An improved procedure, involving mass spectrometry, for *N*-terminal amino acid sequence determination of proteins which are *N*^α-blocked

Keith ROSE and Hans P. KOCHER

Département de Biochimie Médicale, Centre Médicale Universitaire, 9 avenue de Champel, 1211 Geneva 4, Switzerland

and Benjamin M. BLUMBERG and Daniel KOLAKOFSKY

Département de Microbiologie, 64 avenue de la Roseraie, 1205 Geneva, Switzerland

(Received 12 September 1983/Accepted 20 October 1983)

A modification to a previously described procedure [Gray & del Valle (1970) *Biochemistry* 9, 2134–2137; Rose, Simona & Offord (1983) *Biochem. J.* 215, 261–272] for mass-spectral identification of the *N*-terminal regions of proteins is shown to be useful in cases where the *N*-terminus is blocked. Three proteins were studied: vesicular-stomatitis-virus *N* protein, Sendai-virus NP protein, and a rabbit immunoglobulin λ -light chain. These proteins, found to be blocked at the *N*-terminus with either the acetyl group or a pyroglutamic acid residue, had all failed to yield to attempted Edman degradation, in one case even after attempted enzymic removal of the pyroglutamic acid residue. The *N*-terminal regions of all three proteins were sequenced by using the new procedure.

The most commonly used technique for obtaining amino acid sequence information, sequential Edman degradation (e.g. Edman & Henschen, 1975), fails to provide such information in cases where the *N*-terminus of a polypeptide chain is blocked with an acetyl group or is cyclized in the form of a pyroglutamic acid residue: these are the two most commonly encountered blocking groups (Wold, 1981). Fortunately it is frequently possible to remove a pyroglutamic acid residue by means of the enzyme pyroglutamyl aminopeptidase (EC 3.4.11.8) (Doolittle & Armentrout, 1968), but the acetyl group is difficult to remove without concomitant unwanted scission of the polypeptide chain.

A mass-spectrometric procedure has been described, based on the permethylation reaction, which may be used to identify the *N*-terminal few residues of proteins, and neither of the above two blocking groups presented particular difficulties in the examples cited (Gray & del Valle, 1970). In this procedure, the protein is acetylated, digested with

a proteinase, then permethylated. Only the blocked *N*-terminal peptide (blocked naturally or by the prior acetylation step using a ²H-labelled reagent) forms a volatile derivative, which is identified by mass spectrometry. We have recently described a miniaturization of the permethylation reaction which, in conjunction with g.l.c.–mass spectrometry as a sample-introduction method, permits sample sizes of 2–10 nmol to be used for *N*-terminal analyses, general sequence work (Rose *et al.*, 1983a) and for *C*-terminal analyses (Rose *et al.*, 1983b).

We now describe an alternative approach to *N*-terminal sequence determination, involving omission of the prior acetylation step, which is shown to be of value in the case of proteins which present severe problems of solubility and which are thought to be blocked (e.g. failure of the Edman degradation). Vesicular-stomatitis-virus *N* protein, Sendai-virus NP protein and an intractable immunoglobulin λ -light chain are given as examples.

Materials and methods

VSV-N was isolated from a cytoplasmic extract of infected BHK cells by velocity centrifugation on a CsCl gradient, which removes all proteins except

Abbreviations used: VSV-N, vesicular-stomatitis-virus *N* protein; BHK cells, baby-hamster kidney cells; SV-NP, Sendai-virus NP protein; SDS, sodium dodecyl sulphate; Ile, isoleucine or leucine (Rose *et al.*, 1983a); < Glu, pyroglutamic acid.

the nucleocapsid protein N, followed by isopycnic centrifugation in a CsCl gradient containing 3M-guanidinium chloride (B. M. Blumberg & D. Kolakofsky, unpublished work). VSV-N prepared in this way was shown to retain at least part of its biological activity (Blumberg *et al.*, 1983), and was used as a freeze-dried powder free of RNA.

SV-NP was isolated in a similar fashion but, because of the small quantities available, SV nucleocapsids containing only the NP protein were used as a freeze-dried powder still containing the viral RNA. Protein content was estimated by the Bio-Rad dye-binding method, with serum albumin and rabbit γ -globulin as standards.

Isolation of immunoglobulins from the serum of hyperimmunized Basilea rabbits and separation of heavy and light chains was performed as described by Garcia & Jaton (1981). To remove contaminating light chains of the κ -type, the light-chain preparation was incubated in 70% (v/v) formic

acid at 70°C for 1.5h to cleave an aspartic acid-proline bond in the κ -chains. λ -Light chains were separated from the cleavage products by gel filtration on a column (90cm $\times$ 2cm diam.) of Sephadex G-100 in 3M-guanidinium chloride which was 0.2M with respect to NH_4HCO_3 . The λ -light-chain fraction was extensively dialysed against 5% (v/v) acetic acid and was used as a freeze-dried powder.

Digestion conditions, reagent preparation, derivative formation and g.l.c.-mass-spectrometric analysis conditions have been described in full elsewhere (Rose *et al.*, 1983a).

Results and discussion

VSV-N

A sample of the protein (0.4mg, about 10nmol) subjected to acetylation conditions employing a

Fig. 1. Mass spectrum of the N-terminal peptide of VSV-N

For clarity, the conventional three-letter code is used in the Figures and in the relevant discussion in the text to indicate *NO*-permethylated amino acid residues. Background was subtracted by using the DS55 data system; nevertheless, the silicone ion series m/z 281, 341, 355, 415, 429 is due to column bleed. Signals at m/z 59, 86 and 88 inform us of the presence of threonine, valine and serine respectively, without indicating their position in the sequence. A possible interpretation of signals at m/z 301 and 346 would involve charge retention on C-terminal fragments and include loss of methanol (CH_3OH) in the latter case. The sequence assignment is *N*^α-acetyl-Ser-Val-Thr-OCH₃.

mixture of acetic anhydride and $[^2\text{H}_6]$ acetic anhydride (1:1 on a molar basis) and using SDS to improve solubility as described by Rose *et al.* (1983a) gave, upon g.l.c.-mass-spectrometric analysis of a subtilisin digest, evidence of an acetylserine residue. Chemical background due to non-peptide impurities was very high. The g.l.c.-mass spectrometric result was obtained by using half of the derivatized material, corresponding to 5 nmol of sample. Ions at m/z 158, 130 and 126 were present, indicating that the serine residue was naturally acetylated, since corresponding ions at m/z 161, 133 and 129 due to a $^2\text{H}_3$ -labelled acetyl group were not found. A further sample of the protein (0.2 mg, about 5 nmol) was digested directly with subtilisin without prior acetylation. G.l.c.-mass-spectrometric analysis of half of the permethylated digest produced the spectrum shown in Fig. 1. The first two residues of the viral protein are well-established and the third seems clearly assigned; the blocking group is seen to be acetyl, as indicated by the first experiment. The significant improvement over the first experiment is attributed to there being fewer manipulations, and no

SDS. Although SDS may be added in amounts up to 400 μg without greatly affecting results with polypeptides or globular proteins (Rose *et al.*, 1983a), evidence is accumulating that its use in large quantities to assist the solubility of hydrophobic proteins can be detrimental: lower yields of peptide derivatives and a higher background of non-peptide impurities result. Lower yields may result from incomplete enzymic digestion of a protein to which SDS is bound and which has been acetylated and so is less soluble, and the non-peptide impurities may be liberated from the hydrophobic core upon addition of SDS; at the moment this is a speculative suggestion, and no control experiments involving 'spiking' with known compounds have been performed. The sequence found:

N^α -acetyl-Ser-Val-Thr-

is in agreement with that expected on the basis of the VSV-N-gene DNA sequence (Gallione *et al.*, 1981) and shows that the methionine is processed off. Serine, which follows methionine in the

Fig. 2. Mass spectrum of the N -terminal peptide of SV-NP

Background was subtracted by using the DS44 data system. The origin of the signal at m/z 358 is unknown, but it does not confuse the assignments made, since it does not give rise to interpretable mass differences. The signals at m/z 384 and 313 may be due to charge retention on C -terminal fragments. The sequence N^α -acetyl-Ala-Gly-Ilx-Ilx- OCH_3 is deduced.

sequence deduced from the DNA data, is, along with alanine, one of the residues most commonly found to be acetylated (Wold, 1981).

SV-NP

Subtilisin digestion of 1 mg of SV-NP (about 20 nmol by weight, about 1 nmol by dye-binding assay) performed in the absence of SDS left an insoluble residue, which was removed by centrifugation. G.l.c.-mass-spectrometric analysis of one-quarter of the permethylated supernatant produced the mass spectrum shown in Fig. 2. The first three residues are well established and the fourth seems clearly assigned. The sequence found:

is in agreement with that deduced from DNA sequence data (Met-Ala-Gly-Leu-Leu-; C. Giorgi & D. Kolakofsky, unpublished work) with the loss of the initiating methionine residue, followed by acetylation of the succeeding residue, which is one commonly acetylated (Wold, 1981).

λ -Chain

A first experiment with this polypeptide, involving 0.2 mg (about 8 nmol) of material, proceeded as described above for VSV-N: addition of SDS, acetylation, digestion with subtilisin, permethylation and g.l.c.-mass-spectrometric analysis. The sequence <Glu-Pro-Val- was deduced. In a second experiment, again involving 0.2 mg of material, the sample was digested directly with subtilisin in the presence of SDS without prior acetylation. G.l.c.-mass-spectrometric analysis of half of the derivative produced the mass spectrum shown in Fig. 3. The tetrapeptide sequence <Glu-Pro-Val-Ilx- is well-established. Comparison of this sequence with published *N*-terminal λ -light-chain sequences of different species (Kabat *et al.*, 1983) reveals that valine and leucine/isoleucine residues are common in positions 3 and 4 respectively. The spectrum is more intense than the one obtained with the acetylated sample. The blocking group, pyroglutamyl-, was resistant to pyroglutamyl aminopeptidase treatment as described by Podel & Abraham (1978), a method

Fig. 3. Mass spectrum of the *N*-terminal peptide of the λ -chain

Scans 317 and 371, either side of the sample which maximized in scan 345, were averaged and subtracted from scan 345 by using the DS55 data system. The loss of CO from the first sequence ion at m/z 126, always very pronounced in the case of pyroglutamyl-peptides (Morris *et al.*, 1974), is here seen to be total. The sequence found is <Glu-Pro-Val-Ilx-OCH₃.

which has previously been used successfully in our laboratory to deblock quantitatively various proteins (Kühn *et al.*, 1983; H. P. Kocher, unpublished work). These conditions sufficed to remove it quantitatively from a human μ -heavy chain which had been added to a portion of the sample (H. P. Kocher, unpublished work). However, the inability of the enzyme to cleave pyroglutamic acid when the next residue is proline has been previously documented (Uliana & Doolittle, 1969).

Conclusion

The three proteins discussed above had each been the subject of several unsuccessful attempts to obtain sequence information using the Edman degradation technique. In one case, that of the λ -light chain, an unsuccessful attempt had been made to remove enzymically a suspected blocking group. All three proteins yielded to g.l.c.-mass-spectrometric investigation, and in the two cases tried, a more intense spectrum was obtained from material which had been digested without prior acetylation. Of course, had the N -terminal peptides contained a lysine residue this would have become quaternized during the permethylation reaction, and no result would have been obtained at all. Residues of cysteine and arginine also require treatment prior to permethylation (Rose *et al.*, 1983a). Nevertheless, we recommend that, when one performs N -terminal analysis by mass spectrometry by the method of Rose *et al.* (1983a) of proteins thought to be N^α -blocked, an attempt be made to analyse the sample without prior acetylation in cases where acetylation leads to unsatisfactory results, and without SDS, where high background interferes. Besides N^α -acetyl- and pyroglutamyl-, N^α -formyl- should also be amenable to this direct approach, since experiments with N^α -formylmethionyl-peptides were successful (K. Rose, unpublished work). We believe that the techniques reported here complement those de-

scribed previously (Rose *et al.*, 1983a) and should facilitate the study of those proteins whose N -termini are thought to be blocked and which have not yielded to other methods.

This work was supported by the 'Fonds National Suisse de la Recherche Scientifique'. We thank Mlle. Colette Pasquier, Mlle. Rosette Bandelier, Mme. Chantal Dumais and Dr. Marco G. Simona for expert technical assistance. K. R. thanks Biogen S.A. for financial support.

References

- Blumberg, B. M., Giorgi, C. & Kolakofsky, D. (1983) *Cell* **32**, 559-567
- Doolittle, R. F. & Armentrout, R. W. (1968) *Biochemistry* **7**, 516-521
- Edman, P. & Henschel, A. (1975) *Mol. Biol. Biochem. Biophys.* **8**, 232-279
- Gallione, C. J., Greene, J. R., Iverson, L. E. & Rose, J. K. (1981) *J. Virol.* **39**, 529-535
- Garcia, I. & Jaton, J.-C. (1981) *Biochem. J.* **197**, 177-183
- Gray, W. R. & del Valle, U. E. (1970) *Biochemistry* **9**, 2134-2137
- Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest*, U.S. Department of Health and Human Services, National Institutes of Health, Bethesda
- Kühn, L. C., Kocher, H. P., Hanly, W. C., Cook, L., Jaton, J.-C. & Kraehenbühl, J.-P. (1983) *J. Biol. Chem.* **258**, 6653-6659
- Morris, H. R., Williams, D. H., Midwinter, G. G. & Hartley, B. S. (1974) *Biochem. J.* **141**, 701-713
- Podel, N. D. & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun.* **81**, 176-185
- Rose, K., Simona, M. G. & Offord, R. E. (1983a) *Biochem. J.* **215**, 261-272
- Rose, K., Simona, M. G., Offord, R. E., Prior, C. P., Thatcher, D. R. & Otto, B. (1983b) *Biochem. J.* **215**, 273-277
- Uliana, J. A. & Doolittle, R. F. (1969) *Arch. Biochem. Biophys.* **131**, 561-565
- Wold, F. (1981) *Annu. Rev. Biochem.* **50**, 783-814