Journal of Pharmaceutical & Biomedical Analysis
Vol. 2, No. 2, pp. 281–287, 1984
Printed in Great Britain

0731–7085/84 $3.00 + 0.00

Characterization of small amounts of peptides and proteins*

JANUSZ WIDEMAN†, YU-CHING E. PAN, HANS-WERNER LAHM, MAY CHANG and STANLEY STEIN

Biopolymer Research Department, Roche Research Center, Hoffmann–La Roche Inc., Nutley, NJ 07110, USA

Abstract: The use of microchemical methods for the characterization of both natural and recombinant proteins of biomedical importance is discussed. The methods include gel electrophoresis, high-performance liquid chromatography, protein fragmentation, amino acid analysis and automated Edman degradation. Each procedure is applicable at the picomole level.

Keywords: *Microsequencing; amino acid analysis; reversed-phase chromatography; interferon; interleukin; fluorescent detection.*

Introduction

While recombinant DNA technology has had a great impact on protein research, its optimal application has been aided by the development and use of micromethods for the isolation and characterization of proteins. Since many of the peptides and proteins of current interest, such as interferons and interleukins, are highly active biologically, only small amounts are available from natural sources. With micromethods amino acid sequence data for a new protein can be readily obtained; thus the need for purification of that protein on a massive scale is obviated. This allows the preparation of specific oligodeoxynucleotide probes that hasten the successful cloning of the particular protein. Once the protein is expressed in recombinant organisms, these microtechniques enable the structure of the protein product to be confirmed. In addition, since proteins are often processed post-translationally to their biologically active forms, gene analysis cannot substitute for direct protein analysis. In this article the methodology set up in the authors' laboratory is reviewed and its utility is demonstrated through specific examples concerning both natural and recombinant proteins.

Experimental and Results

The first step in the determination of the primary structure of a protein is to evaluate

* Presented at the Symposium on Liquid Chromatography in the Biomedical Sciences, June 1984, Ronneby, Sweden.

† To whom correspondence should be addressed.

its homogeneity. With certain exceptions, a mixture of proteins cannot readily be analysed. The situation is worse when the biologically active protein is a minor (chemically undetectable) component mixed with a pure biologically irrelevant protein that has almost identical physico-chemical properties. The preliminary evaluation of protein homogeneity is by polyacrylamide gel electrophoresis according to the method of Laemmli [1]. The proteins are visualized by silver staining, which permits about 10 ng of protein to be detected [2]. A single stained band on the gel indicates homogeneity. If possible, it should also be shown that the biological activity ascribed to the protein is associated with that band. This is illustrated in Fig. 1 for human fibroblast interferon [3]. Gel electrophoresis also provides an estimate of the molecular weight of the protein; it is an invaluable technique for the protein laboratory.

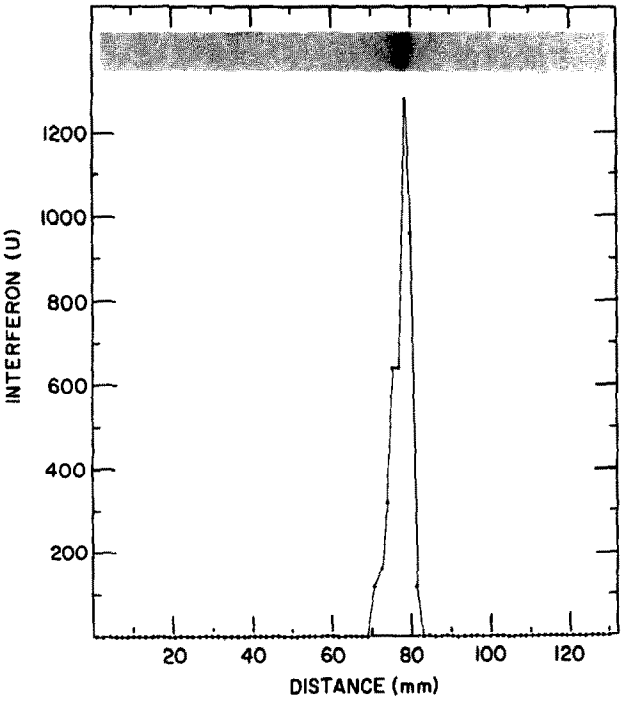

Figure 1
Sodium dodecylsulphate/polyacrylamide gel electrophoresis of homogeneous human fibroblast interferon. A 5–15% polyacrylamide gradient slab gel was used; the gel was stained with Coomassie blue. Antiviral activity was determined on extracts from an unstained track. The points plotted as zero activity on the figure represent <40 units of interferon. The titres of interferon units shown represent units/ml in the 0.1-ml portions that were eluted from the homogenized gel fraction (from [3]).

An additional, and often more reliable approach for evaluating homogeneity is by high-performance liquid chromatography (HPLC). Generally, a C_8 or C_{18} reversed-phase column (5-μm particle size, 30 nm porosity, 50 × 4.6 mm i.d.) is eluted with a gradient of increasing acetonitrile in 0.1% trifluoroacetic acid [4] (approximately pH 2) at a flow rate of about 1 ml/min. This technique provides high resolving power and is usually applicable to polypeptides with molecular weights up to 50 kilodaltons. Reversed-phase HPLC can also be used as a preparative method for achieving homogeneity since a high recovery of protein may be obtained.

Proteins eluted from the HPLC column may be monitored in several ways. In one configuration, the authors use a 3-component detection system (Fig. 2). The first component involves UV-detection at 206 nm, which permits the relatively sensitive detection of peptides and proteins. One drawback is that this method is not specific for the peptide bond alone, since it will also detect many organic compounds. HPLC-grade acetonitrile, trifluoroacetic acid and water are transparent at this wavelength.

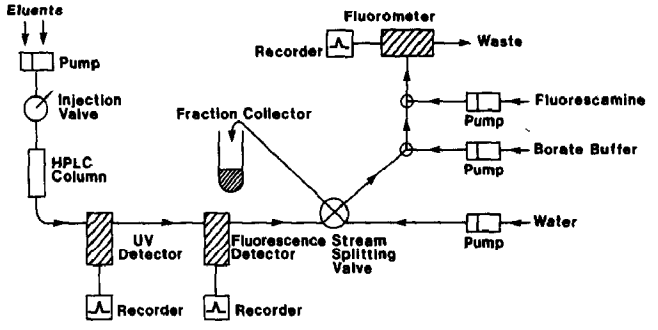

Figure 2
Schematic diagram of the HPLC apparatus, showing the three monitoring systems serially connected. The arrows denote the direction of liquid flow. In the fluorescamine amino acid analyser, the column effluent is mixed directly with the borate buffer and fluorescamine.

The utility of the next component, the fluorescence detector, has been known for many years; however, it is seldom used in other laboratories for peptide and protein monitoring, as compared with UV-absorption. The native fluorescence of tryptophan residues is specifically monitored by excitation at 288 nm with a 340 nm cut-off filter on the emission side. Detection of tryptophan is important for the following reasons. First, this amino acid is difficult to determine by amino acid analysis owing to its lability during acid hydrolysis. Furthermore, tryptophan occurs at a low frequency in proteins and is represented by only one codon; both these circumstances are ideal for providing a highly specific oligodeoxynucleotide probe.

The third detection component is based on fluorescamine-labelling [5]. The column effluent, after passage through the UV-detector and fluorometer, is split into two streams with a valve (Fig. 2). One stream, which usually represents about 3% of the column effluent, is adjusted to pH 9 with borate buffer and then reacted with fluorescamine in continuous-flow manner. Fluorescamine forms a highly fluorescent derivative with ϵ-amino groups of lysine residues and with the α-amino group at the peptide terminal [6]. The fluorescence is then measured using a filter fluorimeter (390 nm excitation and 475 nm emission maxima) and the stream then goes to waste. The remainder of the column effluent (*ca* 97%) is collected, and the protein is then recovered in unmodified form.

The triple detection system described above gives a wealth of information about each component in the chromatogram, together with additional criteria for evaluating homogeneity. The use of a photodiode array detector, which can produce spectral data on each eluted component as well as providing detection of tryptophan residues, is being investigated.

Figure 3 represents the chromatographic analysis of recombinant leukocyte interferon at an intermediate stage of purification, as monitored simultaneously by UV-detection at

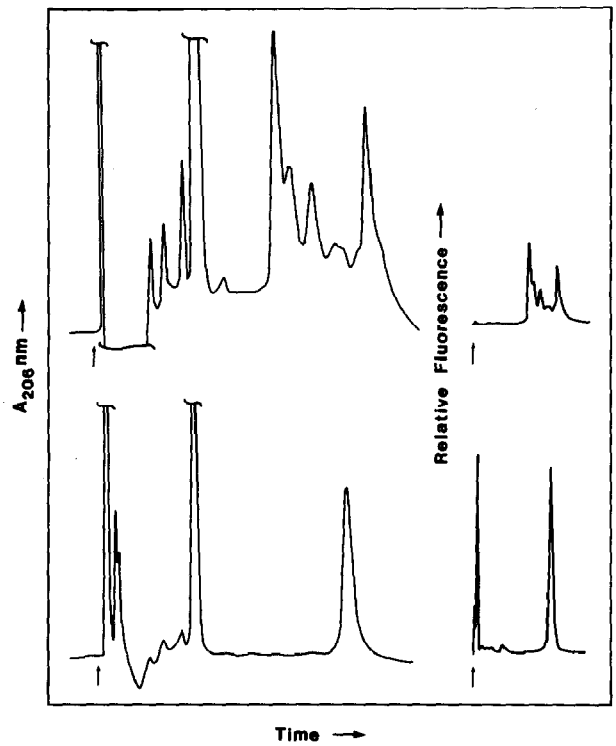

Figure 3
HPLC analysis of recombinant human leukocyte interferon. In each panel the left chromatogram represents UV-absorption at 206 nm (0.1 AUFS) and the right chromatogram represents fluorescamine monitoring, simultaneously recorded during the same HPLC run. Interferon, purified on an antibody-affinity column, was analysed on a reversed-phase column using a gradient of acetonitrile in aqueous trifluoroacetic acid (upper panel). After chemical reduction with 2-mercaptoethanol in guanidine hydrochloride, the mixture of forms was converted to a single, completely reduced protein (lower panel).

206 nm and fluorescamine-labelling. Both detection systems show essentially the same pattern, which indicates extensive heterogeneity (upper panel). It can be seen that the early peaks in the UV trace are not present in the fluorescamine trace; this observation indicates that they represent extraneous solutes in the sample, rather than proteins. Upon chemical reduction, the protein components merge into a single peak (lower panel). Incomplete or improper disulphide bond formation in this recombinant protein is responsible for the phenomenon. This technique is one of the analytical tools used to ensure that only the proper conformational form is present in the final purified preparation [7].

The amino acid analyser is the fundamental instrument in any protein analysis laboratory, and is useful in numerous aspects of protein characterization. This analyser used in the authors' laboratory is based on post-column reaction with fluorescamine [8], and is considered to have the best combination of features required. It is versatile, reliable and has sensitivity in the picomole range. An example of its application is for monitoring the progress of cyanogen bromide cleavage. This reagent causes peptide bond hydrolysis at the carboxy-side of methionine residues and in the process methionine is converted to homoserine. As seen in Fig. 4, the cleavage of natural human interleukin-2 (IL-2) was essentially complete after 24 h [9]. Only 100 ng of protein was consumed for

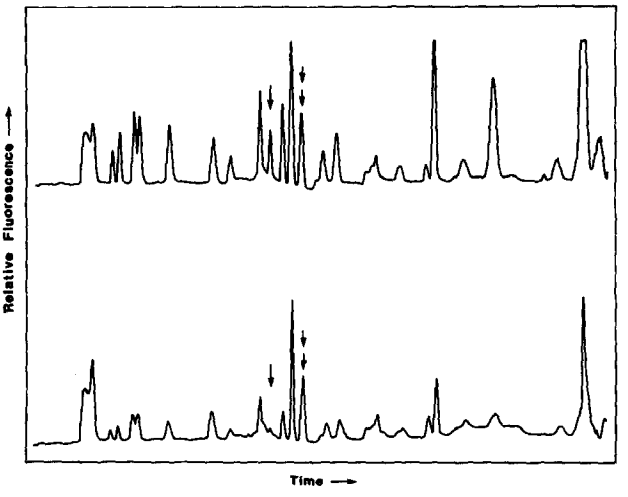

Figure 4
Fluorescamine amino acid analysis. A sample of natural human interleukin-2 was hydrolysed in acid to its constituent amino acids, which were resolved on a sulphonated polystyrene HPLC column (upper panel). Another sample of interleukin-2 was digested with cyanogen bromide in 70% v/v formic acid. After 24 h an aliquot of this digest (about 100 ng) was removed, dried, acid-hydrolysed and analysed (lower panel). The single arrow denotes the position of methionine, the disappearance of which correlates with the extent of cleavage. The double arrow denotes the position of the internal standard, norleucine.

analysis of the digest. Amino acid analysis is used extensively in order to minimize any possible errors in interpretation of data.

Preliminary characterization of a protein, as described above, is followed by sequencing. Before the actual sequencing, proteins are usually reduced and carboxymethylated using a procedure recently developed in the authors' laboratory [10]. This procedure achieves three goals: (i) it converts cysteine to a derivative that is more readily identified during HPLC analysis (see later); (ii) it serves to denature the protein, and thus makes it more amenable to further analysis; and (iii) it includes another separation by reversed-phase HPLC, which allows re-evaluation of the homogeneity of the derivatized proteins. A radiolabel may be introduced via the alkylating reagent; this provides a convenient method for following the protein through other analytical procedures. As described earlier, the completeness of the carboxymethylation reaction can be followed by amino acid analysis.

Protein sequencing is accomplished by automated Edman degradation on a gas-phase instrument [11] from Applied Biosystems (Foster, CA). The amino acids are released from the amino-terminus sequentially during each cycle of the degradation, converted to their respective phenylthiohydantoin (PTH) derivatives and collected in individual tubes. The derivatives are identified by reversed-phase HPLC with UV-absorption at 254 nm. PTH-amino acid analysis is illustrated in Fig. 5 for the first three cycles of degradation of recombinant, human interleukin-2. The experimental details are given in the legend to Fig. 5. Two sequences were found. That is, in each cycle there was a major PTH-derivative peak (about 90%) and a minor peak (about 10%). The minor sequence corresponded to the natural protein (Ala-Pro-Thr . . .) whereas the major sequence revealed the presence of an additional methionine at the amino-terminus (Met-Ala-Pro . . .). In *E. coli* protein biosynthesis, all proteins are initiated with a leader sequence terminated by methionine, then followed by the mature protein sequence. For reasons

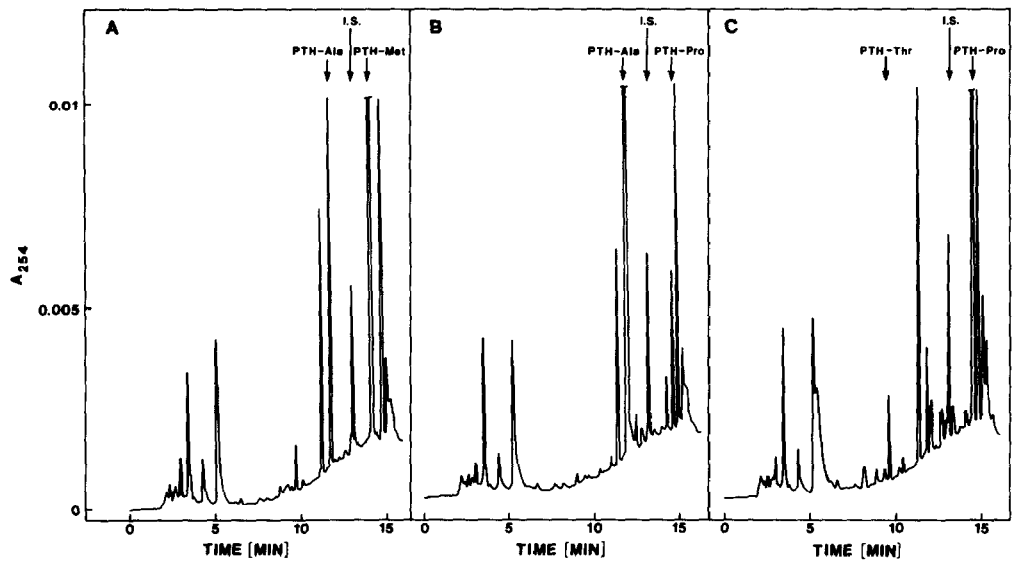

Figure 5
HPLC separation from an *N*-terminal amino acid sequence analysis of 800 pmol of recombinant human interleukin-2. The first three cycles are shown. Aliquots, representing half of each sample, were injected and separated on a 250 × 4.6 mm i.d. Ultrasphere ODS-column (5 μm) at 50°C. Flow rate: 1 ml/min. Solvent A: 20 mM trifluoroacetic acid and 1 mM acetic acid in 10% v/v acetonitrile, pH 5.1. Solvent B: 20 mM trifluoroacetic acid in 75% v/v acetonitrile, pH 3.7. The gradient profile adopted was: $t = 0$, %B = 10%; 7 min, 65%; 10 min, 65%; 14 min, 10%. The expected PTH-amino acid derivatives and the internal standard (IS; PTH-D, L-α-aminobutyric acid) are marked by arrows. Major artefact peaks are *N,N*-dimethyl-*N'*-phenylthiourea (11.25 min) and *N,N'*-diphenylthiourea (14.75 min).

that are not yet clear, the intracellular processes for removing the leader sequence methionine do not always function completely, as previously observed with recombinant proteins [12].

It is possible to obtain a partial amino-terminal sequence using 100 picomoles of peptide. With a few hundred picomoles or more, only the first 20–40 residues are usually elucidated. Some proteins have a 'blocked' amino terminus and then no sequence data is obtained. For these last two reasons, it is necessary to fragment the protein and to purify each fragment by HPLC for sequence analysis. About 1.5 nmole (20 μg) of natural human interleukin-2 [9] was available for sequencing, but it was found to have a blocked amino-terminus by the use of a few hundred picomoles for analysis. The remainder was reduced, alkylated and fragmented by cyanogen bromide (see above). The fragments were separated by reversed-phase HPLC (Fig. 6). Amino acid analysis and sequencing confirmed the structure of the protein, as previously predicted by the sequence of the gene [13].

With the advent of recombinant DNA technology, the structures of proteins, such as interleukin-2, are often first determined by molecular biologists who identify, clone and then sequence the gene. Even when the protein is isolated first, it is no longer necessary to sequence the protein extensively in order to elucidate the complete primary structure. A cDNA probe is synthesized, based on some protein sequence data. This probe is then used to screen recombinant organisms for those which specifically contain the DNA encoding that protein. Sequencing the DNA for deducing the complete primary structure of the protein is rapid and accurate. Thus, the advances in protein microchemistry and recombinant DNA technology are mutually supportive.

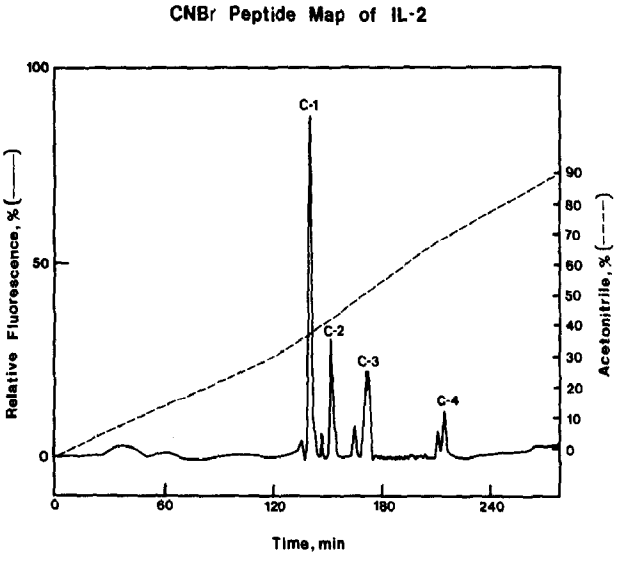

Figure 6
Cyanogen bromide peptides derived from human interleukin-2. Carboxymethylated IL-2 (900 pmol) was cleaved with cyanogen bromide and the resultant peptides were separated by HPLC on a 150 × 4.6 mm i.d. Supelcosil LC-18-DB column (3 μm). Elution was carried out using a linear gradient of acetonitrile (– – –) in 0.1% trifluoroacetic acid at a flow rate of 20 ml/h. Five per cent of the column effluent was diverted to the fluorescamine detection system. Cyanogen bromide peptides are numbered according to their elution position (from [9]).

References

[1] U. K. Laemmli, *Nature (Lond.)* **227**, 680 (1970).
[2] J. H. Morrissey, *Anal. Biochem.* **117**, 307 (1981).
[3] S. Stein, C. Kenny, H.-J. Friesen, J. Shively, U. Del Valle and S. Pestka, *Proc. Natl. Acad. Sci. USA* **77**, 5716–5719 (1980).
[4] H. P. J. Bennett, C. A. Browne, D. Goltzman and S. Solomon, in *Proc. of the 6th American Peptide Symp.* (E. Gross and J. Meienhofer, Eds.), pp. 121–124. Pierce Chem. Co., Rockford, IL (1979).
[5] P. Bohlen, S. Stein, J. Stone and S. Udenfriend, *Anal. Biochem.* **67**, 438–445 (1975).
[6] S. Udenfriend, S. Stein, P. Bohlen, W. Dairman, W. Leimgruber and M. Weigele, *Science* **178**, 871–872 (1972).
[7] S. J. Tarnowski and R. A. Liptak, in *Advances in Biotechnological Processes*, Vol. 12 (A. Mizrachi and A. L. Wetzel, Eds). Alan R. Liss, New York (1983).
[8] S. Stein and L. Brink, in *Methods in Enzymology* (S. Pestka, Ed.), Vol. 79, pp. 20–25. Academic Press, New York (1981).
[9] A. S. Stern, Y.-C. E. Pan, D. L. Urdal, D. Y. Mochizuki, S. DeChiara, R. Blacher, J. Wideman and S. Gillis, *Proc. Natl. Acad. Sci. USA* **81**, 871–875 (1983).
[10] Y.-C. E. Pan, J. Wideman, R. Blacher, M. Chang and S. Stein, *J. Chromatogr.* (in press).
[11] R. M. Hewick, M. W. Hunkapiller and W. J. Dryer, *J. Biol. Chem.* **256**, 7990–7997 (1981).
[12] T. Staehelin, D. S. Hobbs, H.-F. Kung and S. Pestka, *J. Biol. Chem.* **256**, 9750–9754 (1981).
[13] T. Taniguchi, H. Matsui, T. Fugita, C. Takaoka, N. Kashima, R. Yoshimoto and J. Hamara, *Nature (Lond.)* **302**, 305–310 (1983).

[Received for review 20 June 1984]