

High-sensitive protein sequence analysis

Brigitte Wittmann-Liebold

Max-Planck-Institut für Molekulare Genetik Berlin
D-1000 Berlin 33, Ihnestr. 73, Germany

Abstract - The primary structure determination of proteins and peptides has made rapid progress in the recent years and became an increasingly important tool of basic and applied research in various areas of biochemistry, molecular biology, medicine and biotechnology. Direct amino acid sequence analysis is frequently applied in combination with mass spectrometry or DNA-sequencing of the protein's gene. Summarized are the essential techniques which facilitate protein purification and sequence analysis in the picomole scale, such as HPLC chromatography, blotting for direct amino acid analysis and N-terminal sequencing. Modern sequencer design is outlined.

INTRODUCTION

Proteins consist of one or more linear chain(s) of polypeptides that are composed of 20 different amino acids and folded in an unique three-dimensional structure which is characteristic for each protein and enables the high specificity of the molecule. Proteins are the main constituents of the biomolecule mass in the cell; they comprise about 55% of its dry weight [1] and perform many essential functions. They catalyze all processes in the cell, carry out molecule transport, respond to antigens and constitute receptor sites; they are involved in processing, gene regulation and signal transduction; together with lipids and carbohydrates they build walls and laminar structures or compartments; they are main constituents of organelles, viruses and phages. Sequence analysis of peptides and proteins is the basis for any three-dimensional structure analysis and hence, for a deeper understanding of their role in the various important processes on a molecular level. The cytoplasm of cells contains more than 100,000 individual proteins, of diverse sizes, properties and amino acid compositions. The proteins may show modifications, such as phosphorylations, methylations, acetylations or can be associated with lipids and carbohydrates [2]. These many differences in the primary structure cause not only the high variety of proteins, but also reflect the difficulties in their structure analysis. Hence, approaches for their purification and structure determination must be adapted to each type of protein and to certain research topics. Each individual protein or polypeptide mixture may cause specific problems. It needs much experimental skill and know-how to select the optimal techniques in a certain case which is not easy and straightforward. This is one of the reasons why protein structure analysis is so challenging.

In this article techniques are discussed which are applicable to proteins and peptides available in only picomole quantities. Emphasis is laid on methods which are implemented in recent protein-chemical research studies in many fields in biochemistry, molecular biology, medicine and biotechnology. Most recently, these methods are applied to study differences in proteins of normal cell extracts and that of dis-

eases. Other topics are domain or subunit assignment in multi-enzyme complexes, studies on the folding states of proteins, on DNA-, RNA- and tRNA-protein interaction, determination of receptor sites, design of new enzymes or peptide hormones by genetic engineering, or the investigation of brain proteins and inhibitor binding to proteases. On the other hand, protein purifications in large scale are still necessary where physico-chemical investigations, e.g., 2D- and 3D-NMR studies or crystallization of proteins are approached. In these cases at least 50 to 100 mg of purified protein or peptide are requested in order to gain information on their secondary and tertiary structures. Hence, bigger scale isolation methods have to be applied for these studies.

PROTEIN AND PEPTIDE ISOLATION

HPLC methods

Conventional purification techniques, such as gel filtration, ion exchange chromatography, separation by affinity column chromatography or adsorption and ammonium sulfate precipitation are still very common. For small quantities however, in the pico- or femtomole range, all methods have to be adapted to this scale. This is not trivial since small amounts of proteins or peptides tend to stick to glass, column supports and any plastic material. They are frequently lost in dialysis tubes, on membranes, trapped in columns or disappear in lyophilization. The smaller the amounts available the more losses occur in the various purification steps.

For control of losses activity tests may be performed at the different isolation steps. However, purification under native conditions implies usually strong salts the remainder of which disturb at the later sequence analysis of the protein. In addition, the polypeptide chain may become N-terminally blocked, and this type of blockage increases drastically with decreasing sample amounts due to Schiff-base formation or reaction with cyanate in the presence of aldehydes, urea, radicals and other contaminants in the buffers. Many highly purified protein samples subjected to sequence analysis are blocked as a result of the isolation. Since this blockage is irreversible little or no N-terminal sequence information can be obtained. Therefore, the separation conditions applicable for a certain protein have to be carefully selected in order to arrive at a high percentage of amino-terminal "open structure". In reverse, naturally occurring blockage may be removed by the use of enzymes (see H. Hirano et al., this issue).

Most suitable for the separation of minute protein and peptide amounts is high performance liquid chromatography (HPLC) [3] and here especially chromatography on the reversed phase since it allows to avoid salts in the buffers; the chromatography is fast, has high resolving power and yields sharp separations. Salt containing samples may be desalted in the same chromatography step, and volatile polypeptide fractions are obtained suitable for concentration by vacuum centrifugation and direct application to sequence analysis. Minute protein and peptide amounts are separated by microbore HPLC in columns of 1 to 2 mm inner diameter at low flow rates (below 200 μ l) [see R.J. Simpson et al., this issue].

Studying the properties and sequences of protein components of various ribosomes for many years [4] we experienced that protein isolation of complex protein mixtures as well as their final purification in amounts suitable for sequence analysis caused lots of problems, due to low resolution and recoveries as result of their extreme isoelectric points as well as their hydrophobic interactions. Furthermore limitations in the sequence determinations derived from repetitive and monotonous sequence regions [5,6]. On the other hand, the characterization of ribosomal proteins of diverse organisms by sequence analysis was so challenging that we adopted reversed phase HPLC for the fast separation of these complex protein mixtures and the peptide fragments thereof [7,8].

However there are many proteins that cannot be purified by reversed phase HPLC. These are large proteins with molecular masses of more than 30 to 50 kDa or rather hydrophobic polypeptides, such as membrane proteins. In these cases HPLC size exclusion chromatography may be applicable, but usually the resolution of proteins that have similar molecular masses is very low by this method. Therefore conventional column chromatography or affinity column separations are in use. Alternatively, these proteins have to be isolated by genetic engineering and overproduction in bacteria.

Electrophoresis and blotting for microsequence analysis

One-dimensional SDS- [9] and two-dimensional polyacrylamide gels (urea/SDS [10] or IEF/SDS gels [11]) which resolve according to charge and mass have excellent resolving power for small protein quantities. These techniques are now frequently in use to blot proteins from the gels onto membranes that are suitable for direct microsequencing, e.g. onto polyvinylidene difluoride (PVDF) membranes [12] in good efficiency [13-15]. It became possible to employ Western blotting from high resolution 2D-polyacrylamide gels onto various glass fibers or PVDF membranes for direct microsequencing even for highly complex protein mixtures which contain up to 2,000 proteins deriving from unfractionated cell or tissue lysates and to perform a computerized analysis [16]. Recently, a preparative electrophoresis system of high performance has been designed which provides an alternative to the above mentioned methods [17]. It may be applied where the other isolation techniques fail or for rechromatography of unresolved HPLC fractions.

Purity check

Protein structure determinations require isolates of high purity. If this is not the case and the isolate contains more than one polypeptide chain or other contaminants (such as salts) ambiguous results in the amino acid composition as well as in the N-terminal sequence analysis will be obtained. Controls of the purity have to be made, by i) chromatography in at least two to three different separation systems, ii) by one-dimensional and two-dimensional polyacrylamide electrophoresis, respectively, or by isoelectric focussing. The gels may be performed in stamp size for nanogram protein amounts [18]; another new sensitive tool is capillary electrophoresis [19]. Although these tests are all very useful they may not be sufficient when similar sized or charged components are in the mixture. Therefore in addition the N-terminal sequence should be established by the Edman chemistry [20]. Since a contaminating peptide may be N-terminally blocked mass spectrometry is further applied. The mass ion(s) of the sample by time-of-flight or laser desorption mass spectrometry (see below) may be determined.

Only if the polypeptide is tested by various of these methods one may arrive at a final proof of purity. These techniques are of extreme importance in the pharmaceutical industry where the purity of synthetic peptides or proteins that are produced by overexpression have to be proven.

In summary, to decide whether an isolate is purified, takes much effort and time, and the needs for sophisticated equipment are high.

AMINO ACID ANALYSIS

Diverse amino acid analyzers are commercially available which allow a quantitative determination of the amino acid content of a protein. These are based on post-column derivatization by ninhydrin after separation of the unmodified amino acids, or alternatively, on pre-column derivatization and HPLC separation of the modified components, e.g. after ortho-phthaldialdehyde (OPA) reaction, derivatization with phenylisothiocyanate (PITC), with dansyl- or dabsyl chloride or by the FMOC reagent. In recent articles these techniques are described and their advantages and limitations are discussed [21,22]. Yet, none of the present techniques available to date allow high accuracy (with deviations of less than 5%) for low picomole to femtomole protein amounts.

MICRO SEQUENCING TECHNIQUES

Manual methods

Sequence determinations may be performed manually by employing phenylisothiocyanate (PITC) directly [23] or by applying the combined dansyl-Edman degradation scheme [24]. More sensitive is the use of the dimethylaminoazobenzene isothiocyanate (DABITC) in the double coupling routine [25] which releases red-coloured phenylthiohydantoin amino acid derivatives (DABTH-amino acids) easily visible by eye in the low picomole range. A further increase in sensitivity may be gained by the application of fluorescein isothiocyanate (FITC) [26] or by the more reactive dansyl-amino PITC [27] which enables the detection of femtomoles of fluorescent amino acid derivatives.

Sequencers

An enormous increase in protein sequence data was achieved when Pehr Edman and Geoffrey Begg invented the first sequencer which allows automatically to perform extended sequence analysis into 60-70 residues of the polypeptide chain [28]. The first machines performed the degradation in a spinning cup reactor in the liquid phase (in the Edman machine and commercialized in the Beckman sequencer). Secondly, Richard Laursen designed the solid phase sequencer [29] which degrades the polypeptide chain after covalent attachment to a solid support in a column [30-33]. In this mode also the DABITC/PITC doubling method has been successfully applied [34]. In the last year MilliGen/Millipore have developed a solid-phase machine which is based on covalent attachment of the polypeptide to PVDF membranes [35,36].

We improved the performance of the Beckman sequencer considerably [37,38] by introducing the automatic conversion [39], by inserting a straightforward on-line detection system for the released PTH-amino acid derivatives [40,41] and by construction of a dead-volume-free valve system [42,37,38] in order to reach high repetitive yields and to reduce the necessary peptide amounts. These three inventions in combination with the introduction of the cartridge reactor [43] led to automated microsequencing in the picomole range in the gas phase or pulsed-liquid-gas phase sequencer [44] (commercialized by Applied Biosystems). The world-wide application of these types of sequencers in the various research fields led to the rapid increase of completed protein sequences in the last decade.

More recently, a flow-through reactor of minute inner volume was designed [45] and assembled into sequencers of modular construction (Berlin sequencers [46] and Knauer sequencer [47]) in order to allow degradation in the wet phase technique [48,49] and to enable easier servicing and updating of the machine. By this means we arrived at higher average repetitive yields (97-98%) as compared to the gas phase or pulsed-liquid-gas phase sequencer (94%) [49] and could degrade peptides to their C-terminal end by complete attachment of their C-termini to arylamino PVDF [50].

The new sequencer parts are now also in use in automates for peptide and oligonucleotide synthesis and DNA sequence analysis (Applied Biosystems, Biotronik, Abimed). Furthermore, the essential sequencer parts allowed the construction of machines also for other purposes, such as for polypeptide sequencing from the carboxyl-terminal end [51-53].

Recent and new developments

Updated sequencer design follows the tendency to further miniaturize the essential parts and to increase the sensitivity to low picomole/-femtomole quantities by i) the construction of miniaturized flow-through type reactors, ii) the direct use of PVDF membranes for sample application or of PVDF-blot without the need to apply polybrene-treated glass fibers, iii) arrangement of the delivery valve system in a circular fashion around the reactor and converter [51] in order to allow short connections between the delivery lines and the reactor

vessels, the complete omission of oxygen penetration, and fast injection of the released amino acid derivatives into the in-machine on-line HPLC-detection system, iv) use of narrow-bore columns in the HPLC-system, v) on-line identification with an advanced detector for higher sensitivity, vi) application of fluorescent Edman-type reagents or derivatization of the released amino acids with a fluorescent reagent, vii) introduction of new chemistries. Further approaches are feasible, namely by the direct injection of the released anilino-thiazolinone-(ATZ)-amino acids into an on-line detection system and on-line sequencer connections to others than microbore HPLC, namely to capillary electrophoresis, gas chromatography or to LC/MS mass spectrometers. Such developments are under way.

Alternative approaches

Protein sequence determinations benefit largely from alternative techniques, e.g. DNA-sequencing of the protein's gene and from recent developments in mass spectrometry [54-59]. Several new mass spectrometers are available (e.g., double focussing sector field instrument, ^{252}Cf -plasma desorption time-of-flight MS, electrospray quadrupole MS/MS, laser desorption technique) which enable high mass ion determination of proteins with masses of more than 50 kDa, and allow to gain sequence information of fragments with only picomole quantities of substance. In combination with GC or HPLC (by an electrospray or ionic spray split system) these instruments provide alternative protein structural analysis. Mass spectrometry allows to detect amino acid modifications and unusual or crosslinked peptide structures. The recent mass spectrometers open new perspectives for sequence analysis of peptides and proteins. However, all the various approaches have their advantages and imply limitations. Best is to use combinations of the different techniques in order to obtain more sequence informations in short times, in order to save costly peptide, and to arrive at more accurate sequence data.

SUMMARY

In the recent years sensitive and advanced techniques have been developed in protein chemistry to meet the needs of the researchers. The methods serve as sensitive tools in protein-chemical characterization. Progress has been achieved in small scale purification methods, such as HPLC narrow-bore techniques, capillary electrophoresis, blotting after one- and two-dimensional polyacrylamide electrophoresis, as well as in advanced microsequencing. In addition, mass spectrometry of polypeptides has become an important tool for both, the correct assignment of the mass of a protein and for sequence analysis of blocked and modified peptides. By this array of methods peptides can be well characterized in the picomole range. The enhanced technical developments in sequencer construction allowed us to complete the sequence determinations of all proteins from the *E. coli* ribosome until 1980, and since, facilitated the elucidation of many protein sequences from ribosomes of *Bacillus stearothermophilus*, *Halobacterium marismortui*, and of diverse other organisms enabling studies on the evolution of the ribosome [60,61]. The methods described are essential for quality control of synthetic peptides or proteins obtained by genetic engineering in the chemical and pharmaceutical industry. Direct protein sequencing must prove sequences which are deduced from DNA-sequence work, their correct reading frame and the length of the nucleotide message. However, protein analysis in low pico- to femtomole amounts still faces many limitations and may embarque less reliable results. Therefore, efforts will have to be made to further improve our methodology to meet this challenge.

REFERENCES

1. D.S. Goodsell, TIBS 16, 203 (1991).
2. F. Wold, Ann. Rev. Biochem. 50, 783 (1981).
3. F.E. Regnier and U.M. Gooding, Anal. Biochem. 130, 1 (1980).

4. L. Giri, W.E. Hill, H.G. Wittmann and B. Wittmann-Liebold, *Adv. in Prot. Chem.* **36**, 1 (1984).
5. H.G. Wittmann, in: *Ribosomes*, Nomura et al., eds., Cold Spring Harbor Laboratory, p. 115 (1974).
6. H.G. Wittmann and B. Wittmann-Liebold, in: *Ribosomes*, Nomura et al., eds., Cold Spring Harbor Laboratory, p. 93 (1974).
7. R.M. Kamp and B. Wittmann-Liebold, *Methods Enzymol.* **164**, 542 (1988).
8. R.M. Kamp, in: *Advanced Methods in Protein Microsequence Analysis*, B. Wittmann-Liebold, J. Salnikow and V.A. Erdmann, eds. Springer Verlag, Berlin, Heidelberg, p. 8 and 21 (1986).
9. U.K. Laemmli, *Nature* **227**, 680 (1970).
10. E. Kaltschmidt and H.G. Wittmann, *Anal. Biochem.* **36**, 401 (1970).
11. P.H. O'Farrell, *J. Biol. Chem.* **250**, 4007 (1975).
12. P. Matsudaira, *J. Biol. Chem.* **262**, 10035 (1987).
13. J. Vandekerckhove, G. Bauw, M. Puype, J. Van Damme and M. Van Montagu, *Eur. J. Biochem.* **152**, 9 (1985).
14. R. Aebersold, D. Teplow, L.E. Hood, and S.B.H. Kent, *J. Biol. Chem.* **261**, 4229 (1986).
15. T. Choli and B. Wittmann-Liebold, *Electrophoresis* **11**, 562 (1990).
16. J.I. Garrels and B.R. Franza, Jr. *J. Biol. Chem.* **264**, 5283 (1989).
17. D.J. Rose and J.W. Jorgenson, *J. Chrom.* **438**, 23 (1988).
18. J. Brockmoeller and R.M. Kamp, *Biol. Chem. Hoppe-Seyler* **366**, 901 (1985).
19. J.W. Jorgenson and K.D. Lucacs, *Anal. Chem.* **53**, 1298 (1981).
20. P. Edman, *Acta Chem. Scand.* **4**, 283 (1950).
21. P.E. Hare, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 2 (1989).
22. G.J. Hughes and S. Frutiger, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 191 (1989).
23. P. Edman and A. Henschen, in: *Protein Sequence Determination*, S.B. Needleman, ed., Springer Verlag, Berlin, Heidelberg, New York, p. 232 (1975).
24. C.J. Bruton and B.S. Hartley, *J. Mol. Biol.* **52**, 165 (1970).
25. J.Y. Chang, D. Brauer and B. Wittmann-Liebold, *FEBS Lett.* **93**, 205 (1978).
26. K. Muramoto, H. Kamiya and H. Kawauchi, *Anal. Biochem.* **141**, 446 (1984).
27. S.-W. Jin, G.-X. Chen, Z. Palacz and B. Wittmann-Liebold, *FEBS Lett.* **198**, 150 (1986).
28. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
29. R.A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
30. R.A. Laursen, *Methods Enzymol.* **47**, 277 (1977).
31. W. Machleidt, E. Wachter, M. Scheulen, J. Otto, *FEBS Lett.* **37**, 217 (1973).
32. R.A. Laursen and W. Machleidt, *Meth. Biochem. Anal.* **26**, 375 (1980).
33. W. Machleidt, U. Borchart, and A. Ritonja, in: *Advanced Methods in Protein Microsequence Analysis*, B. Wittmann-Liebold, V.A. Erdmann, J. Salnikow, eds., Springer-Verlag, Berlin, Heidelberg, New York, p. 91 (1986).
34. J. Salnikow, A. Lehmann and B. Wittmann-Liebold, *Anal. Biochem.* **117**, 433 (1981).
35. J.M. Coull, J.D. Dixon, R.A. Laursen, H. Köster and D. Pappin, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed., Springer-Verlag, Berlin, Heidelberg, New York, p. 69 (1989).
36. R.A. Laursen, J.D. Dixon, S.-P. Liang, D.M. Nguyen, T. Kelcourse, L. Udell and D. Pappin, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed., Springer Verlag Berlin, p. 61 (1989).
37. B. Wittmann-Liebold, in: *Polypeptide Hormones*, R.F.J Beers and E.G. Bassett, eds. Raven Press, New York, p. 87 (1980).
38. B. Wittmann-Liebold, in: *Chemical Synthesis and Sequencing of Peptides and Proteins*, T.Y. Liu, R.L. Schechter, R.L. Heinrikson, and P.G. Condliffe, eds., Elsevier North-Holland Biomedical Press, Amsterdam, New York, p. 76 (1981).

39. B. Wittmann-Liebold, H. Graffunder and H. Kohls, *Anal. Biochem.* 75, 621 (1976).
40. B. Wittmann-Liebold and K. Ashman, in: *Modern Methods in Protein Chemistry*, vol. II, H. Tschesche, ed., deGruyter Verlag Berlin, p. 303 (1985).
41. K. Ashman, in: *Advanced Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, J. Salnikow and V.A. Erdmann, eds., Springer Verlag Berlin, p. 219 (1986).
42. B. Wittmann-Liebold, in: *Modern Methods in Protein Chemistry*, vol. I, H. Tschesche, ed., deGruyter Verlag, Berlin, New York, p. 229 (1983).
43. M.W. Hunkapiller and L.E. Hood, *Biochem.* 17, 2124 (1978).
44. R.M. Hewick, M.W. Hunkapiller, L.E. Hood and W.J. Dreyer, *J. Biol. Chem.* 256, 7990 (1981).
45. F. Reimann, H. Graffunder and B. Wittmann-Liebold, in: *Methods in Protein Sequence Analysis*, K.A. Walsh, ed., Humana Press, Clifton, NJ., p. 409 (1987).
46. B. Wittmann-Liebold, in: *Methods in Protein Characterization*, J.E. Shively, ed., Humana Press, Clifton, NJ., p. 249 (1986).
47. S. Fischer, F. Reimann and B. Wittmann-Liebold, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed., Springer-Verlag, Berlin, p. 98 (1988).
48. B. Wittmann-Liebold, *J. Prot. Chem.* 7/106, 224 (1988).
49. E. Herfurth, U. Pilling and B. Wittmann-Liebold, *J. Prot. Chem.* 9, 267 (1990).
50. E. Herfurth, H. Hirano and B. Wittmann-Liebold, *Biol. Chem. Hoppe-Seyler*, in press.
51. B. Wittmann-Liebold, L. Matschull, U. Pilling, H.-A. Bradaczek and H. Graffunder, in: *Methods in Protein Sequence Analysis*, H. Jörnvall, J.-O. Höög and M.-A. Gustavsson, eds., Birkhäuser Verlag, Basel p. 9 (1991).
52. A.S. Inglis, R.L. Moritz, G.S. Begg, G.E. Reid, R.J. Simpson, H. Graffunder, L. Matschull and B. Wittmann-Liebold, in: *Methods in Protein Sequence Analysis*, H. Jörnvall, J.-O. Höög and M.-A. Gustavsson, eds., Birkhäuser Verlag, Basel, p. 23 (1991).
53. A.S. Inglis, *Anal. Biochem.* 195, 183 (1991).
54. P. Roepstorff, K. Klarskov and P. Hojrup, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 191 (1989).
55. K. Biemann, R.S. Johnson, J.A. Hill and S.A. Martin, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 175 (1989).
56. D.F. Hunt, J. Shabanowitz, J.R. III. Yates, P.R. Griffin, and N.Z. Zhu, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 183 (1989).
57. H.R. Morris, M. Panico, R. McDowell and A. Chatterjee, in: *Methods in Protein Sequence Analysis*, B. Wittmann-Liebold, ed. Springer Verlag, Berlin, Heidelberg, p. 206 (1989).
58. T. Covey, B. Shushan and R. Bonner, in: *Methods in Protein Sequence Analysis*, H. Jörnvall, J.-O. Höög and M.-A. Gustavsson, eds., Birkhäuser Verlag, Basel, p. 249 (1991).
59. R.M. Caprioli, *Biochem.* 27, 513 (1988).
60. B. Wittmann-Liebold, in: *Structure, Function and Genetics of Ribosomes*, B. Hardesty and G. Kramer, eds., Springer Verlag, Berlin, New York, p. 326 (1986).
61. B. Wittmann-Liebold, A.K.E. Köpke, E. Arndt, W. Krömer, T. Hatakeyma and H.G. Wittmann, in: *Ribosomes*, W. Hill, Dahlberg, A., R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner, eds., *Am. Soc. Microbiol.*, Washington, DC., p. 598 (1990).