Contemporary Methodology for Protein Structure Determination

Michael W. Hunkapiller, James E. Strickler, Kenneth J. Wilson

Summary. The techniques used for the characterization of protein and peptide structure have undergone great changes that have improved the speed, reliability, and applicability of the process. High-performance liquid chromatography and gel electrophoresis have made the purification of proteins and peptides a routine procedure, even when the compound of interest is a minor component of a complex biological mixture. The chemistry and instrumentation used in amino acid analysis and amino acid sequencing now permit the analysis of as little as 5 to 50 picomoles of samples. This represents an increase in sensitivity of more than a thousandfold over the last 10 years and has made possible the structural analysis of a wide variety of scarce but important compounds.

One of the primary goals of protein chemists is to relate the function of a polypeptide to its structure. In the past, the initial phase of protein characterization studies has involved painstaking trial-and-error efforts to purify an active substance from a complex mixture, to conduct a compositional analysis to determine the gross amino acid, carbohydrate, metal ion, and lipid content, and to determine its amino acid sequence. At the present time, the speed at which chromatography and electrophoresis can be performed, the simplicity, reliability, and separation efficiencies of these purification methods, and the generally high recoveries of both sample mass and active component from the supports allow for the preparative isolation of numerous proteins available only in microgram amounts. Similarly, purified samples can be analyzed and sequenced with corresponding speed, accuracy, and sensitivity. As a result, the structural analysis of proteins is now possible with a few micrograms rather than several milligrams of starting material.

As the methodology for protein structure determination has improved, the uses of the structural information have also evolved. Partial structural information obtained by protein microsequence analysis is frequently invaluable in obtaining the DNA probes necessary to clone the genes encoding for the protein and in confirming the DNA sequences obtained from the cloned genes. These cloning technologies in turn form the heart of the recombinant DNA field with their ability to provide unlimited amounts of any protein that may be used for further study or applications in biochemistry, medicine, agriculture, or other areas.

Even though DNA sequencing can be used to infer protein sequences, only direct sequence analysis can address the presence of post-translational processing (glycosylation, phosphorylation, proteolytic cleavage, and so on) of the gene-derived polypeptide, events that are often essential to the biological activity of the protein. Moreover, because the modern structural methods are quantitative as well as qualitative in nature, they can be used to answer questions concerning authenticity and purity of protein preparations (essential when they are being considered, for example, for therapeutic uses) and stoichiometry of multisubunit proteins that often perform the more complex enzymatic or structural functions. In this article we review these structural methods and describe their nature and capabilities.

Protein and Peptide HPLC

The steadily growing variety of high-performance liquid chromatography (HPLC) media is quickly replacing the classical supports used for the separation of proteins. The entire range of chromatography methods—including reverse phase (RPC) separations based on hydrophobicity differences between samples, ion exchange (IEC) separations based on charge differences, and size exlusion (SEC) separations based on size differences—is supported by HPLC. Progress in biochemical and immunological fields has often been dependent on the use of HPLC methodology for the isolation and characterization of various components of a biological system, such as hormones, lymphokines, proteases, and viral and plasma proteins. The speed, versatility, and sensitivity of HPLC make it an ideal method for protein purification, and it has been especially useful for the isolation of proteins that maintain or recover their biological activity after chromatography.

The most widely used HPLC columns are packed with RPC supports. Various covalently bonded surface phases, particle sizes, and pore diameters are available, and the columns come in several lengths (5 to 25 cm) and widths (1 to 10 mm) suitable for analytical and preparative applications. The RPC column that seems best suited for proteins is silica with covalently bonded, short alkyl chains (C_3, C_4, or C_8), 10-μm particles, and 300-Å pores (*1*). The numerous factors affecting gradient elution RPC separations of "standard" proteins with these packings have been described. In general, the columns provide the highest yields and resolution when used with low *p*H aqueous phases such as 0.1 percent trifluoroacetic acid (TFA) or phosphoric acid combined with organic modifiers such as acetonitrile and 1- or 2-propanol. The simple aqueous TFA and acetonitrile combination is generally used unless the relatively harsh acidic conditions give prohibitively low recovery of material or biological activity. The volatility of these solvents makes it possible to recover chromatographed samples, in most instances, simply by vacuum drying the column fractions. Thus, the samples are recovered in a state compatible with amino acid analysis or protein sequencing techniques; that is, they are not contaminated with buffer salts and have not undergone significant chemical modification. Resolubilization can be a problem, however, and different combinations of sodium dodecyl sulfate (SDS), high *p*H (ammonium hydroxide and triethylamine), and low *p*H (acetic acid and TFA) are often required to effect solution of the dried samples.

Three examples of protein purification

Michael W. Hunkapiller is a member of the research staff at Applied Biosystems, Inc., Foster City, California 94404. Kenneth J. Wilson (present address, Applied Biosystems) and James E. Strickler are members of the research staff at Cetus Corporation, Emeryville, California 94608.

by RPC illustrate the method's usefulness. Figure 1 shows that chromatography of interleukin-2 from peripheral blood lymphocytes separated two biologically active components which were chemically indistinguishable by amino acid sequence analysis (*2*). Because interleukin-2 is a glycoprotein, the two components may contain differences in the extent of glycosylation sufficient to effect separation. RPC was the only chromatographic mode needed to isolate neuronal parvalbumin (not shown) (*3*). In this purification neutral *p*H buffer systems were used, and the procedure could be carried out in a single working day with minimum sample loss due to proteolysis and denaturation. Figure 2 shows the final steps in the purification of human fibroblast interferon (*4*). Here, three different RPC columns were used: octyl-, cyanopropyl-, and diphenylsilyl silica. The pyridine–formic acid buffer used in this protocol precluded low ultraviolet (210 to 220 nm) monitoring; instead, a fluorescent detection system based on postcolumn derivatization with fluorescamine was used. This detection scheme is quite sensitive, but it has several drawbacks; for example, it requires additional instrumentation (stream splitter and pumps), detects only peptides with free primary amino groups, and requires removal and purification of the buffer components.

The introduction of several cation- and anion-exchange packings and their reported general utility (*5*) has provided an alternative to the use of RPC. The most important attribute of these IEC supports is their compatibility with mild elution conditions: they can be used at physiological *p*H levels without the organic modifiers required in RPC which disrupt the hydrophobic interactions between solute and bonded phase. The IEC technique is often used in conjunction with RPC as the first step so that the IEC buffer components can be removed from the protein during the RPC step (the hydrophilic buffer components pass through the RPC columns with little retention even under strictly aqueous conditions).

Figure 3 illustrates anion-exchange separation of the three tryptic fragments from diphtheria toxin. Sufficient quantities of each component (differing by a single positive charge) were prepared so that comparisons of their enzymatic activities could be made. Although the fragments can also be separated by RPC with recovery of activity, IEC gives a degree of recovery of activity under mild elution conditions such that small, although significant, differences in the specific activities of the three fragments can be measured (*6*). The same set of toxin fragments was separated by chromatofocusing (not shown); in this experiment, the fragments were applied to the support equilibrated at *p*H 6.1 and eluted with an ampholite mixture adjusted to *p*H 4.0. The average *p*I values for each of

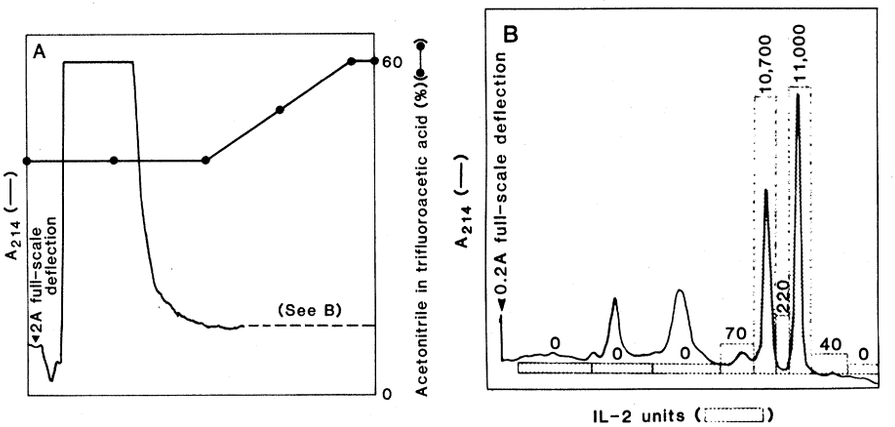

Fig. 1. Isolation of interleukin-2 (IL-2) by RPC. (A) A partially purified fraction from peripheral blood lymphocytes which contained significant amounts of biological activity was loaded onto an RPC-8 column. (B) After most of the material that absorbed ultraviolet light was eluted, a gradient from 40 to 60 percent acetonitrile in 0.1 percent TFA was started. The peaks that eluted late contained approximately 45 percent of the applied biological activity.

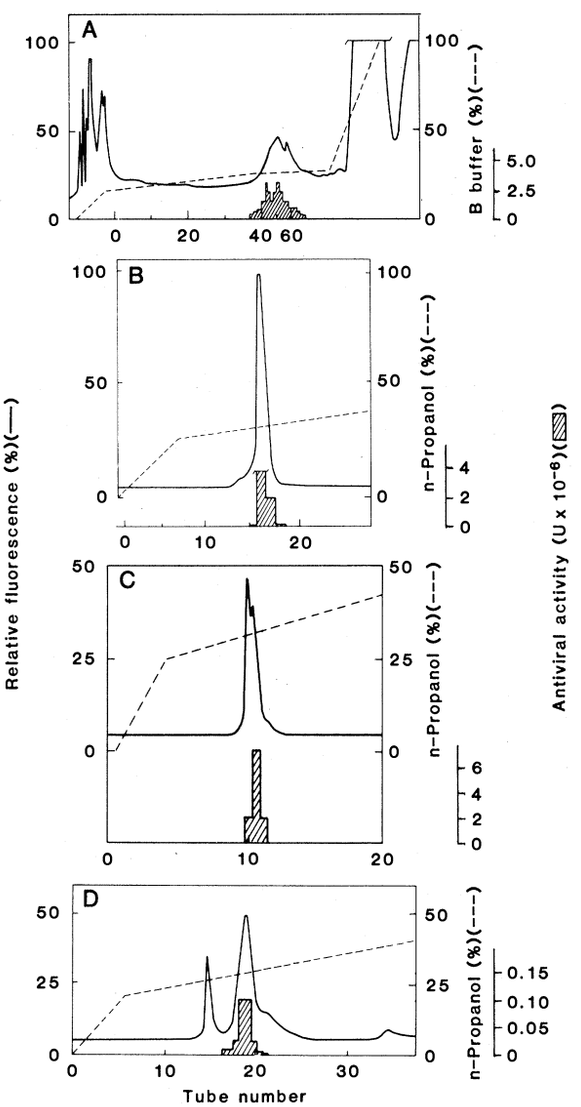

Fig. 2. Isolation of fibroblast interferon by RPC. Chromatography was performed on an RPC-18 (A), cyanopropyl (B and C), and diphenyl support (D). The buffer systems used were pyridine, formic acid, 2-propanol, and *n*-butanol combinations. The chromatograms were taken from (*4*).

the peaks were within 0.05 *p*I units of those determined by isoelectric focusing.

The various HPLC media have become indispensable to protein sequence determination because of their utility in peptide isolation. The generation and purification of peptide fragments derived by enzymatic or chemical cleavage of large proteins has been the rate-limiting step in sequence analysis of these proteins, and HPLC has greatly improved this process in several respects. Overall peptide recoveries (50 to 100 percent), speed (<1 hour per separation), and resolution (more than 12,000 theoretical plates per column) are generally very good, and RPC has for all practical purposes replaced the chromatographic and electrophoretic techniques used 5 to 10 years ago. Peptides differing by as little as a single amino acid residue can be resolved, and the absence of any significant structure within most short polypeptide fragments (less than 30 residues) allows them to interact with the RPC support in such a manner that their retention times in a particular buffer system can be estimated on the basis of the gross amino acid composition (*7*).

Peptide maps are used both for comparing closely related proteins and for isolating peptides generated by various fragmentation procedures. Two-dimensional peptide mapping increases the resolution of HPLC even further. Cation-exchange with subsequent RPC, RPC at two *p*H values, and RPC with two different ion-pairing agents have been used for this purpose.

The RPC system is also useful for the isolation and characterization of peptides or proteins containing specifically

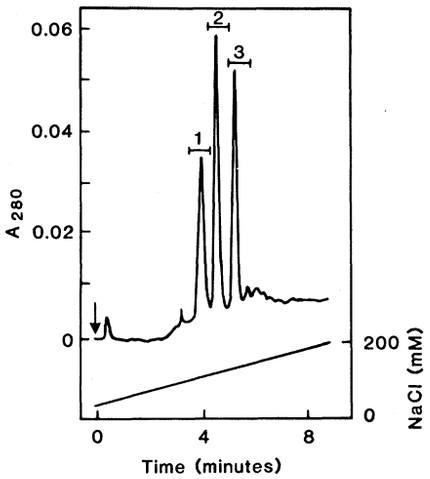

Fig. 3. Ion exchange separations of diphtheria toxin A fragments: anion exchange separation. See (*6*) for details.

modified sites. Differences in elution profiles were observed with enzymatically derived fragments of glycosylated (Fig. 4, A and C) and nonglycosylated (Fig. 4, B and D) ovine prolactin (*8*). Similar investigations have been reported for the glycopeptides of hen ovalbumin, human cerulopasmin and immunoglobulin D, and those generated by the reductive lactosylation of thymosin-α1 and ribonucleases A and B (*9*). Selective modification of some amino acid side chains can introduce significant hydrophobicity changes, and RPC can distinguish not only modified from unmodified species but also those containing modified sites at different positions. Examples of chemical reactions that have been used are *S*-methylation and demethylation of methionyl residues, α and ε amino group methylation, and the hydrogen peroxide oxidation of methionine to methionine sulfoxide (*10*).

As the sensitivities of the numerous techniques employed in protein sequencing increase, the amounts of a particular peptide or protein that can be chromatographed and detected can become a serious limitation. One approach to improving detection limits has been to introduce either a pre- or postcolumn chemical modification of the peptides to label one or more amino acid side chains. A second approach has been to modify the chromatographic conditions so that the actual concentration of the eluting substance is increased by reducing its dilution during separation. The development of small-particle (3 μm) packings and microbore (1 to 2 mm) columns shows the success of this method. Even though HPLC instrumentation has not been optimized for microbore work, by means of a standard HPLC pumping system and ultraviolet detector with a microbore column mapping can be done on as little as 1 to 5 pmol of a protein digest (Fig. 5).

Table 1. Selected examples of proteins recovered from SDS-PAGE by electrophoretic elution of Coomassie-stained gel slices.

Protein or peptide	Molecular weight (thousands)	Amount recovered (micrograms)	Recovery (percent)
Water-soluble proteins			
Human serum albumin	68	8	80
Human β-interferon	18	0.5	85
Mouse interleukin-2	22	2	
^{3}H-labeled basic pancreatic trypsin inhibitor	6.4	(1 to 5)	70
^{3}H-labeled methylated carbonic anhydrase	31	(1)	74
Integral membrane proteins			
Mouse immune response antigen A_e^k	25	0.5	50
Acetylcholine receptor (*Torpedo* subunit)	65	4	71
Diol dehydratase subunit	29	100	80
Rat liver gap junction	28	20	60
Photosynthetic reaction center (subunits of *Rhodopseudomonas capsulata*)	25 to 30	(1 to 5)	>75
Protein fragments			
Torpedo acetylcholine receptor subunit	9	20	
(V8 protease–derived)	2.5	2	
MM–creatine kinase (proteinase K–derived)	41	(1 to 5)	
Diptheria toxin A subunit	21	(1 to 5)	

Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis (PAGE) has been widely used for the analytical separation of proteins. The PAGE separations based on intrinsic charge-to-mass ratio (*11*), apparent molecular weight in the presence of SDS (*12*), and isoelectric focusing (*13*) have remarkable resolving power with amounts of protein ranging from 5 pg to several milligrams. When two of these techniques, molecular sieving with SDS and isoelectric focusing with ampholytes, are combined in two-dimensional separations (*14*), complex mixtures can be resolved into as many as several thousand components in a single analysis. Many of the electrophoretic techniques also permit the separation of medium to large peptides, and the electrophoretic analysis of peptide fragments generated by either limited enzymatic or chemical cleavage of proteins is especially useful for examining similarities between samples. Many of these cleavage techniques can be performed while the protein is inside the gel, thus facilitating the use of two-dimensional gel mapping with its enhanced resolution.

The PAGE protocol also provides a powerful preparative technique for small quantities of proteins if they can be removed from the gel once they have been resolved. Recovery of the proteins can be accomplished by (i) elution from gel pieces after stopping the electrophoresis and (ii) continuous elution as the pro-

teins emerge from the end of the gel during the electrophoresis.

Various postelectrophoresis methods have been used successfully to prepare proteins for subsequent chemical or structural analysis. Passive elution of many proteins by simply soaking the gel slices, often minced into small pieces, in a dilute detergent solution was used to prepare isotopically labeled histocompatibility antigens for microsequencing (*15*). Recoveries with this method are generally low, especially if the gel must be stained with dye to locate the protein bands, and contamination of the sample with bits of acrylamide can interfere with subsequent experiments. Electrophoretic elution techniques generally give higher recoveries and less contamination, and they can be used to concentrate the protein into a small volume. While many devices have been described, most rely on electrophoretic movement of the protein out of the gel slices into a buffer solution where it is retained by a dialysis membrane or other semipermeable barrier (*16*). Using this technique, we have obtained good recoveries of a wide variety of proteins after staining the gels with Coomassie blue dye under acidic conditions for visualization (Table 1).

Continuous elution has not been widely used as a microscale preparative method because the apparatus available commercially, along with most of those described, is designed for rather large-scale work. The resolution of eluted peaks is generally much poorer than that obtained in stained slab gels, and the proteins are eluted into large volumes of buffer which complicates recovery. However, these limitations can be overcome if the size of the apparatus is scaled down to tube gels 1 to 2 mm in diameter, a size sufficient to handle the few micrograms of protein now usable for structural analysis. Proteins differing in molecular weight by only 2 percent can be resolved with the use of Laemmli (*12*) buffers in such a system (*17*), and peak volumes can be kept to a few hundred microliters (Fig. 6). Eluting proteins can be monitored by ultraviolet absorption or intrinsic isotopic label.

Chemical and Enzymatic Fragmentation of Proteins

Existing implementations of the Edman degradation cannot provide the complete primary structure of proteins with only the intact molecules. Most of the sequence must be determined from peptides derived from the parent protein by chemical or enzymatic cleavage, and the generation and separation of these fragments has usually been the rate-limiting step in the structural analysis of proteins (*18*). While most techniques used to fragment large quantities of a protein are applicable to microscale work, the presence of impurities in the reagents can be much more detrimental when working with small amounts of sample either by causing side reactions, such as amino terminal or side chain modification, or by interfering with identification of the peptides during separation after cleavage. The HPLC method, as discussed above, is ideal for the separation of peptides, but SDS-PAGE is also suitable for preparing peptides with molecular sizes greater than a few thousand daltons.

The most widely used chemical method for fragment generation is also one of the oldest: cyanogen bromide (CNBr) cleavage at methionyl residues. Other established chemical methods include hydroxylamine cleavage at Asn-Gly bonds, mild acid cleavage at Asp-Pro bonds, and cleavage at tryptophanyl residues by means of *o*-iodosobenzoic acid. Other less widely used chemical methods include a novel cleavage of a Gly-Phe bond at *p*H 10 with sodium borohydride (used in the sequencing of bacteriorhodopsin), controlled acid hydrolysis at aspartyl residues, and cleavage at cysteinyl residues with Raney nickel.

The most widely used enzymatic method is trypsin cleavage at lysyl and arginyl residues. Trypsin is useful for producing a large number of small peptides, and its specificity can be limited to either arginyl residues (after blockage of the lysyl side chains by reaction with succinic anhydride) or lysyl residues (after blockage of the arginyl side chains by reaction with cyclohexanedione). Sever-

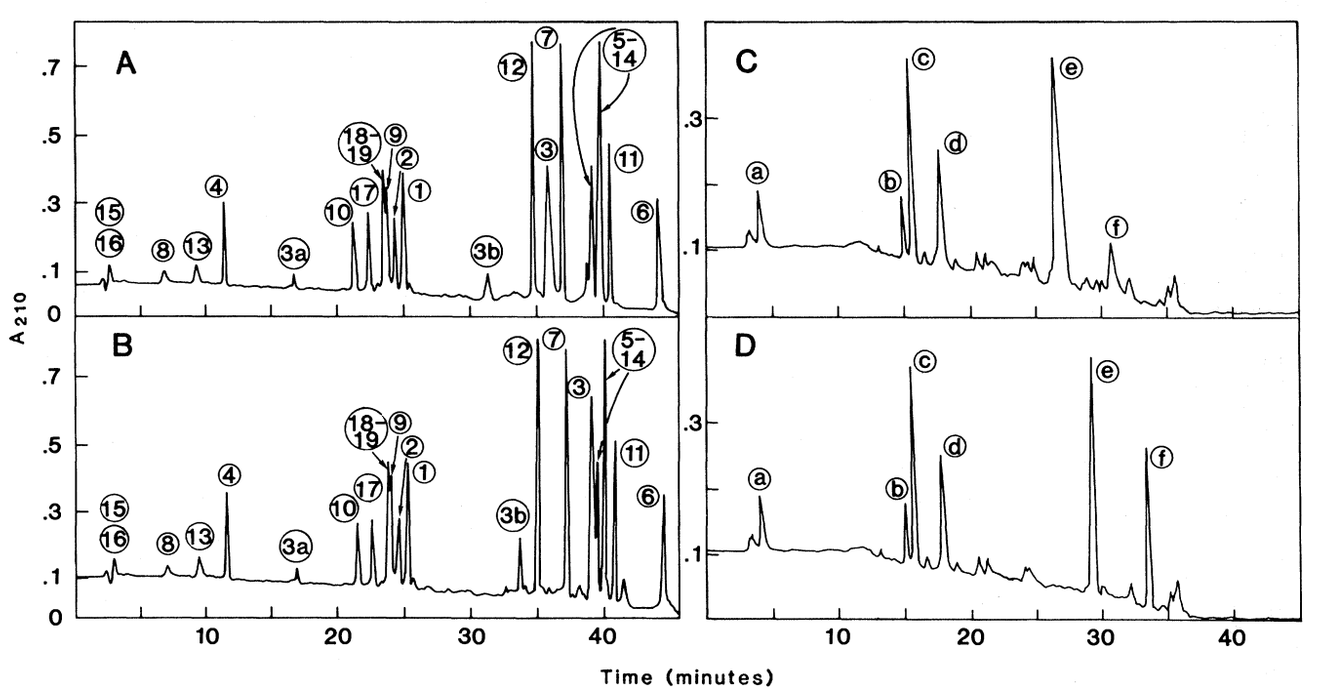

Fig. 4. Peptide mapping by RPC: separation of the tryptic (A and B) and chymotryptic (C and D) peptides from glycosylated and nonglycosylated ovine prolactin. The chromatographic differences in tryptic peptide 3 [compare (A) and (B)] were also seen in the chymotryptic peptides [compare peaks e and f in (C) and (D)]. These maps made possible the isolation and identification of the single glycosylation site in one of a number of possible forms of this growth hormone. The chromatograms were taken from (*8*).

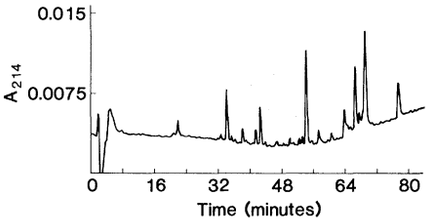

Fig. 5. High-sensitivity peptide mapping by means of microbore RPC columns. A sample (30 pmol) of a tryptic digest of carboxymethylated bovine pancreatic trypsin inhibitor was injected onto a column (1 by 150 mm) packed with a 3-μm Ultrasphere support equilibrated in 0.1 percent TFA. Immediately after the sample was applied, pumping of the limit buffer (60 percent acetonitrile in 0.1 percent TFA) was initiated (50 μl/min), and the gradient was allowed to form in the buffer-mixing chamber.

al less specific proteases, including chymotrypsin, elastase, and pronase, have also been used for some time but their use is generally being replaced by proteases which have much greater cleavage specificity because the generation of a relatively small number of fragments simplifies the task of peptide separation after the cleavage. Two enzymes—*Staphylococcus aureus* endoprotease V8, which cleaves primarily at glutamyl residues, and *Lysobacter enzymogenes* endoprotease lys-C, which cleaves at lysyl residues—have proven to be excellent for producing larger fragments. Their specificity is enhanced by use of mild, nondenaturing conditions in which only a limited number of exposed glutamyl or lysyl residues are cut rapidly, and both enzymes are active in the presence of SDS.

Amino Acid Analysis

In their pioneering work on the development of amino acid analysis, Stein and Moore (*19*) used cation-exchange chromatography on sulfonated polystyrene resins to resolve the amino acids and reaction of the eluted compounds with ninhydrin to detect them. The technique has undergone several significant alterations that resulted in faster analysis, improved resolution, and greatly enhanced sensitivity, with increased sensitivity being most important for studying proteins available only in limited amounts. Currently, the two major uses for high-sensitivity amino acid analysis are (i) to determine an approximate composition and (ii) to determine approximate sample concentration.

The ion-exchange separation of amino acids and postcolumn derivatization still form the basis of a number of commercially available instruments. However, the development of high-performance RPC media and the limitations placed on sensitivity by postcolumn derivatization have led to the development of several interesting alternatives to the Stein and Moore methodology. Some of these (Table 2) are quickly becoming well established.

The most widely used of the new techniques employs precolumn derivatization of the amino acids with orthophthaldehyde (OPA) in the presence of a thiol and detergent (*20*) (Fig. 7A). Sensitivity limits based on fluorescent detection are in the low picomole (1 to 10), range and separation times (including column reequilibration) are less than 60 minutes.

Table 2. Recent methodology for HPLC-based amino acid analysis.

Chromatographic (detection) conditions	Analysis		
	Sensitivity (picomoles)	Time (minutes)	Reference
Free amino acids			
NH_2-bonded phase (200 nm ultraviolet)		30	(*40*)
RPC-18 (postcolumn OPA)	<40	30	(*41*)
Cation-exchange (postcolumn OPA)	10	45	(*42*)
OPA–amino acid derivatives			
RPC-8 (fluorescence)	10	55	(*43*)
RPC-18 (fluorescence)	0.5	30	(*23*)
RPC-18 (electrochemical)	0.5	30	(*21*)
DNS–amino acid derivatives			
RPC-8 (fluorescence)	<40	50	(*24*)
RPC-18 (fluorescence)		40	(*44*)
RPC-18 (254 nm ultraviolet)	100	30	(*45*)
*DABS–amino acid derivatives**			
RPC-18 (436 nm visible)	5	45	(*46*)
FMOC–amino acid derivatives†			
RPC-18 (fluorescence)	<1	30	(*47*)
PITC–amino acid derivatives			
RPC-8, RPC-18 (254 nm ultraviolet)	>1	45	(*25*)

*DABS, dimethylazobenzene sulfonyl chloride. †FMOC, fluorenylmethyloxycarbonyl chloride.

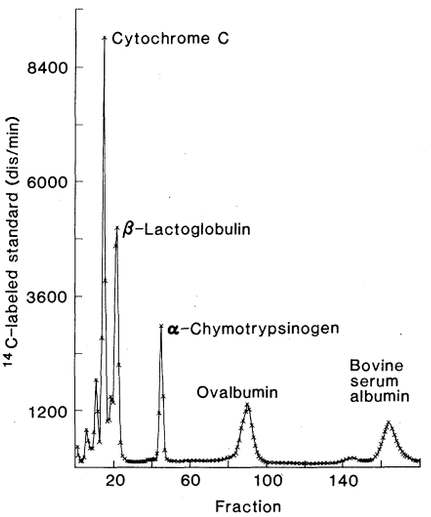

Fig. 6. Profile of the amount of label from continuous elution of ^{14}C-labeled protein standards from 15 percent SDS-PAGE. Fraction size, 50 μl; buffer flow rate, 15 μl/min. See (*17*) for details.

Electrochemical detection has been substituted for fluorescent detection and shown to be useful into the mid-femtomole (about 500) range with *tert*-butylthiol–OPA derivatives (*21*). Other studies with precolumn OPA derivation have shown that hydrolysis plus analysis can be performed in less than 3 hours, that analyses can be carried out after chloramine T oxidation to allow analysis of imino acids (*22*), and that biological amines present in physiological fluids can be analyzed (*23*). The instability of the thiol–OPA–amino acid product requires that the derivatization be carried out under controlled conditions of time and temperature and that the subsequent HPLC analyses be performed as quickly as possible. A simplified procedure used in our laboratories has been to carry out the derivatization in the HPLC unit where the sample is being analyzed; that is, the sample is "sandwiched" between two equal portions of the OPA reagent in the HPLC injection loop immediately before the mixture is subjected to chromatography.

Amino acid analysis by means of a precolumn reaction with 5-dimethylaminonaphthalene-1-sulfonyl chloride (dansyl; DNS-Cl) has also been used successfully (Fig. 7B). In this method, the sample reacts with an excess of DNS-Cl (*24*), excess reagent is quenched by adding a primary amine, and the DNS–amino acids are separated by RPC and detected by their fluorescence. The two main advantages of the DNS-Cl method are the stability of the DNS–amino acids and their utility as sequencing reagents. Because the DNS derivatives are stable to

mild acid hydrolysis (6*N* HCl, 2 to 6 hours), a peptide's amino terminus can be labeled and the labeled amino acid identified after a short hydrolysis. This technique enables rapid determination of peptide homogeneity and concentration after peptide mapping.

The newest derivatization reagent is phenylisothiocyanate (PITC) (Fig. 7C). This is the same reagent used in the sequencing of proteins by the Edman degradation, and the phenylthiocarbamyl amino acids it produces are analyzed with the same HPLC systems used for the analysis of phenylthiohydantoin (PTH) amino acids produced in sequencing. Even though detection is by ultraviolet absorbance, sensitivity with this method is in the range of 1 to 5 pmol (*25*).

Protein Microsequencing

Almost 30 years after its introduction, the Edman degradation continues to be the most widely used method for the direct determination of the primary structure of proteins and peptides. This chemical process is a repetitive reaction in which the α-amino group of the polypeptide chain is coupled to PITC under alkaline conditions, and after removal of excess coupling reagent by appropriate solvent extractions the terminal amino acid is cleaved from the remainder of the sample with acid. The resulting amino acid derivative (an anilinothiazolinone) is converted to the more stable PTH amino acid, which is analyzed by HPLC. Repeated degradation cycles on the remaining sample generate its amino acid sequence. Although Edman's method was originally designed for analysis of several hundred nanomoles of protein, it has shown remarkable versatility and with modern instrumentation can be used with as little as 5 to 50 pmol of sample. It is practiced in both manual and automated forms, and although modifications in both technique and equipment have been made in the last few years, the Edman chemistry remains largely unchanged.

Two signal developments have made possible much of the increase in the performance, particularly the sensitivity, of the Edman chemistry during the past few years. The first was the emergence of HPLC analysis of PTH amino acids (*26*). This rapid, sensitive, quantitative method is capable of resolving the PTH's of all the common amino acids with a single injection, and it has all but replaced the many previously used techniques that individually were capable of identifying only a few of the amino acids. The second was the discovery that proteins and peptides could be held firmly on the glass surfaces of the reaction vessels used in the Edman degradation when mixed with a polymeric quaternary ammonium salt, Polybrene (*27*). The retention of sample during the repeated cycles of chemistry—particularly during the solvent extractions used to remove excess reagents, reaction by-products, and anilinothiazolinones—is the key to all the Edman methods and is especially crucial to microscale work where even small losses of sample can leave the investigator with too little sample to continue the degradation.

Manual sequencing methods patterned after Edman's early work are still practiced and are especially useful for preliminary screening of large numbers of small peptides to assess their purity and suit-

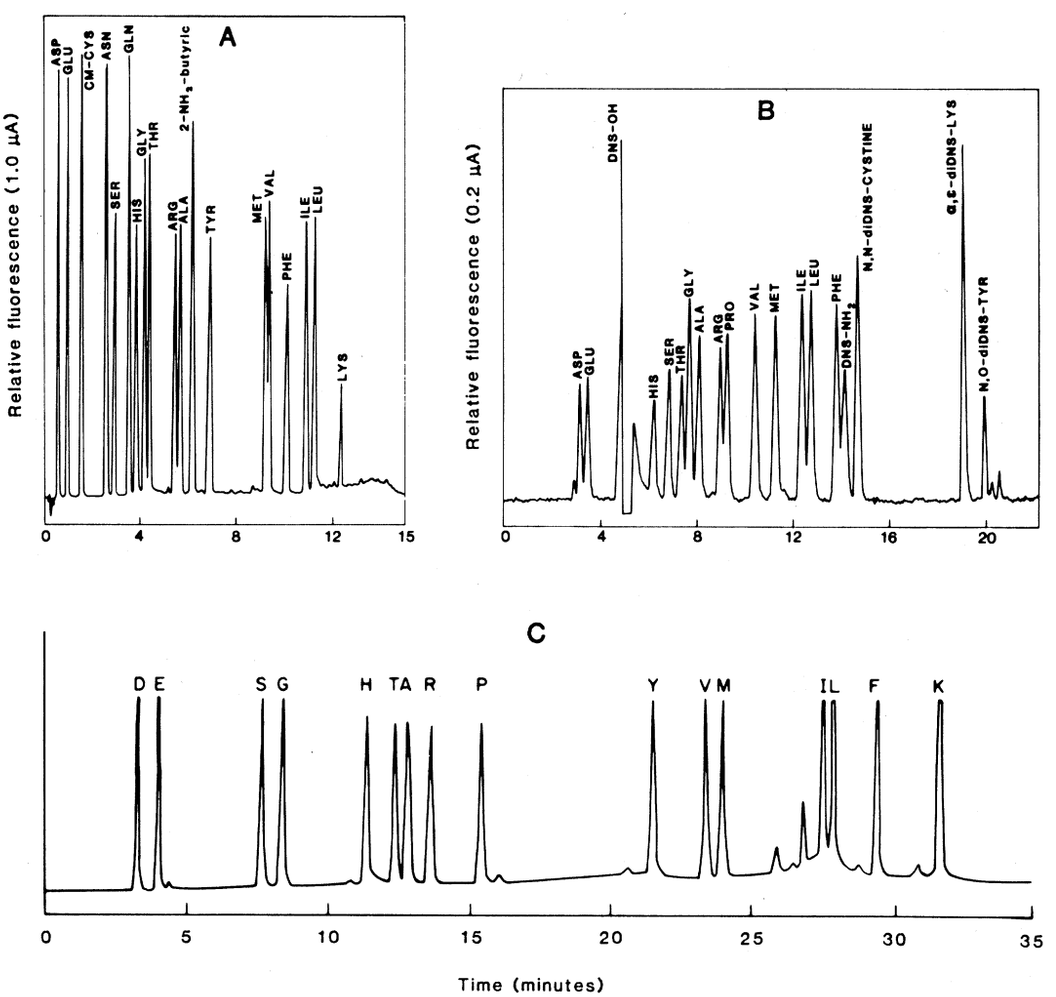

Fig. 7. Separation of amino acid derivatives arising from the precolumn reaction of amino acid standard solutions with OPA (A), DNS-Cl (B), and PITC (C). The analyses were performed on RPC supports, with the derivatives' fluorescence (A and B) or absorbance (C) used for detection. Separations were performed essentially as described for OPA (*23*) and DNS-Cl (*24*). The chromatogram illustrated for PITC was taken from (*25*). A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methiomine; P, proline; R, arginine; T, threonine; V, valine; Y, tyrosine.

ability for more extensive sequence analysis. They are relatively inexpensive, although labor intensive, and can be used with Polybrene to sequence samples of small peptides in the nanomole range (*28*). Intensely colored isothiocyanates such as the dimethylaminoazobenzene derivative can be used to enhance the detectability of the amino acids cleaved from the peptides (*29*).

In 1967, Edman and Begg (*30*) described the first automated instrument capable of performing the Edman degradation on proteins. Their protein sequencer, designed around a spinning cup reaction vessel, quickly became a standard tool for protein structural analysis. Even after the introduction of this instrument, however, the major limitations of Edman's technique were the large amounts of sample required (ten to several hundred nanomoles) and difficulty in sequencing short (<40 amino acids) peptides. The limitations were mainly due to sample loss during repeated cycles of chemistry, side reactions promoted by contaminants in the Edman reagents that interfered with the chemical reactions, and contamination of the PTH's that obscured their analysis. The introduction of Polybrene, more attention to the purification of reagents and the maintenance of their integrity while they are stored in the sequencer, and improved mechanical design of the spinning cup assembly and its associated valving and vacuum system have resulted in a great improvement in the sensitivity of the spinning cup technology (*31*). Several investigators have reported obtaining sequence information using just under 1 nmol of sample with this type of instrument, although most efforts to work below this sample size on a routine basis have been hindered by the large size of the spinning cup assembly and the inefficiency of the solvent extraction mechanism. Together, these factors usually give unacceptable backgrounds during HPLC of the PTH's which interfere with the analysis of samples of less than 1 nmol.

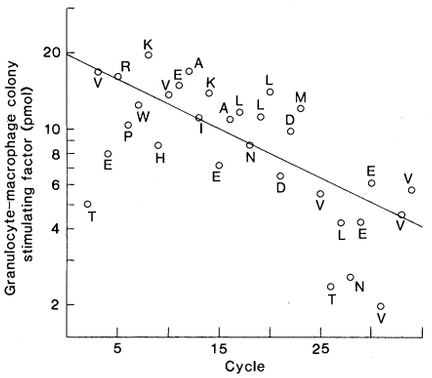

Fig. 8. PTH amino acid yields from gas-phase sequencer analysis of 20 pmol of granulocyte-macrophage colony stimulating factor. Yields were normalized to 100 percent injection onto the HPLC. The repetitive yield, calculated by a linear least-squares fit of all values except threonine, was 95.5 percent. The *y*-intercept (initial yield) from this calculation was 19.9 pmol [see (*39*) for details]. Amino acid notation as in the legend to Fig. 7, with N, asparagine and W, tryptophan.

To minimize the physical loss of peptides from the reaction vessel, Laursen developed the solid-phase sequencer in which the sample was chemically attached to a solid support such as glass or polystyrene beads before starting the Edman degradation (*32*). This technique was theoretically attractive because it eliminated the mechanical complexity of the spinning cup and provided a simple means of miniaturizing the system (the reaction chamber consists of a simple column through which the reagents and solvents flow). However, despite successful application by those skilled in the somewhat complicated and often inefficient chemical attachment methods, this method has not proved to be as generally useful as the spinning cup (liquid-phase) method.

The newest type of Edman sequencer combines the miniaturization and efficient chemical flow system of the column system employed in the solid-phase sequencer with the ease of sample application of the spinning cup sequencer. In this gas-phase sequencer, the sample is embedded in a film of Polybrene on the surface of a thin disk of glass microfiber filter mounted inside a small glass column (*33*). Because the sample is not covalently attached to the support, the coupling base and cleavage acid must be delivered as vapors (carried by a stream of argon) which readily diffuse into the Polybrene film to effect the Edman chemistry. Solvent extraction of the film is very efficient (chemical consumption is only 10 to 20 percent that of the spinning cup sequencer), and the PTH's produced in the sequencer are only minimally contaminated with compounds that interfere with their analysis. The gas-phase sequencer requires as little as 5 to 10 pmol of many proteins and 5 to 50 pmol of many peptides for useful sequence analysis (Fig. 8), with sequence analysis requiring less than 100 pmol being routine (Table 3). Moreover, the sequencer conveniently handles a variety of samples—including glycoproteins, integral membrane proteins, short hydrophobic peptides, large proteins (>100,000 daltons), and peptides covalently attached to solid supports—and is suitable for samples prepared by either HPLC or PAGE. Its sensitivity is currently limited by the sensitivity of the HPLC-based PTH analysis, but if HPLC column technology improves or if the PITC can be replaced with an analog containing a more easily detected chromophore or flurophore (*34*), then the gas-phase sequencer should be usable for analysis of samples of less than 1 pmol.

Mass spectrometry is emerging as an alternative to the Edman degradation for sequencing certain types of proteins and peptides. Its initial successes were in analyses of small peptides, especially those containing unusual amino acids or those with modified terminal or side chain groups not handled well by Edman methods (*35*). Electron-impact ionization after formation of *N*-acetyl-*N*,*O*-permethyl derivatives of the peptides to increase their volatility was used in these early studies, and as little as 2 to 10 nmol of sample was required. A new form of ionization, fast-atom bombardment, has expanded the capabilities of mass spectrometry in sequence analysis (*36*). Fast-atom bombardment also requires only a few nanomoles (1 to 10) of sample, does not require prior derivatization of the sample for it to achieve volatility, and can handle peptides with molecular

Table 3. Selected examples of proteins sequenced by means of the gas-phase sequencer.

Protein or peptide	Reference
Human platelet-derived growth factor	(*48*)
Porcine dynorphin	(*49*)
Calf acetylcholine receptor subunits	(*50*)
Rat transforming growth factor type 1	(*51*)
α and β chains of HLA-DR1 and HLA-DR2 antigens	(*52*)
Human melanoma-associated antigen p97	(*53*)
Mouse granulocyte-macrophage colony stimulating factor	(*39*)
Human growth hormone-releasing factor	(*54*)
Saccharomyces kluyveri mating pheromone	(*55*)
Rat natriuretic peptides	(*56*)
Human epidermal growth factor receptor	(*57*)
Human interleukin-2	(*58*)

weights up to a few thousand. It is particularly well suited to rapid confirmation of protein sequence deduced from DNA sequencing (*37*) and for use in conjunction with Edman degradation for sequencing unknown proteins (*38*).

References

1. K. J. Wilson, E. Van Wieringen, S. Klauser, M. W. Berchtold, *J. Chromatogr.* **237**, 407 (1982); M. J. O'Hare, M. W. Capp, E. C. Nice, N. H. C. Cooke, B. G. Archer, *Anal. Biochem.* **126**, 17 (1982).
2. K. J. Wilson and M. Doyle, in preparation.
3. M. W. Berchtold, K. J. Wilson, C. W. Heizmann, *Biochemistry* **21**, 6552 (1982).
4. H. J. Friesen, S. Stein, M. Evinger, P. C. Familletti, J. Moschera, J. Meienhofer, J. Shiveley, S. Pestka, *Arch. Biochem. Biophys.* **206**, 432 (1981).
5. M. Flashner, H. Ramsden, L. J. Crane, *Anal. Biochem.* **135**, 340 (1983); Y. Kato, K. Nakamura, T. Hashimoto, *J. Chromatogr.* **256**, 143 (1983); M. A. Rounds and F. E. Regnier, *ibid.* **283**, 37 (1984).
6. M. J. Bjorn and K. J. Wilson, in preparation.
7. J. L. Meek and L. Rossetti, *J. Chromatogr.* **211**, 15 (1981).
8. V. J. Lewis, R. N. P. Singh, L. J. Lewis, B. K. Seavey, Y. N. Sinha, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 385 (1984).
9. V. K. Dua and C. A. Bush, *Anal. Biochem.* **137**, 33 (1984); D. Tetaert, N. Takahashi, G. W. Putnam, *ibid.* **123**, 430 (1982); H. Morehead, P. McKay, R. Wetzel, *ibid.* **126**, 29 (1982); K. J. Wilson *et al.*, in preparation.
10. T. Sasagawa, K. Titani, K. A. Walsh, *Anal. Biochem.* **134**, 224 (1983); J. W. Marsh, A. Nahum, D. F. Steiner, *Int. J. Peptide Protein Res.* **22**, 39 (1983); A. L. Frelinger III and J. E. Zull. *J. Biol. Chem.* **259**, 5507 (1984).
11. L. Ornstein, *Ann. N.Y. Acad. Sci.* **121**, 321 (1964).
12. U. K. Laemmli, *Nature (London)* **227**, 680 (1970).
13. N. Catsimpoolas, *Sep. Sci.* **5**, 523 (1970).
14. P. H. O'Farrell, *J. Biol. Chem.* **250**, 4007 (1975).
15. J. Silver and L. W. Hood, *Anal. Biochem.* **67**, 392 (1975).
16. S. Spiker and I. Isenberg, *Methods Enzymol.* **91**, 214 (1983); M. W. Hunkapiller, E. Lujan, F. Ostrander, L. E. Hood, *ibid.*, p. 227.
17. M. W. Hunkapiller, E. Lujan, L. E. Hood, in preparation.
18. B. Keil, in *Methods in Protein Sequence Analysis*, M. Elzinga, Ed. (Humana, Clifton, N.J., 1982), pp. 291–304; M. A. Hermodson, in *ibid.*, pp. 313–324.
19. D. H. Spackman, W. H. Stein, S. Moore, *Anal. Chem.* **30**, 1190 (1958).
20. W. S. Gardiner and W. H. Miller, *Anal. Biochem.* **101**, 61 (1980).
21. L. A. Allison, G. S. Mayer, R. E. Shoup, *Anal. Chem.* **56**, 1089 (1984).
22. J. D. H. Cooper, M. T. Lewis, D. G. Turnell, *J. Chromatogr.* **285**, 484 and 490 (1984).
23. B. N. Jones and J. P. Gilligan, *ibid.* **266**, 471 (1983).
24. C. De Jong, G. J. Hughes, E. Van Wieringen, K. J. Wilson, *ibid.* **241**, 345 (1982); Y. Tapuhi, N. Miller, B. L. Karger, *ibid.* **205**, 325 (1981).
25. R. L. Heinrikson and S. C. Meredith, *Anal. Biochem.* **136**, 65 (1984).
26. C. L. Zimmerman, E. Appella, J. J. Pisano, *ibid.* **77**, 569 (1977).
27. G. E. Tarr, J. F. Beecher, M. Bell, D. J. McKean, *ibid.* **84**, 622 (1978).
28. G. E. Tarr, in *Methods in Protein Sequence Analysis*, M. Elzinga, Ed. (Humana, Clifton, N.J., 1982), pp. 223–232.
29. J. Y. Chang, *Methods Enzymol.* **91**, 455 (1983).
30. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
31. M. W. Hunkapiller and L. E. Hood, *Biochemistry* **17**, 2124 (1978).
32. R. A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
33. R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
34. J. J. L'Italien and S. B. H. Kent, *J. Chromatogr.* **283**, 149 (1984).
35. H. R. Morris, *Philos. Trans. R. Soc. London Ser. A* **293**, 39 (1979).
36. M. Barber, R. S. Bordoli, R. D. Sedgwick, A. N. Tyler, *Chem. Commun.* **101**, 325 (1981).
37. T. Takao, T. Hitouji, Y. Shimonishi, T. Tanable, S. Inouye, M. Inouye, *J. Biol. Chem.* **259**, 6105 (1984).
38. T. S. Anantha Samy *et al.*, *ibid.* **258**, 183 (1983).
39. N. M. Gough *et al.*, *Nature (London)* **309**, 763 (1984); L. Sparrow, D. Metcalf, M. W. Hunkapiller, L. E. Hood, A. W. Burgess, in preparation.
40. R. Schuster, *Anal. Chem.* **52**, 617 (1980).
41. M. K. Radjai and R. T. Hatch, *J. Chromatogr.* **196**, 319 (1980).
42. G. J. Hughes, K. H. Winterhalter, E. Boller, K. J. Wilson, *ibid.* **235**, 417 (1982).
43. G. L. Lookhart, B. L. Jones, D. B. Cooper, S. B. Hall, *J. Biochem. Biophys. Methods* **7**, 15 (1982).
44. Y. Tapuhi, D. E. Schmidt, W. Lindner, B. L. Karger, *J. Chromatogr.* **205**, 325 (1981).
45. N. B. Levina and I. V. Nazimov, *ibid.* **286**, 207 (1984).
46. J.-Y. Chang, R. Knecht, D. G. Braun, *Biochem. J.* **203**, 803 (1982).
47. S. Einarsson, B. Josefsson, S. Lagerkvist, *J. Chromatogr.* **282**, 609 (1983).
48. R. F. Doolittle *et al.*, *Science* **221**, 275 (1983); M. D. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983).
49. W. Fischli, A. Goldstein, M. W. Hunkapiller, L. E. Hood, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5435 (1982).
50. B. M. Conti-Tronconi, C. M. Gotti, M. W. Hunkapiller, M. A Raftery, *Science* **218**, 1227 (1982).
51. H. Marquardt, M. W. Hunkapiller, L. E. Hood, G. J. Todaro, *ibid.* **223**, 1079 (1984).
52. L. E. Walker, R. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, R. A. Reisfeld, *Biochemistry* **22**, 185 (1983).
53. J. P. Brown, R. M. Hewick, I. Hellstrom, K. E. Hellstrom, R. F. Doolittle, W. J. Dreyer, *Nature (London)* **296**, 171 (1982).
54. F. S. Esch *et al.*, *Biochem. Biophys. Res. Commun.* **109**, 152 (1982).
55. A Sakurai *et al.*, *FEBS Lett.* **166**, 339 (1984).
56. M. G. Currie *et al.*, *Science* **223**, 67 (1984).
57. A. Ullrich *et al.*, *Nature (London)* **309**, 418 (1984).
58. R. Robb, R. Kutny, M. Tanico, H. Morris, W. DeGrado, C. Chowdhry, *Biochem. Biophys. Res. Commun.* **116**, 1049 (1983).