium phosphate buffer, pH 6.0, at 37°C overnight. The glycosidase digestion products were subjected to N-linked oligosaccharide profiling gel electrophoresis for mobility shift analysis. Other lanes: 2 = oligosaccharide DP 5.3; 3 = neuraminidase (4 mU) digestion product; 4 = neuraminidase (4 mU) and β-galactosidase (10 mU) digestion product; 5 = neuraminidase (4 mU) + β-galactosidase (10 mU) + N-acetyl β-glucosaminidase (80 mU) digestion product; 6 = neuraminaidase (4 mU) + β-galactosidase (10 mU) + N-acetyl β-glucosaminidase (80 mU) + α-mannosidase (20 mU) digestion product; 7 = a standard ladder of glucose polymers in which the lowest band is G_5. A comparison of the mobility change of each glycosidase digestion and the DP value of each digestion product with the standards in Tables 2 and 3 identifies the sequence of oligosaccharide DP 5.3 as Neu5Acα2 - 6Galβ1 - 4GlcNAcβ1 - 2Manα1 - 6(Neu5Acα2 - 6Galβ1 - 4GlcNAcβ1 - 2Manα1 - 3)Manβ1 - 4GlcNAcβ1 - 4GlcNAc.

Sequencing of complex branched oligosaccharides by FACE is more difficult and is currently under development.

7. Interactions of proteins and carbohydrates

Binding of carbohydrates to proteins is vital for the function of many important biological systems. Lectins and selectins are two well-known examples of carbohydrate binding proteins. The former play critical roles in initiating viral, bacterial, mycoplasmal and parasite infections; cell–cell adhesion; fertilization and development; proliferation and differentiation; and metastasis [77–82]. The latter have been recently identified as endothelial-cell-surface carbohydrate binding proteins that function in the recruitment of leukocytes to injured tissues [23–25].

Historically, carbohydrate–protein interactions have been investigated by a variety of techniques including quantitative precipitation, hapten inhibition, equilibrium dialysis and affinity chromatography on immobilized ligands. Although each of these techniques provides some unique answer, all either require large amounts of material or must use radioactive carbohydrate derivatives. Consequently, we have developed a gel retardation assay using FACE technology [83] and demonstrated its feasibility both for screening and identifying carbohydrate ligands of

proteins and for the determination of apparent dissociation constants of native and ANTS-labeled carbohydrate ligands. The ANTS-labeled carbohydrate or carbohydrate mixture is incubated with its binding protein(s), subjected to gel electrophoresis, visualized and quantified by the FACE system. A side-by-side comparison between protein-present and protein-absent samples easily identifies the carbohydrate ligand of the protein. Protein-bound carbohydrate remains at the top of the gel, whereas free carbohydrate migrates into the gel and is thereby separated. Fig. 7 illustrates a gel retardation assay of lectin–carbohydrate binding. Three ANTS-labeled oligosaccharides are separated from each other by gel electrophoresis. Each is seen to be retarded by its specific lectin. By this method, the apparent dissociation constants for binding of the ANTS-labeled and unlabeled oligosaccharide Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3)-Manβ1-4GlcNAcβ1-4GlcNAc to the lectin *Sambucus nigra* agglutinin (SNA) have been determined to be 0.74 and 0.22 μM, respectively [83]. This is apparently the first example in which dissociation constants for labeled and unlabeled ligand have been measured separately. Usually, in order to measure the dissociation constant, it must be assumed that the binding of labeled and unlabeled ligand to a protein are the same. This is not always valid since labeling can frequently alter the structure of the ligand. This assumption is not required in the determination of binding affinities by gel retardation/FACE, constituting a unique advantage of this technique.

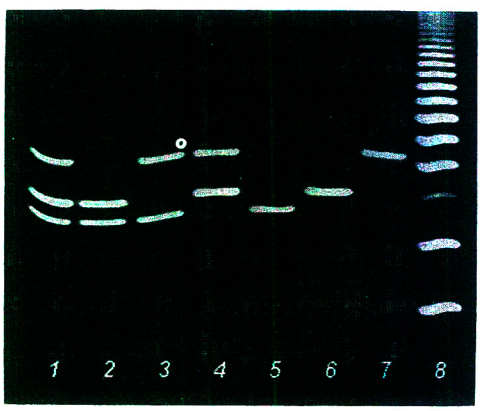


Fig. 7. Gel retardation assay of lectin-carbohydrate binding (from [83]). A mixture of three ANTS-labeled oligosaccharides, each 10 μM, was incubated with 20 μM *Sambucus nigra* agglutinin (SNA, lane 2), 20 μM *Galanthus nivalis* agglutinin (GNA, lane 3), 20 μM *Maackia amurensis* agglutinin (MAA, lane 4), 20 μM SNA + 20 μM GNA (lane 5), 20 μM SNA + 20 μM MAA (lane 6), and 20 μM GNA + 20 μM MAA (lane 7), respectively, in 50 mM Tris–HCl, pH 7.5, 0.15 M NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$ and 1 mM $CaCl_2$ at room temperature for 30 min at a final volume of 10 μl. Lane 1 is the oligosaccharides control incubated without any of the lectins. The three bands in lane 1 are, from top to bottom, Neu5Acα2 - 6Galβ1 - 4GlcNAcβ1 - 2Manα1 - 6(Neu5Acα2 - 6Galβ1 - 4GlcNAcβ - 2Manα1 - 3)Manβ1 - 4GlcNAcβ1 - 4GlcNAc, Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc and Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc, respectively. Lane 8 is a standard ladder of glucose polymers in which the lowest band is G_2.

The gel retardation/FACE assay can also be used to study the interaction between proteins and glycosaminoglycans. The latter are negatively charged polysaccharides composed of repeating disaccharide units [84]. They play important biological roles in a variety of cell processes including cell-cell interactions, cell adhesion on the extracellular matrix, ligand and receptor binding, etc. [85]. All types of glycosaminoglycans can be fluorescently labeled by ANTS and appear as smeared bands on oligosaccharide profiling gels. Fig. 8 depicts how this method can detect the binding of ANTS-labeled N-acetylheparin to angiogenin, a potent inducer of neovascularization [86].

8. Potential applications of FACE in clinical diagnosis

It is known that changes in glycosylation of some glycoproteins are associated with various pathological conditions [31]. Thus, liver disease can be accompanied by reduced sialylation and increased glycan branching whereas cancer can be accompanied by increased fucosylation [87]. Table 4 lists some examples of carbohydrate diagnostic targets. Among them, FACE diagnosis may benefit the lysosomal storage diseases, a large group of genetically determined or ac-

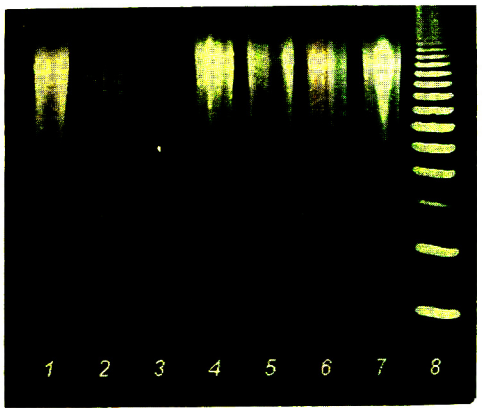


Fig. 8. Binding of proteins to glycosaminoglycans. N-Acetylheparin was labeled by ANTS under the conditions described in the legend to Fig. 3. ANTS-labeled N-acetylheparin, 500 pmol, was incubated with 1 nmol of protein in 10 μl PBS at room temperature, and after 30 min, the mixture was applied to an N-linked oligosaccharide profiling gel. Lane 1, control; lane 8, a standard ladder of glucose polymers in which the lowest band is G_2. The low fluroescence intensities clearly indicate that bovine and human angiogenins (lanes 2 and 3) bind to N-acetylheparin, whereas bovine angiogenin-like protein, ribonuclease A, lysozyme and actin (lanes 4–7) do not.

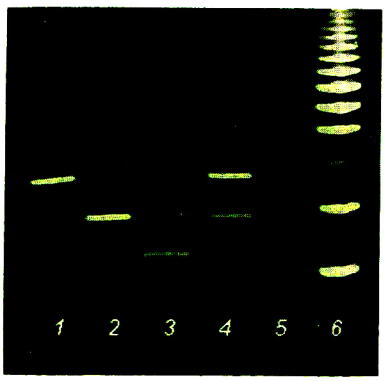


Fig. 9. FACE analysis of glycolipids. A 10-μg sample of monosialoganglioside(s) was incubated with 0.2 ml ceramide glycanase in 0.5 *M* NaAc, pH 5.0, and 75 μg/ml cholate in a final volume of 50 μl at room temperature overnight. After ethanol precipitation, the supernatant—which contains the glycans released from the glycolipid—was dried in a centrifugal vacuum evaporator and labeled with ANTS as described in Fig. 3. Lanes: 1–3 = the glycans released from monosialogangliosides GM_1, GM_2 and GM_3, respectively; 4 = a combination of the samples from lanes 1–3; 5 = the glycan of glycolipid in 5 μl normal human serum; 6 = a standard ladder of glucose polymers in which the lowest band is G_2.

quired glycolipid metabolic disorders [88,89], since glycolipids are less heterogeneous than glycoproteins and therefore easier to identify as diagnosis target(s). In addition, the carbohydrates or glycolipids that accumulate in the diseases are often present in urine so that sampling is convenient. For instance, patients with GM_1 gangliosidosis excrete GM_1 ganglioside in their urine. Fig. 9 shows that the glycans from

Table 4
Potential carbohydrate diagnostic targets

Pathological conditions	Potential carbohydrate markers
Bacterial infection	Bacterial polysaccharides
Cartilage tunor and sarcoma	Keratan and chondroitin sulfate
Choriocarcinoma	Chorionic gonadotropin glycosylation
Colorectal cancer	Glycosaminoglycans
Coronary heart disease	Apolipoprotein glycosylation
Hepatoblastoma	Sialic acids
Leprosy	Phenolic glycolipids
Lymphocytic leukemia	GM_1 glycolipids
Lysosomal storage diseases	Oligosaccharides, glycosaminoglycans, glycolopids
Melanoma	GD_2, GD_3 and GM_2 glycolipids
Mesothelioma	hyaluronic acid
Pancreas and urinary bladder tumor	Lewis antigens
Parasitic diseases	Protozoa-specific carbohydrates
Rheumatoid arthritis	Galactose, N-acetyl glucosamine in IgG

GM_1, GM_2 and GM_3 gangliosides can be separated and identified readily after ceramide glycanase digestion and ANTS labeling. Therefore, the abnormal appearance of GM_1 or GM_2 bands on a FACE gel analysis of urine or serum could be a marker of GM_1 gangliosidosis or Tay–Sach's disease, respectively. There is a large area of unexplored application of FACE in clinical diagnosis but the potential is clear.

9. Conclusions

FACE has been developed to fulfill the need for a rapid, simple and cost-effective method of carbohydrate analysis. The basic advantages of this technique are its simplicity and sensitivity. It separates carbohydrates by PAGE and thus utilizes the high resolving power of this method which is technically familiar and accessible for most laboratories. The carbohydrates are fluorescently labeled and their fluorescence is readily observed and recorded which allows detection and quantification at the subpicomolar level. Moreover, multiple samples can be analyzed in parallel and the results are visible and easy to interpret. A series of applications based on the use of FACE technology has been developed. Various kits now available include monosaccharide composition analysis, N-linked and O-linked oligosaccharide profile analysis, oligosaccharide sequence analysis, glycosaminoglycan analysis and several others. FACE is a significant addition to the panoply of techniques currently available for research in glycobiology. Its full potential, including applications in clinical diagnosis, has yet to be exploited.

Acknowledgements

I thank Drs. Jack T. Johansen of Millipore Corp. for providing the FACE system, Christopher M. Starr of Glyko, Inc. for providing reagents and supplies, Drs. James F. Riordan and Bert L. Vallee for valuable discussions and advice. This work was supported by the Endowment for Research in Human Biology, Inc.

References

[1] T.W. Rademacher, R.B. Parekh and R.A. Dwek, *Annu. Rev. Biochem.*, 57 (1988) 785.
[2] L.O. Narhi, T. Arakawa, K.H. Aoki, R. Elmore, M.F. Rohde, T. Boone and T.W. Strickland, *J. Biol. Chem.*, 266 (1991) 23022.
[3] C.F. Goochee, M.J. Gramer, D.C. Andersen, J.B. Bahr and J.R. Rasmussen, *Biotechnology*, 9 (1991) 1347.
[4] I. Carlstedt, J.K. Sheehan, A.P. Corfield and J.T. Gallagher, *Essays Biochem.*, 20 (1985) 40.
[5] A. Kobata, *Eur. J. Biochem.*, 209 (1992) 483.
[6] E. Larsen, T. Palabrica, S. Sajer, G.E. Gilbert, D.D. Wagner, B.C. Furie and B. Furie, *Cell*, 63 (1990) 467.
[7] M.G. Malaise, P. Franchimont, F. Gomez, C. Bouillenne and P.R. Mahieu, *Clin. Immunol. Immunopathol.*, 45 (1987) 1.
[8] N. Ratner, A.D. Elbein, M.B. Bunge, S. Porter, R.P. Bunge and L.J. Glaser, *J. Cell Biol.*, 103 (1986) 159.
[9] J.D. Bleil and P.M. Wasserman, *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 6778.
[10] S.D. Rosen and T.A. Yednock, *Mol. Cell Biochem.*, 72 (1986) 153.
[11] H.C. Krivan, D.D. Roberts and V. Ginsburg, *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 6157.
[12] T.J. Prichett, R. Brossmer, U. Rose and J.C. Paulson, *Virology*, 160 (1987) 502.
[13] A. Erslev, *New Engl. J. Med.*, 316 (1990) 101.
[14] R.B. Parekh, R.A. Dwek, C.J. Edge and T.W. Rademacher, *Trends Biotechnol.*, 7 (1989) 117.
[15] D.A. Cumming, *Glycobiology*, 1 (1991) 115.
[16] P. Stanley, *Glycobiology*, 2 (1992) 99.
[17] J.R. Rasmussen, *Curr. Op. Struct. Biol.*, 2 (1992) 682.
[18] H. Lis and N. Sharon, *Eur. J. Biochem.*, 218 (1993) 1.
[19] N.W. Barton, R.O. Brady, J.M. Dambrosia, A.M. Di Besceglie, S.H. Doppelt, S.C. Hill, H.J. Mankin, G.J. Murray, R.I. Parker, C.E. Argot, R.P. Grewal, K.-T. Yu et al., *New Engl. J. Med.*, 324 (1991) 1464.
[20] E. Bentler, A. Kay, P. Garver, D. Thurston, A. Dawson and B. Rosenbloom, *Blood*, 78 (1991) 1183.
[21] E. Tsuda, M. Goto, A. Murakami, K. Akai, M. Ueda, G. Kawanishi, N. Takahashi, R. Sasaki, H. Chiba, H. Ishihara, M. Mori, S. Tejima, S. Endo and Y. Arata, *Biochemistry*, 27 (1988) 5646.
[22] M.W. Spellman, J.L. Basa, C.K. Leonard, J.A. Chakel, J.V. O'Connor, S. Wilson and V.H. Halbeck, *J. Biol. Chem.*, 246 (1989) 14100.
[23] L.M. Stoolman, *Cell*, 56 (1989) 907.
[24] B.K. Brandley, S.J. Swiedler and P.N. Robbins, *Cell*, 63 (1990) 861.
[25] T.A. Springer and L.A. Lasky, *Nature*, 349 (1991) 196.

[26] J. Hodgson, *Biotechnology*, 9 (1991) 609.
[27] S. Borman, *Chem. Eng. News*, 7 Dec. (1992) 25.
[28] T.W. Rademacher, *Trends Biotechnol.*, 10 (1992) 27.
[29] T. Feizi and R.A. Childs, *Trends Biochem. Sci.*, 10 (1985) 24.
[30] T. Feizi and R.A. Childs, *Biochem. J.*, 245 (1987) 1.
[31] J.S. Martinez and C. Barsigian, *Lab. Invest.*, 57 (1987) 240.
[32] M.N. Fukuda, *Glycobiology*, 1 (1990) 9.
[33] A. Mackiewcz, M.J. Dewey, F.G. Berger and H. Baumann, *Glycobiology*, 1 (1991) 265.
[34] H. Stibler, *Clin. Chem.*, 37 (1991) 2029.
[35] R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 54 (1985) 631.
[36] A.L. Tarentino, C.M. Gomez and T.H. Plummer, Jr., *Biochemistry*, 24 (1985) 4665.
[37] T. Mizuochi, J. Amano and A. Kobata, *J. Biochem. (Tokyo)*, 95 (1984) 1209.
[38] T.H. Plummer, Jr., J.H. Elder, S. Alexander, A.W. Phelan and A.L. Tarentino, *J. Biol. Chem.*, 259 (1984) 10700.
[39] J. Umemoto, V.P. Bhavandan and E.A. Davidson, *J. Biol. Chem.*, 252 (1977) 8609.
[40] H.T. Sojar and O.P. Bahl, *Methods Enzymol.*, 138 (1987) 341.
[41] J. Amano and A. Kobata, *Methods Enzymol.*, 179 (1989) 261.
[42] A.S.B. Edge, C.R. Faltynek, L. Hof, L.E. Reichert, Jr. and P. Weber, *Anal. Biochem.*, 118 (1981) 131.
[43] R.B. Parekh, R.A. Dwek, J.R. Thomas, G. Opdenakker, T.W. Rademacher, A.J. Wittwer, S.C. Howard, R. Nelson, N.R. Siegel, M.G. Jennings, N.K. Harakas and J. Feder, *Biochemistry*, 28 (1989) 7644.
[44] T. Patel, J. Bruce, A. Merry, C. Bigge, M. Wormald, A. Jaques and R. Parekh, *Biochemistry*, 32 (1993) 679.
[45] M.F. Chaplin, in M.F. Chaplin and J.F. Kennedy (Editors), *Carbohydrate Analysis: A Practical Approach*, IRL, Oxford, 1986, p. 1.
[46] C.A. White and J.F. Kennedy, in M.F. Chaplin and J.F. Kennedy (Editors), *Carbohydrate Analysis: A Practical Approach*, IRL, Oxford, 1986, p. 37.
[47] K. Yamashita, T. Mizuochi and A. Kobata, *Methods Enzymol.*, 83 (1982) 105.
[48] R.D. Cummings, R.K. Merkle and N.L. Stuits, *Methods Cell Biol.*, 32 (1989) 141.
[49] R.R. Townseud, M.R. Hardy and Y.C. Lee, *Methods Enzymol.*, 179 (1989) 65.
[50] G.N. Misevic, *Methods Enzymol.*, 179 (1989) 95.
[51] S. Pelkonen and J. Finne, *Methods Enzymol.*, 179 (1989) 104.
[52] S.C. Churms, *J. Chromatogr.*, 500 (1990) 555.
[53] S. Honda, S. Susuki, S. Zaiki and K. Kakechi, *J. Chromatogr.*, 523 (1990) 6687.
[54] S. Honda, K. Yamamoto, S. Suzuki, M. Ueda and K. Kakechi, *J. Chromatogr.*, 588 (1990) 327.
[55] M. Taverna, A. Baillet, D. Boiu, M. Schluter, R. Werner and D. Ferrier, *Electrophoresis*, 13 (1992) 359.
[56] H. Debray and J. Montreuil, *Adv. Lectin Res.*, 4 (1991) 51.
[57] R.A. Laine, *Methods Enzymol.*, 193 (1990) 539.
[58] J.K. Welply, *Trends Biotechnol.*, 7 (1989) 5.
[59] A. Varki, *FASEB J.*, 5 (1991) 226.
[60] P. Jackson, *Anal. Biochem.*, 216 (1994) 243.
[61] P. Jackson, *Anal. Biochem.*, 196 (1991) 238.
[62] P. Jackson, *Biochem. Soc. Trans.*, 21 (1993) 121.
[63] P. Jackson, *Biochem. J.*, 270 (1990) 705.
[64] P. Jackson and G.R. Williams, *Electrophoresis*, 12 (1990) 94.
[65] P. Jackson, *Methods Enzymol.*, 230 (1993) 150.
[66] R.J. Stack and M.T. Sullivan, *Glycobiology*, 2 (1992) 85.
[67] K.-B. Lee, A. Al-Hakim, D. Loganathan and R.J. Linhardt, *Carbohydr. Res.*, 214 (1991) 155.
[68] U.K. Laemmli, *Nature*, 227 (1970) 680.
[69] M.R. Hardy, R.R. Townsend and Y.C. Lee, *Anal. Biochem.*, 170 (1988) 54.
[70] E.Y.J. Kang, R.D. Coleman, H.J. Pownall, A.M. Gotto, Jr. and C.Y. Yang, *J. Protein. Chem.*, 9 (1990) 31.
[71] S. Honda, E. Akao, S. Suzuki, M. Okuda, K. Kakehi and J. Nakamura, *Anal. Biochem.*, 180 (1989) 351.
[72] A. Dell, *Adv. Carbohydr. Chem. Biochem.*, 45 (1987) 19.
[73] J.F.G. Vliegenthart, L. Dorland and H. van Halbeek, *Adv. Carbohydr. Chem. Biochem.*, 41 (1983) 209.
[74] T. Mizuochi, K. Yonemasu, K. Yamashita and A. Kobata, *J. Biol. Chem.*, 253 (1978) 7404.
[75] A. Kobata, *Anal. Biochem.*, 100 (1979) 1.
[76] C.J. Edge, T.W. Rademacher, M.R. Wormald, R.B. Parekh, T.D. Butters, D.R. Wing and R.A. Dwek, *Proc. Natl. Acad. Sci. U.S.A.*, 89 (1992) 6338.
[77] K. Yamashita, K. Umetsu, T. Suzuki and T. Ohkura, *Biochemistry*, 31 (1992) 11647.
[78] B.A. Williams, M.C. Chervenak and E.J. Toone, *J. Biol. Chem.*, 267 (1992) 22907.
[79] G. Ashwell and J. Harford, *Annu. Rev. Biochem.*, 51 (1982) 531.
[80] K.A. Karlsson, *Annu. Rev. Biochem.*, 58 (1989) 309.
[81] N. Sharon and H. Lis, *Science*, 246 (1989) 227.
[82] L.A. Lasky, *Science*, 258 (1992) 964.
[83] G.-F. Hu and B.L. Vallee, *Anal. Biochem.*, 218 (1994) 185.
[84] M. Höök, L. Kjellen, S. Johansson and J. Robinson, *Ann. Rev. Biochem.*, 53 (1984) 847.
[85] M. Yanagishita and V.C. Hascall, *J. Biol. Chem.*, 267 (1992) 9451.
[86] J.F. Riordan and B.L. Vallee, *Br. J. Cancer*, 57 (1988) 587.
[87] G.A. Turner, *Clin. Chim. Acta*, 208 (1992) 149.
[88] E.F. Neufeld, *Annu. Rev. Biochem.*, 60 (1991) 257.
[89] J. Alroy, R.D. Gasperi and C.D. Warren, *Carbohydr. Res.*, 213 (1991) 229.

ELSEVIER

Journal of Chromatography A, 705 (1995) 105–114

JOURNAL OF CHROMATOGRAPHY A

Review

Host cell contaminant protein assay development for recombinant biopharmaceuticals

Leslie C. Eaton

Control Biotechnology Analytical Assay Development, OU4861-233-075, The Upjohn Company, Kalamazoo, MI 49001, USA

Abstract

The efficiency and consistency of a biopharmaceutical purification process determines drug quality, including which specific types and concentrations of residual host cell or process contaminants may remain. Commercial reagents and generic analytical methods are available for quantitating most of these contaminants. However, no generic assay is available for quantitation of the specific contaminant host cell proteins (HCPs) which are unique to a novel purification process. Because of this, proprietary reagents and assays must be developed for the quantitation of process-specific HCPs in each biopharmaceutical drug. The need to develop proprietary reagents which are both sensitive to, and specific for, potentially complex mixtures of unique contaminant proteins has defined what is acceptable methodology for development of quantitative HCP assays. Within the biopharmaceutical industry this need is most often satisfied by the development of multi-analyte HCP immunoassays based upon the null cell mock purification model. Confidence in the quantitative nature of a given HCP assay, and the validity of analytical measurement obtained by the assay, is dependent upon empirical demonstration of the unique stoichiometry of the HCP assay reagents. In conjunction with other analytical and validation methods, an HCP immunoassay may be thought of as a necessary quantitative tool for the optimization and validation of biopharmaceutical purification process efficiency and consistency, rather than as an end in itself.

Contents

0021-9673/95/$29.00
SSDI 0021-9673(94)01249-0

1. Introduction

Recombinant biopharmaceuticals are treated like biologicals with respect to quality assurance for purity [1,2], in that the efficiency and consistency of the purification process determines the quality of the product. The design and operation of a given process [3], therefore, will define the specific types, and amounts, of residual host cell or process contaminants that will remain in the bulk drug. These potential bioprocess contaminants may include items as diverse as: cell culture medium serum proteins; immunoglobulin affinity ligands; Protein A or Protein G affinity ligands; viruses; endotoxin; DNA; non-protein cell wall constituents; and host cell proteins (HCPs). Removal of each of these process and host cell contaminants to acceptably safe levels must be assured both by development of specific contaminant assays and by rigorous process validation [4–23]. In a broad sense, minimization of all forms of contamination in a therapeutic biopharmaceutical is a matter not only of safety, but of pharmaceutical elegance.

Commercial reagents and standardized analytical methods are available for detecting and quantitating most contaminants including: cell culture media proteins (H. Merrick, personal communication) [24]; immunoglobulin [25]; affinity ligand proteins [25]; non-protein cell wall constituents [26]; endotoxin [27]; DNA [28–31]; or viruses [32]. In contrast, in the presence of a vast excess of recombinant protein product, no generic protein assay for quantitation of undefined, complex mixtures of process-specific HCPs is feasible [4,10,13,33]. Because of this, proprietary reagents and assays must be developed and validated for the quantitation of process-specific HCPs. The evolution of this type of HCPs analytical development has been predominantly industry driven, because it has been essential for the development, optimization and validation of proprietary biopharmaceutical purification processes. These coordinated activities are a natural extrapolation of total quality principles [16,18]. Because of their proprietary nature, publication of relatively detailed descriptions of very recent developments has often been significantly delayed.

2. HCP assay development is specification driven

In the final analysis, process design, process optimization, process validation, and contaminant assay development, are specification driven. Thus, it is imperative to identify the approximate specification range expected for HCPs (or any other contaminant), based upon the prospective use of the product [3], as well as current analytical capability. Ultimately, acceptance of a given HCP specification (or the absence of a specification) by regulatory authorities is determined on a "case-by-case" basis [7,9,20], and depends upon a scientifically sound, multifaceted, quality assurance rationale [2,4,5, 10,11,14].

Given the current regulatory environment, and the current state of analytical capability, it is always prudent to approach contaminant specifications conservatively. Thus, the acceptability of an HCP specification for any given biopharmaceutical will vary depending upon combinations of a variety of mitigating factors including: dose size; acute versus chronic administration; use as an analytical or diagnostic reagent, rather than as a therapeutic drug; use as a vaccine; the nature of the expression organism (e.g. prokaryotic versus eukaryotic, untransformed versus transformed cells); and any unique "risk versus benefit" considerations [1–15,20–22,40].

For example, a therapeutic immunoglobulin G (IgG) expressed by a non-recombinant murine hybridoma cell line may require no specification assay for HCPs, although analytical and process validation assurance of the removal of typical cell culture medium proteins like BSA, bovine IgG, insulin, or transferrin is certainly expected (A. Lawton, personal communication). This rationale is based upon many years of manufacturing experience using stringently validated, standardized, immunoglobulin purification methods.

In contrast, a recombinant biopharmaceutical with a novel purification process will require HCP assay development. The required level of quality assurance, and the stringency of the specification, will depend upon one or more of

the factors cited above. A recombinant vaccine which is expressed in yeast, and administered in only one or two doses, may have a relatively high HCP content, e.g. 0.22% [1,34]. In contrast, a recombinant growth hormone produced in bacteria, and destined for chronic administration, will require a much lower HCP content, e.g. less than 10 parts-per-million (ppm) [4,35].

During the earliest stages of purification process design and evolution (pre-IND and into early Phase I [19]) it is advisable to meet informally with regulatory authorities to address the issue of acceptable specification ranges for a novel biopharmaceutical, with special emphasis upon the nature of the purification process, and the intended clinical use of the product.

3. The consensus HCP analytical triad

A consensus quality assurance analytical "triad" has evolved for the control of HCPs in recombinant therapeutic biopharmaceuticals. This strategy includes: sensitive silver staining (and immunoblotting) of electrophoretic gels; rigorous process validation clearance studies; and the development of quantitative, process-specific HCP assays [4,5,10,11,13,16–18,36–40]. The electrophoresis-based characterization methods [40–43], and clearance studies [17], are methodologically generic. In contrast, the specific methodological requirements for development of quantitative HCP assays have been constrained by the need to develop proprietary reagents which are both sensitive to, and specific for, potentially complex mixtures of unique contaminant proteins at very low concentrations [4,44]. In this sense, the quantitative nature of the multi-analyte HCP assay, and the validity of analytical measurement obtained by it, is dependent upon empirical demonstration of the unique stoichiometry of the HCP assay reagents.

4. The requirement for sensitivity, specificity, and stoichiometry

Demonstration of assay sensitivity is not arbitrarily a matter of mere analytical capability [45]. There is a practical requirement to quantitate residual HCPs at a level which is "suitable to obtain regulatory approval". There is no a priori presumption that residual HCPs from one expression organism are more likely to be deleterious to the recipient than are HCPs from another expression organism. However, the intent has always been to avoid "unsafe" levels of residual HCPs which might lead to toxic or immunological sequelae [44], and this concern is not entirely trivial. During early clinical administration of recombinant human growth hormone, it had been observed [46,47] that unacceptable levels of residual *Escherichia coli* (*E. coli*) HCPs may not only elicit anti-HCP antibody, but may also induce an "adjuvant effect" which results in elicitation of undesirable antibody against the biopharmaceutical protein itself [48]. This problem was ameliorated by improved purification methods which significantly decreased the bacterial HCP content of the hormone. No specific anti-HCP antibody elicitation was observed in recipients of recombinant human growth hormone of mammalian cell origin [49].

What is an acceptably safe level for residual HCPs? A detection range of 1–100 ppm of residual HCPs has been quoted as a regulatory (and analytical) benchmark for therapeutic proteins [20]. However, as will be seen later, leading biopharmaceutical firms have designed processes which consistently limit residual HCPs to a range of 1 to 10 ppm, and correspondingly have implemented assays which will consistently quantitate HCPs within, or below, that range [10]. Therefore, domestically, and internationally, for either human or veterinary pharmaceuticals, a useful working target for HCP content is 10 ppm [40]. The sensitivity and specificity of any unique HCP assay which is used to support such a target should be demonstrated accordingly.

Specificity must be defined by the empirical assignment of proprietary HCP detection reagent stoichiometry. These detection reagents are most often polyclonal IgG preparations which have been elicited against a process-specific immunogen that represents the "most probable" [44,52] contaminant HCP which consistently result from a unique purification process.

Thus, combined analytical sensitivity and

specificity has been achieved by employing these proprietary, process-specific HCP immunogens, and their corresponding proprietary, polyclonal, anti-HCP immunoreagents, in the development of "multi-analyte" (multi-antigen) immunoassays [50,51]. These methods generally employ a "sandwich" format [52] and include, but are not limited to: radioimmunoassay (RIA); immunoradiometric assay (IRMA); enzyme-linked immunosorbent assay (ELISA); antigen selected ELISA (ASIA); immunoligand assay (ILA); and electrophoretic, or dot, immunoblots.

5. HCP immunoassay rationale

A quantitative, process-specific, HCP assay may initially be intended for use as a highly sensitive specification assay for profile of the bulk drug. However, it is essential to recognize that HCP assay development is not mutually exclusive of the other parts of the HCP quality assurance analytical triad. By judicious timing of purification process development [76], the anti-HCP immunoreagents which are elicited for use in the specification HCP assay can have a positive, iterative, impact upon optimization of an evolving purification process [77]. This might include qualitative probing for HCPs in electrophoretic immunoblots, or quantitation of HCPs in purification process aliquots. In the later stages of process development HCP assays can ultimately be employed to support definitive process validation studies at full manufacturing scale [76,77]. In this way, a process-specific HCP assay is a means for validating that a process is consistently able to meet an acceptable HCP specification. If a process is designed appropriately for specific removal of contaminants [78] and validation is done properly, it may adequately demonstrate that a specification assay per se will not be required. However, timing is critical. Prolonged timeframes must be anticipated when planning for the manufacture and characterization of complex HCP immunogens, the elicitation, purification and qualification of anti-HCP immunoreagents, and the development and validation of multi-analyte HCP immunoassays. In this sense it is advisible to plan HCP immunoassay development as if it were required for a specification assay.

6. Process-specific HCP immunogen preparation

Within the biopharmaceutical industry the generally accepted paradigm for quantitative, process-specific, HCP assay development is the null cell mock purification model [4,10,13,16, 33,44,51–53]. A null cell (or blank cell) is a version of the expression strain that cannot express product. For example, for a prokaryotic expression system, a bacterial null cell is the expression host cell strain which contains a plasmid vector that has been cured of the expression gene.

For some eukaryotic expression systems the null cell may be more difficult to define. For instance, an acceptable null cell for a recombinant myeloma which expresses IgG as a product might be a "non-secretor null parental cell" (M.A.C. Costello, personal communication). In any event, the genotypic and phenotypic relationship of the null cell strain to the expression strain, and the analytical rationale for specifically selecting it, must be clearly defined and documented as early as possible during purification process development.

Mock fermentation of the null cell strain, and purification of null cell HCPs, should be executed under normal production conditions, and "acceptably close" to normal production scale. Regulatory input may be required in order to determine what production level is acceptable, in that full scale HCP manufacture for some eukaryotic expression systems may be impracticable, or possibly prohibitively expensive (V.R. Anicetti, personal communication). Ideally, subsequent isolation and purification of the null cell HCP should proceed to that point in the purification process at which the biopharmaceutical would normally approach 95–99% purity. This potentially complex mixture of null cell HCPs conservatively represents the "most probable" [44,52] downstream HCP contaminants, and is used as an immunogen for eliciting proprietary,

process-specific, HCP immunoreagents. A substantial amount of this preparation (e.g. 200 mg or more) may be required, not only for specific use as an immunogen, but for other purposes such as analytical characterization, possible use as an immunoaffinity matrix [58], and use as an assay standard over the lifetime of the product. Acquisition of this immunogen is usually time-consuming, labor intensive, and expensive.

After this immunogen is isolated, the purification process steps upstream of the immunogen selection point, and the expression organism itself, should not be significantly altered. To do so may invalidate all previously completed HCP assay development activities. After assay development is initiated, any potential changes or "improvements" to the purification process *upstream* of the immunogen selection point, *however minor they may appear to be*, require the approval of the purification process developer, quality assurance professionals, *and* the HCP assay developer. Identifying the difference between a "major" process change and a "minor" process change is sometimes a matter of serious contention. In many instances, regulatory agency approval should be obtained, as well.

The need for a decision to designate the purification process upstream of the prospective HCP immunogen selection point as "invariant" should be anticipated by Phase I, and implemented by Phase II, in order to allow enough time to have a process-specific HCP assay in place by Phase III [19]. This timing is particularly critical, since HCP assay development is expected to have a positive, iterative impact on purification process improvement and validation.

Because of the time and expense involved in completing these activities, it is important to avoid having to repeat them. Thus, it is essential that the null cell HCP immunogen preparation not be contaminated with the biopharmaceutical product which is usually purified by the process [51,52,61]. This situation may be avoided by any, or all, of the following: use of dedicated equipment which has not been exposed to the biopharmaceutical of interest; use of equipment which has been exhaustively cleaned and validated; or by contracting with a consultant firm for partial or complete manufacture of the null cell HCP immunogen in an outside facility. Two-dimensional silver-stained gel separations of prospective immunogen HCPs may be compared with duplicate immunoblots probed with a sensitive, and specific, immunoreagent for the biopharmaceutical product. This type of comparison can give reasonable assurance that the HCP immunogen is free of detectable contamination.

7. Anti-HCP immunoreagent development

In one conceptually simple case, S3 ribosomal protein, a unique, major HCP contaminant of recombinant human acidic fibroblast growth factor, was quantitated using specific monoclonal antibodies [54]. However, while this approach may be feasible for unusually simple immunogens, it would obviously be much more difficult to apply as the prospective HCP immunogen becomes more complex. There may be hundreds of potential HCP contaminants in a given immunogen preparation, and the individual HCP will differ widely in relative concentration and relative antigenicity. Thus, the practical goal becomes elicitation of a polyclonal immunoreagent which has demonstrable specificity, and sensitivity, for each of the antigenic proteins in the complex HCP immunogen. Using adjuvants, an efficient sequential immunization protocol, and a nominal boost-bleed cycle of 3 to 4 weeks, it may take 3–6 months to elicit, purify, and characterize an acceptable polyclonal anti-HCP immunoreagent against a very complex HCP immunogen. Any one of a variety of methods for purifying the IgG are acceptable. However, semi-preparative immunoaffinity chromatography (e.g. using Protein G) has been most efficient (this laboratory).

One must select an animal host which gives a strong immune response, and which yields appreciable amounts of immunospecific IgG. Specific pathogen free NZW rabbits are often a good choice. In order to minimize individual variations in immune competance, multiple animals (3–6) of a given species must be used, and

in some instances more than one species may be used.

Pragmatically, one must first assure a suitable immune response to the strong, or major, HCP antigens, and then induce a progressively enhanced response to the weaker, or minor, HCP antigens. The immune response to a limited number of HCP antigens may be relatively rapid, particularly when adjuvants are used to boost the immune response. However, in complex mixtures, differential enhancement of the immune response toward weaker, or lower concentration, antigens requires differential immunosuppression of major antigens. This may be accomplished by a couple of methods, either individually, or sequentially.

One method, passive immunization [55], involves purification of IgG from successive bleeds of immunized animals. The purified IgG is administered to the animals at the time of the next immunization along with the next dose of complete HCP immunogen. This IgG, which will have specificity against major antigens, tends to suppress further recognition of those same antigens in vivo. Immune recognition of, and antigenic response to, weaker antigens is promoted. This procedure may be performed at each boost-bleed cycle. Thus, the progressive immune response against weaker antigens is enhanced.

Another method, cascade immunization [56,57], also requires that immune IgG from successive bleeds be purified. However, this purified IgG is covalently affixed to a chromatographic support, and is used to immunospecifically adsorb major antigens from the HCP immunogen itself. The "adsorbed" HCP immunogen, which has been relatively depleted for major antigens in vitro, and thus relatively enriched for minor antigens, is administered in the next boost. This "antigen selection" facilitates immune recognition of, and response to, minor antigens. This procedure may be performed at each boost-bleed cycle. In this way, the immune response against progressively weaker antigens is enhanced.

To achieve true quantitation in complex, multi-analyte immunoassays, it is essential that the immunoreagent simultaneously specifically recognize, and be in stoichiometric excess to, all of the individual analytes of interest in an assay sample [50]. Demonstration of acceptable stoichiometry, and hence the ultimate acceptability of the anti-HCP immunoreagent, is strictly empirical. The progressive immune response achieved for a given immunization scheme may be monitored by using purified IgG from each sequential bleed to immunospecifically probe blots of HCP immunogen which has been adequately separated by electrophoresis. If the immunogen is simple, mono-dimensional electrophoresis [e.g. sodium dodecylsulphatepolyacrylamide gel electrophoresis (SDS-PAGE)] may be sufficient. If the immunogen is complex, two-dimensional electrophoresis (e.g. isoelectric focusing/SDS-PAGE) may be required to give optimal separation [41–43]. Ideally, all of the proteins detected by sensitive silver stain of an HCP immunogen electrophoretic gel should eventually also be detected in an immunoblot by the progressively enhanced anti-HCP immune IgG [57]. In practice, some minor proteins which elicit a strong immune response may not be visible in the silver stained gel, and some poorly antigenic proteins which are detectable by silver stain may never elicit a detectable immune response.

Anti-HCP immunoreagent specificity and stoichiometry are thus defined by, and limited to, those HCP immunogen proteins which have been demonstrated on blots to be specifically detectable by the immunoreagent. Sensitivity of the immunoreagent will ultimately be defined during assay development and validation, and should permit quantitation at ppm levels. Thus, the practical endpoint of anti-HCP immunoreagent development occurs when the investigator is satisfied that most of the HCP immunogen proteins are detectable by the immunoreagent, at an acceptable level of sensitivity.

Reagent stoichiometry may be further enhanced by immunospecific subpurification of specific anti-HCP IgG away from pre-existing IgG. For example, HCP immunogen may be covalently affixed to a chromatographic support and used as an immunoaffinity purification ma-

trix to obtain immuno-specifically subpurified anti-HCP IgG [58]. Microgram quantities of this pure reagent, used to detect nanograms to picograms of contaminant proteins in an individual assay sample would, by definition, be in stoichiometric excess [50]. In practice, an acceptable region, or range, of antibody excess in the standard curve of a given assay may be demonstrated empirically by spiked-addition (spike-recovery) studies, whether the immunoreagent is subpurified, or not [79].

Demonstration of stoichiometric immunoreagent excess for some assay samples may only be achieved by analyzing a dilution series [52,58,65]. It may be virtually impossible to achieve true antibody excess with the specific immunoreagent IgG *versus all proteins in a complex mixture* when attempting to assay upstream purification process aliquots which contain very high total HCP levels (D.V. Sinicropi; personal communication).

8. Examples of HCP immunoassays

Once suitable stoichiometry of the immunoreagent has been defined, HCP assay development is relatively straightforward. Several immunoassay methodologies have been employed for quantitative (and semi-quantitative) polyclonal HCP immunoassays. Examples include: an ELISA with nanogram level sensitivity for yeast HCPs in recombinant hepatitis B surface antigen vaccine [34]; an ELISA with ppm sensitivity for *E. coli* HCPs in recombinant human interferon gamma [58]; an ELISA with ppm sensitivity for *E. coli* HCPs in recombinant human growth hormone [52,59]; an ELISA for mammalian "cellular" HCPs (not media proteins) in recombinant human clotting Factor VIII:C [60]; an ELISA with ppm sensitivity for mouse fibroblast HCPs in recombinant human erythropoetin [61]; RIA/IRMA [62,63] or ELISA [64] with ppm sensitivity for *E. coli* HCPs in recombinant human insulin; an antigen-selected ELISA (ASIA) with ppm sensitivity for chinese hamster ovary (CHO) cell HCPs in a recombinant human tissue plasminogen activator [65]; and an "ultrasensitive" ELISA for quantitation of CHO HCPs in therapeutic recombinant human pancreatic DNAse I (D.V. Sinicropi; personal communication).

A semi-quantitative, monoclonal antibody immunoblot with 100 ppm discrimination for *E. coli* S3 ribosomal protein in recombinant human acidic fibroblast growth factor was previously mentioned [54]. A semi-quantitative, polyclonal, dot-blot immunoassay with 100 ppm discrimination has been developed for quantitation of *E. coli* HCPs in each of two recombinant malarial vaccine candidates which are purified by the same process [66].

The relatively recent introduction of immunoligand assay (ILA) technology [67] shows promise in shortening individual assay turnaround times per se, as well as for rapidly accelerating the pace of HCP immunoassay development in general. This is due to adaptation of a sensitive detection system [68] originally designed for generic quantitation of DNA at picogram levels [30], and to the introduction of commercially available IgG labeling reagents (Molecular Devices Corp., Menlo Park, CA, USA) which are compatible with the detector technology. Sensitive ILA methods have been developed for quantitation of CHO HCPs at ppm levels in recombinant human erythropoetin [69], for quantitation of *E. coli* HCPs at ppm levels in recombinant human basic fibroblast growth factor [70], for quantitation of *E. coli* HCP at ppm levels in recombinant human alpha interferon (M.A.C. Costello; personal communication), and for quantitation of residual *E. coli* HCP at ppm levels in recombinant bovine somatotropin (this laboratory [79]). ILA methods have been developed for rapid and sensitive quantitation of individual residual media proteins, affinity ligands, and immunoglobulins, as well (H. Merrick, personal communication; [75]).

Many other HCP immunoassays and immunoligand assays for quantitation of host cell or process contaminants have been developed and implemented in the biopharmaceutical industry in recent years. However, the proprietary nature of the purification processes upon which most of the recently developed process-specific HCP

assays are based, continues to preclude detailed descriptions in the open literature. As more biopharmaceuticals reach market this publishing situation should improve. In general, the thrust of current and future HCP immunoassay development is to shorten assay development timeframes, to shorten assay turnaround times, and to continue to improve analytical sensitivity below ppm levels.

9. Alternative assay methods

Currently there is no acceptable alternative to the use of highly specific multianalyte immunoassays to obtain quantitation of process-specific HCP in biopharmaceuticals, especially when ppm sensitivity for mixtures of proteins is required. Purely chromatographic separations have been inadequate except in the case of single impurities [54,71]. However, very recent developments in multidimensional chromatographic/electrophoretic separations, using post column reaction fluorescence detection, show great potential for use in sensitive, non-immunologically based HCP assay development [72–74]. This methodology is being aggressively pursued by leaders in the biopharmaceutical industry (D.V. Sinicropi, personal communication).

10. Conclusions

Quality control of residual HCPs in biopharmaceuticals is currently based upon a consensus triad of quality assurance analytical methods. This "regulatory package" of methods includes: silver-stained electrophoretic gels and immunoblots; process validation clearance studies based upon sound process design; and process-specific host cell protein assays. This triad of assurance, and the specific need to develop sensitive multi-analyte immunoassays for HCP, is specification driven. The desired approximate specification range for HCP for a given biopharmaceutical must be identified early in process development, and is related to the prospective use of the product. Often this specification will fall into a range of 1–10 ppm. It is prudent to include planning for development of a process-specific HCP immunoassay with part-per-million sensitivity into the earliest stages of product development. Timing is critical due to the prolonged time frames associated with the null cell mock purification model currently employed for multi-analyte HCP assay development.

In a practical sense, the triad of quality assurance methods are not mutually exclusive. While each method alone is insufficient to guarantee consistent removal of HCPs to desired levels, all three together give mutually supporting assurance. Anti-HCP immunoreagents generated for HCP assay development may also be used semi-quantitatively for early process optimization, or later quantitatively for process validation. HCP assay development can have a positive, iterative impact upon an evolving purification process, with respect to assuring consistency in achieving a desired specification range for the bulk drug.

Once process consistency for removal of HCPs to desired levels has been rigorously validated using the quality assurance triad, a specification assay for HCPs per se may not be required at all, or perhaps only on an audit basis. In this sense HCP immunoassay development may be thought of as a quantitative tool for validation of a consistent biopharmaceutical purification process, rather than as an end in itself. Combined qualitative and quantitative demonstration of process consistency for removal of contaminants is a natural extension of total quality principles [16,18]. These activities, in conjunction with properly designed pivotal toxicology and safety studies, should be suitable to formulate a persuasive contaminant specification rationale for any biopharmaceutical regulatory submission.

Acknowledgements

It is not possible to list everyone with whom I have discussed these issues over the past several years, but I would like to thank the following individuals for relatively recent productive discussions: Vincent R. Anicetti, Maureen A.C.

Costello, Gregory C. Davis, Ibrahim Ghobrial, Thomas C. Hageman, William S. Hancock, James W. Jorgenson, Ashley Lawton, Holly Merrick, Kenneth Seamon, Gail Sofer and Dominic V. Sinicropi.

References

[1] WHO Consultation, *Bull. World Health Org.*, 61 (1983) 897–911.

[2] R. Baffi, *Standard Guide for Determination of Purity, Impurities, and Contaminants in Biological Drug Products, Book of Standards E1298-89, American Society for Testing and Materials*, Philadelphia, PA, 11.04 (1989) 992–994.

[3] E.W. Leser and J.A. Asenjo, *J. Chromatogr.*, 584 (1992) 43–57.

[4] A.J.S. Jones and J.V. O'Connor, *Dev. Biol. Stand.*, 59 (1985) 175–180.

[5] W.S. Hancock, *Chromatography Forum*, September/October (1986) 57–59.

[6] S.J. Brewer, *J. Chem. Tech. Biotechnol.*, 37 (1986) 367–373.

[7] F.A. Robey, in K.A. Walsh (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1986, pp. 67–78.

[8] R.J. Mattaliano, J.J. Rosa, C. Foeller, J.P. Woodard and M.J. Bertolini, in K.A. Walsh (Editor), *Methods in Protein Sequence Analysis*, Humana Press, Clifton, NJ, 1986, pp. 79–95.

[9] D.T. Liu, N. Goldman and F. Gates III, in M. Phillips, S.P. Shoemaker, R.D. Middlekauff and R.M. Ottenbrite (Editors), *The Impact of Chemistry on Biotechnology: Multidisciplinary Discussions*, Vol. 362, American Chemical Society, Washington, DC, 1988, pp. 162–171.

[10] V.R. Anicetti, B.A. Keyt and W.S. Hancock, *Trends. Biotechnol.*, 7 (1989) 342–349.

[11] R.L. Garnick, N.J. Solli and P.A. Papa, *Anal. Chem.*, 60 (1988) 2546–2557.

[12] *Biotech Headings for Notice to Applicants, part II. Draft No. 7.* Commission of the European Communities, Brussels, April 1991, p. 5.

[13] J. Briggs and P.R. Panfili, *Anal. Chem.*, 63 (1991) 850–859.

[14] V.P. Garg, M.A.C. Costello and B.A. Czuba, *Bioprocess. Technol.*, 12 (1991) 29–54.

[15] C. Schoneich, S.K. Kwok, G.S. Wilson, S.R. Rabel, J.F. Stobaugh, T.D. Williams and D.G. Vander Velde, *Anal. Chem.*, 65 (1993) 67R–84R.

[16] H.L. Levine, S.J. Tarnowski, M. Dosmar, D.M. Fenton, J.N. Gardner, T.C. Hageman, P. Lu, G. Sofer and B. Steineger, *J. Parent. Sci. Technol.*, 46 (1992) 87–97.

[17] T.C. Hageman, *BioPharm*, 4 (1991) 39–41.

[18] T.J. Schoemaker, *Pharm. Engineer.*, 11 (1991) 23–27.

[19] N.J. Chew, *BioPharm*, 9 (1991) 16–20.

[20] D.T. Liu, *Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology*, Department of Health and Human Services, OBRR, NCDB, FDA, November 18, 1983.

[21] T. Hoffman, *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*, Department of Health and Human Services, OBRR, CDB, FDA June 1, 1987.

[22] K.C. Zoon, *Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals*, Department of Health and Human Services, CBER, FDA, May 17, 1993.

[23] P. Gavit, M. Walker, T. Wheeler, P. Bui, S.-P. Lei and J. Weickmann. *BioPharm*, 5 (1992) 28, 29, 32–33, 58.

[24] L.S. Hanna, P. Pine, G. Reuzinsky, S. Nigam and D.R. Olmstead, *BioPharm*, 9 (1991) 33–37.

[25] C. Lucas, C. Nelson, M.L. Peterson, S. Frie, D. Vetterlein, T. Gregory and A.B. Chen, *J. Immunol. Methods*, 113 (1988) 113–122.

[26] J.-P. Himanen, M. Sarvas and I.M. Helander, *Vaccine*, 11 (1993) 970–973.

[27] *USP XXII/NF XVII, Bacterial Endotoxin Test (85), New Supplement 7*, US Pharmacopeial Convention, January 1, 1990.

[28] L.C. Eaton, *J. Pharm. Biomed. Anal.*, 7 (1989) 633–638.

[29] J. Briggs, V.T. Kung, G. Pontis, P. Nagainis, P. Panfili, C. Chow, T. Smith, E. Sheldon, D. Ross, R. Smethers, B. Gomez, M. Lacy, V. Drbal, R. King and R. Zuk, *Amer. Biotechnol. Lab.*, February, 1989.

[30] V.T. Kung, P.R. Panfili, E.L. Sheldon, R.S. King, P.A. Nagainis, B. Gomez, D.A. Ross, J. Briggs and R.F. Zuk, *Anal. Biochem.*, 187 (1990) 220–227.

[31] S. Mizusawa, R. Homma, Y. Miura, Y. Kobayashi, K. Misu and K. Komuro, *Pharm. Tech. Japan*, 7 (1991) 1–9.

[32] E.M. White, J.B. Grun, C.-S. Sun and A.F. Sito, BioPharm, 4 (1991) 34–39.

[33] A.J.S. Jones, in M. Phillips, P. Shoemaker, R.D. Middlekauff and R.M. Ottenbrite (Editors), *The Impact of Chemistry on Biotechnology, Multidisciplinary Discussions*, Vol. 362, American Chemical Society, Washington, DC, 1988, pp. 193–201.

[34] T. Ohmura, A. Ohmizu, A. Sumi, W. Ohtani, Y. Uemura, H. Arimura, M. Nishida, Y. Kohama, M. Okabe, T. Mimura and C.M. Heldebrant, *Biochem. Biophys. Res. Commun.*, 149 (1987) 1172–1178.

[35] P. Gellerfors, G. Eketorp, K. Fholenhag, B. Pavlu, S. Johansson and L. Fryklund, *J. Pharm. Biomed. Anal.*, 7 (1989) 173–183.

[36] K. Veeraragavan, *FEMS Microbiol. Lett.*, 61 (1989) 149–152.

[37] R.A. Hart, U. Rinas and J.E. Bailey, *J. Biol. Chem.*, 266 (1990) 12728–12733.

[38] U. Rinas, T.C. Boone and J.E. Bailey, *J. Biotechnol.*, 28 (1993) 313–320.

[39] U. Rinas and J.E. Bailey, *Appl. Microbiol. Biotechnol.*, 37 (1992) 609–614.
[40] R.P. Gooding and A.F. Bristow, *J. Pharm. Pharmacol.*, 37 (1985) 781–786.
[41] E.R. Tovey and G. Bers, in B.A. Baldo, E.R. Tovey and N.S.W. St. Leonards (Editors), *Protein Blotting: Methodology, Research and Diagnostic Applications*, Karger, New York, 1989, pp. 5–42.
[42] C.R. Merrill and D. Goldman, in J.E. Celis and R. Bravo (Editors), *Two-Dimensional Electrophoresis: Methods and Applications*, Academic Press, New York, 1984, pp. 4–36.
[43] B.S. Dunbar, *Two-Dimensional Electrophoresis and Immunological Techniques*, Plenum Press, New York, 1987, pp. 1–76, 123–159.
[44] V.R. Anicetti, in Cs. Horváth and J.G. Nikelly (Editors), *Analytical Biotechnology: Capillary Electrophoresis and Chromatography (ACS Symposium Series*, No. 434), American Chemical Society, Washington, DC, 1990, pp. 127–140.
[45] W.R. Griswold, *Molec. Immunol.*, 24 (1987) 1291–1294.
[46] J. Tyllstrom, B. Karlen and O. Guilbaud, in R.D.G. Milner and H. Flodh (Editors), *Somatonorm (Somatrem): Immunological aspects—Immunological Aspects of Human Growth Hormone*, Oxford, 1985, pp. 61–68.
[47] J.R. Bierich, *Acta Paediatr. Scand.*, 325 (suppl) (1986) 13–18.
[48] L. Fryklund, J. Brandt, M. Hagerman, B. Pavlu, B. Skoog and A. Wichman, in S. Raiti and R.A. Tolman (Editors), *Methionyl Human Growth Hormone (Somatonorm): Purity and Properties—Human Growth Hormones*, Plenum, New York, 1986, pp. 257–266.
[49] H.J. Zeisel, A. Lutz and W.v. Petrykowski, *Horm. Res.*, 37 (suppl) (1992) 47–55.
[50] R.P. Ekins, *J. Pharm. Biomed. Anal.*, 7 (1989) 155–168.
[51] United States Pharmacopeial Convention, *Biotechnology Articles: In-Process Revision, Pharmacopeial Forum*, May–June, 1990, pp. 490–517.
[52] V.R. Anicetti, E.F. Fehskens, B.R. Reed, A.B. Chen, P. Moore, M.D. Geier and A.J.S. Jones, *J. Immunol. Meth.*, 91 (1986) 213–224.
[53] I.S. Johnson, *Science*, 219 (1983) 632–637.
[54] D.O. O'Keefe, P. DePhillips and M.L. Will, *Pharm. Res.*, 10 (1993) 975–979.
[55] J. Thalhammer and J. Freund, *J. Immunol. Meth.*, 80 (1985) 7–13.
[56] J. Thalhammer and J. Freund, *J. Immunol. Meth.*, 66 (1984) 245–251.
[57] V.R. Anicetti, M.A. Simonetti, L.L. Blackwood, A.J.S. Jones and A.B. Chen, *Appl. Biochem. Biotechnol.*, 22 (1989) 151–168.
[58] A.B. Chen, A.A. Chempionsmith, J. Blanchard, J. Gorrell, B.A. Niepelt, M.M. Federici, J. Formento and D.V. Sinicropi, *Appl. Biochem. Biotechnol.*, 36 (1992) 137–152.
[59] US Pharmacopeial Convention, *Somatrem, In-process revision, Pharmacopeial Forum*, January–February, 1993, pp. 4518–4525.
[60] J.E. Brown, *Eur. Pat. Appl.* EP 379887 A2, August 1, 1990.
[61] J.A. Pauly, B. Siebold, R. Schulz, W. List, G. Luben and F.R. Seiler, *Behring Inst. Mitt.*, 86 (1990) 192–207.
[62] R.S. Baker, J.R. Schmidtke, J.W. Ross and W.C. Smith, *The Lancet*, ii (1981) 1139–1142.
[63] I.S. Johnson, *Diabetes Care*, 5 (suppl) (1982) 4–12.
[64] United States Pharmacopeial Convention, *Purity: Insulin related substances, USP Open Conference, Biotechnology-derived Products/Drafts*, 1989, pp. 110–112.
[65] A.J.S. Jones and R.L. Garnick, in A.S. Lubiniecki (Editor), *Large Scale Mammalian Cell Culture Technology*, Marcel Dekker, New York, 1990, pp. 543–566.
[66] B.J. Takacs and M.-F. Girard, *J. Immunol. Meth.*, 143 (1991) 231–240.
[67] *Threshold Application Note, Host cell-derived protein (HCP) assays using the Threshold Immunoligand Assay (ILA) system. Immunoligand Assay Series*, Molecular Devices, Menlo Park, CA, March, 1991.
[68] D.G. Hafeman, J.W. Parce and H.M. McConnell, *Science*, 240 (1988) 1182–1185.
[69] I. Ghobrial, D. Wong and B. Sharma, presented at the International Biotechnology Expo and Scientific Conference, San Francisco, CA, 11–13 October, 1992, poster.
[70] C.C. Ting, presented at the *199th ACS Meeting, 22–27 April* 1990, Division of Analytical Chemistry, American Chemical Society, Washington, DC, abstract 52.
[71] D.O. O'Keefe and M.L. Will, in S. Ahuja (Editor), *Chromatography of Pharmaceuticals (ACS Symposium Series*, No. 512), American Chemical Society, Washington, DC, 1992, pp. 121–134.
[72] A.V. Lemmo and J.W. Jorgenson, *J. Chromatogr.*, 633 (1993) 213–220.
[73] A.V. Lemmo and J.W. Jorgenson, *Anal. Chem.*, 65 (1993) 1576–1581.
[74] J.P. Larmann, A.V. Lemmo, A.W. Moore, Jr. and J.W. Jorgenson, *Electrophoresis*, 14 (1993) 439–447.
[75] J. Briggs, V.T. Kung, B. Gomez, K.C. Kasper, P.A. Nagainis, R.S. Masino, L.S. Rice, R.F. Zuk and V.E. Ghazarossian, *BioTechniques*, 9(5) (1990) 598–606.
[76] J. Akers, J. McEntire and G. Sofer, *BioPharm*, 7(2) (1994) 54–56.
[77] G. Sofer, *BioPharm*, 7(4) (1994) 46–50.
[78] M. Hansson, S. Stahl, R. Hjorth, M. Uhlen and T. Moks, *Bio/Technology*, 12 (1994) 285–288.
[79] M.L. Schafer and L.C. Eaton, in preparation.

ELSEVIER

Journal of Chromatography A, 705 (1995) 115–127

JOURNAL OF CHROMATOGRAPHY A

Review

Moisture content in proteins: its effects and measurement

John K. Towns

Biopharmaceutical Analytical Development, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA

Abstract

Residual moisture content has a significant impact on the solid-state stability of biopharmaceutical products. Protein degradation due to residual moisture is minimal at or below the monolayer level of hydration owing to low availability of water and limited dynamic activity of the protein. However, residual moisture content beyond a monolayer generally results in increased rates of decomposition due to the enhanced conformational flexibility of the protein and the ability of the less tightly bound water to mobilize reactants. In addition to moisture content, the temperature and the composition of the lyophilized plug are important variables dictating the stability of proteins in the amorphous solid state. Water can act as a plasticizer to reduce the glass transition temperature, T_g, of the amorphous polymer, thus an increase in temperature or plasticizer level can result in a phase transition from a dynamically constrained state to a dynamically relaxed state. The selection of excipients can have a large impact on water–protein interactions as small ions and/or crystallization of excipients can redistribute water available to the protein. Owing to the key role that water content plays in the conformational and/or chemical state of the protein, an accurate and precise moisture determination is essential in resolving stability issues. A wide variety of techniques for the determination of moisture content have been utilized, with special attention being placed on sample handling to minimize atmospheric moisture contamination.

Contents

0021-9673/95/$29.00
SSDI 0021-9673(94)01257-1

1. Introduction

Recombinant DNA techniques have resulted in great advances in biotechnology, making it possible to produce a variety of proteins for biopharmaceutical use. Many proteins, however, have poor stability in aqueous solution and the freezing of the solution or the addition of preservatives fails to provide the required product stability [1]. To prevent excessive loss of potency or excessive increase in the level of decomposition, pharmaceutical protein products are often freeze-dried (lyophilized). Freeze-drying, however, is still usually not sufficient, as many products must be refrigerated at 5°C to provide reasonable shelf-life (1–2 years).

The process of lyophilization removes water from a system based on the principle of ice sublimation at reduced pressure. When a protein solution is lyophilized, the bulk water that resides in ice matrices of the frozen solution sublimes first. The multilayer water surrounding the protein molecule is then removed, leaving a residual monolayer of water on the surface of the protein. This operation allows drying of heat-labile materials to low residual moisture content under moderate temperature conditions. If the product is intended for parenteral use, then the protein is usually lyophilized in a final container such as a flame-sealed glass ampoule or a glass container with rubber closure that is usually sealed under vacuum or nitrogen. The water content of the freeze-dried material in the final container may vary depending on the freeze-drying process and may increase during storage [2].

The amount of water present in the protein has a significant impact on stability and is a concern for both bulk solid and lyophilized formulations. Residual moisture refers to the low level of surface water, ranging from less than 1% to 5%, remaining in a freeze-dried biological product after the bulk of the aqueous solvent has been removed [3]. Residual moisture should not compromise the potency and integrity of the product. The appropriate level of residual moisture to optimize stability is largely dependent on the particular protein's decomposition pathway. The generally accepted view is the drier the better. However, levels of residual moisture for certain products should not be so low that overdrying adversely affects product stability. Water retention varies with the type of water present, bound, surface and/or trapped, and is different for each product. Because more than one type of water may exist in a freeze-dried biological product, different moisture results may be found when different methods are employed in determining the moisture content of the sample [3].

Regulations of the Center for Biologics Evaluation and Research (CBER) require that each lot of freeze-dried biological product be tested for residual moisture [4]. The regulations pertaining to residual moisture are published in Title 21 of the Code of Federal Regulations for Food and Drugs and require that moisture levels meet and not exceed established limits as specified by an approved method on file in the product license application [5]. Guidelines for the determination of residual moisture in dried biological products have been issued under 21 CFR 10.90 to describe residual moisture test methods and procedures used to set product residual moisture limits [6]. For most products, levels of residual moisture should range from less than 1.0% to 3.0% so that the chemical and/or conformational stability, and therefore potency of the product, are not compromised over time.

This paper addresses the importance of residual moisture on protein stability and the need to control and measure water content accurately. An overview of the binding of water and its effects on protein stability is included. Examples illustrate how the amount of water present in the lyophilized plug can have a significant impact on solid state stability. Many of the examples emphasize that a specific degradation pathway is contingent on the increased flexibility of the protein to expose amino acid residues to the surrounding environment. The methods employed in the measurement of residual moisture content and the importance of controlling atmospheric moisture contamination are discussed.

2. Water sorption and protein mobility

The importance of water sorption on the solid-state stability of proteins can be addressed through an understanding of the binding of the sorbed water. The binding of water to proteins is a result of numerous molecular and intermolecular interactions due to hydrophilic, hydrophobic and ionized groups on the protein. The water content is determined following equilibration at a given relative humidity. A sorption isotherm is generated by assessing equilibrium moisture content at varying relative humidities. The sorption isotherm for proteins can be roughly separated into three regions [7]. The first region is binding of water to highly active sites such as charged and highly polar groups. The second region is a transition region from monolayer to multilayer coverage. It occurs with the binding of water to weaker sorption sites such as the peptide backbone and polar surface groups. Additional water binding occurs via clustering at or near charged and highly polar groups and through filling of voids created by swelling of the polymer. The last region, or multilayer region, occurs with condensation of water at very weak binding sites and layering of loosely held water [8]. It is in this region, hydration of 30–40%, that the true monolayer coverage of a protein actually occurs [9].

The behavior of a system can be predicted using isotherms and existing mathematical models to identify several areas of the isotherms that represent distinctly different characteristics of the solid. The uptake of water by proteins occurs with penetration into the disordered structure of the solid and is not limited to surface adsorption [10]. At levels of water below the Brunauer–Emmet–Teller (BET) monolayer level, generally 5–9% water content, the mobility of the absorbed water and flexibility of the protein are limited [8]. At or near this level of hydration, there is some indication that secondary relaxation of the amorphous system may occur and many decomposition pathways become observable over a reasonable time frame. The dynamic mobility of the system then slowly increases with increasing hydration. The dynamic flexibility of a protein in the solid state is dependent on interrelated variables of temperature, hydration and composition of the amorphous protein.

The adsorbed water may act as a plasticizer, resulting in an increase in free volume and greater macroscopic mobility [11]. The outcome of this plasticization is a lowering of the glass transition temperature, T_g, below the temperature of the surroundings. The phase transition from the dynamically constrained "glassy" state to the dynamically relaxed viscoelastic state is possible when the temperature is raised above T_g [12]. This phase transition can also occur when sufficient water has been absorbed in the amorphous solid to lower T_g below the temperature of the surroundings.

Reactions dependent upon the mobility of the water and/or the protein are therefore greatly enhanced at temperatures above T_g [13]. Increases in both the mobility of the water and protein as a result of the plasticizing effect of water are generally related to the monolayer moisture content. An increase in plasticization of the protein due to water sorption has been observed to result in a rapidly decreasing T_g with increasing moisture up to the BET monolayer region [14]. The onset of internal protein flexibility correlates reasonably well with the attainment of the BET monolayer level of water [9]. Significant increases in internal motions of the protein begin at hydration levels slightly greater than the BET monolayer. For proteins, the monolayer approximates the amount of water vapor necessary to cover the highly active heterogeneous sorption sites. Reasonable estimates of the BET monolayer were found for methionyl human growth hormone (met-hGH) and rt-PA by considering one water molecule per polar group [1].

It is a concern that the BET equation may not be adequate for determining the monolayer water content in proteins, owing to the unequal affinity of water associated with the weak and strong binding sites [8]. The chemical composition of the protein may dictate the level of residual moisture adsorbed on to the surface of the protein. As an example, met-hGH BET

monolayer results are in good agreement with the theoretical calculation employing only strong polar groups [1]. It is likely that the weakly polar and non-polar groups do not contribute significantly to the formation of the monolayer adsorbed on the strong polar groups of met-hGH.

3. Water–protein interactions

Water in the monolayer is thought to have low thermodynamic activity owing to its strong interaction with the protein molecule. This is in contrast to water in the multilayer that is thought to retain more free water activity depending on its distance from the surface of the protein. It is not the absolute residual water content that is important for maintaining protein stability, but rather the way in which residual moisture is adsorbed and how it is available, i.e., its water activity. The hydration of proteins results from coulombic, hydrogen bonding, Van der Waals and hydrophobic interactions between water and specific functional groups of the protein [8]. The early stage of hydration involves water adsorption and its predominant interaction with charged groups. The strength of these interactions can vary significantly as the amino acid residues provides large differences in ionic, polar and non-polar sites. These interactions are further complicated by neighboring amino acids in the folded protein. Unless all of the decomposition pathways are similarly affected by moisture, the sensitivity of long-term stability will be dependent on the sensitivity of the relevant pathway to increased moisture content levels.

At high residual moisture content, the possibility of chemical reactions occurring (cleavage, oxidation, deamidation, denaturation, aggregation, etc.) is increased. This is due to conformational flexibility of the protein molecule, the availability of water for hydrolytic reactions and the increased mobility of reactants. Owing to high protein mobility and flexibility, the protein backbone segments and/or amino acids functional groups can have appropriate orientation and energy to participate in chemical reactions. When the moisture content is decreased, the probability of reactions occurring is consequently reduced. The appropriate level of moisture to optimize stability varies from product to product and is dependent on the protein and its particular degradation pathway(s) [1,15–23]. Protein reactions can be minimized by reducing the moisture content to the monolayer water level or less. However, an optimum moisture content is necessary to maintain protein activity during storage. Overdrying will remove water from the monolayer, leading to increased exposure of the protein surface and to various unwanted reactions. Too low residual moisture levels may cause aggregation, loss of activity and/or inadequate reconstitution.

The effect of water content and/or water activity on the solid-state stability of proteins results from either (1) changes in dynamic activity or conformational stability of the protein or (2) participation of water as a reactant or medium for mobilization of reactants. Studies on the hydration of proteins have identified several critical levels of hydration at which significant changes in properties of the protein and the bound water occur [9]. Most decomposition reactions are minimal at or below the monolayer level of hydration owing to the low availability of the water and limited dynamic activity of the protein. Residual water in excess of monolayer coverage increases molecular mobility in the solid protein, thereby increasing general reactivity, resulting in increased rates of decomposition.

Dynamic motions within proteins impose a significant degree of flexibility in the conformational structure. The impact of water on the conformational stability of proteins depends on its flexibility and the ease of conformation structure variations. The outcome of such unfolding can be irreversible physical aggregation induced by exposure of hydrophobic amino acid residues [2,16]. Increases in this flexibility are reasonably well correlated with decreases in conformational stability. Decomposition pathways related to covalent bond formation or cleavage are also expected to require flexibility, of at least a localized segmental motion, in order for orientational constraints to be overcome and chemical reactions to occur.

A high water content may decrease the protein stability in freeze-dried solids via several mechanisms. Chemical modification generally results in changes to the primary sequence and may or may not have a subsequent effect on conformational structure. Chemical modifications can result from both intra- and inter-molecular reactions or involve reactions with other components of a heterogeneous system. These increased rates are primarily due to the increased conformational flexibility of the protein and the ability of the less tightly bound water to mobilize reactants. This is observed in studies of the Maillard reaction, where reaction rates are increased due to increased reactant mobilization [24]. The reaction is initiated by Schiff base formation between amino nucleophiles of the protein, such as lysine, and the carbonyl groups of the reducing sugars. This leads to the "browning reaction" as the Schiff bases undergo further rearrangement with degradation to unsaturated carbonyls and eventually polymerization. At water levels below the BET monolayer level, the reaction is minimal owing to the low mobility of reactant and the limited mobility of the protein side-chains containing nucleophilic species. At water levels above the BET monolayer level, water is sufficiently mobile to solubilize reactants, that has been determined to correlate directly with the observation of the browning reaction [16].

The increased flexibility of the protein in the solid state due to the presence of water can lead to both reducible and non-reducible cross-linking reactions. Rates of reducible cross-linking of proteins containing free sulfhydryls are increased with increasing water content [25]. The pH of the solution to be freeze-dried is an important point to consider with this reaction, as an increase in the rate of cross-linking is observed with an increase in pH. Rates of non-reducible cross-linking reactions have been observed in the covalent dimerization of somatotropins at high water content [18]. Initial zero-order rates for covalent dimerization of bovine somatotropin in lyophilized formulations stored at 47°C were impacted by residual moisture. Levels of residual water of 1–5% resulted in significant increases in the amount of rBST lost due to non-reducible dimerization as compared with water levels below the BET monolayer level.

Enzymatic reactions are also impacted by the increased flexibility of the protein and the mobilization of reactants in the solid state. Self-proteolysis, however, is of minimal concern when formulating proteases, because the intermolecular reactions require a significant mobility to the segmental portions of the protein backbone [9]. Water is a reactant in the deamidation reaction, and high levels of water should increase the rate of deamidation [10]. Rates of acid-catalyzed [26] and base-catalyzed [17,27,28] deamidation reactions are impacted by residual moisture in the solid state. These reactions require either an increase in the mobility of the side-chain (acid-catalyzed) or a large change in the flexibility of the protein segment(base-catalyzed).

The effect of water on oxidation reactions is varied; it can have both anti- and pro-oxidant effects depending on the system. Generally, water acts as an antioxidant to facilitate the recombination of free radicals [29]. Free radicals are stable at water levels below the monolayer, but decay rapidly through recombination as the water content increases. This can result in an increase in oxidation reactions even though the moisture content is below the BET monolayer. At water levels exceeding the monolayer, the pro-oxidant effects of water become important as it mobilizes catalysts and increases protein flexibility, resulting in increased exposure of the reactions sites.

4. Examples of protein stability

Human growth hormone (hGH) is a good example of a protein that may be freeze-dried without significant degradation, but where the resulting lyophilized protein is potentially unstable [2,17]. Chemical decomposition via oxidation of methionine residues and deamidation of asparagine residues occurs in the solid state of hGH, in addition to aggregation to dimer and high order aggregates.

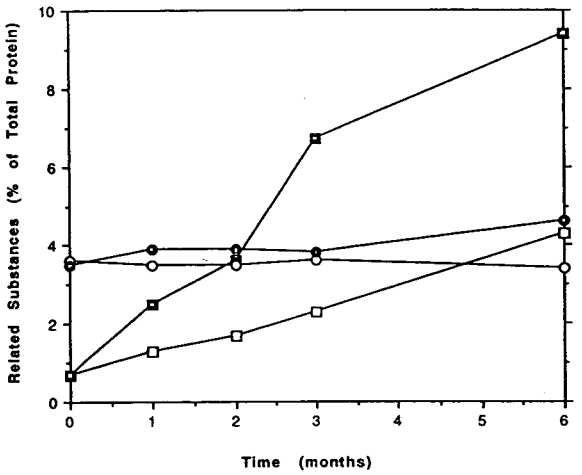


Fig. 1. Percentage of total protein for two major related products of hGH formulation stored for 6 months at 25°C. Squares represent the oxidation product (Met-14). Circles represent the deamidation product (Asn-149). Closed squares and circles represent the high-moisture content (3%) lot and open squares and circles represent the low-moisture content (1%) lot.

Fig. 1 shows the increase (as a percentage of total protein) of the two major degradation products of an hGH formulation. Two lots containing different amounts of moisture content were stored for 6 months at 25°C. The "high" and "low" lots contain approximately 3% and 1% moisture content, respectively, as determined by Karl Fischer titration measurements. The formulation contains a glycine–mannitol excipient system that offers optimum stability at low moisture levels when stored in a nitrogen headspace [2]. In the lyophilized plug, mannitol is thought to be crystalline, with glycine remaining amorphous [17]. The major oxidation product (at Met-14) and the major deamidation product (at Asn-149) have initial amounts of approximately 0.7% and 3.5%, respectively. Over the 6 months at 25°C, an increase is observed in the oxidation product for both lots, with a dramatic increase in the high-moisture lot. The deamidation product, on the other hand, showed little increase in degradation over the 6-month period for either lot. The large difference in degradation rates for the two amino acid residues points to the fact that the effect of moisture on protein stability is sensitive to the individual pathway. In this example, the oxidation of Met-14 is sensitive to moisture content, whereas deamidation at Asn-149 is unaffected.

A second oxidation product (Met-125) also shows an increase in oxidation for both the 1% and 3% moisture lots, with a much larger increase in the 3% moisture lot. The degradation, however, is much less than is observed for Met-14. The third methionine present in hGH(Met-171) shows no degradation over the 6-month storage period at 25°C. The difference in degradation among the three methionine groups is consistent with the observation that the extent of chemical degradation depends on the availability and sensitivity of individual amino acid residues to moisture content. Met-14 and Met-125 are located in relatively inflexible helices, leading to the conclusion that the mobility of the side-chain is only required for the oxidation reaction [8].

The dramatic increase observed in oxidation of Met-14 in the high-moisture lot is consistent with what would be expected at moisture levels above the BET monolayer moisture content. The monolayer water content for this hGH formulation has been previously determined to be 1.69% [17], and therefore the low- and high-moisture lots are on either side of the monolayer moisture content. Assuming that all the mannitol is crystalline, the water content of the amorphous phase at BET monolayer is 5.5%, which is consistent with the monolayer moisture content calculated for dry met-hGH(1). At 25°C, an increase in moisture content above the BET monolayer level showed a dramatic increase in the rate of chemical decomposition for Met-14, as opposed to moisture content below the monolayer level. This observation is consistent with the greater reactivity one would expect above the BET monolayer level due to increased protein flexibility and water mobility. In addition to chemical degradation, the extent of aggregation of freeze-dried human growth hormone after 10 days at 40°C has been found to be about five times greater at a moisture level of 3% than at less than 0.5% [1].

A separate oxidation pathway was observed in the increased level of hemoglobin to methemoglobin (oxidation of the heme group) in a freeze-dried sucrose-containing formulation as a result of an increase in residual moisture. The oxidation is twice as high for the formulation with 8% residual moisture as the same formulation containing 2% moisture content after 4 years of storage at room temperature [18]. In a separate study, the influence of hydration on the oxidation of lyophilized carbon monoxyhemoglobin to methemoglobin was determined [19]. The percentage of methemoglobin increased slightly over the moisture content range 0–12%. The rate of oxidation to methemoglobin then increased dramatically, reaching a maximum at approximately 18% water. The rate of oxidation at 18% moisture content is almost 35 times faster than in solution. Above 18% moisture content, the rate of oxidation decreases with increasing moisture content to the point where the oxidation rates are comparable to those obtained at very low water levels. In addition to oxidation, at water levels above 18% moisture content, the carbon monoxide ligand becomes replaced by the oxygen ligand. This oxygenation is possibly due to an increase in flexibility of the conformational structure, as the exchange is not observed below 12% moisture content. The oxygenation may prevent the oxidation of hemoglobin to methemoglobin, thus explaining the decreased rates of oxidation observed at high moisture content levels.

Maintaining low residual moisture is well illustrated for five rDNA cytokine preparations (three interleukins and two colony-stimulating factors) that were ampouled and lyophilized for use as internal standards [30]. The preparations were analyzed under accelerated storage conditions (56°C) in order to assess the procedure used to prepare cytokine standards and to assess the long-term stability of the preparations for use as standards. Each of the cytokines was ampouled in microgram amounts with milligram amounts of bulking agent such as albumin and carbohydrate. The moisture content after freeze-drying of a sample batch of any one cytokine preparation was between 1.1% and 0.64%. A low residual moisture content could be reproduced for the same or different cytokine preparations with a similar formulation. The cytokine preparations showed good retention of activity under the accelerated storage conditions as measured by flow cytometry. A sufficiently low moisture content was achieved such that the activity of the cytokines was not adversely affected by water-induced hydrolysis. The freeze-drying of cytokines under these conditions, for use as standards, enabled identical samples to be supplied without affecting the integrity or stability of the preparation.

Biological products are more likely to be adversely affected by over-drying than low-molecular mass drugs. Excessive drying may allow for the removal of structural water and disruption of the conformational integrity [31]. Studies to determine the optimum residual water level in formulations of tissue-type plasminogen activator have shown that the widely held view that the drier the better may not be appropriate for protein formulations [1]. The data suggest that the protein was mainly altered during the lyophilization process since subsequent degradation during storage was minimal. Drying the rt-PA protein to below the calculated monolayer water levels would appear to have an immediate and detrimental effect on the physical lots stability of the protein. Higher residual moisture lost more activity, suggesting that higher residual water levels permit more chemical degradation reactions to take place. The lower moisture content has an apparent higher level of aggregation that is largely due to the greater amount of aggregates initially generated during the lyophilization process. Results suggest that each protein may have a minimum moisture content that is necessary to shield the polar groups. This water may exist as clusters associated with the strong polar groups on the surface of the protein, rather than a continuous monolayer. Results of the stability on rt-PA show that it is desirable to achieve a balance between too little moisture (causing physical instability) and too much moisture (causing biological instability) [1].

5. Formulation and container considerations

Solid-state stability is especially complex in heterogeneous systems such as protein formulations, where the extent of product stability is sensitive to both formulation variables and the level of residual water in the dry solid. The residual water content varies with formulation largely because, while the percentage of water is based on the total sample, mass, the crystalline components do not retain significant amounts of water during freeze-drying, nor do crystalline components adsorb water from the stopper during storage. Formulations with excipients that adsorb very little moisture and contain a low ratio of protein to excipient are especially sensitive to small differences in water content. A small change in the moisture content of the lyophilized formulation can have a large effect on the amount of water associated with the protein and thus adversely affect stability. As an example, the greater stability of the hGH–mannitol–glycine formation described earlier is consistent with the concept that stability enhancement depends on the excipient system being at least partially amorphous to allow molecular interaction with the protein and/or to act as a "sink" for residual water [2]. The use of amorphous excipients provides more reproducible formulations from a stability standpoint, which is a necessity on the production scale owing to the difficulty in precisely controlling residual moisture levels.

In amorphous solids, the excipients can directly impact the stability by altering the effective protein concentration and pH. Adding excipients to the protein dilutes the effective concentration of the protein, increasing stability at high levels of hydration. The addition of buffer salts can also affect the stability of the protein. It is expected that low-molecular-mass species act as plasticizers to decrease the T_g of the amorphous systems and increase protein mobility. The addition of small ions can result in a lower moisture content needed to mobilize reactants. The ions preferentially bind to the proteins, thus excluding the water from the protein surface. This allows the water to be available to increase the mobility of the system again. The increasing availability of water as a medium for mobilization is also observed with the crystallization of excipients. This redistribution of water increases the moisture content in the remaining amorphous phase, leading to a decrease in stability through the lowering of the T_g.

In heterogeneous systems such as protein formulations, the rates of decomposition are influenced by additives, such as glycerol, propylene glycol or other polyhydric alcohols, which are commonly used in protein formulations as cryoprotectants during freezing and lyophilization. The presence of additives can increase water content at constant relative humidity or water activity. These additives act as plasticizers and the presence of such a liquid in the amorphous solid will decrease the water level necessary for reactant mobilization and protein flexibility. However, if the water content of the formulation is constant, the presence of these additives will decrease the water activity in the formulation. The presence of soluble salts will also increase water uptake at high relative humidities by dissolving in the loosely bound water and decreasing the vapor pressure, causing increased condensation. At low relative humidities, such electrolytes may actually decrease the water uptake by occupying proteins binding sites for water. Sugars such as trehalose and lactose can serve to satisfy partially the hydrogen-bonding requirements of the polar groups in dried proteins, and thus serve as water substitutes for dried proteins [31–33].

A low residual moisture content in the dried product upon freeze-drying is essential to maintain the stability of compounds. Low residual moisture after manufacture, however, does not ensure low moisture throughout the shelf-life of the product. The rubber stoppers used in the container hold a measurable amount of water that can transfer to the freeze-dried product, eventually coming to equilibrium with the water in the stopper [34,35]. The use of amorphous polymers protects the protein by serving as a moisture "sink" for the small amounts of mois-

ture that may transfer from the stopper. The use of amorphous excipients provides more reproducible formulations from a stability standpoint. During validation studies of each product and changes such as vial size, manufacturers should test several samples from several positions on each shelf in the freezer-dryer to determine that the lot meets the residual moisture specification of the product.

6. Moisture content determination

Owing to the key role water content plays in the solid-state stability of protein formulations, an accurate and precise moisture determination is essential in resolving stability issues. Accurate moisture determination will facilitate a more complete understanding of the role of moisture content on product stability and the setting of meaningful specifications and criteria for acceptable moisture content. Guidelines for the determination of residual moisture in dried biological products have been issued to describe residual moisture test methods and procedures used to set product residual moisture limits. These guidelines have been issued under 21 CFR 10.90 stating the principles and practices of general applicability. These are not legal requirements but can be relied upon by the user with the assurance of its acceptability to the FDA [6].

A number of chromatographic, spectroscopic, electrochemical, thermal and wet chemical methods have been used to determine moisture [6,36,37]. The most common of these are loss on drying (LOD), thermogravimetric analysis (TGA), gas chromatography using a thermal conductivity detector and the Karl Fischer titration. In addition to these methods where the sample is not recovered after analysis, the non-destructive technique of near-infrared spectroscopy has also been utilized [37]. In this technique a fiber-optic diffuse-reflectance probe measures reflectance through intact glass vials. These water content measurements, however, require that an accurate reference moisture value be obtained as a guide, with the Karl Fischer titration being the method of choice.

6.1. Gravimetric method (loss on drying)

The procedure for loss on drying involves dispensing the substance to be tested into a tared, glass-stoppered, shallow weighing bottle [38]. The optimum sample size is approximately 200 mg and may require that the contents from several final product containers be pooled. The particle size of the test substance may need to be reduced (down to 2 mm) to allow for proper drying if large crystals are present. The test specimen is prepared for analysis in a low-humidity glove-box and is weighed and evenly distributed along the bottom of the bottle for even drying. The test is performed in a temperature- and humidity-controlled environment to prevent ambient humidity from interfering in the test procedure. The unstoppered bottle and contents are then loaded in the drying chamber and the test specimen dried at a given temperature and time. The approved test method for residual moisture (Code of Federal Regulations, 21 CFR 610.13) in freeze-dried biological products measures the maximum loss in mass of a weighed sample equilibrated to constant mass over anhydrous phosphorus pentoxide at a pressure of not more than 1 mmHg and a temperature of 20–30°C for as long as it has been established is sufficient to results in a constant mass. After drying, the bottle is stoppered promptly and allowed to come to room temperature in a desiccator before weighing.

This method relies on the removal of water from the cake when exposed to heat in a vacuum oven. Measurements of the opened vial and cake are made before and after the addition of heat. The difference in mass is used to calculate the amount of water present. The gravimetric method measures surface moisture and loosely bound water of hydration [39]. Surface moisture is the classical definition of residual moisture. The major disadvantage of this method is that is very difficult to remove all the residual water, which may remain tightly bound to either the protein or the excipients used in the formulation buffer. Therefore, this method may underestimate the actual residual moisture level [40]. The general recommendation for most products is that the

residual moisture should not exceed 1.0% by the gravimetric method [3].

6.2. Thermogravimetric

In thermogravimetric analysis, the mass of a sample is recorded continuously as its temperature is increased linearly from ambient to as high as 1200°C [41]. This method can provide more useful information than LOD at a fixed temperature for a fixed time. The essential features of the equipment are a recording balance and programmable heat source. Variations on the equipment employed include the sensing of the specimen temperature, size of the sample holder and range of atmosphere control. In the determination of water content, the TGA will show a decrease in mass from ambient to 100°C and then plateau. The mass loss from the initial sample value to this plateau is due to the loss of water.

In determining the loss of mass upon drying, it is necessary that the method differentiates between measuring the "moisture" and not "total volatiles" content of the sample. Depending on the sample and test conditions, the amount of mass lost from the sample may include substances in addition to water (methanol, etc.). Care must be taken to choose the proper method to insure that the mass loss is due solely to moisture and does not include other species. The thermogravimetric method measures both surface and bound moisture in freeze-dried biological products. The TGA method can determine the moisture content in freeze-dried samples as small as two milligrams. CBER used TGA methodology as a second method to confirm Karl Fischer test results, especially when the Karl Fischer test results indicate that the sample has a falling moisture content [3].

In complex thermograms, the transitions attributed to residual moisture are verified for samples by thermogravimetry–mass spectrometry (TG–MS). This technique provides precise TG heating conditions and mass loss information along with mass spectral identification of volatiles evolved during the mass loss process. Mass spectra are taken of the TG off-gases continuously with ion intensities of mass peaks 18 and 44 being monitored to show the changes in the amounts of water and carbon dioxide. The mass spectral ion intensities verify the transition caused by moisture in the freeze-dried sample. This is done by differentiating between the water content of the sample and the water evolved from thermal decomposition of the sample, which coincides with the evolution of carbon dioxide.

6.3. Gas chromatography

In addition to the thermal methods of analysis, gas chromatography (GC) with a thermal conductivity detector has been used for the determination of water in proteins. In this method the sample is dissolved in an appropriate solvent and injected on-column along with an internal standard (i.e., methanol). In a method used in the European Pharmacopoeia monograph for hGH, 1.0 mg of the preparation is suspended in 0.1 ml of 2-propanol using anhydrous methanol as the internal standard [42]. The chromatographic procedure is carried out using a stainless-steel column packed with styrene–divinylbenzene copolymer using helium as the carrier gas. The water content of the sample is calculated taking into account its density (0.997 g/ml at 20°C) and any water detectable in the internal standard solution. Care must be taken to assure that the entire contents of the sample are well dissolved and that any contamination of the solvent(s) due to water is corrected for.

Gas chromatography may be utilized if the need exists to perform residual moisture determinations on a micro scale. GC analysis, however, is complicated to perform and is prone to overestimating the amount of moisture content in the sample. Precautions are needed, such as sample preparation in a glove-box purged with a dry inert gas, to prevent rehydration of the sample with moisture from the atmosphere [40].

6.4. Near-infrared spectroscopy

In addition to methods where the sample is not recovered after analysis, the non-invasive,

non-destructive technique of near-infrared spectroscopy has also been utilized [37]. In this technique, a fiber-optic diffuse-reflectance probe was used to measure reflectance from 1100 to 2500 nm through the bottom of intact glass vials. The correlation of the method with results obtained by Karl Fischer analysis was very good. The method is rapid, as analysis times are short (typically 20-s analysis times) and little sample preparation is required. In addition, the sample container remains closed during the analysis, removing the threat of erroneous results from atmospheric moisture. The water content measurements, however, require that an accurate reference moisture value be obtained as a guide, with the Karl Fischer titration being the method of choice.

6.5. Karl Fischer methodology

Of all these methods, the Karl Fischer titration, originally described in 1935 [43], is the approach most widely used in the determination of water content. The Karl Fischer method gives a better estimate of the total residual moisture, is very reproducible and can be automated. The titration can be run in either protic or aprotic media with the protic medium seeing wider use owing to the higher sensitivity of the titer to sample and solvent composition [44]. The reaction in protic media (i.e., alcohol) involves sulfur dioxide reacting with the alcohol to produce an alkyl sulfite in a buffered medium using an appropriate base to maintain the solution at the optimum pH. The method of adding iodine to the reaction differs according to the type of experiment, volumetric or coulometric. In the volumetric experiment, the iodine is contained in a buret and metered out as required. The amount of iodine per unit volume is determined empirically during a standardization step and the amount of water in the sample is calculated from this titer. In a coulometric experiment, the iodine is generated electrochemically from iodide present in the cell. Iodine together with pyridine, sulfur dioxide and methanol form the Karl Fischer reagent, which reacts quantitatively with water. The cake is dissolved or suspended in methanol and a coulometric oxidation of iodine is made titrimetrically. Alternative "pyridine-free" Karl Fischer reagents substitute another amine for pyridine [45]. The electrochemical efficiency of this method is generally 100%, and the amount of water in the sample is calculated from the number of moles of electrons used in the iodine generation.

The ability of the coulometric Karl Fischer method to measure residual moisture in about 10 mg of a freeze-dried biological sample makes it the most practical Karl Fischer method for the determination of residual moisture in freeze-dried biological products in single dose final containers. Freeze-dried biological products cannot be analyzed by the Karl Fischer methodology when (1) materials are present in the matrix of the biological product that interfere with the Karl Fischer reagents, (2) the sample does not dissolve adequately in the Karl Fischer reagent and (3) the sample moisture does not adequately extract into these solvents.

An important point to consider with the Karl Fischer titration is the possibility of erroneous results due to water contamination during sample handling. Water content of a protein at a given relative humidity can be influenced by its prior process and storage conditions. In studies involving the freeze-dried formulation of human growth hormone, an increase in water content arose from absorption of water from the stopper, not transmission of water through the stopper [33]. Titration of small amounts of water in coulometric systems requires correction for atmospheric moisture entering the system. Sample handling can have a significant impact on the results of a titration as a result of gain or loss of moisture between sampling and analysis. This becomes increasingly frustrating when one considers the increased variability of the assay due to wide fluctuations in relative humidity throughout the year [46]. Atmospheric moisture contamination can be partly circumvented by using a dry-box to minimize interference from ambient humidity, but care must be taken when setting the relative humidity in the glove-box so as not to remove or add moisture to the sample. An alternative to the glove-box is to add a known

amount of anhydrous methanol with a syringe to the freeze-dried biological product in a final container [3]. The methanol should dissolve the sample. Known amounts of sample and methanol are withdrawn and added to the Karl Fischer titration vessel for moisture determination. The moisture content of the anhydrous methanol is determined as the blank.

7. Conclusions

The relationship between water and protein is important in the understanding of the shelf-life stability of lyophilized formulations of biopharmaceuticals. The flexibility of a protein in the solid state is increased by temperature and moisture content. Water has the potential to act as a plasticizer in the amorphous polymer to reduce the glass transition temperature, T_g. If the T_g is decreased below the temperature of the surroundings, a phase transition from a glassy state to a dynamically relaxed state may occur. The level of moisture content resulting in a more mobile system can be extrapolated using isotherms to help predict the BET monolayer. At or near the BET monolayer, there is increased mobility of the absorbed water and the flexibility of the protein. The increased protein flexibility and water mobility may adversely affect the chemical and/or conformational stability of the protein. The observed decomposition pathways are sensitive to the flexibility of particular segments of the protein backbone and the orientation of exposed amino acid residues. In addition to temperature and moisture content considerations, care must be taken in the selection of the proper excipients. Excipients that exclude water from the protein surface (such as small ions) or that increase the water content in the amorphous phase (those that crystallize) can adversely affect protein stability. Strategies employed to maintain the solid-state stability of proteins need to minimize protein flexibility and water mobility by maximizing the T_g through low moisture content and careful selection of excipients. Owing to the close interplay between water and protein stability in the amorphous system, the accurate determination of moisture content in the lyophilized plug is imperative. A number of techniques have been utilized to determine moisture content, with each having its own set of advantages and limitations. The one common limitation, however, is the ability to prevent atmospheric moisture contamination during sample handling.

Acknowledgements

I thank the following individuals for the many helpful and stimulating discussions regarding the effects on stability and determination of moisture content in lyophilized proteins; Gregory C. Davis, Michael J. Hageman, Lee E. Kirsch, Michael J. Pikal and Ralph M. Riggin.

References

[1] C.C. Hsu, C.A. Ward, R. Pearlman, H.M. Nguyen, D.A. Yeung and J.G. Curley, *Dev. Biol. Stand.*, 74 (1991) 255.
[2] M.J. Pikal, K.M. Dellerman, M.L. Roy and R.M. Riggin, *Pharm. Res.*, 8 (1990) 427.
[3] *Guideline for the Determination of Residual Moisture in Dried Biological Products*, Center for Biologics Evaluation and Research (CBER), Food and Drug Administration, 1990.
[4] *Federal Resister (1990), General Biological Products Standard: Test for Residual Moisture* Vol. 55, No. 133, US Government printing office, Washington, DC, 1990.
[5] *Code of Federal Regulations*, 21 CFR 610.13, 1989, pp. 52–53.
[6] J.C. May, R.M. Wheeler, N. Etz and A. Del Grosso, *Dev. Biol. Stand.* 74 (1991) 153.
[7] J.D. Leeder and I.C. Watt, *J. Colloid Interface Sci.*, 48 (1974) 339.
[8] M.J. Hageman, in T.J. Ahern and M.C. Manning (Editors), *Part A: Chemical and Physical Pathways of Protein Degradation*, Plenum Press, New York, 1992, pp. 273–309.
[9] J.A. Rupley, E. Gratton and G. Careri, *Trends Biochem. Sci.*, 8 (1983) 18.
[10] M.J. Hageman, *Drug Dev. Ind. Pharm.*, 14 (1988) 2047.
[11] L. Slade, H. Levine and J.W. Finney, *Food Sci. Technol.* 29 (1989) 9.
[12] F. Franks, R.H.M. Hatley and S.F. Mathias, *Biopharm* 4, No. 9 (1991) 38.
[13] F. Franks, *Cryo. Lett.* 11 (1990) 93.

[14] H. Levine and L. Slade, in F. Franks (Editor), *Water Science Reviews*, Vol. 3, Cambridge University Press, Cambridge, 1987, pp. 79–185.
[15] D. Greiff, *Cryobiology* 8 (1971) 145.
[16] M.L. Roy, M.J. Pikal, E.C. Rickard and A.M. Maloney, *Dev. Biol. Stand.*, 74 (1992) 323.
[17] M.J. Pikal, K.M. Dellerman and M.L. Roy, *Dev. Biol. Stand.*, 74 (1991) 21.
[18] M.J. Hageman, P.L. Possert and J.M. Bauer, *J. Agric. Food. Chem.*, 40 (1992) 342.
[19] T.I. Pristoupil, M. Kramlova, H. Fortova and S. Ulrych, *Haematologia*, 18 (1985) 45.
[20] R. Sanches, W.L.B. Melo and M.F. Colombo, *Biochem. Biophys. Acta*, 874 (1986) 19.
[21] M.W. Townsend and P.P. DeLuca, *J. Pharm. Sci.* 80 (1991) 63.
[22] S.L. Nail and W. Johnson, *Dev. Biol. Stand.* 74 (1991) 137.
[23] M.J. Pikal, *Biopharm* 3, No. 8 (1990) 18.
[24] R.B. Duckworth, in L.B. Rockland and G.F. Stewart (Editors), *Water Activity: Influcences of Food Quality*, Academic Press, New York, 1981, pp. 297–317.
[25] W.R. Liu, R. Langer and A.M. Klibanov, *Biotechnol. Bioeng.*, 37 (1991) 177.
[26] D.A. Gearhart and L.E. Kirsch, *Pharm. Res.*, 5 (1989) 244.
[27] G.W. Becker, P.M. Tackitt, W.W. Bromer, D.S. Lefeber and R.M. Riggin, *Biotechnol. Appl. Biochem.*, 10 (1988) 326.
[28] R. Pearlman and T.H. Nguyen, in D. Marshak and D. Liu (Editors), *Therapeutic Peptides and Proteins: Formulations, Delivery, and Targeting*, Communications in Molecular Biology, Cold Spring Harbor Laboratory, 1989, pp. 23–30.
[29] M. Karel, in R.B. Duckworth (Editor), *Water Relations in Foods*, Academic Press, New York, 1975, pp. 435–453.
[30] P.J. Dawson, *Dev. Biol. Stand.*, 74 (1991) 273.
[31] O. Fennema, in J.R. Whitaker and S.R. Tannenbaum (Editors), *Food Proteins*, AVI Publishing, Westport, CT, 1977, pp. 50–90.
[32] J.F. Carpenter, T. Arakawa and J.H. Crowe, *Dev. Biol. Stand.*, 74 (1991) 225.
[33] H. Levine and L. Slade, *Biopharmaceutics* 5 (1992) 36.
[34] M.J. Pikal and S. Shah, *Dev. Biol. Stand.*, 74 (1991) 165.
[35] F. DeGrazio and K. Flynn, *J. Parenter. Sci. Technol.* 46 (1992) 54.
[36] *United States Pharmacopeia, XXIII Revision*, US Pharmacopeial Convention, Rockville, MD, 1995, pp. 1840–1843.
[37] M.S. Kamat, R.A. Lodder and P.P. DeLuca, *Pharm. Res.*, 6 (1989) 961.
[38] *United States Pharmacopeia, XXIII Revision*, US Pharmacopeial Convention, Rockville, MD, 1995, pp. 1801–1802.
[39] J.C. May, *J. Biol. Stand.*, 10 (1982) 249.
[40] R.A. Baffi and R.L. Garnick, *Dev. Biol. Stand.*, 74 (1991) 181.
[41] *United States Pharmacopeia, XXIII Revision*, US Pharmacopeial Convention, Rockville, MD, 1995, pp. 1837–1838.
[42] *PharmEuropa*, 5 (1993) 243.
[43] K. Fischer, *Angew. Chem.*, 48 (1935) 394.
[44] S.K. MacLeod, *Anal. Chem.*, 63 (1990) 557A.
[45] J. Lindquist, *J. Pharm. Biomed. Anal.*, 2 (1984) 37.
[46] Lilly Research Laboratories, unpublished data.

ELSEVIER

Journal of Chromatography A, 705 (1995) 129–134

JOURNAL OF CHROMATOGRAPHY A

Review

Processing of C-terminal lysine and arginine residues of proteins isolated from mammalian cell culture

Reed J. Harris

Analytical Chemistry Department, Genentech, Inc. (#62), 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

Abstract

C-terminal Lys or Arg residues whose presence was expected based on gene sequence information are often absent in proteins isolated from mammalian cell culture. This discrepancy is believed to be due to the activity of one or more basic carboxypeptidases. Internal Arg/Lys residues that become C-terminal upon proteolysis or zymogen activation, such as in the two-chain form of tissue plasminogen activator, may also be removed from the mature protein. Charge heterogeneity results when this type of processing is incomplete; such heterogeneity can be detected by isoelectric focusing or ion-exchange chromatography. The absence of C-terminal basic residues is not usually a regulatory concern, as plasma-derived proteins are often similarly processed.

Contents

1. Introduction

In theory, the characterization of recombinant proteins is a straightforward matter, as the requisite genetic engineering provides an expected amino acid sequence, with potential sites of modification identified based on known consensus sequences [1]. In practice, however, a number of variations from the expected structure can be found. Variants may result from either known or novel types of in vivo (posttranslational) modification [2] or from spontaneous (nonenzymatic) protein degradation, such as methionine oxidation [3], diketopiperazine for-

0021-9673/95/$29.00
SSDI 0021-9673(94)01255-5

mation [4], aspartate isomerization and deamidation of asparagine residues [5], or succinimide formation [6,7].

This review will cover a type of posttranslational processing that is becoming a common analytical protein chemistry experience: the removal of Lys or Arg residues from the C-terminus of a protein obtained through mammalian cell culture. Several examples of this type of processing have been reported, and some successful approaches for identifying such variants are reported herein.

2. Experimental conditions

2.1. CNBr/C4 assay

Samples were exchanged into 0.1% formic acid by dialysis, then 88% formic acid was added to bring the samples to a final solution of 20% formic acid. CNBr (Pierce) was dissolved in 20% formic acid at a concentration of 50 mg/ml, then added to sample(s) at a 5:1 (CNBr:protein) weight ratio. After stirring 20 h at room temperature in the dark, the CNBr was removed under a nitrogen stream. Samples were reconstituted in 20 μl 88% formic acid, then diluted with water to 300 μl final. C-terminal CNBr fragments were resolved using a Vydac C4 (250 × 2.1 mm) column. The system was equilibrated for 20 min at 100% solvent A [0.1% trifluoroacetic acid (TFA) in water], then, 6 min after sample injection, a linear gradient to 25% solvent B (0.1% TFA in acetonitrile) was developed over 50 min by a Hewlett-Packard 1090 system. The flow-rate was 0.20 ml/min, with a constant temperature of 40°C.

2.2. Cation-exchange chromatography

A Pharmacia MonoS column (50 × 5 mm) was equilibrated with 95% solvent A (20 m*M* sodium phosphate, pH 6.9)–5% solvent B (solvent A + 100 m*M* NaCl) at 40°C with a flow-rate of 1.0 ml/min. Upon injection of 72 μg from three lots of rhuMAb HER2, a gradient from 5 to 40% solvent B was developed over 40 min to elute peak fractions.

2.3. Materials

Tissue plasminogen activator (tPA) purified from Chinese hamster ovary (CHO) cells transfected with the human tPA gene was produced at Genentech (Activase). Bowes melanoma tPA was provided by Desire Collen (University of Leuven). rhuMAb HER2 is a recombinant humanized antibody produced in transfected CHO cells [8]. TFA was purchased from Pierce, while acetonitrile was from Burdick and Jackson. Vydac C4 (214TP52) and C18 (218TP58) columns were purchased from The Separations Group.

3. Results

3.1. Antibodies and antibody-related proteins

rCD4-IgG is a recombinant chimeric homodimeric protein, secreted from transfected CHO cells, with the C_H1 and C_H2 regions of a human IgG_1 heavy chain replaced by residues 1–180 of the human CD4 receptor [9]. A CNBr cleavage/C4 RP-HPLC method was developed to identify the C-terminus of the mature CHO-expressed protein; in that study, the expected C-terminal Lys residues were found to be completely absent [10]. This approach can be used for any protein with human IgG_1 heavy chains such as rhuMAb HER2 [11], as shown in Fig. 1. In this example, CNBr cleavage after Met^{431} liberated C-terminal peptides that were isolated by RP-HPLC and characterized. Peak A contains the expected heavy chain C-terminal CNBr peptide (residues 432–450: HEALHNHYTQKSLSLSPGK), but the major product (peak B) is a des-Lys^{450} peptide (residues 432–449: HEALHNHYTQKSLSLSPG). Minor additional peaks were obtained that resulted from formylation of the peptide during the CNBr/formic acid incubation.

We and others have reported this type of processing with antibody and antibody-like proteins from a variety of sources (Table 1). The

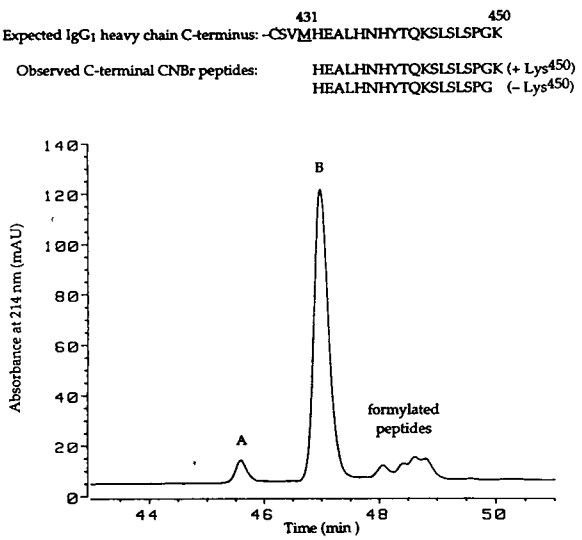


Fig. 1. Isolation of C-terminal CNBr peptides from 25 mg of a human IgG_1 antibody by C4 RP-HPLC. Peak A: residues 432–450 (+ Lys^{450}). Peak B: residues 432–449 (– Lys^{450}).

absence of the C-terminal Lys residues from these proteins is not due to cloning errors, but, rather appears to be due to the action of carboxypeptidase(s). The penultimate residues (Gly) do not appear to be removed, suggesting that the responsible carboxypeptidase(s) are specific for basic residues. Plasma-derived antibodies also typically lack the heavy chain C-terminal Lys residues [12,13].

Incomplete removal of C-terminal Lys residues from IgG heavy chains causes charge heterogeneity, as forms with 0, 1 or 2 Lys residues will result. Such charge variants can be resolved by cation-exchange chromatography [14,15]. For example, rhuMAb HER2 shows five charge species (Fig. 2). The main peak (peak 3) has no Lys^{450} residues, while the more basic peaks 4 and 5 have one or two Lys^{450} residues, respectively (data not shown). The more acidic peaks 1 and 2 are deamidated at Asn^{30} in one light chain; peak 1 has no Lys^{450} residues, while peak 2 has one Lys^{450} residue.

3.2. Two-chain tPA

The activation of serine protease zymogens, including many of the coagulation/fibrinolytic proteins, occurs by proteolysis of arginyl bonds, converting the zymogen to a two-chain form whose polypeptide chains remain associated by disulfide bonds. Tissue plasminogen activator (tPA) is synthesized as a single-chain 527-residue polypeptide; depending on the cell culture conditions employed, proteolytic cleavage can occur between Arg^{275} and Ile^{276}, converting tPA to a two-chain form. Two CHO-expressed recombinant forms were produced at Genentech, a primarily two-chain product and a primarily single-chain product (the latter is licensed as Activase). tPA isolated from the Bowes melanoma cell line is largely a two-chain product.

Table 1
Examples of C-terminal Lys/Arg processing

Protein	Susceptible residue	Cell line/source	Ref.
rCD4-IgG	Lys	Transfected CHO	[10]
rhuMAb HER2	Lys	Transfected CHO	[11]
OKT3 MAb	Lys	Hybridoma/ascites	[14]
OKT3 MAb	Lys	Hybridoma/cell culture	[14]
CEM231 MAb	Lys	Hybridoma/ascites	[15]
CEM231 MAb	Lys	Hybridoma/cell culture	[15]
Hu-anti-Tac MAb	Lys	Transfected SP2/0	[16]
2-Chain tPA	Arg	Transfected CHO	
2-Chain tPA	Arg	Bowes melanoma	
huEPO	Arg	Human urine	[17]
rhuEPO	Arg	Transfected CHO	[17]

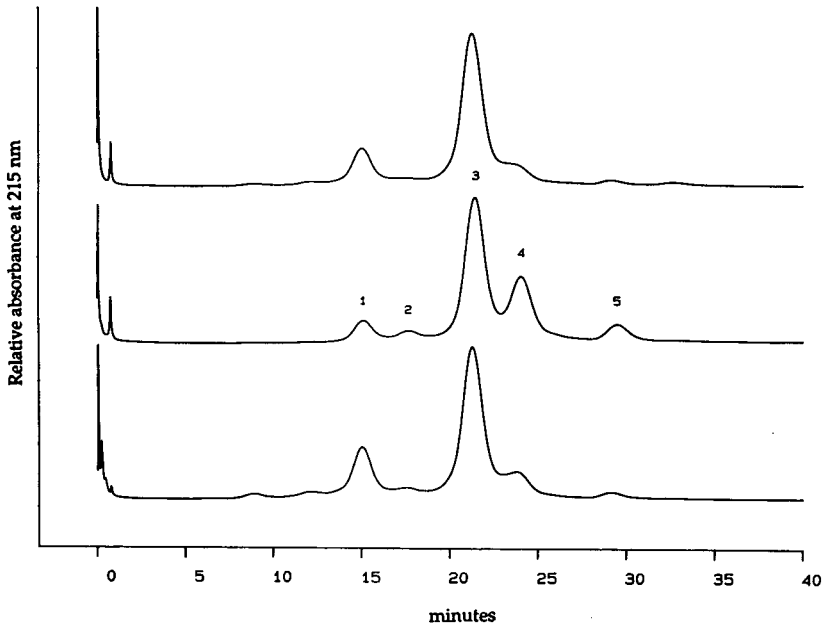


Fig. 2. Cation-exchange chromatography of three lots of rhuMAb HER2. Chromatographic conditions are given in Section 2.2.

Tryptic peptide mapping of the two CHO-expressed forms showed peptides with or without Arg^{275} (residues 268–275: QYSQPQFR and 268–274: QYSQPQF, respectively) as shown in Fig. 3. The 268–274 (des-Arg^{275}) peptide predominates in the two-chain form, whereas the single-chain material gave the 268–275 peptide almost exclusively. Evidently, cleavage of the Arg^{275}–Ile^{276} bond in the two-chain forms exposes Arg^{275} to basic carboxypeptidase(s) in the cell culture fluid. No evidence for further processing at the C-terminus of the heavy chain (residues 1–274/275) was evident in peptides from the tryptic map. In both single-chain and two-chain tPA, the light chain C-terminus (-Met-Arg-Pro^{527}) is unprocessed (unpublished data). The predominantly two-chain melanoma-derived tPA shows roughly equivalent levels of the ± Arg^{275} peptides (Fig. 4). Figs. 3 and 4 differ slightly because different RP-HPLC columns (albeit from the same vendor) were used.

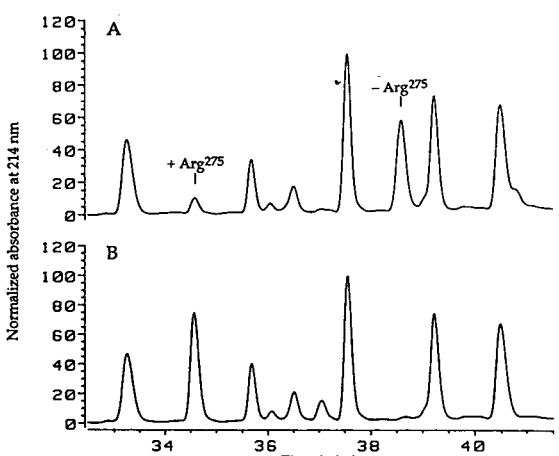


Fig. 3. Detail showing differences between single-chain and two-chain CHO-expressed tPA tryptic maps. Chromatographic details are given in Ref. [18]. (A) Two-chain tPA digest. (B) Single-chain tPA digest. The peaks marked $+Arg^{275}$ and $-Arg^{275}$ contain residues 268–275 (QYSQPQFR) and 268–274 (QYSQPQF), respectively.

4. Discussion

When derived from mammalian cell culture, proteins that might be expected to terminate with Arg or Lys residues may be processed such that these residues are absent in the purified product. A number of antibody structural studies

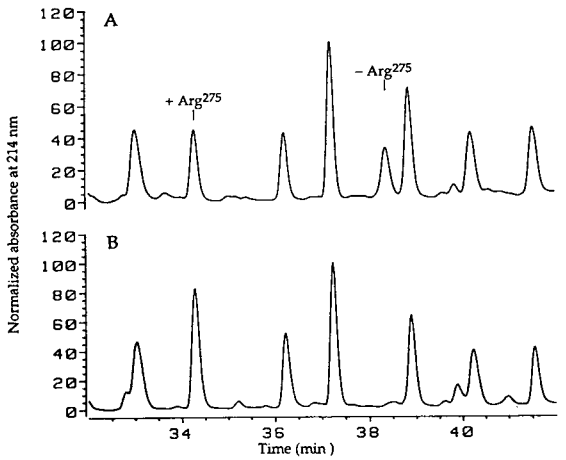


Fig. 4. Detail showing differences between Bowes melanoma tPA and single-chain CHO-expressed tPA tryptic maps. Chromatographic details are given in Ref. [17]. (A) Melanoma tPA digest. (B) Single-chain CHO tPA digest. The peaks marked $+Arg^{275}$ and $-Arg^{275}$ contain residues 268–275 (QYSQPQFR) and 268–274 (QYSQPQF), respectively.

have demonstrated the absence of heavy chain C-terminal Lys residues [11,14–16]. The absence of an expected C-terminal Arg residue from urinary and recombinant (CHO-expressed) human erythropoeitin (huEPO) has also been reported [17]. Similarly, as shown for tPA, conversion of serine protease zymogens to two-chain forms upon cleavage after Arg or Lys residues during cell culture may allow basic carboxypeptidase processing to occur at the newly exposed Arg/Lys C-terminus. Incomplete processing of basic residues will cause charge heterogeneity that can be detected by ion-exchange chromatography or isoelectric focusing [14], although additional factors (e.g., deamidation, phosphorylation or sialic acid variability) may also contribute to the overall charge heterogeneity.

C-terminal processing may also be detected by detailed characterization of peptide maps using the LC–MS approach (peptide digestion followed by RP-HPLC separation with on-line mass spectrometric detection) [19]. In electrospray mass spectrometry, the major observed ion roughly corresponds to the number of basic groups [20]; the N-terminus and a C-terminal Lys/Arg side-chain usually provide two basic groups for tryptic peptides. The number of basic groups is reduced by one in peptides that result from basic carboxypeptidase processing; as a consequence, the processed form(s) of C-terminal peptides may be overlooked unless the investigators actively look for any potentially minor des-Arg/Lys singly protonated form. Automated methods for C-terminal analysis [21] may also assist in C-terminal processing studies.

Basic carboxypeptidases are known to regulate peptide hormonal activity (e.g. bradykinin, the enkephalins and anaplylatoxins) [22]. However, no biological effect(s) have been reported for the C-terminal processing of the proteins described in this report. The presence or absence of C-terminal Lys residues on antibody heavy chains is not likely to influence antigen binding, which is mediated by the distant complementarity-determining regions [23], nor are binding to the Fcγ receptor or complement C1q likely to be affected, as these involve residues in the hinge-C_H2 region and C_H2 domains, respectively [24,25].

In general, the absence of C-terminal Lys or Arg residues is unlikely to be considered a cause for concern, as similar processing affects plasma- or urinary-derived proteins such as antibodies [12,13] and huEPO [17]. Variation in the extent of C-terminal processing can lead to production lots with different charge distributions. The charge variants generated by incomplete processing of the Lys/Arg residue(s) may be isolated by cation-exchange chromatography; assaying these fractions for potency or clearance will help assess the appropriate level of concern for this type of heterogeneity.

The responsible carboxypeptidase(s) have not yet been identified. The lack of processing beyond the Lys/Arg residue(s) suggests that an enzyme similar to one or more of several known basic carboxypeptidases may be responsible; such basic carboxypeptidases include pancreatic carboxypeptidase B, plasma carboxypeptidase N, membrane-bound (extracellular) carboxypeptidase M, and carboxypeptidase H of secretory granules [22]. Plasma plasminogen-binding and urinary basic carboxypeptidases have also been

identified [26,27]. Basic carboxy-peptidase activity has been reported for hybridoma cell culture supernatants and ascites fluid [14,15]; antibodies purified from ascites tend to be completely processed. It is also possible that the activity seen during cell culture results from release of a cytosolic enzyme from damaged cells, as is the case for the sialidase isolated from CHO cell culture [28]. Isolation and characterization of the responsible carboxypeptidase(s) should allow susceptible proteins to be cultured in the presence of inhibitors, generating unprocessed material that could be used for investigations as to the effect(s) of the C-terminal processing.

Acknowledgements

The author thanks Madelyn Marino and Brent Larsen for developing the cation-exchange method shown in Fig. 2, and Karen Wagner for contributing to the characterization of the ion-exchange peak fractions.

References

[1] A. Aitken, *Identification of Protein Consensus Sequences*, Ellis Horwood, UK, 1990.
[2] F. Wold, *An. Rev. Biochem.*, 50 (1981) 783–814.
[3] M. Kunitani, P. Hirtzer, D. Johnson, R. Halenbeck, A. Boosman and K. Koths, *J. Chromatogr.*, 359 (1986) 391–402.
[4] J.E. Battersby, W.S. Hancock, E. Canova-Davis, J. Oeswein and B. O'Connor, *Int. J. Peptide Protein Res.*, 44 (1994) 215–222.
[5] T. Geiger and S. Clarke, *J. Biol. Chem.*, 262 (1987) 785–794.
[6] G.T. Teshima, J.T. Stults, V. Ling and E. Canova-Davis, *J. Biol. Chem.*, 266 (1991) 13544–13547.
[7] M. Kwong and R.J. Harris, *Protein Sci.*, 3 (1994) 147–149.
[8] P. Carter, L. Presta, C.M. Gorman, J.B.B. Ridgway, D. Henner, W.L.T. Wong, A.M. Rowland, C. Kotts, M.E. Carver and H.M. Shepard, *Proc. Natl. Acad. Sci. USA*, 89 (1992) 4285–4289.
[9] D.J. Capon, S.M. Chamow, J. Mordenti, S.A. Marsters, T. Gregory, H. Mitsuya, R.A. Byrn, C. Lucas, F. Wurm, J.E. Groopman, S. Broder and D.H. Smith, *Nature*, 339 (1989) 525–531.
[10] R.J. Harris, K.L. Wagner and M.W. Spellman, *Eur. J. Biochem.*, 194 (1990) 611–620.
[11] R.J. Harris, A.A. Murnane, S.L. Utter, K.L. Wagner, E.T. Cox, G. Polastri, J.C. Helder and M.B. Sliwkowski, *Bio/Technology*, 11 (1993) 1293–1297.
[12] G.M. Edelman, BA. Cunningham, W.E. Gall, P.L. Gottleib, U. Rutishauser and M.J. Waxdal, *Biochemistry*, 63 (1969) 78–85.
[13] J.W. Ellison, B.J. Berson and L.E. Hood, *Nucleic Acids Res.*, 10 (1982) 4071–4079.
[14] P. Rao, A. Williams, A. Baldwin-Ferro, E. Hanigan, D. Kroon, M. Makowski, E. Meyer, V. Numsuwan, E. Rubin and A. Tran, *BioPharm*, 4 (1991) 38–43.
[15] J.P. McDonough, T.C. Furman, R.M. Bartholomew and R.A. Jue, *US Pat.* 5 126 250 (1992).
[16] D.A. Lewis, A.W. Guzzetta, W.S. Hancock and M. Costello, *Anal. Chem.*, 66 (1994) 585–595.
[17] M.A. Recny, H.A. Scoble and Y. Kim, *J. Biol. Chem.*, 262 (1987) 17156–17163.
[18] R.J. Harris, C.K. Leonard, A.W. Guzzetta and M.W. Spellman, *Biochemistry*, 30 (1991) 2311–2314.
[19] H.A. Scoble and S.A. Martin, *Methods Enzymol.*, 193 (1990) 519–536.
[20] S.K. Chowdhury, V. Katta and B.T. Chait, *Biochem. Biophys. Res. Commun.*, 167 (1990) 686–692.
[21] J.M. Bailey, N.R. Shenoy and J.E. Shively, in R.H. Angeletti (Editor), *Techniques in Protein Chemistry III*, Academic Press, San Diego, 1992, pp. 11–22.
[22] R.A. Skidgel, *Trends Pharm. Sci.*, 9 (1988) 299–304.
[23] T.E. Wu and E.A. Kabat, *J. Exp. Med.*, 132 (1970) 211–250.
[24] M.S. Chappel, D.E. Isenman, M. Everett, Y.Y. Xu, K.J. Dorrington and M.H. Klein, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 9036–9040.
[25] A.R. Duncan and G. Winter, *Nature*, 332 (1988) 738–740.
[26] D.L. Eaton, B.E. Malloy, S.P. Tsai, W. Henzel and D. Drayna, *J. Biol. Chem.*, 266 (1991) 21833–21838.
[27] R.A. Skidgel, R.M. Davis and E.G. Erdos, *Anal. Biochem.*, 140 (1984) 520–531.
[28] T.G. Warner, J. Chang, J. Ferrari, R. Harris, T. McNeary, G. Bennett, J. Burnier and M.B. Sliwkowski, *Glycobiol.*, 3 (1993) 455–463.

ELSEVIER

Journal of Chromatography A, 705 (1995) 135–154

JOURNAL OF CHROMATOGRAPHY A

Capillary electrophoresis of S. nuclease mutants

Franka Kálmán[a], Stacey Ma[a], Robert O. Fox[b], Csaba Horváth[a,*]

[a]*Department of Chemical Engineering, Yale University, New Haven, CT 06520, USA*

[b]*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA*

Abstract

The electrophoretic migration behavior of 12 S. nuclease variants from *Staphylococcus aureus* with small but well defined structural differences from site directed mutation was investigated in free solution capillary electrophoresis at pH 2.8 to 9.5. The nucleases are basic proteins; the p*I* and the M_r of the wild type are 10.3 and 16.811 kd, respectively. With specially selected oligoamino buffers and with an inert, hydrophilic wall coating in 75 μm I.D. quartz capillary tubes, most of the proteins could be separated by CZE without interference by wall adsorption even at pH 9.5 where the selectivity was the highest. At pH 2.8, 4.1 and 7.0, S. nucleases are known to be in the random coil, "swollen" and the tight native state. Assuming that in a given state, i.e., at a certain pH, the molecular radii of the nucleases are the same, their hydrodynamic radii were calculated from their pertinent electrophoretic mobilities. The respective radii of 50.1, 26.8, and 25.0 Å thus obtained agreed very well with the corresponding radii of gyration obtained from X-ray scattering. In fact, from the electrophoretic mobilities at pH 9.5, the existence of a hitherto unknown swollen basic state of the nuclease having a hydrodynamic radius of 30.5 Å was postulated. In addition, a method was described to evaluate the valence of the proteins at different pH from their pertinent electrophoretic mobilities. A general advantage of this method is that only the differences between the valences of the mutants and the wild type are needed; and for none of the proteins is required the knowledge of the actual valence. The results of the methods allowed the construction of a pH profile of the protein's valence. For the wild type, this profile was compared to the H^+ titration curve and the agreement was excellent. Both methods employed some novel structure–electrophoretic mobility relationships and the predicted protein properties compared remarkably well to the values obtained by exoelectrophoretic methods such as pH titration and X-ray scattering. Surprisingly, certain S. nucleases having the same valence could also be readily separated by CZE in some cases under the same conditions used for the others. Close examination of appropriate X-ray crystallography and/or NMR data indicated subtle differences in the molecular structure of these proteins that could be responsible for slight alteration in their hydrodynamic radii. The notion of virtual mobilities was used to assess the effect of such changes; and it is shown that for the native form of the S. nucleases a 2 Å change in the molecular radius is equivalent to a unit change in valence in bring about the same variation in the electrophoretic mobility.

1. Introduction

Capillary zone electrophoresis (CZE) has been used for the analysis of complex protein mixtures such as human serum proteins [1–5] and bovine brain proteins [5] as well as for the assay of purified biosynthetic proteins, *e.g.*, human insulin [6,7] and recombinant human

* Corresponding author.

0021-9673/95/$09.50
SSDI 0021-9673(94)00549-O

growth hormone [8]. Despite the success of CZE in protein analysis as demonstrated also by the separation of variants of transferrin [9] and immunoglobulin G [10], there is still a lacuna in understanding the dependence of electrophoretic mobility on molecular structure although the observed electrophoretic behavior of small peptides has been related to their molecular properties in a relatively straightforward fashion [6,11–15].

With large peptides or proteins only in a few studies were quantitative relationships developed between electrophoretic mobility and molecular structure. For instance, an attempt was made to relate the electrophoretic mobility of deamidated variants of immunoglobulin G to their amino acid composition and the pH of the medium in order to predict the optimal pH range for the separation of these similarly sized proteins [10]. In many cases only a qualitative interpretation of the experimental data was presented. For example, the higher mobility of the unfolded form of bovine α-lactalbumin has been attributed to an increase in its characteristic charge upon thermal denaturation [16]. The separation of recombinant human growth hormone from its deamidated variant was also interpreted by a charge difference in this case between the two similarly sized proteins [8].

In this work, we investigated the conditions for the separation of 11 mutants and the wild type of the enzyme Staphylococcal nuclease (S. nuclease) [ribonucleate(deoxyribonucleate)-3′-nucleotido-hydrolase, EC 3.1.4.7], produced by r-DNA technology in *E. coli*. This protein has served in several model studies of enzyme function and thermodynamic stability [17–19] as well as that of protein folding and unfolding [20]. The crystal structure of S. nuclease [19] is known at 1.7 Å resolution and information is available on its hydrodynamic radius in the native state [21] and on the radii of gyration of its different conformations [23]. Renewed interest in S. nuclease is due to the ready availability of variants obtained by site directed mutagenesis. They have been used in studies on protein stability [21] and enzymatic activity [25]. The 3D structure of several mutants is known from X-ray diffraction with 1.9 Å resolution [19,22,26–28] and NMR measurements [24,29]. The wild type and mutants of S. nuclease under investigation comprise a set of well characterized proteins with small but well defined structural differences as a result of site-specific mutation at only one amino acid residue, as illustrated in Fig. 1.

Despite the difficulties encountered earlier in the separation of nuclease isoforms from *Serratina marcescens* [30], the S. nucleases represent a set of well characterized and closely related proteins suitable for establishing quantitative structure and mobility relationships in CZE. Similar efforts in HPLC had only limited success probably due to denaturation of the proteins under conditions of reversed-phase chromatography [31].

In this study, we examine the conditions for the separation of 12 nucleases and relate their electrophoretic behavior to molecular features. Particular attention is paid to the chemical properties of the buffers employed to cover the wide pH range investigated. For the separation of these very basic proteins and thus for the successful completion of this study, the choice of the buffers was decisive.

2. Experimental

2.1. Instruments

CZE experiments were performed using a 2100 Model P/ACE capillary electrophoresis unit with UV detection at 214 nm at a temperature setting of 25°C. Only a few experiments were carried out at elevated temperatures. A PowerMate SX/20 from NEC Technologies (Boxborough, MA, USA) with P/ACE 2000 Series Microsoft for Windows version 2.02 (Beckman Instruments, Fullerton, CA, USA) was used for control of the instrument and data processing. Fused-silica capillary tubes of 375 μm O.D. and 75 μm I.D. with polyimide outer coating were obtained from Quadrex (New Haven, CT, USA). Their inner wall was coated with a polyacrylamide layer by a method adapted from the literature [32,33]. In each experiment, the length of the capillary and the pertinent migration distance were 470 mm and 400 mm, respectively.

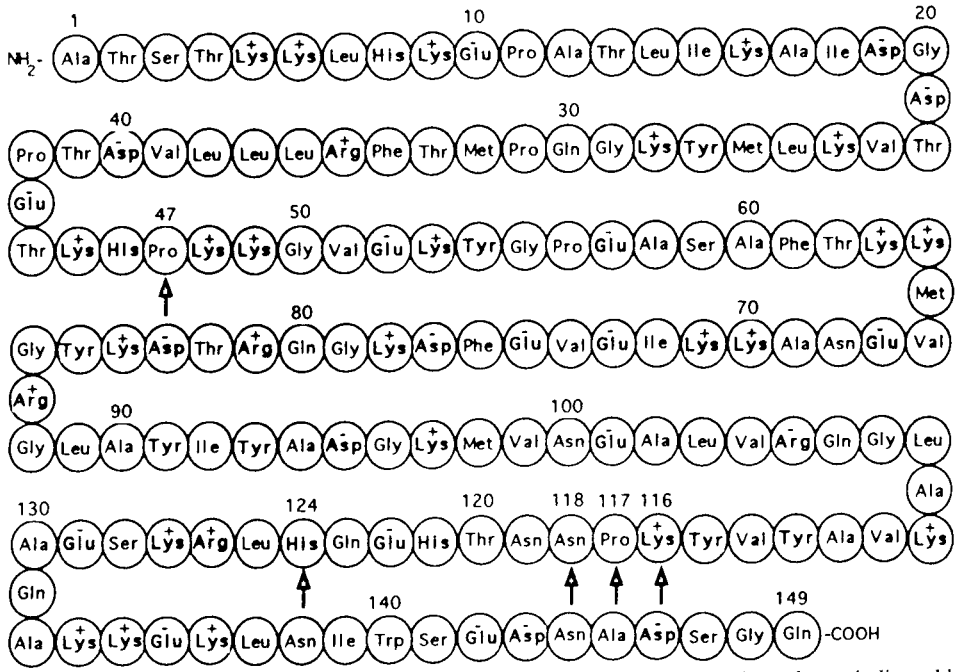


Fig. 1. Covalent structure of S. nuclease [20]. The positions of point mutation pertinent to this study are indicated by the arrows. At neutral pH positively and negatively charged amino acid residues are marked with + and − respectively.

2.2. Proteins

Staphylococcal nuclease A, the wild type nuclease from *Staphylococcus aureus*, and its mutants were obtained by oligonucleotide directed mutagenesis [34] and expressed in E. coli. The detailed procedures as well as the purification of the protein and the subsequent NMR and crystallographic experiments are described for the wild type in the literature [19,35], whereas the pertinent information on the mutants are given in Table 1. Ribonuclease A from Bovine Pancreas Type II-A was purchased from Sigma (St. Louis, MO, USA).

2.3. Chemicals

Trimethylphenylammonium iodide was purchased from ChemService (West Chester, PA, USA). Diethylenetriamine (DIEN) and piperazine (PIP) were purchased from Fluka (Ronkonkoma, NY, USA). Triethylenetetramine (TETA) was obtained from Aldrich (Milwaukee, WI, USA). Reagent grade phosphoric acid (85%), boric acid and sodium hydroxide were supplied by Fisher (Pittsburgh, PA, USA), sodium borate ($Na_2B_4O_7 \cdot 10H_2O$) from Mallinckrodt (Paris, KT, USA), trifluoroacetic acid (TFA) from J.T. Baker (Phillipsburg, NJ, USA) and citric acid was obtained from Sigma (St. Louis, MO, USA). Volumetric solutions of 0.1 *M* HCl and 0.1 *M* KOH were purchased from J.T. Baker. Potassium chloride was supplied by Fisher. Buffer solutions were obtained for the calibration of the pH electrode at pH 10.00 from Brand-Nu Laboratories (Meriden, CT, USA), at pH 4.01 (potassium biphthalate) and at pH 7.00 (phosphate) from J.T. Baker. Deionized water was prepared by the NanoPure purification system from Barnstead (Boston, MA, USA) and used throughout the experiments.

2.4. Buffers

Citrate buffer, pH 2.8 was prepared by adjusting the pH of 30 m*M* aqueous citric acid solution with 1 *M* NaOH. A series of buffers composed of phosphate and/or trifluoroacetic acetate salts

Table 1
S. nucleases under investigation.

Protein	Group	Code	Ref.	$\Delta(M_r)$[a]	$(z_{wt} - z_m)$	Valence
wild type	1	0	[19]	–	–	7.49
Asn118Gly	1	1a	[24]	−57	0	7.49
Pro117Thr	1	1b	[22,39,40]	+4	0	7.49
Pro117Gly	1	1c	[22,39,40]	−40	0	7.49
Pro117Ala	1	1d	[22]	−26	0	7.49
Pro47Ala	1	1e	[27]	−26	0	7.49
His124Leu	2	2a	[40]	−66	1 (pH ≪ 5.9[b])	
					0 (pH ≫ 5.9[b])	7.49
Lys116Gly	2	2b	[26]	−71	1	6.63
Lys116Ala	2	2c	[26]	−57	1	6.63
Lys116Met	2	2d	[26]	+3	1	6.63
Lys116Asp	3	3a	[27]	−13	1 (pH ≪ 3.6[c])	
					2 (pH ≫ 3.6[c])	5.60
Lys116Glu	3	3b	[27]	+1	1 (pH ≪ 4.3[c])	
					2 (pH ≫ 4.3[c])	5.49

The changes in molecular weight, $\Delta(M_r)$, and in valence, $(z_{wt} - z_m)$ upon substituting the appropriate amino acid residue in the wild type are shown for each mutant. The proteins are divided into three groups 1, 2 and 3 depending on the valence change upon mutation and given a letter code for identification, see Figs. 3 and 13 as well as Tables 4 and 5. Also shown is the valence of the proteins at pH 8.89 as calculated from experimental data by Eqs. 4 and 6

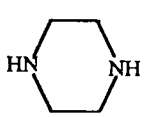
[a] The molecular weight of S. nuclease based on its amino acid composition is 16 811.

[b] pK_a value of His124 in S. nuclease at 25°C [29].

[c] pK_a values of Asp and Glu in aqueous solution at 25°C [52].

of selected aliphatic di-, tri- and tetra-amines [36] having pK_a values in the pH range from 3.25 to 9.84 as shown in Table 2 was employed. Phosphoric acid (85%) or trifluoroacetic acid was added to the 12.5 m*M* or 25 m*M* aqueous amine solution to reduce the pH close to the pK_a value of the amine. Borate buffer, pH 8.4 was prepared by adjusting the pH of 25 m*M* aqueous sodium borate with 100 m*M* boric acid.

Table 2
Chemical structure and pK_a values at 25°C of amines used for the preparation of buffers [36]

Amine	Chemical structure	pK_a
DIEN	$H_2NC_2H_4NHC_2H_4NH_2$	4.23 9.02 9.84
TETA	$H_2NC_2H_4NHC_2H_4NHC_2H_4NH_2$	3.25 6.56 9.08 9.74
PIP		5.76 9.72

2.5. Electrophoresis

The operating conditions are listed in Table 3. Proteins were dissolved in water to obtain a concentration of ca. 1 mg/ml. The samples were injected by applying 0.5 p.s.i. (1 p.s.i. = 6,894.76 Pa) pressure for 1 to 2 s, and the respective sample volumes were estimated as 6 nl and 12 nl. Between runs the capillary was flushed for 2 min with water and for 3.5 min with the background electrolyte at an inlet pressure of 20 p.s.i. Data concerning the measurement of the electrophoretic mobility of S. nuclease are shown in Table 3. The electrophoretic mobilities of the mutants were measured with the wild type as an "internal standard". The precision of the relative electrophoretic mobilities was better than 0.23% R.S.D.

The capillaries were tested according to the following protocol. Newly prepared capillaries

Table 3
Experimental conditions and precision of electrophoretic mobility measurements

pH	Temp. (°C)	Buffer [a]	Ionic strength (mmol/kg)	Double layer thickness (10^{-9} m)	Electric field (V/cm)	Current (μA)	Precision of electrophoretic mobillity	
							RSD [b]	n [c]
2.8	25	30 m*M* Citric acid–1M NaOH	5.5	2.94	426	28	0.3	9
4.1	25	25 m*M* DIEN–H_3PO_4	55.0	0.92	319	57	1.1	3
5.7	25	25 m*M* PIP–TFA	36.0	1.13	426	60	2.3	6
6.8	25	12.5 m*M* TETA–TFA	25.5	1.34	426	65	2.2	7
8.9	25	12.5 m*M* TETA–TFA	14.0	1.82	426	44	1.9	12
9.5	25	12.5 m*M* TETA–TFA	6.5	2.66	426	28	1.8	6

[a] The acronyms are explained in Section 2.3 and in Table 2.
[b] Percent relative standard deviation of the mobility for the wild type.
[c] Number of runs.

were checked with aqueous acrylamide solution (0.05%, w/v) and only those capillaries were used which exhibited no electroendoosmotic flow at 30 kV in either the cathodic or the anodic direction with 100 m*M* borate buffer, pH 8.4, as the background electrolyte. In the course of the experiments with proteins, the capillaries were tested under the conditions given in Table 3 by injecting a 0.06% (w/v) solution of trimethylphenylammonium iodide after every 5 runs. The mobility of the trimethylphenylammonium tracer was $(3.43 \pm 0.024) \times 10^{-8}$ m^2/Vs and did not change with the pH. This allowed a rapid diagnosis of the electrophoretic system since any malfunction manifested itself in broad peaks and relatively high apparent mobility of the tracer.

2.6. H^+ titration of the S. nuclease

The H^+ titration of S. nuclease was carried out in a fashion similar to that described in Refs. [37,38]. The experimental set-up consisted of a Model pHM82 pH meter with a No. GK 473901 LL2 glass electrode, a Model ABU80 10 ml auto burette and a Model III80 titrator from Radiometer America (Cleveland, OH, USA). The titration curve was recorded with a Kipp and Zonen (Delft, Netherlands) Model BD41 strip chart recorder at a chart speed of 1 mm/s. The electrode was calibrated at 25°C and pH of 4.01, 7.00 and 10.00. The deionized water used was boiled ad hoc to remove carbon dioxide. First highly purified ribonuclease A was titrated in the pH range from 2.50 to 12.00 according to Ref. [38] and the titration curve thus obtained was identical to that published in the literature [38]. For the H^+ titration of S. nuclease a 10-mg/ml stock solution of lyophilized and salt free protein was prepared in 0.05 *M* KCl and the pH of the solution was 6.61. The titration was carried out in the pH range from 2.50 to 12.00 with 0.1 *M* KOH and 0.1 *M* HCl above and below pH 6.61, respectively. The protein titration curve was corrected by the results of a direct titration of the 0.05 *M* KCl solution. Every experiment was repeated three times and the precision of the measurements was ± 0.02 pH units.

3. Results and discussion

3.1. The proteins under investigation

The S. nucleases are single chain globular proteins of 149 amino acid residues without sulfhydryl or disulfide groups. The amino acid sequence of the S. nuclease wild type is shown in Fig. 1. It contains 5 arginine, 23 lysine, 4 histidine, 7 tyrosine, 12 glutamic and 8 aspartic

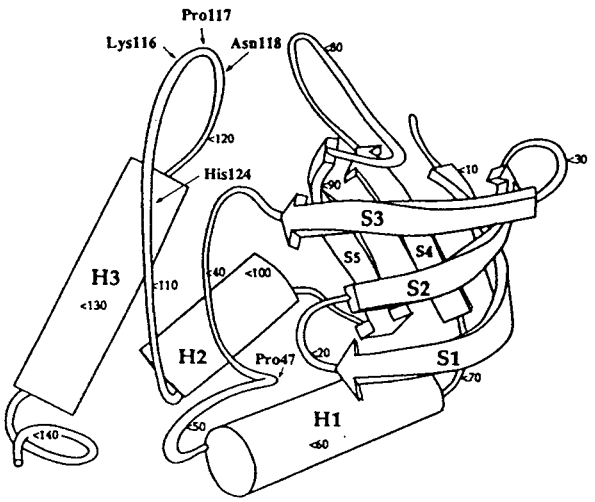


Fig. 2. Tertiary fold of native S. nuclease. The five β-strands S1–S5 that form a β-barrel are indicated. H1–H3 denote the three α-helices and the places of point mutation pertinent to this study are indicated by small arrows (adapted from [19]).

acid residues and has an isoelectric point of 10.3. The tertiary structure of S. nuclease is shown in Fig. 2. It is composed of a highly twisted, five stranded β-barrel labeled S1–S5, and three α-helices labeled H1–H3. The backbone structure and side chain orientation of the wild type were extensively studied by X-ray and NMR measurements [19,29,35]. The radius of gyration of the native tight tertiary fold is 16.2 Å as determined by X-ray scattering [23]. The mutants under investigation are very similar as far as their molecular shape and dimensions are concerned [22,24,26,27,39,40].

The S. nuclease mutants in this investigation fall into three main groups according to their valence, z_m, with respect to that of the wild type, z_{wt}, as shown in Table 1 and discussed below.

(1) Five mutants are obtained by replacing one neutral amino acid in the wild type by another. The mutants and the wild type have the same valence; the two mutants Pro47Ala and Pro117Ala have identical amino acid compositions.

(2) Four mutants are the result of the replacement of a basic amino acid (Lys116 or His124) in the wild type by a neutral amino acid (Ala, Gly, Leu or Met). As a result the valence of these mutants is $(z_{wt} - 1)$ at acidic pH where the basic amino acids are fully protonated.

(3) Two mutants are obtained by the replacement of a basic amino acid (Lys116) in the wild type by an acidic amino acid (Asp or Glu). Their valence is $(z_{wt} - 1)$ at low pH where the glutamic or aspartic acid are fully protonated or $(z_{wt} - 2)$ at pH values were the acidic side chains are fully dissociated.

3.2. Development of the separation method

The first goal of our work was to evaluate the potential of CZE for the rapid separation of basic proteins that are closely related both in charge and size. The set of S. nucleases was selected not only for the availability of the mutants but also for their suitability as model proteins to study conditions for the separation of basic proteins by free solution CZE over a wide pH range.

To overcome the problem of wall adsorption, one strategy involves the employment of capillaries with inert coating which is hydrolytically stable in the operational pH range [32,33,41–46]. The most common coating imparts to the tube a hydrophilic inner surface via a polyacrylamide layer and allows the use of buffers having low ionic strength and covering a wide pH range. In our experiments such capillaries were used in the pH range from 2.5 to 10 for several hundred runs without deterioration.

Selection of the buffer system

The most common buffers in CE are based on phosphate that has three pK_a values of 2.1, 7.2 and 12.3. Therefore, it can be used in a wide pH range, but between pH 3 to 5.5 and from pH 8.5 to 11 it buffers poorly. Carboxylic buffers are well suited in the pH range from 2.8 to 5.6 but they absorb light rather strongly below 230 nm. Citrate buffer has a uniform buffering capacity in the pH range from 2.6 to 6.9, however, the mobility of S. nucleases above pH 4 was an order of magnitude lower in citrate than in other buffers at the same pH. Consequently, citrate was suitable for rapid separation of the strongly

basic proteins only below pH 4. The use of glycine–HCl buffer at pH 2.8 resulted in a plurality of peaks [47] upon injecting any of the S. nucleases or ribonuclease A and this phenomenon is subject pending investigations. Since Good's buffers [48] are zwitterionic and have low conductivity, they are popular in CZE. However, their pK_a values are between 6.1 and 10.4 so that they do not cover a sufficiently wide pH range. Furthermore, we observed with Tris, HEPES and MES buffers that the mobilities of S. nucleases in the pH range from 6.5 to 8.5 were a magnitude lower than with phosphate under otherwise identical conditions.

In our quest for an appropriate buffer system, we explored the use of buffers based on aliphatic di-, tri- and tetra-amines listed in Table 2 that have been originally introduced in reversed-phase chromatography [36]. Their pK_a values range from 3.25 to 9.84 and they provide a series of buffers covering a wide pH range and having a sufficient high transparency in the low UV. These buffers are also potential masking agents for silanolic adsorption sites at the surface [36]. Since these amines have a relatively high heat of protonation, the pH of the buffers is strongly temperature dependent [36] and this mandates precise temperature control of the capillaries. As seen from Table 3 the greatest relative standard deviation was 2.3% for the mobility of the wild type at different pH and buffer environments. This suggests that the temperature control of the instrument was satisfactory even under such demanding conditions. By using such buffers in CZE at pH 6.8, the separation of the standard proteins cytochrome *c*, lysozyme and trypsin was carried out with high efficiency and high selectivity in 10 min. The number of theoretical plates per meter is higher than 500 000, a respectable efficiency with basic proteins at neutral pH.

Operating conditions

In our experiments, the capillary length and inner diameter were fixed at 47 cm and 75 μm, respectively. The pH ranged from 2.8 to 9.5 so that it was below the p*I* of the S. nuclease. In each experiment, the pH was close to one of the pK_a values of the buffer, *cf.* Table 2, in order to obtain high and uniform buffering capacity at relatively low concentrations. This allowed the use of high voltage without untoward increase in the current so that high separation efficiency could be obtained. The applied voltage was kept below 20 kV as to operate in the Ohm's law regime. The voltage range was kept narrow in order to facilitate a comparison of the separation efficiencies at all different pH conditions. Consequently, the current was not the same in all experiments as seen in Table 3.

In order to investigate the effect of elevated temperature on the efficiency of separation, experiments were carried out at pH 6.8 and temperatures up to 50°C. The efficiency and selectivity of the separation of S. nucleases at 25 and 30°C were comparable but decreased dramatically above 35°C. In the temperature range from 40 to 50°C, excess peak broadening and distortion of peak shape were observed for most of the S. nucleases. This behavior is explained by thermally induced conformational changes of the proteins [16]. Indeed, the temperatures where the highly irregular elution profiles were obtained, correspond to the transition temperatures 49, 50 and 54°C [26] of the wild type, Lys116Ala and Lys116Gly, respectively.

Effect of pH on the mobilities and separation

The dependence of the electrophoretic mobilities of all S nucleases on the pH is shown in Table 4 and the corresponding electropherograms of 5 representative proteins in Fig. 3. The results show that under the conditions employed in this study relatively sharp peaks were obtained over the whole pH range investigated and that at intermediate pH such closely related proteins can be separated with high efficiency and resolution. From Table 4 and Fig. 3 we gather that the mobilities of all S. nucleases decrease with increasing pH and no mutant has a mobility higher than the wild type because of the limitations on the types of available mutants. Although all 12 S nucleases could not be separated from each other, all mutants with the exception of Pro47Ala could be separated from the wild type.

Table 4
Electrophoretic mobilities of S. nucleases at different pH values

Protein	Code [a]	Electrophoretic mobility ($m^2\ V^{-1}\ s^{-1} \times 10^8$) at pH					
		2.1	4.1	5.7	6.8	8.9	9.5
Wild type	0	2.45	2.04	1.85	1.81	1.31	0.81
Asn118Gly	1a	2.45	2.00	1.81	1.78	1.28	0.77
Pro117Thr	1b	2.45	2.04	1.85	1.81	1.28	–
Pro117Gly	1c	2.45	2.04	1.85	1.81	1.28	–
Pro117Ala	1d	2.45	2.04	1.85	1.81	1.28	–
Pro47Ala	1e	2.45	2.04	1.85	1.81	1.31	–
His124Leu	2a	2.45	2.04	1.85	1.81	1.31	–
Lys116Gly	2b	2.37	1.94	1.71	1.65	1.14	0.67
Lys116Ala	2c	2.37	1.94	1.73	1.68	1.16	0.69
Lys116Met	2d	2.37	1.94	1.73	1.68	1.16	–
Lys116Asp	3a	2.37	1.86	1.62	1.55	0.98	–
Lys116Glu	3b	2.37	1.82	1.61	1.53	0.96	0.51

The experimental conditions and precision of measurements are given in Table 3.
[a] Codes for the mutants are listed in Table 1.

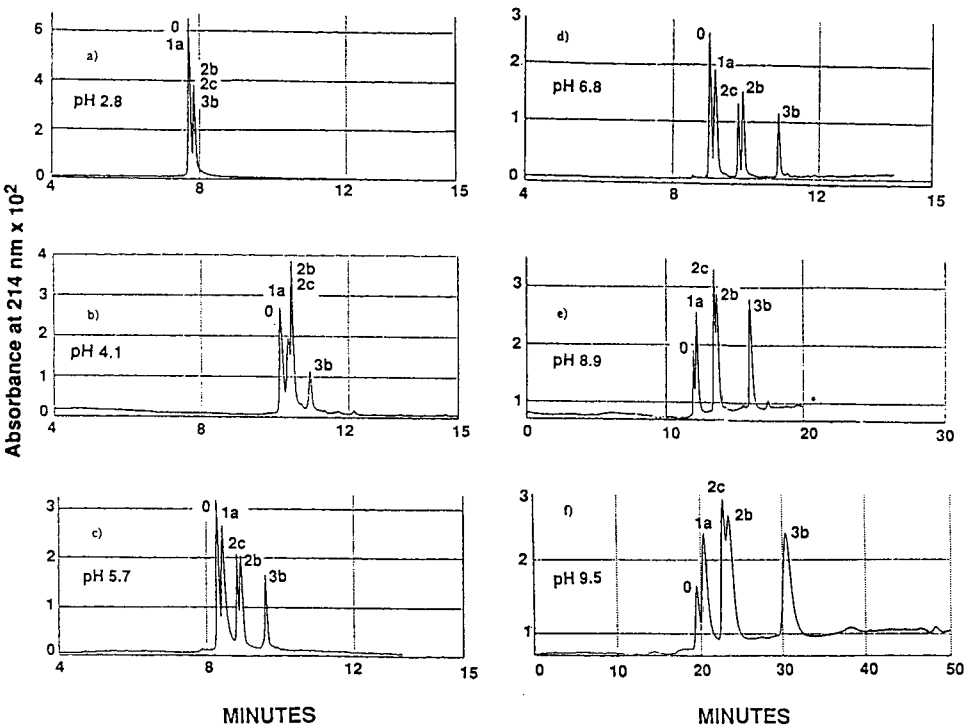


Fig. 3. Electropherograms of the S. nucleases obtained at various pH values. Peaks: 0 = wild type, 1a = Asn118Gly, 2c = Lys116Ala, 2b = Lys116Gly, 3b = Lys116Glu. The experimental conditions are in Table 3.

The data shown in Fig. 3 and Table 4 indicates that the valence differences between the protein at the pH of the experiment are mainly responsible for the separation. For instance, at pH 2.8 the proteins migrate in two groups, whereas at higher pH they form three main groups in agreement to the classification in Table 1 according to the $(z_{wt} - z_m)$ values. However, as seen from Table 4 and Fig. 3, in the pH range from 4.1 to 9.5, several S. nucleases that carry the same net charge and therefore belong to the same Group in Table 1 were also separated. Although the resolution of such variants is much lower than that of mutants having different valences, the separation of such very similarly sized proteins is quite unexpected and must be related to minute differences in the tertiary fold of the variants. This is in agreement with the observation that at pH 2.8 where the proteins are denatured with the loss of the subtle differences in their tertiary structure, only variants having different $(z_{wt} - z_m)$ values could be separated.

Fig. 4 shows the pH dependence of the relative electrophoretic mobilities of the mutants with respect to that of the wild type. It is seen that the proteins migrate in three groups in the pH range from 4.1 to 9.5 and in two groups at pH 2.8 according to their $(z_{wt} - z_m)$ values shown in Table 1. For mutants that differ in their $(z_{wt} - z_m)$ values, the relative mobility increases with the pH due to an increase in the relative valence differences when the pH approaches the p*I* of the proteins that is in the neighbourhood of 10.3. For instance, the electrophoretic mobility of the Lys116Ala mutant relative to that of the wild type increases by a factor of about 5 when the pH is raised from 2.8 to 9.5. Of course, the concomitant decrease in the electrophoretic mobility results in protracted separation time as seen in Fig. 3 and Table 4.

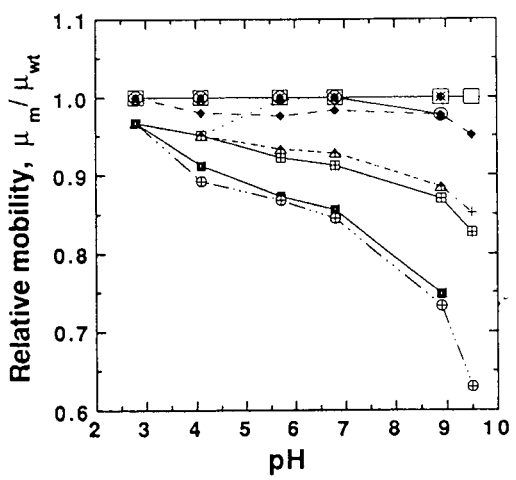


Fig. 4. The effect of pH on the relative electrophoretic mobilities (μ_m/μ_{wt}) of the S. nuclease mutants with that of the wild type as the reference. The experimental conditions are in Table 3. The symbols for the proteins are: □ = wild type, ◆ = Asn118Gly, ○ = Pro117Thr, ● = Pro117Gly, ▲ = Pro117Ala, * = Pro47Ala, ◇ = His124Leu, ⊞ = Lys116Gly, + = Lys116Ala, △ = Lys116Met, ◘ = Lys116Asp, ⊕ = Lys116Glu.

3.3. Combined effect of molecular weight and valence on the electrophoretic mobility of the S. nucleases

The electrophoretic mobility, μ, of a spherical particle such as a protein molecule in an electrolyte solution can be expressed [49] as

$$\mu = \frac{ze\ \phi(\kappa R)}{6\pi\eta R\ (1 + \kappa R)} \tag{1}$$

where z is the valence, e is the electronic charge (1.602×10^{-19} coulombs), R is the hydrodynamic radius, η is the viscosity, κ is the Debye screening parameter and $\phi(\kappa R)$ is Henry's function [50].

At fixed experimental conditions Eq. 1 can be simplified as

$$\mu = \frac{z}{\gamma R} \tag{2a}$$

where

$$\gamma = 6\pi\eta\ (1 + \kappa R)\ /\ e\ \phi(\kappa R) \tag{2b}$$

γ is constant under a given set of experimental conditions.

In an attempt to use the experimentally obtained electrophoretic mobility, μ, for the development of quantitative structure–mobility relationships, we make use of the geometrical argument that the radius R of spherical protein

molecules such as S. nucleases [23] is proportional to the one third power of the molecular volume or molecular mass (M_r) [51]. By substituting $(M_r)^{1/3}/\rho$ for the hydrodynamic radius R in Eq. 2 we obtain

$$\mu = \frac{\rho}{\gamma} \frac{z}{(M_r)^{1/3}} \quad (3)$$

where the factor ρ is inversely proportional to the cubic root of the partial specific volume of the spherical protein molecule and is constant as long as the molecular size of the protein remains the same.

In Fig. 5a, the experimentally obtained mobilities of the protein variants at different pH are plotted against their $z/(M_r)^{1/3}$ ratios according to Eq. 3. In a first approximation the z values of the proteins were calculated as the sum of the valences of the amino acids (see Fig. 1) in free solution as obtained by using the Henderson–Hasselbach equation with pK_a data from the literature [52] for the pH environment of interest.

In Eq. 3, ρ is considered constant and independent of the pH as long as the conformation of the proteins, and therefore, their molecular sizes remain the same. Under such conditions, the electrophoretic mobility should be a linear function of $z/(M_r)^{1/3}$ with a constant slope of ρ/γ. As seen in Fig. 5a, the relationship is linear for all proteins in the pH environments investigated. However, the slopes are pH dependent. The slope is significantly smaller at pH 4.1 than in the pH range from 5.7 to 9.5 and even smaller at pH 2.8. If γ is fairly constant, changes in ρ with pH must be responsible for the slope differences. The findings in Fig. 5a suggest that the molecular radii of the S. nucleases are pH dependent and fall into three groups.

In the literature, pH dependent conformational states of S. nucleases have been described in detail. The radius of gyration of the tight native tertiary fold is 16.2 Å as measured at neutral pH by X-ray scattering [23]. It is known that in a mildly acidic environment, around pH 4, the S. nucleases are in a "swollen" state which is sometimes referred to as "molten globule". The radius of gyration of the wild type in the "swollen" state is 21.2 Å as determined by X-ray scattering of a large S. nuclease fragment as the model molecule [23]. For the random coil, which is present at strongly acidic pH, the radius of gyration was estimated 40–50 Å [23]. In view of the very close structural similarity between the wild type and the mutants under investigation, we may assume that the three radii of gyration

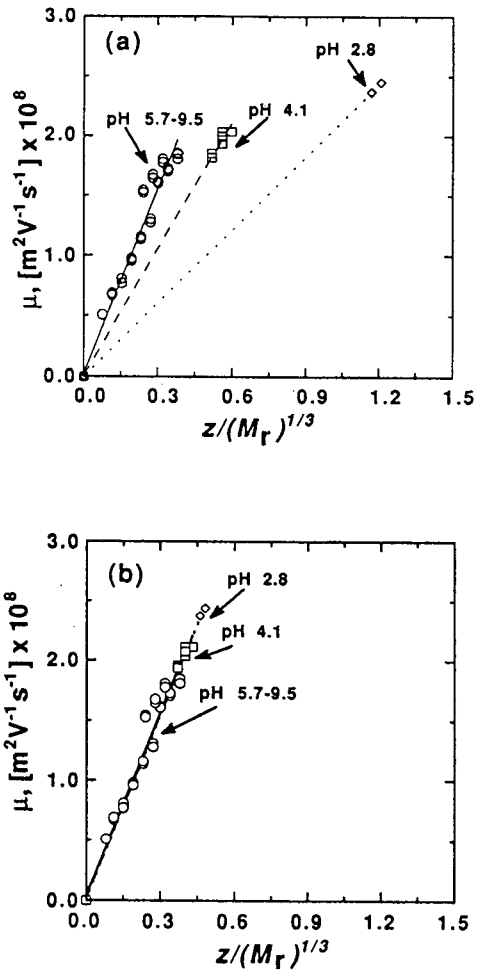


Fig. 5. Plots of the electrophoretic mobility and $z/(M_r)^{1/3}$ according to eqn. 3. (a) electrophoretic mobility measured at different pH values. (b) electrophoretic mobility adjusted for the size differences based on the slopes. The symbols are: for pH 2.8 (◇); for pH 4.1 (□); for pH 5.7, 6.8 and 8.9 and 9.5 (○). The valence for the proteins was calculated based on the Henderson–Hasselbach equation and pK_a values of the amino acid side chains taken from [52].

discussed above apply with reasonable accuracy to the appropriate conformations of the S. nucleases used in this study.

It is intriguing to consider that the three lines in Fig. 5a correspond to these three different states of the S. nucleases. In order to examine this possibility, we calculated the ratios of the radii of gyration from the above mentioned literature data. The radius of gyration ratios of the denatured and "swollen" S. nuclease molecules to the tight native tertiary fold were determined as 2.5 and 1.3, respectively. These values compare very favorably to 2.5 and 1.4, the respective slope ratios of the straight lines shown in Fig. 5a. The close correlation of the ratios of the radii estimated on the basis of protein structural studies on one hand and from electrophoretic mobilities on the other is encouraging. Indeed, by using the ratios of the literature data above we adjusted the experimental data for pH induced changes in the molecular sizes of these proteins and obtained a single straight line upon plotting μ versus $z/(M_r)^{1/3}$ as shown in Fig. 5b.

3.4. Estimation of the valence of the proteins

The approach described above offers a very simple method for screening the various conformational states of S. nucleases or other closely related proteins. Nevertheless, the full potential of this method can not be exploited until the valences of the proteins are known more accurately. The present use of the Henderson–Hasselbach equation is impaired by a shortfall of the pK_a values appropriate for the amino acid residues present in the protein molecule [53]. The calculation of valence from mobility according to Eq. 1 would require the knowledge of the hydrodynamic radius, which at best can only be crudely approximated [10,54], therefore, the use of this approach for the determination of the valences of the proteins is beset with uncertainties.

Valence of the wild type from CZE data

In the following we shall use the subscripts wt and m to denote properties of the wild type and the mutants, respectively. The mobility of the wild type, μ_{wt}, is the highest among the proteins investigated (see Table 4) and it will be used as the reference. If we assume that the mutants have the same hydrodynamic radius as the wild type, i.e., $R_{wt} \approx R_m \approx R$, then we obtain from Eq. 1 that

$$\frac{\mu_{wt} - \mu_m}{\mu_{wt}} = \frac{1}{z_{wt}}(z_{wt} - z_m) \quad (4)$$

where the mobilities and valences are at a given pH and buffer environment. The LHS of Eq. 4 can be readily evaluated from experimental data, whereas the valence difference $(z_{wt} - z_m)$ can be calculated for each mutant from the change in the valence upon substituting a given amino acid by another. According to Eq. 4 plots of the two terms against each other should yield straight lines with slopes given by the reciprocal valence, $1/z_{wt}$, of the wild type at the pH of the experiment. The merit of Eq. 4 is that the valence, the drag coefficient and the double layer thickness of the protein do not have to be known, only the knowledge of the valence difference between the wild type and the mutants is required. Evidently, the change in valence upon substituting a single amino acid can be estimated with a higher accuracy than the valence of the protein proper. Of course, Eq. 4 is applicable only if the proteins have nearly the same hydrodynamic radius.

The application of the method requires at least two mutants having different valences. In general, site directed point mutation could result in the following $(z_{wt} - z_m)$ values: -2, $+2$, -1, $+1$ and 0. In our restricted set of mutants, however, only point mutation of the lysine116 side chain was utilized to bring about valence changes and, as shown in Table 1, valence differences of only 2, 1 and 0 were obtained. Nevertheless, in the plots, all available data points were used and the average mobility of the mutants in Group 1, 2 and 3 were employed in further calculations.

According to Eq. 4, data obtained for the wild type and proteins in Group 1, 2 and 3, in the pH range from 4.1 to 9.5, are depicted in Fig. 6, which shows that the experimental data handsomely conforms to the linear behavior predicted by Eq. 4. From the slopes of the plots, the

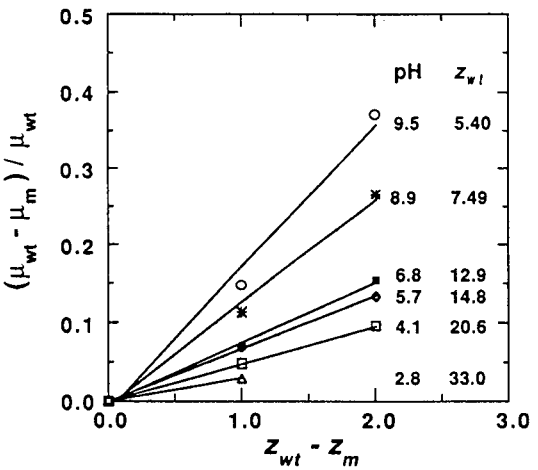


Fig. 6. Difference plots according to eqn. 4 to evaluate the valence, z_{wt}, of the wild type.

valences of the wild type at different pH have been determined and they are also shown in Fig. 6. It is seen that the valence of the wild type decreases by a factor of 6 from 33.0 to 5.4 when the pH is increased from 2.8 to 9.5.

Valence of the wild type from H^+ titration curve

In order to compare the above results to extra-electrophoretic data, the wild type was subjected to H^+ titration in the pH range from 2.5 to 12.0 as described in the experimental part. Titration curves for the study of the reaction of proteins with hydrogen ions can be obtained experimentally with considerable accuracy and the underlying theory is well established [55]. The resulting number of hydrogen ions bound per molecule of S. nuclease was calculated assuming that the protein reacts only with H^+ and OH^- and all charged groups are at the surface of the protein [56]. The results of blank titration were used to correct the S. nuclease titration data for the interaction of H^+ or OH^- with other substances present in the titration mixture. The valence of the protein was obtained by the relationship

$$z_{wt} = [H_S^+]_{pH} - [H_S^+]_{pI} \quad (5)$$

where $[H_S^+]_{pH}$ and $[H_S^+]_{pI}$ is the number of hydrogen ions bound per S. nuclease molecule at the experimental pH and at the p*I* of S. nuclease, respectively.

Comparison of the titration curves of S. nuclease

From the H^+ titration curve the valence of the wild type was calculated as a function of pH by Eq. 5 and the results are illustrated by the dashed line in Fig. 7. This was then compared to the valence calculated by Eq. 4 previously and the results are shown by the solid line in Fig. 7. It should be once more mentioned, that the valence obtained according to Eq. 4 for the wild type was determined independently from its hydrodynamic radius. It is shown in Fig. 7 that the two independent methods for constructing the titration curve of S. nuclease yield very similar results. This supports our earlier assumption that the wild type and the mutants have essentially the same hydrodynamic radii.

Valence of the mutants

Again, assuming that the mutants have the same molecular size as the wild type, from Eq. 1, we can express the valence of a mutant as follows

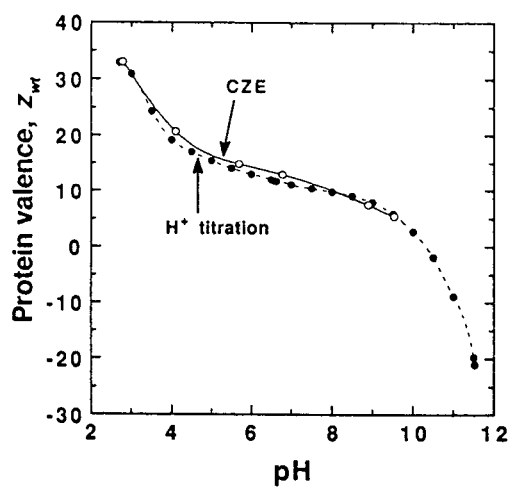


Fig. 7. Comparison of S. nuclease "titration curves" obtained directly by H^+ titration (●) or from the pH dependency of the electrophoretic mobilities calculated by using eqn. 4 (○).

$$z_m = \frac{\mu_m}{\mu_{wt}} z_{wt} \quad (6)$$

The valences of the mutants were determined at pH 8.9, where the S. nucleases are present in their native state, from their relative mobilities with respect to that of the wild type according to Eq. 6 with z_{wt} calculated by Eq. 4 and the results are shown in Table 1. It is seen in Tables 1 and 4 that a few mutants have slightly different mobilities despite their identical valences. Such mobility differences will be discussed later and assigned to minute differences in the tertiary fold of the variants.

3.5. Estimation of the hydrodynamic radius of the proteins

A similar approach is employed to estimate the pH dependence of the hydrodynamic radius, R, of our proteins. Although their radii may change with the pH, they are assumed to be the same at a given pH value. Therefore, we obtain from Eq. 1 the following relationship.

$$\mu_{wt} - \mu_m = \frac{1}{R} \frac{e}{6\pi\eta} \frac{\phi(\kappa R)}{(1+\kappa R)} (z_{wt} - z_m) \quad (7)$$

Eq. 7 allows us to evaluate R without the need of knowing the valences of the proteins at the experimental pH because this approach requires only the valence difference upon mutation. The corresponding data obtained for the wild type and the mutants of the Group 1, 2 and 3 (see Table 1) are plotted according to Eq. 7 with pH as the parameter in Fig. 8 that shows that the dependence of the mobility difference linearly depends on the valence difference at all pH values investigated.

The slopes of the straight lines in Fig. 8 depend on the hydrodynamic radius, R, the double layer thickness, $1/\kappa$, and the viscosity, η, according to Eq. 7. Since the buffers used here are fairly dilute, $\eta = 0.89 \times 10^{-3}$ Ns/m^2 [57], the viscosity of water at 25°C was taken for the calculations. We used an iteration procedure for the evaluation of the hydrodynamic radius, R, of the wild type and the mutants from the slopes in

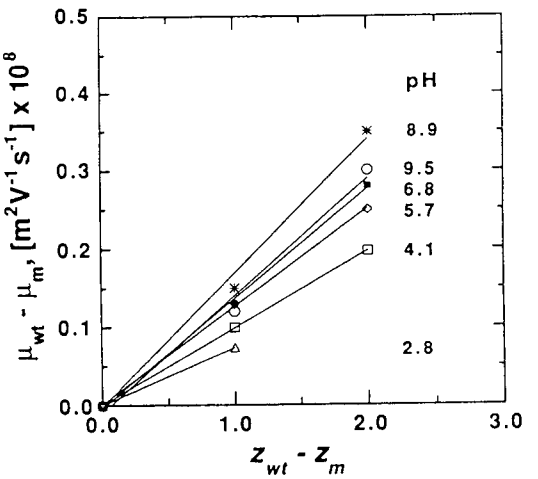


Fig. 8. Difference plots according to eqn. 7 to evaluate the hydrodynamic radii of S. nucleases.

Fig. 8 at an iteration tolerance less than 0.1%. For the double layer thickness the values shown in Table 3 were used. The radii of gyration, R_g, for S. nuclease at each pH were taken from the literature to start the iteration. First for each pH the κR_g values were obtained, then the appropriate values of Henry's function [49] were determined. Finally the appropriate hydrodynamic radii were calculated.

The hydrodynamic radii obtained by this procedure are shown in Fig. 9. It is seen that they strongly depend on the pH and fall into three groups. The largest radius, 50.1 Å is found in strongly acidic medium at pH 2.8 and the smallest radius is obtained under non denaturing conditions in the pH range from 5.7 to 8.9 with an average value of 25.0 Å. This is reasonably close to 21 Å, the hydrodynamic radius of the native S. nuclease as determined by gel filtration [21]. At pH 4.1 and 9.5 the radius had an intermediate value of 26.8 and 30.5 Å, respectively. The results lend further support to the results of protein structural studies [21,23,58]. They also confirm the existence of three different conformational states of S. nuclease as suggested by Fig. 5: the tight native tertiary fold, the swollen form (molten globule) and the random coil. However, we have found no reports to explain the observed 12% increase in the hydro-

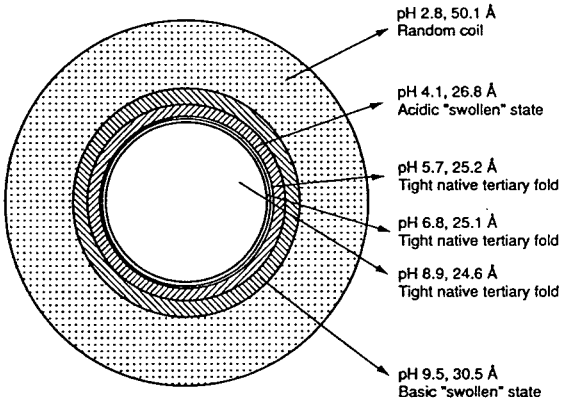


Fig. 9. Hydrodynamic radii of S. nucleases in different pH environments. The radii were calculated from the differences in the electrophoretic mobilities of the mutants and the wild type and from the valence differences upon mutation according to eqn. 7.

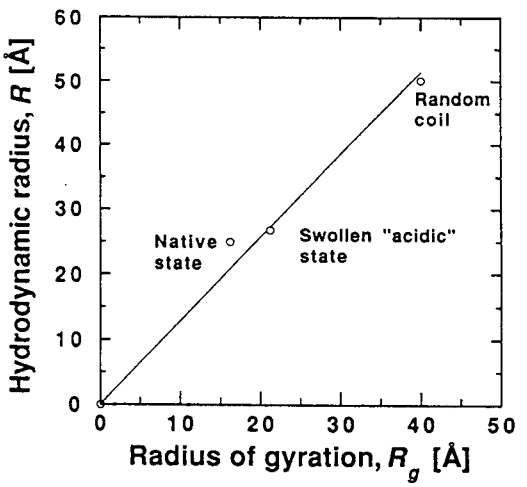


Fig. 10. Plot of the hydrodynamic radii against radii of gyration for S. nuclease in various conformational states. The radii of gyration were obtained from X-ray and NMR data [22], [21], [58], [23], whereas the hydrodynamic radii were calculated from electrophoretic mobilities by using eqn. 7.

dynamic radius of the S. nucleases at pH 9.5 with respect to that in the pH range from 5.7 to 8.9 as shown in Fig. 9. Therefore, the tentative existence of a fourth state of S. nuclease requires further investigation.

Hydrodynamic radius and radius of gyration

We recall that according to Eq. 1 the electrophoretic mobility depends on the hydrodynamic radius, R, that reflects the effective size and shape including solvent molecules clustering around it. On the other hand, the radius of gyration, R_g, is given by the average root mean square distance of the atoms from the center of the gravity [59]. For a sphere of radius, R, the theoretical relationship [60] between R and R_g is given by the proportionality

$$R = 1.291\, R_g \tag{8}$$

In Fig. 10 values of R taken from Fig. 9 were plotted against R_g taken from the literature [22,23]. There is an excellent linear correlation between the two kinds of radii of these conformations of S. nuclease for which R_g values were available. The slope of the plot in Fig. 10 gives a proportionality constant 1.286, which is very close to that in Eq. 8. This supports the notion that the S. nucleases are spherical in the different conformational states.

3.6. Dependence of the electrophoretic mobility at zero ionic strength on the molecular weight and valence of the proteins

Earlier in this report, notably in Eq. 3 as well as in Fig. 5, the effect of ionic strength on the mobility of the protein was not taken into account explicitly. With reliable estimates at hand of the R values shown in Fig. 9, we can evaluate the electrophoretic mobility of the proteins at zero ionic strength, μ^0, from the experimentally obtained mobility, μ, by using the following relationship

$$\mu^0 = \frac{\mu\,(1+\kappa R)}{\phi(\kappa R)} \tag{9a}$$

Furthermore, we expect that by using μ^0 instead of μ a more accurate correlation between the electrophoretic mobility and $z/(M_r)^{1/3}$ than that given by Eq. 3, which is rewritten for this case as

$$\mu^0 = \frac{\rho}{\gamma^*} \frac{z}{(M_r)^{1/3}} \quad (9b)$$

where

$$\gamma^* = 6\pi\eta / e \quad (9c)$$

and is constant for a given set of experimental conditions. Further improvement of the correlation can be expected from the use of more accurate z values calculated by Eqs. 4 and 6. Fig. 11 shows plots of μ^0 against $z/(M_r)^{1/3}$ for our proteins with data obtained at different pH values. The results presented in Fig. 11 are similar to those depicted in Fig. 5a, but a few important differences have to be noted.

The slope of the plot pertinent to pH 9.5 is different from that applicable for the pH range from 5.7 to 8.9, but similar to that which corresponds to pH 4.1. This finding is in agreement with the differences in the hydrodynamic radii shown in Fig. 9 and argues for the existence of another molecular state of S. nuclease at basic pH that is very similar in size to that observed at weakly acidic pH. It should also be noted that the data points measured at pH 6.8 were outliers in Fig. 5 whereas in Fig. 11 they are right on the line. This can be explained by the fact that the values of the electrophoretic mobilities in Fig. 5 at the various pH environments were established at different ionic strengths and the valences were only roughly estimated from valences of the amino acid residues by using the Henderson–Hasselbach equation whereas the valences used in Eq. 9 were obtained experimentally according to Eqs. 4 and 6.

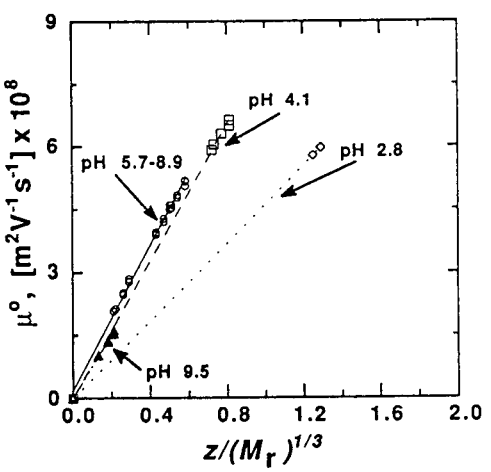


Fig. 11. Plots of the electrophoretic mobilities, $\mu°$, of all S. nucleases investigated against their $z/(M_r)^{1/3}$ value in the pH range from 2.8 to 9.5. The valences for the proteins were taken from Fig. 6. The least square correlation coefficient is 0.999 in all cases. The symbols are: for pH 2.8 (◇); for pH 4.1 (□); for pH 5.7, 6.8 and 8.9 (○) and for pH 9.5 (▲).

We used two methods for the evaluation of the hydrodynamic radii (or their ratios) of the S. nuclease forms. The first approach allowed us to calculate the appropriate R values by using Eq. 7 from the mutational valence difference without knowledge of the species' valences. The results are shown in Fig. 9. In the second method Eq. 9b was employed with z values calculated from Eqs. 4 and 6, i.e., independently from the hydrodynamic radii and the ionic strength of the medium. Plots of the results are shown in Fig. 11. The slopes are ρ/γ^*, where γ^* is constant and ρ equals $(M_r)^{1/3}/R$ for spherical molecules. In order to compare the results obtained by the two methods, we calculated the ratios of the hydrodynamic radii of the denatured, the "swollen" and the "basic" state S. nuclease forms to the tight native tertiary fold. The respective ratios were 2.0, 1.1 and 1.2 according to the first and 1.9, 1.0 and 1.2 according to the second method. The agreement between the two sets of ratios is gratifying as it demonstrates the internal consistency of this treatment.

3.7. Electrophoretic mobility and H^+ titration of S. nuclease

The pH profile of electrophoretic mobilities

Fig. 12 shows the experimentally obtained electrophoretic mobility of S. nuclease, μ_{wt}, as a function of pH in the range from 2.8 to 9.5. It is seen that the directly measured mobility of the S. nuclease, μ_{wt}, is not strongly influenced by the pH; it decreases only by a factor of 3 upon increasing the pH from 2.8 to 9.5. Since the valence of the protein decreases by a factor of 6 in this pH range, as shown in Fig. 7, one would expect a much larger change in μ_{wt}.

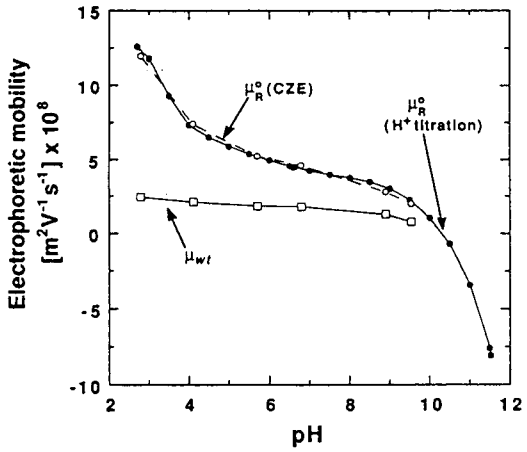


Fig. 12. The pH profiles of the various electrophoretic mobilities of S. nuclease. The pH dependence of the measured electrophoretic mobility, μ_{wt}, (□) was corrected for pH induced size changes and ionic strength effects by using eqn. 10a to obtain μ_R^0, (○) that is compared to the theoretical pH profile of S. nuclease calculated by using eqn. 10b (●) with valences obtained from H^+ titration and the hydrodynamic radius of S. nuclease at native state taken from Fig. 9.

In light of the previous discussion, the mobility data for S. nuclease were corrected for pH mediated changes in the molecular radius and for ionic strength effects. The fully corrected electrophoretic mobility, μ_R^0, was calculated by the relationship

$$\mu_R^0 = \mu_{wt} \frac{(1+\kappa R_{wt})}{\phi(\kappa R_{wt})} \frac{R_{wt}}{R_n} \quad (10a)$$

where R_n is the radius of the wild type in its native state and R_{wt} is its actual radius at the pH of the experiment. In Fig. 12 the pH profile of μ_R^0 for S. nuclease is shown by the dashed line.

The corrected mobility μ_R^0 is expected to be directly proportional to the valence of the protein at any pH

$$\mu^0 R = \frac{z_{wt}\, e}{6\pi\eta\, R_n} \quad (10b)$$

The theoretical pH profile of the S. nuclease mobility, μ_R^0, was calculated by using Eq. 10b with valences obtained from H^+ titration and with the hydrodynamic radius of the native S. nuclease from Fig. 9. The two pH profiles of μ_R^0 obtained from CZE and H^+ titration are very similar and the results suggest that valences obtained from H^+ titration can be used for *a priori* calculation of electrophoretic mobilities provided the hydrodynamic radii of the proteins under investigation are available.

3.8. Origin of electrophoretic selectivity for S. nuclease mutants of the same valence

Examination of X-ray and NMR data

As expected, mutants of the same valence generally exhibited the same mobility as shown in Table 4. Yet, certain S. nuclease mutants could be separated despite their nearly identical valence and molecular dimensions, as shown in Table 5 and in Figs. 3 and 13. As mentioned above, small differences in the tertiary fold may affect the relative electrophoretic mobilities of

Table 5
Examples for the separation of S. nucleases with the same valence

Protein pair [a]		pH [b]	Electropherogram	$\Delta\mu$ [c] $[m^2V^{-1}s^{-1}] \times 10^{10}$
Protein I	Protein II			
Wild type (0)	Asn118Gly (1a)	4.1 to 9.5	Fig. 3 (pair 1)	3
Lys116Ala (2c)	Lys116Gly (2b)	5.7 to 9.5	Fig. 3 (pair 2)	2
Pro47Ala (1e)	Pro117Ala (1d)	8.9	Fig. 13a	3
Lys116Asp (3a)	Lys116Glu (3b)	8.9	Fig. 13b	2

[a] The faster and slower migrating S. nuclease variants are protein I and protein II respectively, protein codes see Table 1.
[b] The pH of the experiment.
[c] Difference between the mobilities of the two proteins.

such similarly sized and charged proteins. For mutants of different valences, differences in the relative electrophoretic mobilities increase with the pH as seen from the data in Figs. 3 and 4. In contrast, for Asn118Gly and the wild type the relative electrophoretic mobility is invariant in the pH range from 4.1 to 8.9 as shown in Fig. 4. Similarly, the relative electrophoretic mobility of Lys116Gly and Lys116Ala is invariant in the pH range from 5.6 to 8.9. The relative electrophoretic mobilities of these mutants having the same valence are close to unity in the pH range of the experiment and this suggests that their relevant molecular dimensions are nearly the same. Indeed, X-ray and NMR measurements [22,24,26,27] indicate only subtle differences in the tertiary folds of the S. nuclease variants. Nonetheless these differences are sufficient to affect their electrophoretic mobilities so that they can be separated. This subject has received detailed treatment in Ref. [61].

The case of mutations at Pro117 or Asn118

At residues No. 117 and 118 located in the loop between helices H2 and H3, see Fig. 2, mutation changes the loop conformation. Upon replacing Asn118 by Gly, the loss of hydrogen bond destabilizes the linkage between the two loops forming the nucleotide binding pocket in the wild type and the protein structure may become less compact. This may result in a slight increase in the molecular size of Asn118Gly with respect to that of the wild type [24,26]. According to the expectation, the mutant Asn118Gly has a smaller electrophoretic mobility than the wild type as seen in Fig. 3 and Table 5. The difference in electrophoretic mobility is 3×10^{-10} V/m^2s at pH 8.9. Similar increase in size is expected for the Pro117 mutants due to conformation changes brought about by the replacement of the proline residue at the pertinent β-turn present in the wild type [22,27,39,40]. The Pro117 mutants can be separated from the wild type at pH 8.9 as seen from the difference in electrophoretic mobility of 3×10^{-10} V/m^2s in Table 4. For Pro47Ala, the same three-dimensional fold as for the wild type is assumed [27,39,40]. Therefore, the separation between

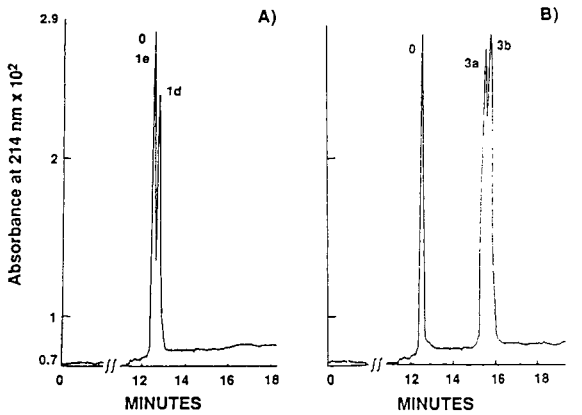


Fig. 13. Separation of the mutant pairs (A) Pro117Ala (1d) and Pro47Ala (1e) and (B) Lys116Asp (3a) and Lys116Glu (3b) at pH 8.9. The wild type (0) is also indicated on the electropherograms. The experimental conditions are in Table 3.

Pro47Ala and Pro117Ala, which is shown in Fig. 13A, stems from the slightly different sizes of those two proteins; the mutant with the slightly larger size (Pro117Ala) has the smaller electrophoretic mobility.

The case of mutation at Lys116

The conformation of the peptide bond between residues No. 116 and 117 in the loop changes from *cis* in the wild type to *trans* in the mutants Lys116Gly [26] and Lys116Asp [27] according to X-ray studies. As a result the molecular size of these mutants is greater than that of the wild type, Lys116Ala or Lys116Glu, which have the same loop conformation. In case of the separation between Lys116Gly and Lys116Ala, shown in Fig. 3 (pair No. 2) the Lys116Gly mutant with the smaller molecular weight has a smaller electrophoretic mobility than Lys116Ala. This unexpected migration order, however, is in agreement with the observation from X-ray crystallographic data that due to conformational changes in the mutation loop, Lys116Gly has a slightly larger size than Lys116Ala [26]. Nonetheless, the difference in the electrophoretic mobility between the Lys116Ala and Lys116Gly mutants is rather small, 2×10^{-10} V/m^2s, at pH 8.9 as seen in Table 5.

Analysis of mobility data

As mentioned above, the observed slight mobility differences for mutants having the same valence are attributed to minute differences in their electrophoretically relevant molecular dimensions. In the following we wish to estimate what differences in the hydrodynamic radii would bring about the observed mobility differences.

For a mutant and the wild type having the same valence, Eq. 1 takes the following form

$$\frac{R_m(1+\kappa R_m)}{\phi(\kappa R_m)} = \frac{\mu_{wt}}{\mu_m}\frac{R_{wt}(1+\kappa R_{wt})}{\phi(\kappa R_{wt})} \quad (11)$$

Eq. 11 allows us to evaluate R of a mutant, μ_m and μ_{wt}, as well as the hydrodynamic radius of the wild type, R_{wt}, are known. At pH 8.9, the reference pH where most of the mutants are best separated, the hydrodynamic radius of the equivalent mutant was evaluated by trial and error, since $\phi(\kappa R_m)$ is a nonlinear function of the radius.

Based on the hydrodynamic radii thus obtained, the experimentally observed electrophoretic mobility differences of 3×10^{-10} and 2×10^{-10} m^2/Vs at pH 8.9 correspond to the respective radius differences of 0.39 and 0.26 Å. Since the differences in the radius are very small in comparison to the hydrodynamic radius of the wild type, it is assumed that the electrophoretic mobility difference is directly proportional to the difference in the radii. Based on the data shown above for the differences in mobilities and hydrodynamic radii, the relationship between the mobility difference and the radius difference for an equivalent mutant and S. nuclease at pH 8.9 is expressed as

$$R_{wt} - R_m = 0.13\times(\mu_m - \mu_{wt}) \quad (12)$$

where the unit of the proportionality factor is Vs/m.

"Virtual" mobility of the first kind

In order to gain further support for the notion that minute differences in the tight tertiary fold of certain equivalent mutants are responsible for their separation, we wish to estimate the effect of small changes in the hydrodynamic radius alone on the electrophoretic mobility. For this reason, we define virtual mobility of the first kind, μ_ν^R, of a protein as the hypothetical mobility of an otherwise identical protein which has a 1 Å greater hydrodynamic radius. It is expressed as

$$\mu_\nu^R = \frac{ze}{6\pi\eta(R+1\text{ Å})}\frac{\phi[\kappa(R+1\text{ Å})]}{[1+\kappa(R+1\text{ Å})]} \quad (13)$$

The mobility, μ_ν^R, calculated for S. nuclease according to Eq. 13 at pH 8.9, the reference pH where most of the mutants are best separated, is given in Table 6. By comparing this mobility to the experimentally obtained electrophoretic mobility of the wild type, μ_{wt}, at pH 8.9 the hypothetical value of 7.7×10^{-10} m^2/Vs was obtained for the change in mobility due to a 1 Å change in its radius. For such a small change, we may again assume that the electrophoretic mobility difference is directly proportional to the radius difference. Then, in a way similar to Eq. 12, we calculate the proportionality factor be-

Table 6
Experimental mobility, μ_{wt}, of the wild type and the corresponding two virtual mobilities, μ_ν^R, and, μ_ν^z, at pH 8.9.

Electrophoretic mobility, $(m^2V^{-1}s^{-1})\times10^8$		R (Å)	z
μ_{wt}	1.31 [a]	24.6 [b]	7.49 [c]
μ_ν^R	1.21 [d]	25.6 [e]	7.49
μ_ν^z	1.12 [f]	24.6	6.49 [g]

The actual and virtual hydrodynamic radii, R, and valences, z are also listed.

[a] Experimentally obtained electrophoretic mobility, μ_{wt}, of the wild type at pH 8.9.

[b] Hydrodynamic radius, R_{wt}, of the wild type at pH 8.9 as established according to Eq. 7.

[c] Valence, z_{wt}, of the wild type at pH 8.9 as established according to Eq. 4.

[d] Virtual mobility, μ_ν^R of S. nuclease calculated with its valence and with a radius 1 Å greater than the actual hydrodynamic radius, R, at pH 8.9 according to Eq. 13.

[e] Virtual radius.

[f] Virtual mobility, μ_ν^z of the wild type calculated with its hydrodynamic radius, R, and with a valence smaller by one than the actual valence at pH 8.9 according to Eq. 14.

[g] Virtual valence.

tween the mobility change due to unit change in the radius and the unit radius difference as 0.1299 that is in excellent agreement with the corresponding value in Eq. 12. Consequently the difference in the electrophoretic mobilities of 3×10^{-10} and 2×10^{-10} m^2/Vs should corresponds to 0.38 and 0.26 Å differences in the hydrodynamic radius, respectively. These two radius differences compare favorably to those calculated by Eq. 12 from the measured electrophoretic mobilities. Thus, subtle differences in the tertiary structure of the S. nucleases that are associated with changes in the hydrodynamic radius less than 0.5 Å can already account for the observed separation of certain mutants despite their equal valences. In any case these size differences are very small, for instance, a 0.39 Å increase accounts only for 1.6% of the protein radius. Therefore, our assumption that the hydrodynamic radii of all S. nucleases are essentially the same is justifiable on this count also.

"Virtual mobility" of the second kind

After examining the changing selectivity of the electrophoretic system by unit change in the hydrodynamic radius of S. nuclease at pH 8.9, it is of interest to compare it also to that occurring by unit change in the valence. In order to facilitate the estimation of the mobility difference between two proteins, which differ only in their valences, we introduce the virtual mobility of the second kind, μ_ν^z. It is defined for a given protein as the mobility of a hypothetical protein that is identical except its valence is less by one so that

$$\mu_\nu^z = \frac{(z-1)\, e\, \phi(\kappa R)}{6\pi\eta R\,(1+\kappa R)} \qquad (14)$$

The mobility, μ_ν^z, calculated for S. nuclease at pH 8.9 according to Eq. 14 is also given in Table 6. By comparing μ_ν^z and μ_{wt}, at pH 8.9, we find a mobility difference of 17×10^{-10} m^2/Vs due to unit valence difference. This value has strong experimental support, since protein variants having a unit valence difference exhibit a mobility difference of 16.5×10^{-10} m^2/Vs on the average, as shown in Table 4 and Fig. 3.

It was shown before that the mobility difference upon unit change in the hydrodynamic radius of S. nuclease at pH 8.9 was 7.7×10^{-10} m^2/Vs. On the other hand the corresponding mobility change due to unit change in valence under the same conditions was 17×10^{-10} m^2/Vs, i.e., about twice as large. This is in agreement with the observation that for S. nucleases a unit valence difference brings about a six times greater increase in the mobility than 0.3 Å radius difference as shown in Figs. 3 and 4.

Acknowledgements

We thank Alec Hodel for protein samples and helpful discussions and the Alexander von Humboldt-Foundation for financial support in form of a Feodor-Lynen Fellowship for F.K. This work was supported by grants GM No. 20993 (Cs. H.) and AI No. 23923 (ROF) from National Institute of Health, US Public Health Service and No. BCS-9014119 from National Science Foundation and by the Howard Hughes Medical Institute (ROF).

References

[1] K.J. Lee and G.S. Heo, *J. Chromatogr.*, 559 (1991) 317.
[2] M.J. Gordon, K.-J. Lee and A.A. Arias and R.N. Zare, *Anal. Chem.*, 63 (1991) 69.
[3] F.T.A. Chen, *J. Chromatogr.*, 559 (1991) 445.
[4] T.T. Lee and E. Yeung, *Anal. Chem.*, 64 (1992) 3045.
[5] H. Yamamoto and T. Manabe and T. Okuyama, *J. Chromatogr.*, 515 (1990) 659.
[6] P.D. Grossman, J.C. Colburn, H.H. Lauer, R.G. Nielsen, R.M. Riggin, G.S. Sittampalam and E.C. Rickard, *Anal. Chem.*, 61 (1989) 1186.
[7] G. Mandrup, *J. Chromatogr.*, 604 (1992) 267.
[8] J. Frenz and S. Wu and W.S. Hancock, *J. Chromatogr.*, 480 (1989) 379.
[9] F. Kilár and S. Hjertén, *J. Chromatogr.*, 480 (1989) 351.
[10] B.J. Compton, *J. Chromatogr.*, 559 (1991) 357.
[11] R.E. Offord, *Nature*, 211 (1966) 591.
[12] R.V. Wenn, *J. Biochem.*, 145 (1975) 281.
[13] P.D. Grossman, J.C. Colburn and H.H. Lauer, *Anal. Biochem.*, 179 (1989) 28.
[14] E.C. Rickard, M.M. Strohl and R.G. Nielsen, *Anal. Biochem.*, 197 (1991) 197.

[15] H.J. Gaus, A.G. Sicker and E. Bayer, *Anal. Chem.*, 65 (1993) 1399.
[16] R.S. Rush, A.S. Cohen and B.L. Karger, *Anal. Chem.*, 63 (1991) 1346.
[17] F.A. Cotton, E.E. Hazen Jr. and M.J. Legg, *Proc. Natl. Acad. Sci.*, 76 (1979) 2551.
[18] P.W. Tucker, E.E. Hazen Jr. and F.A. Cotton, *Mol. Cell. Biochem.*, 22 (1978) 67.
[19] T.R. Hynes and O.F. Fox, *Proteins: Structure, Function and Genetics*, 10 (1991) 92.
[20] C.B. Anfinsen, *Science*, 181 (1973) 223.
[21] D. Shortle and A.K. Meeker, *Proteins: Structure, Function and Genetics*, 1 (1986) 81.
[22] T.R. Hynes, A. Hodel and R.O. Fox, *Biochemistry*, 33 (1994) 5021.
[23] *J.M. Flanagan, M. Kataoka, D. Shortle and D.M. Engelmann, Biochemistry*, 89 (1992) 748.
[24] A. Hodel, R.A. Kautz, D.M. Edelman and R.O. Fox, *Protein Science*, 3 (1994) 549.
[25] E.H. Serpersu, D. Shortle and A.S. Mildvan, *Biochemistry*, 26 (1987) 1289.
[26] A. Hodel, R.A. Kautz, M.D. Jacobs and R.O. Fox, *Protein Science*, 2 (1993).
[27] A. Hodel and R.O. Fox, unpublished data.
[28] T.R. Hynes, R.A. Kautz, M.A. Goodman, J.F. Gill and R.O. Fox, *Nature*, 339 (1989) 73.
[29] P.A. Evans, R.A. Kautz, R.O. Fox and C.M. Dobson, *Biochemistry*, 28 (1989) 362.
[30] J. Pedersen, M. Pedersen, H. Soeberg and K. Biedermann, *J. Chromatogr.*, 645 (1993) 353.
[31] J.C. Ford and J.A. Smith, *J. Chromatogr.*, 390 (1987) 307.
[32] S. Hjertén, *J. Chromatogr.*, 347 (1985) 191.
[33] K.A. Cobb, V. Dolnik and M. Novotny, *Anal. Chem.*, 62 (1990) 2478.
[34] M.L. Zoller and M. Smith, *Methods Enzymol.*, 100 (1983) 468.
[35] R.O. Fox, P.A. Evans and C.M. Dobson, *Nature*, 320 (1986) 192.
[36] W.R. Melander, J. Stoveken and Cs. Horváth, *J. Chromatogr.*, 185 (1979) 111.
[37] A.W. Kenschington, in P. Alexander and R.J. Block (Editors), *Laboratory Manual of Analytical Methods of Protein Chemistry*, Pergamon, Oxford, 1960, p. 353.
[38] C. Tanford and J.D. Hauenstein, *J. Am. Chem. Soc.*, 78 (1956) 5287.
[39] P.A. Evans, C.M. Dobson, R.A. Kautz, G. Hatfull and R.O. Fox, *Nature*, 329 (1987) 266.
[40] R.A. Kautz and R.O. Fox, in J.J. Villafranca (Editor), *Techniques in Protein Chemistry II*, Academic Press, New York, 1991, p. 263.
[41] G.J.M. Bruin, J.P. Chang, R.H. Kuhlman, K. Zegers, J.C. Kraak and H. Poppe, J. *Chromatogr.*, 471 (1989) 429.
[42] J.K. Towns and F.E. Regnier, *J. Chromatogr.*, 516 (1990) 69.
[43] W. Nashabeh and Z. El Rassi, *J. Chromatogr.*, 559 (1991) 367.
[44] M. Gilges, H. Husmann, M. Kleemib, S.R. Motsch and G. Schomburg, *J.High Resolut. Chromatogr.*, 15 (1992) 452.
[45] M. Huang and M.L. Lee, *J. Microcol. Sep.*, 4 (1992) 491.
[46] J.T. Smith and Z. El Rassi, *Electrophoresis*, 14 (1993) 396.
[47] T. Hara, S. Okamura, S. Kato, J. Yokogi and R. Nakajima, *Anal. Sci.*, 7 (1991) 261.
[48] N.E. Good, G.D. Winget, W. Winter, T.N. Connolly, S. Izawa and R.M.M. Singh, *Biochemistry*, 5 (1966) 467.
[49] H.A. Abramson, L.S. Moyer and M.H. Gorin, *Electrophoresis of Proteins*, Reinhold, New York, 1942.
[50] D.C. Henry, *Proc. Roy. Soc. (London) Ser. A.*, 133 (1931) 106.
[51] J.L. Oncley, *Ann. N.Y. Acad. Sci.*, 41 (1941) 121.
[52] H. Jakubke and H. Jeschkeit, *Amino acids, Peptides and Proteins*, Wiley, New York, 1977.
[53] A.S. Yang, M.R. Gunner, R. Sampogna, K. Sharp and B. Honig, *Proteins: Structure, Function and Genetics*, 15 (1993) 252.
[54] S. Beychok and R.C. Warner, *J. Am. Chem. Soc.*, 81 (1959) 1892.
[55] C. Tanford and T. Shedlovsky, *Electrochemistry in Biology and Medicine*, Wiley, New York, 1st ed., 1955.
[56] J. Steinhardt and J.A. Reynolds, *Multiple Equilibria in Proteins*, Academic, New York, 1969.
[57] J.A. Dean, *Lange's Handbook of Chemistry*, Marcel Dekker, New York, 13th ed., 1986.
[58] D. Shortle and A.K. Meeker, *Biochemistry*, 28 (1989) 936.
[59] P.W. Atkins, *Physical Chemistry*, W.H. Freeman, San Francisco, 2nd ed., 1982.
[60] C. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, 1st ed., 1961.
[61] F. Kálmán, S. Ma, A. Hodel, R.O. Fox and Cs. Horváth, *Electrophoresis*, in press.

ELSEVIER

Journal of Chromatography A, 705 (1995) 155–161

JOURNAL OF CHROMATOGRAPHY A

Analysis of recombinant human growth hormone in *Escherichia coli* fermentation broth by micellar high-performance liquid chromatography

Mark A. Strege, Avinash L. Lagu*

Lilly Research Laboratories, a Division of Eli Lilly and Co., Indianapolis, IN 46285, USA

Abstract

A method for the reversed-phase high-performance liquid chromatographic (HPLC) determination of recombinant methionylaspartyl-human growth hormone (MD-HGH) in *Escherichia coli* fermentation broth is described. The technique utilizes mobile phases containing *n*-propanol and the anionic surfactant sodium dodecyl sulfate (SDS) under micellar conditions at pH 6.4. The methodology is directly applicable to the analysis of samples solubilized via sulfitolysis in the presence of SDS, and offers superior resolution in comparison with chromatography in the absence of the surfactant. Using this method, acceptable precision (day-to-day R.S.D. = 4.9%), accuracy, selectivity, range, linearity and ruggedness were achieved.

1. Introduction

Advances in recombinant DNA technology have provided a variety of host–vector systems for the cloning and expression of heterologous proteins in microorganisms. The production and isolation of these recombinant proteins has typically been monitored through the use of analytical techniques such as sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) [1,2], immunoassays [3,4], biological activity assays [4–6] and high-performance liquid chromatography (HPLC) [3,6–9]. However, these methods have been limited by non-specificity, labor-intensive sample manipulation or the use of harsh conditions such as high concentrations of formic acid [8], or are useful for only properly folded, soluble target protein. There have been few reports of the determination of recombinant proteins in fermentation broth matrices, where the target molecules frequently exist inside the microorganism as inclusion bodies in an insoluble denatured state.

The analysis of heterologous proteins in recombinant hosts such as *E. coli* presents many challenges to the analytical biochemist. The cells must be lysed and the inclusion bodies solubilized prior to quantification. Cell lysis and protein solubilization can be accomplished chemically through the use of SDS, an anionic surfactant [9]. As the proteins in inclusion bodies can exist as a distribution of forms, such as covalent and non-covalent polymers, it is crucial to convert the target protein into a single molecular entity prior to analysis. This can be achieved by unfold-

* Corresponding author.

0021-9673/95/$09.50
SSDI 0021-9673(94)01253-9

ing the proteins and disrupting the inter- and intramolecular disulfide bonds via reduction [10] or sulfitolysis [11]. As both the whole cell and the inclusion bodies can also contain nucleic acids, salts, lipids and other host molecules in addition to proteinaceous material, the complexity of the matrix adds to the difficulty of the determination of the recombinant protein.

The employment of ionic surfactants such as SDS has traditionally been avoided in the HPLC of proteins owing to the generation of a strongly denaturing environment. However, for the determination of recombinant proteins in fermentation broth, the characteristics of ionic surfactants are in fact useful for the enhancement of the solubility of the unfolded denatured proteins, the elimination of irreversible adsorption on the stationary phase and the facilitation of unique separation selectivity.

In this paper, a method for the determination of a protein in *Escherichia coli* fermentation broth is described. Recombinant human growth hormone (HGH) is synthesized in *E. coli* as methionylaspartyl-HGH (MD-HGH), a hydrophobic protein which readily polymerizes and is not easily solubilized in its sulfitolyzed form. The method described utilizes isocratic reversed-phase micellar HPLC at near neutral pH to determine sulfitolyzed MD-HGH.

2. Experimental

MD-HGH reference material and fermentation broth were provided by Eli Lilly (Indianapolis, IN, USA). SDS and potassium tetrathionate were purchased from Fluka (Ronkonoma, NY, USA), and *n*-propanol (HPLC grade) from Baxter Scientific Products (McGaw Park, IL, USA) and tris(hydroxymethyl)-aminomethane (Tris) from Fisher Scientific (Springfield, NJ, USA). Human carbonic anhydrase I and bovine carbonic anhydrase were purchased from Sigma (St. Louis, MO, USA). All other chemicals were of at least analytical-reagent grade and were supplied by EM Science (Gibbstown, NJ, USA). Purified water was obtained via a Milli-Q reagent water system (Millipore, Millford, MA).

Chromatography was achieved through the use of a Gilson Medical Electronics (Middleton, WI, USA) two-pump system composed of a Model 305 and a Model 303 pump, a Model 805 manometric module, a Model 811 dynamic mixer, a Model 231 autosampler equipped with a Model 401 diluter and a 20-μl injection loop and a Model 116 variable-wavelength UV detector. The column temperature was maintained at 60°C with a Model CHM column oven equipped with a Model TCM control module (Waters, Millford, MA, USA). A 25 cm × 0.46 cm I.D. Nucelosil C_4 packing (particle size 5 μm, pore size 300 Å) (Phenomenex, Torrance, CA, USA) was employed as an analytical column following a 1.5 cm × 0.32 cm I.D. Brownlee RP-4 guard cartridge (Munhall, Worthington, OH, USA) for all the experiments unless noted otherwise. A saturation precolumn was prepared by packing a 7.5 × 0.75 cm I.D. stainless-steel column with 100–200-μm silica (Universal Scientific, Atlanta, GA, USA), and was incubated inside the column oven in-line between the dynamic mixer and the injection valve to presaturate the mobile phase with dissolved silica to preserve the analytical resin. Mobile phase A [50 m*M* ammonium phosphate (pH 6.4)–20% *n*-propanol–1.0% SDS] was prepared by dissolving 3.2 g of monobasic ammonium phosphate, 1.6 g of dibasic ammonium phosphate and 10.0 g of SDS in 800 ml of water and adding 200 ml of *n*-propanol. The preparation of mobile phase B [50 m*M* ammonium phosphate (pH 6.4)–30% *n*-propanol–1% SDS] was identical with that of A with the exception that the salts were dissolved in 700 ml of water and 300 ml of *n*-propanol. The pH of the mobile phases was 6.3–6.5 (if necessary, adjustments were made using either concentrated phosphoric acid or 6 *M* sodium hydroxide solution). Other parameters included a flow-rate of 0.5 ml/min and UV detection at 214 nm. Analog data (1.0 Hz sampling rate) were collected directly from the detector on an in-house centralized chromatography computer system based on the Hewlett-Packard Model 1000

minicomputer. Retention time, efficiency, normalized peak width, tailing and resolution were calculated by the minicomputer.

The sulfitolysis solubilization reagent was prepared in the following manner. SDS (20.0 g) was dissolved in 700–800 ml of water. To this solution was added 6.1 g of Tris, 12.0 g of anhydrous sodium sulfite, 3.7 g of anhydrous potassium tetrathionate and 0.37 g of disodium ethylenediaminetetraacetate dihydrate. The pH of the solution was adjusted to 8.5–8.7 using 6 *M* hydrochloric acid and the solution was diluted to 1 liter.

Fermentation broth samples were diluted with an equal volume of methanol prior to preparation for analysis. An appropriate aliquot of the broth–methanol suspension was transferred into a 125 mm × 16 mm I.D. centrifuge tube. Following centrifugation at 3000 *g* for 15 min, the supernatant was discarded and the pellet was resuspended in 10.0 ml of solubilization reagent. The sample was cooled on ice while sonicated for 30 s using a Vibra Cell sonicator (Sonics and Materials, Danbury, CT, USA) equipped with a microtip and set at a power setting of 5 in the continuous mode. The solution was then stirred for 10–16 h at room temperature (an investigation of the kinetics of protein sulfitolysis in the presence of SDS at ambient temperature revealed that sulfitolysis under these conditions is complete within 6–8 hours [12]). On completion of stirring, the sulfitolysis reaction was quenched and the solubilized protein stabilized by adjusting the pH of the solution to 5.8–6.2 using 1 *M* maleic acid. The samples were then filtered (0.45-μm nylon membranes; Alltech, Deerfield, IL, USA) prior to analysis.

3. Results and discussion

Micellar mobile phases have been employed successfully for the chromatography of small biomolecules, and these studies have demonstrated the potential advantages offered by systems of this nature, such as enhanced fluorescence [13] and improved selectivity [14]. These advantages may also be characteristic of the micellar chromatography of macromolecules such as proteins. The unique selectivity of the micellar HPLC of MD-HGH is demonstrated in Fig. 1, where the elution profiles of a 467 μg/ml MD-HGH standard are displayed. Fig. 1c, obtained through the use of isocratic elution with *n*-propanol in the absence of SDS (SDS was omitted from the mobile phases, which were otherwise identical with those described under Experimental), revealed the presence of a contaminant in the purified reference material which eluted within the front shoulder of the MD-HGH peak. The contaminant was also visible in the profiles of solubilized fermentation broth samples analyzed under the same conditions (data not shown). This peak did not appear on the shoulder of the main peaks generated by the same samples when they were analyzed using mobile phases containing SDS (see the chromatogram of the standard in Fig. 1b). Approximately 30 μg of the contaminant (determined by dry mass relative to that of an equivalent volume of mobile phase following freeze-drying) was isolated via fraction collection in the absence of SDS and then dried and reconstituted in 1.0 ml of solubilization reagent. Analysis of the resolubilized isolated contaminant by micellar HPLC revealed that the material generated a peak at ca. 540 s, an elution time much earlier than that of MD-HGH (see Fig. 1a). Also, the analysis of a solubilization reagent blank (Fig. 1d) demonstrated that the peak of interest had not been generated by a component present in the reagent. Because it offered superior resolution between MD-HGH and this contaminant, micellar HPLC in SDS was the method of choice for the analysis of the protein in *E. coli* fermentation broth.

Fig. 2b shows a typical chromatogram of MD-HGH in solubilized fermentation broth obtained via micellar HPLC. Under the chromatographic conditions described, the values of the asymmetry and theoretical plate count were 1.2 and 1569 theoretical plates/m, respectively [15]. A test of precision with a fermentation broth sample containing ca. 740 μg/ml of MD-HGH (five injections analyzed on five different days) generated a relative standard deviation (R.S.D.)

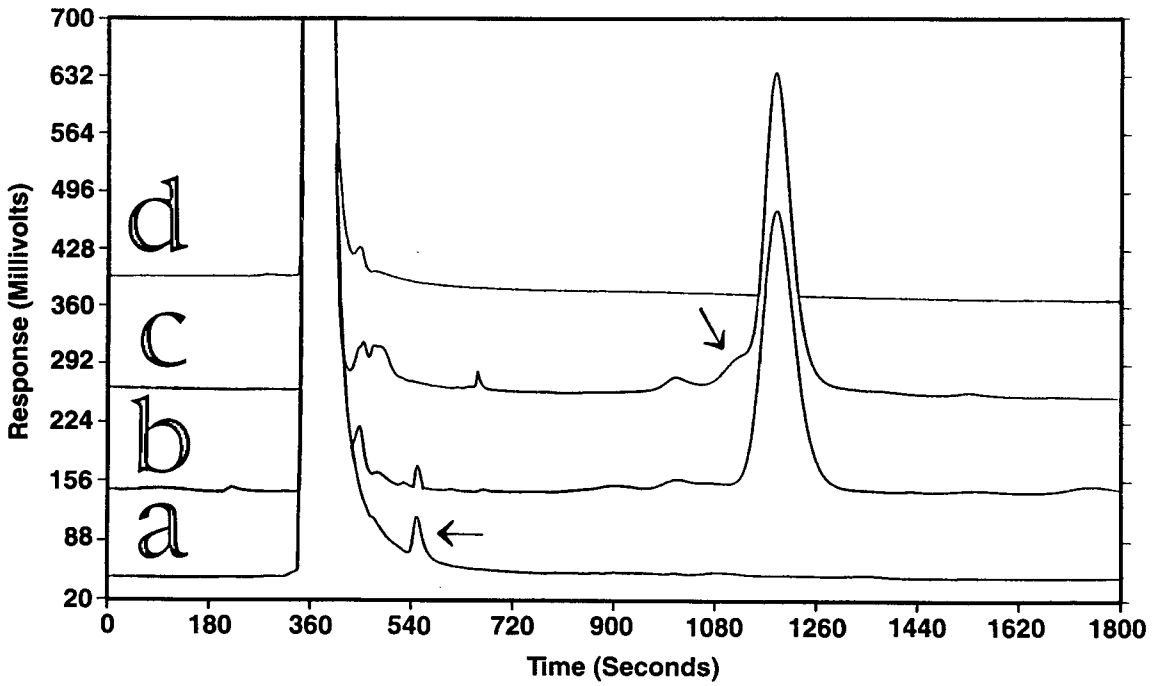


Fig. 1. HPLC profiles of 20-μl injections of (a) a 30.0 μg/ml solution of contaminant peak (isolated using chromatography in the absence of SDS) obtained via isocratic elution at 47.5% B in 50 m*M* ammonium phosphate (pH 6.4)–1.0% SDS, (b) a 467 μg/ml MD-HGH standard obtained using isocratic elution at 47.5% B in 50 m*M* ammonium phosphate (pH 6.4)–1.0% SDS, (c) a 467 μg/ml MD-HGH standard obtained using isocratic elution at 45.0% B in 50 m*M* ammonium phosphate (pH 6.4) in the absence of SDS and (d) a reagent blank obtained using isocratic elution at 47.5% B in 50 m*M* ammonium phosphate (pH 6.4)–1.0% SDS, using a 250 cm × 0.46 cm I.D. Nucleosil C_8 packing, a flow-rate of 1.0 ml/min and UV detection at 214 nm. The contaminant in the elution profiles of the solubilized standard material and the re-chromatographed fraction of the isolated contaminant are indicated by arrows.

of 4.46% (the within-day RSD ranged from 0.50 to 1.88%). The method was determined to be linear over the range 50–800 μg/ml (correlation coefficient = 0.999) using peak area for quantification.

To evaluate the possibility of matrix effects on the assay, a series of standards ranging from 0 to 800 μg/ml MD-HGH were prepared in a control fermentation broth containing cells in which MD-HGH had not been expressed. The slope of a plot of peak area versus concentration for this standard series was 45.06 arbitrary units. The slope of a calibration plot for standards prepared in solubilization reagent alone was 45.20 units, i.e., only a 0.3% difference. In addition, the former plot passed through the origin. Further evidence for the absence of matrix interference was provided by an overlay of the chromatographic profiles of the control broth and a 467 μg/ml standard (see Fig. 2a and c).

The fermentation broth sample recovery was evaluated by spiking a known volume of fermentation broth with a non-sulfitolyzed MD-HGH standard solution at four levels ranging from 100 to 800 μg/ml. The recovery of MD-HGH at these levels averaged 103.5%, indicating complete extraction of the sulfitolyzed protein from the broth.

To determine the magnitude of the sensitivity of the assay to changes in environmental and operational parameters, analyses of an MD-HGH standard, fermentation broth sample and a resolution mixture containing human carbonic anhydrases II and bovine carbonic anhydrase

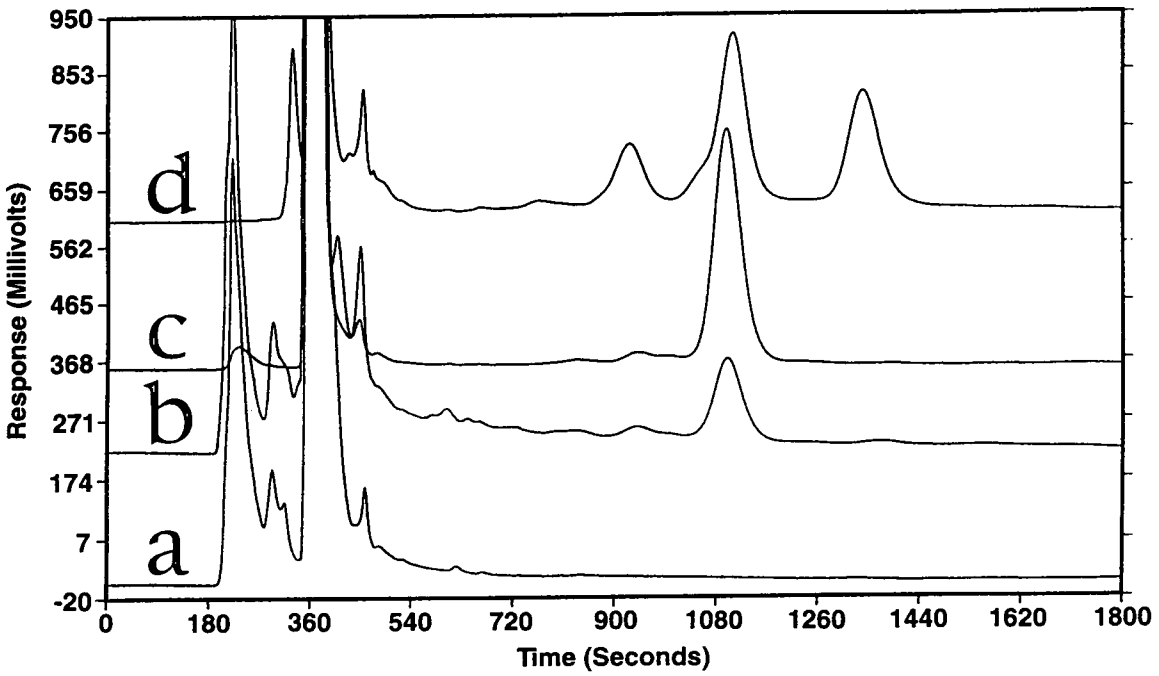


Fig. 2. Micellar HPLC profiles of 20-μl injections of (a) solubilized control fermentation broth, (b) solubilized fermentation broth, (c) 467 μg/ml MD-HGH standard and (d) resolution control containing 500 μg/ml each of bovine carbonic anhydrase and human carbonic anhydrase IIa obtained using isocratic elution at 47.5% B in 50 m*M* ammonium phosphate (pH 6.4)–1.0% SDS using a 250 cm × 0.46 cm I.D. Nucleosil C_8 packing, a flow-rate of 1.0 ml/min and UV detection at 214 nm.

[chosen because of their elution time proximity to MD-HGH in this assay (see Fig. 2d)] were obtained in systems where the organic modifier concentration, column temperature, mobile phase pH and column packing were varied. The two latest eluting peaks within the elution profile of the resolution control were utilized for the calculation of resolution values.

The effects of *n*-propanol concentration in the mobile phase on the isocratic elution of MD-HGH and the resolution control are displayed in Table 1, where the retention time, plate count, normalized peak width, tailing and resolution are listed for analyses obtained at 45.0, 47.5, 50.0, 52.5 and 55.0% B (a 1% change in B corresponded to a 0.1% change in the *n*-propanol

Table 1
Effects of mobile phase organic composition on the retention time, efficiency, normalized peak width, peak tailing and resolution of MD-HGH

B (%)	Retention time (s)	Efficiency (plates)	Peak width (s)	Asymmetry	Resolution
45.0	1319	1170	111.7	1.30	–[a]
47.5	1190	1465	103.9	1.19	1.73
50.0	1085	1569	100.0	1.20	2.64
52.5	1000	1636	98.4	1.19	1.74
55.0	927	1670	98.4	1.16	2.33

All data are mean values from three replicate injections.
[a] Peaks did not elute within a 30-min run time.

Table 2
Effects of column temperature on the retention time, efficiency, normalized peak width, peak tailing and resolution of MD-HGH

Temperature (°C) retention	Time (s)	Efficiency (plates)	Peak width (s)	Asymmetry	Resolution
54.0	1262	1679	99.8	1.12	2.60
57.0	1174	1710	97.8	1.15	2.61
60.0	1099	1934	92.3	1.14	2.75
63.0	1043	2114	88.2	1.14	2.75
66.0	996	2111	88.3	1.14	2.99

All data are mean values from three replicate injections.

concentration in the mobile phase). The retention time, peak width and asymmetry all appeared to decrease in response to an increase in mobile phase organic content. There was no apparent decrease in resolution with these changes in B content.

The effect of temperature on the chromatography of MD-HGH was investigated by performing analyses at five different column oven settings ranging from 54 to 66°C. The chromatographic parameters determined in this set of experiments are listed in Table 2. The effects of temperature on the mass transfer kinetics appear to be evidenced by the decrease in retention time and peak width and the enhanced efficiency as the column over temperature was adjusted to higher levels. Peak asymmetry and resolution were not influenced by temperature changes within the range studied.

Mobile phase pH can have strong effects on the elution characteristics of proteins. In general, acidic proteins such as MD-HGH ($pI = 5.9$) will display a tendency toward greater hydrophobicity as the pH is lowered, since acidic residues become protonated under these conditions. Evidence for this occurrence is provided by the data in Table 3. The optimum mobile phase pH for the determination of MD-HGH appears to be near 6.4. At pH 5.76 and 6.08, MD-HGH does not elute within 1800 s. As the mobile phase pH was increased from 6.40 to 7.04, the retention time and peak width increased by ca. 10%, while the efficiency decreased by 20%. Also, at pH > 6.40, the carbonic anhydrase resolution control generated a profile of multiple partially resolved peaks, which prevented the calculation of resolution for this sample under these conditions. The extra peaks which appear at the higher pH may represent contaminants present in the commercially available carbonic anhydrase material which are not resolved at pH 6.40. The effect of temperature on mobile phase pH was determined to be minimal, as the pH of 50 m*M* ammonium phosphate, which measured pH 6.4

Table 3
Effects of mobile phase pH on the retention time, efficiency, normalized peak width, peak tailing and resolution of MD-HGH

pH	Retention time (s)	Efficiency (plates)	Peak width (s)	Asymmetry	Resolution
5.76	–[a]	–[a]	–[a]	–[a]	–[a]
6.08	–[a]	–[a]	–[a]	–[a]	–[a]
6.40	990	1563	101.0	1.19	2.18
6.72	1020	1298	108.0	1.25	–[b]
7.04	1058	1223	111.0	1.25	–[b]

All data are mean values from three replicate injections.
[a] Peaks did not elute within a 30-min run time.
[b] Multiple (>4) peaks were generated, which prevented determination of resolution.

Table 4
Effects of column packing on the retention time, efficiency, normalized peak width, peak tailing and resolution of MD-HGH chromatography

Packing	Retention time (s)	Efficiency (plates)	Peak width (s)	Asymmetry	Resolution
Nucleosil	1099	1934	92.3	1.14	2.75
Vydac	1108	1338	111	1.14	2.35

All data are mean values from three replicate injections.

at room temperature (26°C), rose to only 6.5 on being heated in a water bath to 60°C to emulate the environment present inside the column oven.

To identify an alternative column for use in the micellar HPLC determination of MD-HGH, a Vydac C_4 packing [15 cm × 0.46 cm I.D. Vydac C_4 packing, particle size 5 μm, pore size 300 Å (Munhall)] was evaluated. Although the experimental data suggested that the Vydac packing was slightly inferior to the Nucleosil packing for use in this application (see Table 4), the performance of both columns appeared to be acceptable for the determination of the recombinant protein in fermentation broth.

4. Conclusions

A micellar HPLC assay for the determination of MD-HGH in *E. coli* fermentation broth has been described. For this application, micellar chromatography was directly compatible with samples solubilized in SDS sulfitolysis reagent and offered a resolution superior to that achievable via reversed-phase chromatography in the absence of surfactant. The method is rugged in response to variations in mobile phase elution strength, temperature and column packing, but not to mobile phase pH changes. Analytical column lifetimes have been found routinely to correspond to at least 400 injections under the conditions specified in the method. The linear dynamic range of the method is particularly useful for monitoring fermentation processes.

Acknowledgement

The authors thank Mr. Wayne Mascher for technical contributions.

References

[1] R.G. Schoner, L.F. Ellis and B.E. Schoner, *BioTechnology*, 3 (1985) 151.
[2] D.C. Williams, R.M. Van Frank, W.L. Muth and J.P. Burnett, *Science*, 215 (1982) 687.
[3] B.K. Chow, G.W. Morrow, M. Ho, R.A. Pederson, C.H.S. McIntosh, J.C. Brown and R.T.A. MacGillivray, *Peptides*, 11 (1990) 1069.
[4] E.A. Auerswald, G. Genenger, I. Assfalg-Machleidt and J. Kos, *FEBS Lett.*, 243 (1989) 186.
[5] J. Hoppe, H.A. Weich and W. Eichner, *Biochemistry*, 28 (1989) 2956.
[6] K. Ashman, N. Matthews and R.W. Frank, *Protein Eng.*, 2 (1989) 387.
[7] G. Folena-Wasserman, R. Inacker and J. Rosenbloom, *J. Chromatogr.*, 411 (1987) 345.
[8] M. Hummel, H. Herst and H. Stein, *Eur. J. Biochem.*, 180 (1989) 555.
[9] F.W. Putnam, *Adv. Protein Chem.*, 4 (1948) 79.
[10] W.W. Cleland, *Biochemistry*, 3 (1964) 480.
[11] J.L. Bailey and R.D. Cole, *J. Biol. Chem.*, 234 (1959) 1733.
[12] J.S. Patrick and A.L. Lagu, unpublished results.
[13] N.N. Singh and W.L. Hinze, *Analyst*, 107 (1982) 1073.
[14] M.G. Khaledi, *Trends Anal. Chem.*, 7 (1988) 293.
[15] J.P. Foley and J.G. Dorsey, *Anal. Chem.*, 55 (1983) 730.

ELSEVIER

Journal of Chromatography A, 705 (1995) 163–169

JOURNAL OF CHROMATOGRAPHY A

Detection of neu differentiation factor with a biospecific affinity sensor during chromatography

Hsieng S. Lu*, David Chang, David Brankow, Duanzhi Wen

Amgen Inc., Amgen Center, Thousand Oaks, CA 91320, USA

Abstract

A technique using a biospecific affinity sensor, BIAcore, was applied to monitor and determine mammalian cell-derived neu differentiation factor (NDF) in column fractions during chromatography. Specific purified polyclonal antibody against *Escherichia coli*-derived NDF was chemically bound to the surface of BIAcore sensor chips and the derivatized sensor chips were used to detect the specific binding of NDF. The measurement of NDF at very low levels can be assessed by injecting small volumes of the crude media or column fractions into the BIAcore sensor containing antibody-bound sensor chips. This automated procedure performed under computer programming control allows direct measurement of multiple NDF samples in a short period of time and provides excellent quantitative data, which is not possible using other related methods such as Western blotting, sodium dodecyl sulfate–polyacrylamide gel electrophoresis and stimulatory activity assay on receptor autophosphorylation.

1. Introduction

The neu differentiation factor (NDF) or heregulin gene family encodes a number of secreted forms of soluble factors that specifically interact with an M_r 185 000 tyrosine kinase receptor, termed p185^{neu}, HER-2 or c-erbB-2 [1–4]. Upon ligand-receptor binding, NDF can stimulate receptor autophosphorylation and promote subsequent cellular functions [2,4]. Overexpression of p185^{neu} or HER-2 has been found in some neoplastic tissues and also associated with poor prognosis for several types of cancers, including breast cancer [5–8]. The structural and functional aspects of the multiplicity of the human and rat NDF have been investigated [3,4].

Our efforts, aiming at understanding the structural characteristics of soluble NDF forms and their roles in as yet unknown biological functions, led to the production and isolation of rat NDF from rat cell lines in addition to human and rat NDF recombinant forms in Chinese hamster ovary (CHO) cells. As both natural cell lines and the engineered CHO cells can only secrete a small amount of NDF, the detection and isolation of mammalian cell-derived NDF has been difficult [1,9,10]. Although Western blotting, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and tyrosine kinase phosphorylation assay [9] can be adapted to monitor the column effluents during fractionation, these methods, in general, are time consuming, less sensitive and not applicable to samples containing low expression levels of NDF. In this paper, we describe the use of an antibody-based bioaffinity technique (BIAcore) [11,12] as an alternative to measure rat NDF

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(94)00737-3

during column chromatography from media conditioned by natural cell line or by recombinant CHO cells. This automated procedure is capable of handling multiple column fractions with sufficient speed and sensitivity in detection and provides precise NDF quantification for various samples including pooled fractions and purified NDF preparation.

2. Experimental

2.1. Production and isolation of NDF

NDF prepared from rat-1 EJ cells was produced according to our previous papers [1,2]. Expression of rat proNDF-α2 and -β4 forms (encoding 639 and 661 amino acid proNDFs, respectively) in CHO cells was also described in previous papers [4,10]. In the cultured medium, the NDF proteins are secreted as M_r 44 000 soluble molecules. They were isolated by a series of purification steps including concentration of medium by diafiltration and chromatography using heparin-Sepharose, DEAE-Sepharose and phenyl-Sepharose columns [10]. Western blotting and SDS-PAGE were performed according to previously described methods [10].

2.2. Polyclonal antiserum production

Rabbit polyclonal anti-rate NDF antibody was produced against *Escherichia coli*-derived recombinant rat NDF-α2 and purified using an Affigel affinity column coupled with the same rat NDF protein. The partially purified antibody was used in assays with a BIAcore system or Western blotting.

2.3. Real-time biospecific interaction analysis (BIA)

The analysis was performed using a BIAcore instrument (Pharmacia Biosensor, Uppsala, Sweden) [11,12]. To detect the specific binding between rat NDF and the antibody, a sensor chip surface used in the BIAcore system was first immobilized with affinity-purified rabbit antibody against *E. coli*-derived rat NDF-α2. The immobilization was performed using sensor chip CM5 programs according to the procedures described in the instructions of the amine coupling kit (Pharmacia). Samples containing various amounts of NDF were diluted with BIAcore HBS buffer [10 m*M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.4)–150 m*M* NaCl–3.4 m*M* EDTA and 0.05% BIAcore surfactant P20] to a concentration ranging from 5 to 300 ng/ml before injection. Samples were automatically injected and flowed through the sensor chip surface. Injection volumes ranged from 10 to 100 μl depending on sample concentration. After sample injection, the chip surface was washed with HBS buffer to eliminate non-specific binding. Any NDF bound to the surface results in an increase in surface plasma resonance (SPR), which can be detected optically by the equipment. The intensity of this resonance response, expressed as resonance units (RU), is proportional to the amount of NDF bound to the chip surface. After calibration with a standard amount of CHO cell-derived rat NDF-α2 which was determined by amino acid analysis [10], samples containing unknown NDF concentrations can be analysed.

Regeneration of the NDF-bound sensor chips was obtained by a washing step with 10 m*M* HCl (pH 2.2) and an equilibration step with HBS buffer. The regenerated sensor chips can then be used to assay the next sample.

3. Results and discussion

Fig. 1A shows SPR raw data obtained from the analysis of an NDF standard produced in engineered *E. coli* cells using a freshly prepared antibody-bound sensor chip. In these particular analyses, the BIAcore program used a 12-min run time, including a 5-min sample analysis (as shown) and an extra 7-min HCl washing and equilibration (not shown). The SPR was obtained from the RU difference between the sample injection starting point (S) and the equilibration point after sample injection (E). Sensorgram 1 as indicated in Fig. 1A represents a blank run containing only sample buffer and

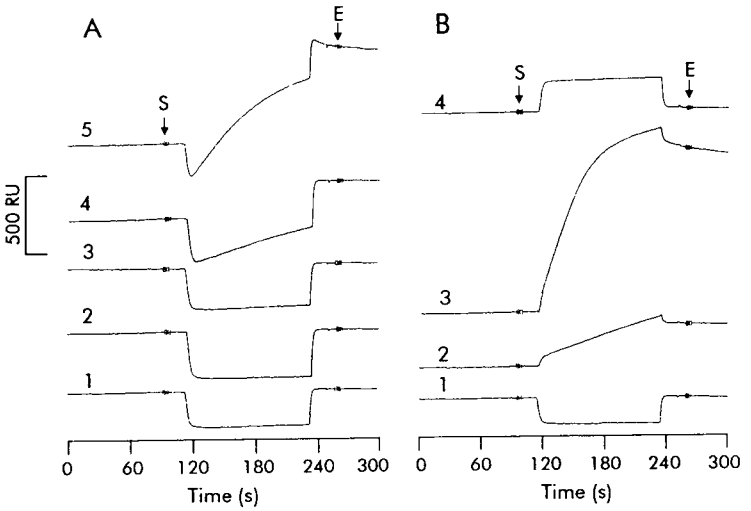


Fig. 1. Real-time BIAcore assay of rat NDF using a sensor chip coupled with anti-NDF antibody. (A) Sensorgrams from analysis of an NDF standard; (B) sensorgrams from analysis of fractions from heparose-Sepharose chromatography (see Fig. 3).

shows almost no response (RU = 5). Sensorgrams 2–5 represent analyses of NDF standard at concentrations of 5, 20, 100 and 300 ng/ml. The background RU for controls ranges from ca. −10 to +10 RU. The data clearly show that the method can detect NDF at concentrations above 20 ng/ml. Samples were always diluted with HBS buffer (minimum dilution = 1:1). Although concentrated culture media can be analysed directly they can potentially cause background interference due to light scattering.

Fig. 2 illustrates typical titration curves for standard rat NDF-α-2 and -β4 recombinantly produced in CHO cells as well as rat NDF-α2 produced in *E. coli*. These data indicate that antibody produced against bacterially derived rat NDF-α2 is able to recognize glycosylated forms of rat NDF-α2 and -β4 derived from CHO cells [10], but with a lesser extent of binding affinity. This lower binding for CHO-derived NDFs may be contributed to by carbohydrates present in the molecule. The difference in binding between CHO-derived NDF-α2 and -β4 may be due to variations in the primary structure at the EGF domains and their C-termini. The difference may also be caused by the pattern of glycosylation, as CHO cell-derived rat NDF-α2 and -β4 contain two Asn-linked and eleven O-linked sugars [10]. Although the responses are different among these NDF forms, the BIAcore response versus NDF concentration is linear at all samples, indicating that the antibody coupled to the sensor chips is capable of recognizing both *E. coli* and CHO cell-derived NDF feeding into the sensor surface in a concentration-dependent manner. The NDF concentrations that fall into the linear quantitation range is approximately 10–200 ng/ml for rat NDF-α2 derived from *E. coli*, 20–300 ng/ml for CHO-derived NDF-α2 and 50–600 ng/ml for CHO-derived NDF-β4.

Precautions should be taken into consideration when the bioaffinity technique is used for sample monitoring and quantification. For example, the accuracy of measurement relies on the recognition of NDF by the coupled antibody being within a linear response range (see Fig. 2); samples containing NDF above the concentration limit require further dilutions with HBS buffer. A gradual decrease in RU occurred on daily use. To compensate for this problem, calibration standards are always established at the beginning and end of runs to ensure consistency of the results within 24-h operation. A 3–5% variation for the observed RU is acceptable and is frequently observed within 24 h. After coupling to the sensor chips, usually anti-

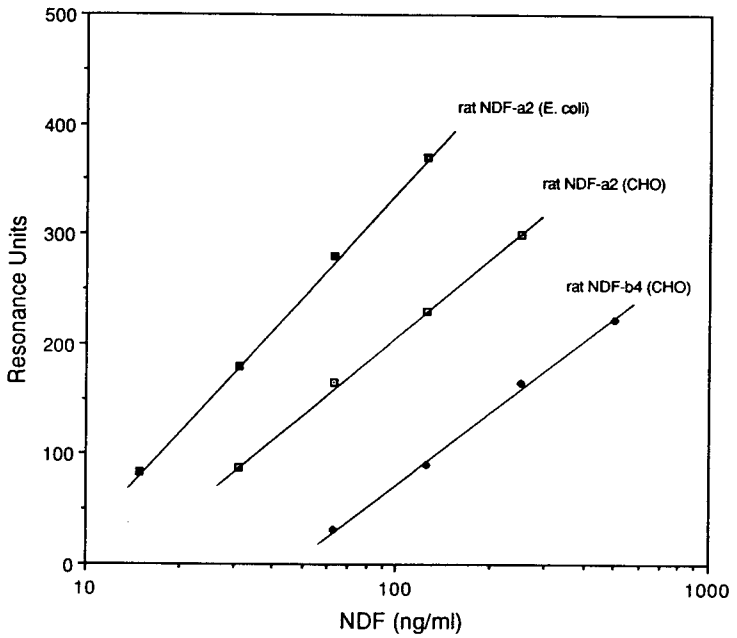


Fig. 2. Calibration graphs for determination of CHO cell-derived recombinant rat NDF-α2 and -β4 in solution versus rat NDF-α2 produced in *E. coli*. The resonance response on the vertical axis is proportional to the amount of the NDF bound to anti-NDF antibody on the sensor chip surface.

bodies are stable and can be used for approximately 2 weeks, at which time the resonance unit may remain less than half of a freshly prepared chip. Chips coupled with antibodies can be repeatedly assayed for approximately 1000 samples. Thus, as a routine and automatic operation, this procedure can continuously assess measurements for 2 weeks using a single chip. However, chips coupled with other protein molecules, especially soluble receptors, may be less stable, and might have a much shorter life span. This decreased SPR or shorter life span may be partly due to washout or acid-induced inactivation of the bound antibodies or proteins during regeneration. The coupling efficiency is usually reproducible with antibodies obtained from the same purification batch. However, significant variations in resonance response did occur using different batches of antibodies. Coupling of lower titer antibodies sometimes yielded an unusable chip.

As shown in Table 1, the linear response characteristic for binding between NDF and antibody allows the reliable quantification of crude NDF preparation concentrated from the cultured medium. The accurate quantification of NDF was not possible by methods using both Western blotting and tyrosine kinase assays [1,2,10]. SDS-PAGE gave multiple overlapping protein bands from the concentrated medium over a wide range of molecular mass, and the method was unable to locate the M_r 40 000–44 000 NDF band after Coomassie Blue staining in samples having low NDF expression.

Fig. 1B shows SPR raw data obtained from analyses of several column fractions from heparin-Sepharose chromatography of rat NDF-α2 produced in CHO cells. In these analyses, fractions were diluted 50–100-fold with HBS buffer before sample loading. Sensorgrams 1–4 are for samples taken from fractions 20, 53, 60 and 80 as indicated in Fig. 3A. Fraction 60 clearly gave a very high reading (RU > 600), indicating that it exceeds the linear response range. Any samples such as this were diluted further and reanalyzed within the linear response

Table 1
Determination of rat NDF by 280 nm and BIAcore detection

Step	Volume (ml)	Protein[a] (280 nm)	NDF[b] (mg)	Purification (-fold)	Yield (%)
(A) CHO cell-derived rat NDF-β4					
(1) Medium	10000	–	–	–	–
(2) Concentrate	1000	1600.0	5.52	1	100
(3) Heparin-Sepharose	110	22.2	3.60	72	65.2
(4) DEAE-Sepharose	78	10.8	2.50	148	45.3
(5) Phenyl-Sepharose and concentration	7	1.10	2.00	1455	36.2
(B) Rat-1 EJ cell-derived NDF					
(1) Medium	115000	–	–	–	–
(2) Concentration	2100	10710.0	0.081	1	100
(3) Heparin-Sepharose	165	49.5	0.042	216	52.9
(4) DEAE-Sepharose	60	3.5	0.028	3060	35.0
(5) Phenyl-Sepharose and concentration	1.7	~0.015[c]	0.025	–[c]	31.3

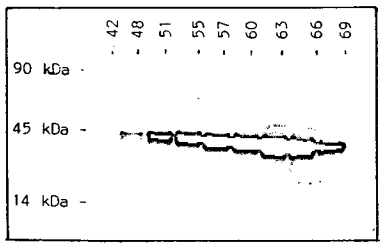
For detailed purification procedure, see Experimental.
[a] Proteins were measured by absorbance at 280 nm. The numbers on this column are the total absorbance readings.
[b] The NDF content was measured by BIAcore (see Experimental).
[c] The 280-nm absorbance is not accurate owing to low concentration, hence the purification fold cannot be precisely calculated.

range. Fig. 3A shows a typical chromatographic profile of the first heparin-Sepharose column step to purify rat NDF-β4 from 1 l of concentrated medium (see also Table 1A). The chromatographic profile shows both the protein elution detected by the 280 nm absorption and BIAcore quantification. The amount of NDF eluted was quantified and expressed in μg per fraction after the detected resonance units were calibrated with a calibration graph (Fig. 2). Clearly, the NDF eluted later as a shoulder by 280-nm measurement and was located at the right portion of the major protein peak by BIAcore detection. This observation matches the detected M_r 44 000 NDF bands observed in Western blots (inset in Fig. 3) where the antibody reacts strongly with the SDS-denatured NDF protein bands After SDS-PAGE.

The NDF elution profile detected by antibody surface binding in BIAcore also works really well for both DEAE- and phenyl-Sepharose chromatography. In these chromatographic separations, detection at 280 nm only picks up NDF protein peak at very low absorption. As shown in Fig. 2B, BIAcore clearly demonstrates its better sensitivity and capability of identifying the NDF fractions during phenyl-Sepharose chromatography. The sensitivity limit for NDF quantification is approximately 20–30 ng/ml.

Table 1A summarizes the purification recovery and step yield for recombinant NDF-β4 determined by absorption measurement and BIAcore. The overall step yield for the recovery of NDF protein is about 36% when measured by BIAcore. Various chromatographic steps yielded approximately 1500-fold purification from the concentrated medium. The NDF content in the final concentrated and purified preparation is calculated to be 2 mg, consistent with the data precisely quantified by amino acid analysis.

Rat-1 EJ cells produce only very small amounts (700 ng/l) of natural NDF. The detection of the NDF fraction using phosphorylation assay has been difficult [1,2]. Assay using BIAcore allowed high-sensitivity detection and quantification of NDF. As shown in Table 1B, NDF can only be detected after substantial concentration. A 2.1-l diafiltrate concentrated from 115 l of medium contained less than 100 μg of NDF. Following similar purification proce-



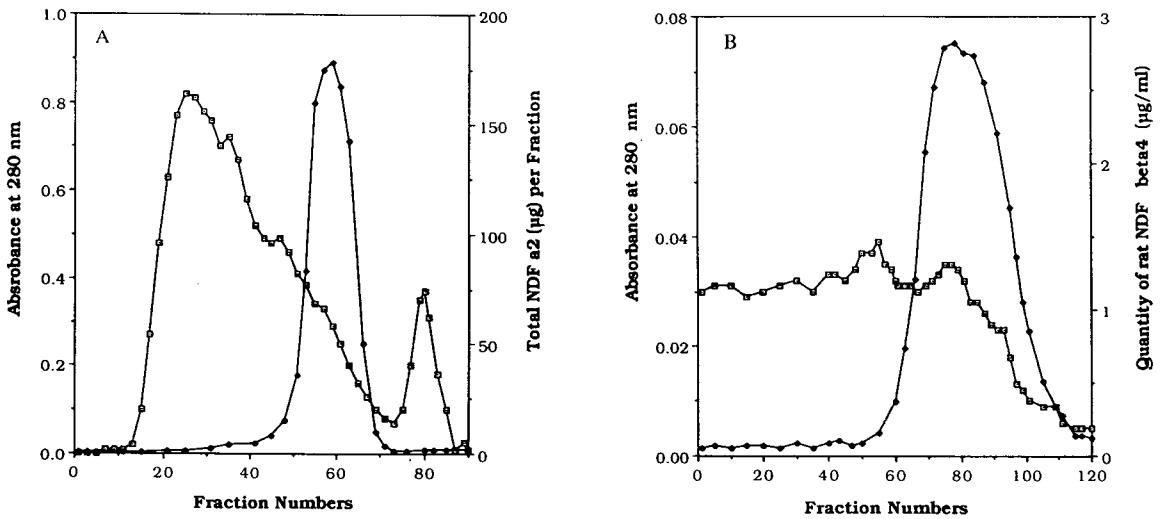


Fig. 3. (A) Heparin-Sepharose chromatography of a crude medium containing recombinant rat NDF-β4 (also see Table 1). The concentrated medium loaded on to the column (20 × 2.5 cm I.D.) contains 5.52 mg of NDF and 1.6 g of total proteins. After sample loading, 95% of contaminating proteins was washed off the column by 0.3 *M* NaCl in PBS buffer (pH 7.2). The column was then eluted with a 0.3–1.0 *M* NaCl linear gradient to fractionate NDF. Fractions (3.0 ml) were collected and detected for 280 nm absorption (□). Aliquots of fractions, after dilution, were also subjected to BIAcore analysis (◆). Inset: Western blotting of fractions shown in (A). (B) Phenyl-Sepharose chromatography of a DEAE-Sepharose pool containing recombinant rat NDF-β4 (also see Table 1). A linear gradient was established from 1.5 to 0 *M* ammonium sulfate in PBS (150 ml each) after sample loading. Fractions (1.8 ml) were collected and subjected to UV absorption measurement (□) and BIAcore assay (◆). kDa = kilodalton.

dures, separation of NDF can be obtained. Fig. 4 illustrates the DEAE-Sepharose chromatographic profile of a natural rat NDF preparation after a heparin-Sepharose column fractionation. A small shoulder peak, detected at 280 nm and eluted earlier than a major contaminating peak, was recognized to contain NDF activity. After final purification, a total of 25 μg of pure NDF was obtained.

The above results support the advantage of BIAcore analysis over phosphorylation assay and Western blotting in sample detection and NDF quantification. As the tyrosine kinase stimulatory activity of NDF was performed using membrane-bound p185^{neu} in whole cell lysates, the assay results displayed phosphorylated receptor in protein blots which was visualized by anti-phosphotyrosine antibody after NDF stimulation. As the assay requires several manual operations, the observed data would inevitably vary from run to run. The phosphorylation assay is thus less applicable in handling multiple chromatographic samples and performing precise quantification. Western blotting is also inherently not suitable

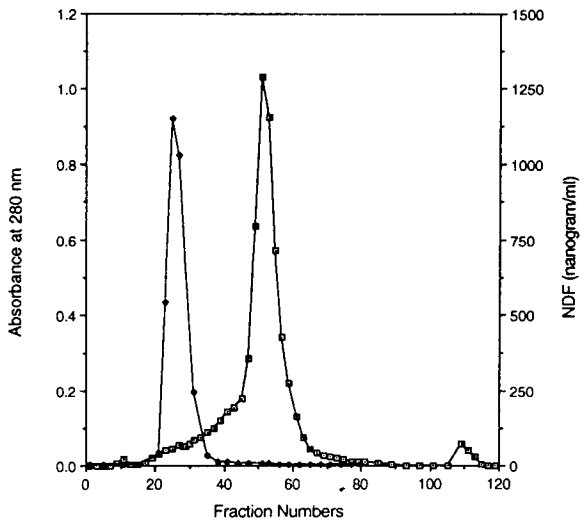


Fig. 4. DEAE-Sepharose chromatography of an NDF sample obtained from heparin-Sepharose chromatography of concentrated media conditioned by rat-1 EJ cells (see Table 1B). A linear gradient from 0 to 0.5 *M* NaCl in 10 m*M* phosphate buffer (pH 7.0) (200 ml each) was applied after loading. Fractions (3 ml) were collected and assayed by 280-nm absorption measurement (□) and BIAcore assay (◆).

for quantification as recognition of a target protein by its specific antibody is performed by an SDS-PAGE–electroblotting procedure.

In summary, the specific bioaffinity technique provides an alternative measurement for the quantification of NDF produced naturally or recombinantly during chromatographic fractionation. This automated method provides sufficient sensitivity for detection and quantification and should be applicable to other proteins using similar antibody binding as described here and other affinity assay techniques such as receptor–ligand binding. A recent paper has reported the use of such an approach to discover the ligand for an ECK receptor of unknown function [13].

Acknowledgement

We are indebted to the Departments of Cell Culture Process Development and Mammalian Cell Culture groups at Amgen for their technical support and to Joan Bennett for typing the manuscript.

References

[1] E. Pales, S.S. Bacus, R.A. Koski, H.S. Lu, D. Wen, S.G. Ogden, R.B. Levy and Y. Yarden, *Cell*, 69 (1992) 205–216.
[2] D. Wen, E. Peles, R. Cupples, S.V. Suggs, S.S. Bacus, Y. Luo, G. Trail, S. Hu, S.M. Silbiger, R.B. Levy, R.A. Koski, H.S. Lu and Y. Yarden, *Cell* 69 (1992) 559–572.
[3] W.E. Holmes, M.X. Sliwkowski, R.W. Akita, W.J. Henzel, J. Lee, J.W. Park, D. Yansura, N. Abadi, H. Raab, G.D. Lewis, H.M. Shepard, W.-J. Kuang, W.I. Wood, D.V. Goeddel and R.L. Vandlen, *Science*, 256 (1992) 1205–1210.
[4] D. Wen, S.V. Suggs, D. Karunagaran, N. Liu, R.L. Cupples, Y. Luo, A.M. Janssen, N. Ben-Baruch, D.B. Trollinger, V.L. Jacobson, S.-Y. Meng, H.S. Lu, S. Hu, D. Chang, W. Yang, D. Yanigihara, R.A. Koski and Y. Yarden, *Mol. Cell. Biol.*, 14 (1994) 1909–1919.
[5] H. Tsuda, S. Hirahashi and Y. Shimosato, *Cancer Res.*, 49 (1989) 3104–3108.
[6] D.J. Slamon, G.M. Clark, S.G. Wong, W.J. Lewin, A. Ullrich and R.L. McGurie, *Science*, 235 (1987) 177–182.
[7] D.J. Slamon, W. Godolphin, L.A. Jones, J.A. Holt, S.C. Wong, D.E. Keith, W.J. Levin, S.G. Stuart, J. Udore, A. Ullrich and M.F. Press, *Science*, 244 (1989) 707–712.
[8] M.S. Berger, G.N. Locher, S. Sauer, W.J. Gullick, M.D. Waterfield, B. Groner and N.E. Hynes, *Cancer Res.*, 48 (1988) 1238–1243.
[9] Y. Yarden and E. Peles, *Biochemistry*, 30 (1991) 3543–3550.
[10] H.S. Lu, S. Hara, L.W.-I. Wong, M.D. Jones, V. Katta, G. Trail, A. Zou, D. Brankow, S. Cole, S. Hu and D. Wen, *J. Biol. Chem.*, submitted for publication.
[11] L. Fägerstam, *Tech. Protein Chem.*, 2 (1991) 65–71.
[12] B. Johnson, S. Löfås and G. Lindquist, *Anal. Biochem.*, 198 (1991) 268–277.
[13] T.D. Bartley, R.W. Hunt, A.A. Welcher, W.J. Boyle, V.P. Parker, R.A. Lindberg, H.S. Lu, A.M. Colombero, R.L. Elliott, B.A. Guthrie, P.L. Holst, J.D. Skrine, R.J. Toso, M. Zhang, E. Fernandez, G. Trail, B. Varnum, Y. Yarden, T. Hunter and G.M. Fox, *Nature*, 368 (1994) 558–560.

END OF SPECIAL ISSUE

PUBLICATION SCHEDULE FOR THE 1995 SUBSCRIPTION

Journal of Chromatography A and *Journal of Chromatography B: Biomedical Applications*

MONTH	1994	J–M	A	M[a]	J	J	
Journal of Chromatography A	Vols. 683–688	689–695	696/1 696/2 697/1 + 2 698/1 + 2	699/1 + 2 700/1 + 2 702/1 + 2 703/1 + 2	704/1 704/2 705/1 705/2	706/1 + 2 707/1 707/2 708/1	The publication schedule for further issues will be published later.
Bibliography Section		713/1			713/2		
Journal of Chromatography B: Biomedical Applications		663–665	666/1 666/2	667/1 667/2	668/1 668/2	669/1 669/2	

[a] Vol. 701 (Cumulative Indexes Vols. 652–700) expected in October.

INFORMATION FOR AUTHORS

(Detailed *Instructions to Authors* were published in *J. Chromatogr. A*, Vol. 657, pp. 463–469. A free reprint can be obtained by application to the publisher, Elsevier Science B.V., P.O. Box 330, 1000 AH Amsterdam, Netherlands.)

Types of Contributions. The following types of papers are published: Regular research papers (full-length papers), Review articles, Short Communications and Discussions. Short Communications are usually descriptions of short investigations, or they can report minor technical improvements of previously published procedures; they reflect the same quality of research as full-length papers, but should preferably not exceed five printed pages. Discussions (one or two pages) should explain, amplify, correct or otherwise comment substantively upon an article recently published in the journal. For Review articles, see inside front cover under Submission of Papers.

Submission. Every paper must be accompanied by a letter from the senior author, stating that he/she is submitting the paper for publication in the *Journal of Chromatography A* or *B*.

Manuscripts. Manuscripts should be typed in **double spacing** on consecutively numbered pages of uniform size. The manuscript should be preceded by a sheet of manuscript paper carrying the title of the paper and the name and full postal address of the person to whom the proofs are to be sent. As a rule, papers should be divided into sections, headed by a caption (*e.g.*, Abstract, Introduction, Experimental, Results, Discussion, etc.). All illustrations, photographs, tables, etc., should be on separate sheets.

Abstract. All articles should have an abstract of 50–100 words which clearly and briefly indicates what is new, different and significant. No references should be given.

Introduction. Every paper must have a concise introduction mentioning what has been done before on the topic described, and stating clearly what is new in the paper now submitted.

Experimental conditions should preferably be given on a *separate* sheet, headed "Conditions". These conditions will, if appropriate, be printed in a block, directly following the heading "Experimental".

Illustrations. The figures should be submitted in a form suitable for reproduction, drawn in Indian ink on drawing or tracing paper. Each illustration should have a caption, all the *captions* being typed (with double spacing) together on a *separate sheet*. If structures are given in the text, the original drawings should be provided. Coloured illustrations are reproduced at the author's expense, the cost being determined by the number of pages and by the number of colours needed. The written permission of the author and publisher must be obtained for the use of any figure already published. Its source must be indicated in the legend.

References. References should be numbered in the order in which they are cited in the text, and listed in numerical sequence on a separate sheet at the end of the article. Please check a recent issue for the layout of the reference list. Abbreviations for the titles of journals should follow the system used by *Chemical Abstracts*. Articles not yet published should be given as "in press" (journal should be specified), "submitted for publication" (journal should be specified), "in preparation" or "personal communication".

Vols. 1–651 of the *Journal of Chromatography*; *Journal of Chromatography, Biomedical Applications* and *Journal of Chromatography, Symposium Volumes* should be cited as *J. Chromatogr.* From Vol. 652 on, *Journal of Chromatography A* (incl. Symposium Volumes) should be cited as *J. Chromatogr. A* and *Journal of Chromatography B: Biomedical Applications* as *J. Chromatogr. B.*

Dispatch. Before sending the manuscript to the Editor please check that the envelope contains four copies of the paper complete with references, captions and figures. One of the sets of figures must be the originals suitable for direct reproduction. Please also ensure that permission to publish has been obtained from your institute.

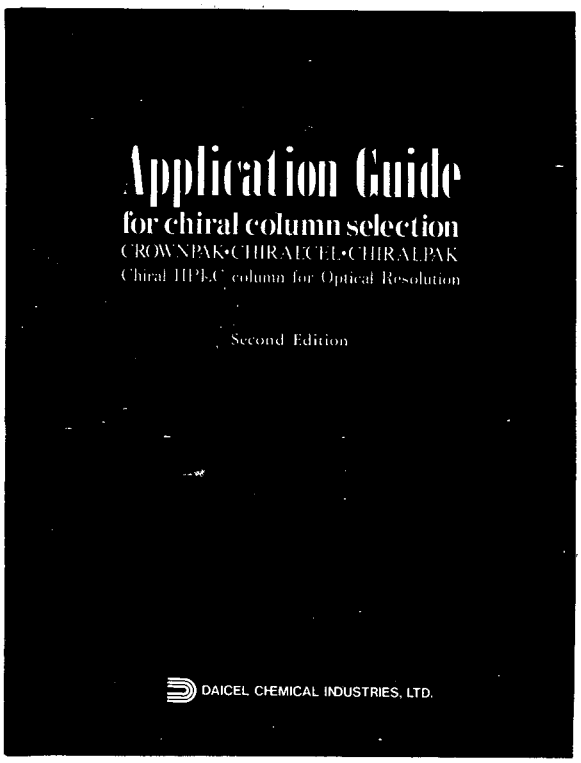
Proofs. One set of proofs will be sent to the author to be carefully checked for printer's errors. Corrections must be restricted to instances in which the proof is at variance with the manuscript.

Reprints. Fifty reprints will be supplied free of charge. Additional reprints can be ordered by the authors. An order form containing price quotations will be sent to the authors together with the proofs of their article.

Chiral HPLC Column

Application Guide for Chiral HPLC Column Selection

SECOND EDITION!



FREE OF CHARGE

The 112-page green book contains chromatographic resolutions of over 350 chiral separations, cross-indexed by chemical compound class, structure, and the type of chiral column respectively. This book also lists chromatographic data together with analytical conditions and structural information. A quick reference guide for column selection from a wide range of DAICEL chiral HPLC columns is included.

To request this book, please let us know by fax or mail.

DAICEL CHEMICAL INDUSTRIES, LTD.

AMERICA

CHIRAL TECHNOLOGIES, INC.

730 Springdale Drive, P.O. Box 564
Exton, PA 19341
Phone: 800-624-4725
Facsimile: 610-594-2325

EUROPE

DAICEL (EUROPA) GmbH

Oststr. 22
D-40211 Düsseldorf, Germany
Phone: +49-211-369848
Facsimile: +49-211-364429

ASIA/OCEANIA

DAICEL CHEMICAL INDUSTRIES, LTD.
CHIRAL CHEMICALS NDD
8-1, Kasumigaseki 3-chome,
Chiyoda-ku, Tokyo 100, JAPAN
Phone: +81-3-3507-3151
Facsimile: +81-3-3507-3193

484 a