Vol. 117, No. 1, 1983 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1983 Pages 299-305

FAB-MAPPING* OF RECOMBINANT-DNA PROTEIN PRODUCTS

Howard R. Morris, Maria Panico and Graham W. Taylor

Department of Biochemistry, Imperial College of Science and Technology
London SW7 2AZ. U.K.

Received October 5, 1983

SUMMARY: A new method is described for screening and establishing the primary structure of proteins. The procedure is both rapid and sensitive and allows study of the C-terminus of the protein with equal facility to the N-terminus. This new strategy which is illustrated here for the polypeptide hormone Insulin, has obvious applications in recombinant DNA biotechnology research, in post-translational modification and site-directed mutagenesis studies, or any other aspect of modern protein chemistry requiring accurate definition of primary structure.

An important problem in current Biotechnology research and practice lies in defining the exact structure, primary, secondary and tertiary, of the recombinant-DNA protein product, either for regulatory or quality control purposes or because low specific activity may suggest proteolytic cleavage, errors of translation, or some other modifications of the sequence anticipated from the nucleotide template.

In collaboration with M-SCAN (3) and consequent upon our original work on tryptic digests of proteins (1,4) we have developed the new concept of FAB-MAPPING described here, specifically to address questions of primary structure fidelity in recombinant protein research but also to be of general use in protein chemistry. Here we illustrate our strategy with the FAB-MAPPING of the polypeptide hormone Insulin.

METHODS. Enzymic digests of the protein (5-50 nanomoles) are carried out in pH 8.5 ammonium bicarbonate buffer (50 millimolar) at an enzyme: substrate ratio of 1:50 for 3h. at 37°C.

Carboxypeptidase B digestion is carried out on the mixture of peptides produced from the primary digest, under identical conditions.

* FAB-MAPPING is a term we have previously introduced to describe the examination or screening of protein digests, subdigests (chemical or enzymatic) or degradation products, by the procedure of High Field Magnet Mass Spectrometry in the Fast Atom Bombardment (FAB) mode (1,2).

0006-291X/83 $1.50

Reduction of the protein is carried out with a four molar excess of dithiothreitol under nitrogen in TRIS buffer pH 8.4 for 30 mins. Alkylation is then carried out using a five fold molar excess over total thiol for 1h. at room temperature, followed by acidification to pH 3.5 using glacial acetic acid. Desalting of the protein is achieved using a C18-SEP PAK (WATERS Associates), and eluting with 4 vols of 5% acetic acid, and then 4 vols of 5% acetic acid containing 20%, 30% and 40% n-propanol consecutively.

Edman degradation is carried out on the digest mixture (tryptic or tryptic and carboxypeptidase B treated peptides) under standard conditions using phenyl isothiocyanate at 45°C for 1h, followed by cleavage with trifluoro-acetic acid in 10 mins at 45°C. The remaining shortened peptides (after n cycles) are studied by FAB.

FAB Atom Bombardment (FAB) is carried out using a VG ZAB HF instrument equipped with an M-SCAN FAB gun using Xenon as the primary ionising beam (current of 20μA at 8KeV). The samples 1-5 nanomoles in 2μl of 5% acetic acid, are loaded onto a glycerol-coated probe tip, followed by 1μl of thioglycerol. Spectra are recorded at 300 secs. per decade in mass.

RESULTS AND DISCUSSION. We have previously described the importance of High Field Magnet Mass Spectrometry (5,6) in the analysis of biopolymers, and demonstrated the production of quasimolecular ion signals for intact non-reduced Insulin by a combination of High Field Magnet and FAB Mass Spectrometry (7). Significant though this numerical and isotopic ratio data can be for diagnostic purposes (total amino acid composition, extent of disulphide or thiol composition) determination of masses in the 6,000 Dalton range may be limited by sample size (if <10 nanomoles) or instrument parameters. In addition, much further information can be gained by digesting the molecule, and confirming structure or pinpointing problems on smaller fragments (perhaps an unexpected degree of formylation or de-formylation etc) at better sensitivity up to 3-4,000 Daltons. This is the effective mass range of current High Field Magnet Spectrometers and it so happens that most tryptic peptides obtained from proteolytic digests are encompassed by this mass range.

With large molecules such as the genetically engineered Interferons (MW 15-20 K Daltons) the observation of the intact molecular species with reasonable accuracy of mass measurement is not practicable by any method. The FAB-MAPPING procedure overcomes these difficulties and at the same time gives valuable information on the nature of the protein by observing the results of a series of simple protein chemical manipulations of the molecule under study.

GLY-ILE-VAL-GLU-GLN-CYS-CYS-ALA-SER-VAL-CYS-SER-LEU-TYR-GLN-

-LEU-GLU-ASN-TYR-CYS-ASN $[M+H]^+$ 2338 A CHAIN

PHE-VAL-ASN-GLN-HIS-LEU-CYS-GLY-SER-HIS-LEU-VAL-GLU-ALA-LEU-TYR-LEU-VAL-

-CYS-GLY-GLU-ARG | GLY-PHE-PHE-TYR-THR-PRO-LYS | ALA $[M+H]^+$ 3397 B CHAIN

2486 3326

SCHEME 1: REDUCED INSULIN

Here we illustrate the principles of the FAB-MAPPING strategy with the example of Insulin, Scheme 1. The total molecule is first brought into the effective routine mass range of the High Field Magnet Spectrometer by a single reduction step demonstrating the presence of the 2 chain nature of the Insulin molecule via signals at m/z 3397 (B chain) and m/z 2338 (A chain), Figure 1. The number of thiol groups present in the molecules may now be

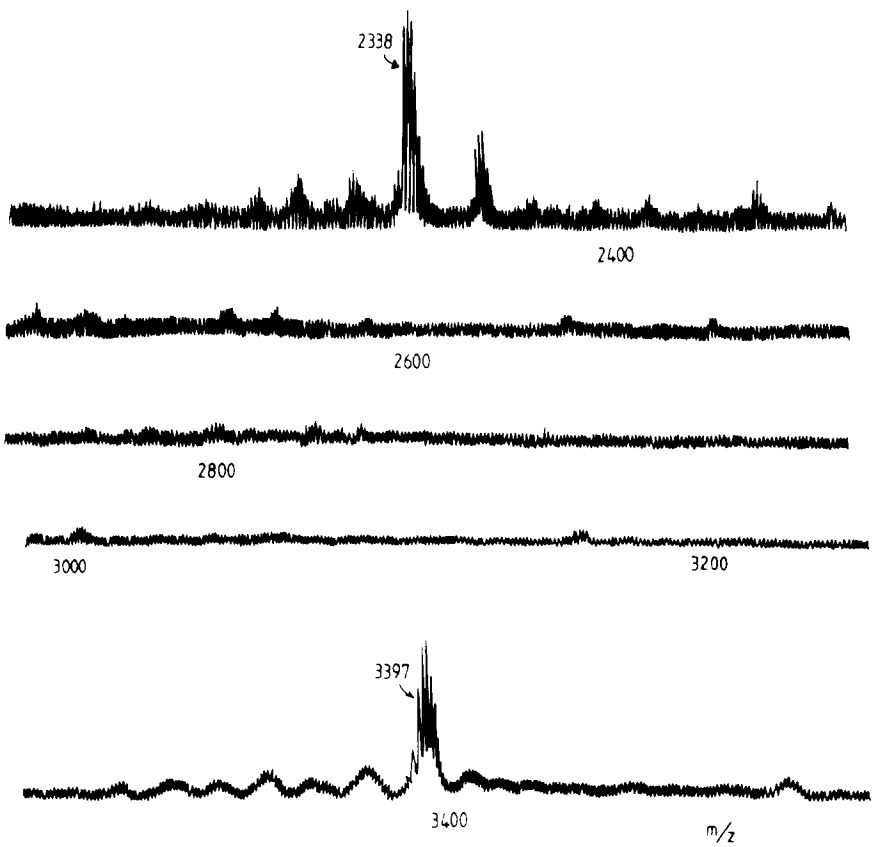

Figure 1. The High Field FAB Mass Spectrum of reduced bovine Insulin.

simply counted, in an unknown case, by reacting with iodoacetic acid (see Methods). The shift in signal is observed as a multiple of 58 mass units (Figure 3 (a)) giving the number of free thiols: in the case of the Insulin B chain, moving the m/z 3397 signal to m/z 3513. With many proteins it is essential to block the thiol groups of cysteine in this way in order to observe peptide signals, and here Insulin is an exception.

Returning to the reduced Insulin spectrum in Figure 1, the FAB-MAPPING strategy now procedes with tryptic digestion to give the spectrum seen in Figure 2.

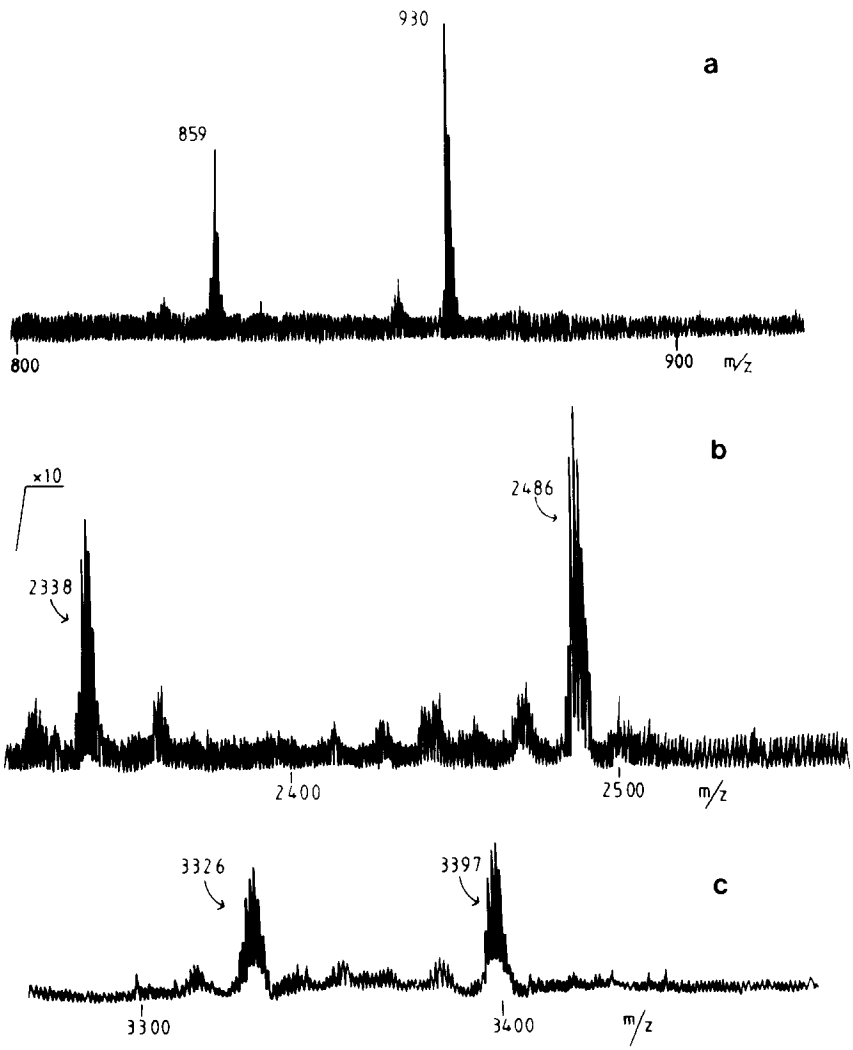

Figure 2. A FAB MAP of the tryptic digest of reduced Insulin.

The numbers themselves together with simple arithmetic give a surprising amount of information on the molecule. We see that the A chain (m/z 2338) remains intact being devoid of tryptic cleavage sites. The B chain, on the other hand, splits as predicted into fragments at m/z 859, 930, 2486 and 3326, with some intact B chain remaining at m/z 3397. The mass difference between the heaviest fragment (3326) and the intact molecule (3397) of 71 u. shows that the C-terminus of the B chain is ALANINE. In this case the complete structure is mapped in the one experiment.

In more complex analyses further subdigests or derivative formations and FAB-MAPS are recommended particularly if the problem demands location of a point mutation, blocked N-terminus, or the C-terminus of a protein. Such problems may be solved by continuing the FAB-MAPPING strategy with, for example, carboxypeptidase B digestion and/or n cycles of Edman degradation.

Examples of this idea are shown in Figure 3 (b) and (c). Figure 3(b) shows the partial FAB-MAP around m/z 2300 after treating the tryptic digest of the reduced Insulin with carboxypeptidase B. The signal at m/z 2338 remains unchanged indicating the absence of a C-terminal LYS or ARG (unless preceded by PRO) in the peptide; since this was derived by tryptic digestion, this is a special case of a C-terminal peptide assignment i/e. the signal remains unchanged on carboxypeptidase B digestion. In contrast m/z 2486 in

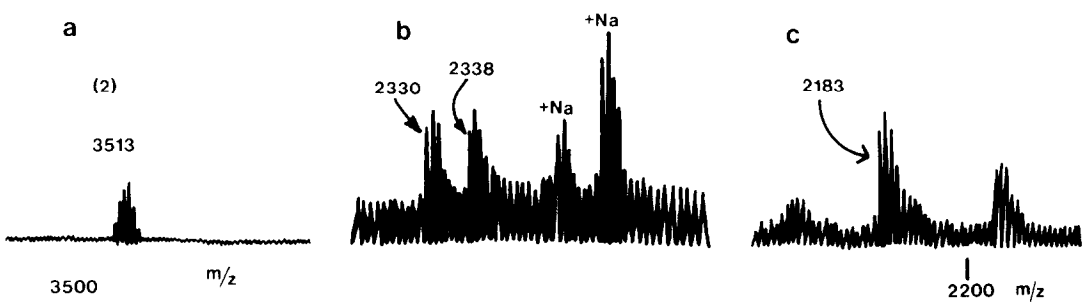

Figure 3. (a) Partial FAB MAP of reduced and carboxymethylated Insulin showing the addition of 2 carboxymethyl units to the B chain.
(b) Partial FAB MAP of the carboxypeptidase B-treated tryptic digest of reduced Insulin. The sodium cationised signals derive from the commercial enzyme preparation.
(c) Partial FAB MAP of the Edman degradation product of the carboxypeptidase B-treated tryptic digest of reduced Insulin.

Figure 2 shifts to m/z 2330 indicating a C-terminal ARGININE in this molecule (156 u.).

Because the experiments are simple and the FAB-MAPPING procedure is reasonably sensitive, a number of operations may be carried out on the same sample. This is illustrated in Figure 3(c) which shows a region of the FAB-MAP produced by subjecting the tryptic and carboxypeptidase B treated digests of reduced Insulin to a step of Edman degradation. The arithmetic simply tells us that m/z 2183 derives from m/z 2330 in Figure 3(b) by loss of an N-terminal PHENYLALANINE (147 u.). This signal cannot derive from m/z 2338 since the mass difference of 155 u. is meaningless; this signal in fact shifts to m/z 2281 confirming the N-terminal GLY residue in the A chain.

CONCLUSION A powerful new method, which we term FAB-MAPPING, is described for the rapid screening of recombinant-DNA or other protein or biopolymer products, with respect to primary structure.

The method combines conventional micro-scale protein chemical procedures with Fast Atom Bombardment analysis on a High Field Magnet Mass Spectrometer giving >3000 mass range.

The method allows rapid confirmation or indeed correction of the primary structure, including amino acid sequence and any blocking groups or other post-translational modifications such as formyl, acetyl, or proteolytic processing etc. In the case of unknown proteins we have demonstrated the potential for complementing existing conventional sequence assignment, and for identifying the C-terminus of the protein. This, together with S-S bridge assignment will be the subject of further communications from this laboratory.

Summarising the overall strategy:

1. Digest the protein, with or without reduction/alkylation, by a 'specific' method, ideally trypsin or CNBr.
2. Examine total mixture by FAB-MS (positive and negative) if the protein is <30,000 or subfractionate by HPLC (8) if the mixture is too complex.
3. (a) Check that the signals are protein-derived by 1:1 acetylation (1)- this will also reveal blocked N-terminals (no shift in mass).

(b) Check C-terminal residues as LYS or ARG by shifts of 128 or 156 mass units after carboxypeptidase B digestion of the mixture - this will also reveal the C-terminus of the protein (except in the unlikely event that it terminates in LYS or ARG) as a signal not shifting after C-terminal digestion. Such a signal will be the free C-terminus of the protein or a -PRO-LYS or -PRO-ARG sequence.

(c) Check sequence if discrepancies are observed by carrying out n steps of Edman degradation or subdigests on the mixture followed by FAB-MAPPING.

The sensitivity of the method is good, requiring approximately 1-5 nanomoles per map on a routine basis i.e. of the order of 10-100 nanomoles of protein depending upon the extent of information required.

Finally it is important to emphasise that this FAB-MAPPING procedure provides information not readily available by conventional or gas-phase sequencing methods since it is (a) applicable to blocked N-terminal proteins and (b) provides an equal probability of N-terminal and C-terminal sequence information, as demonstrated here in Figures 2 and 3(b). The method thus allows the detection of errors of translation, mutation, insertions, deletions or truncations, as well as location and characterisation of the C-terminal sequence of a protein.

ACKNOWLEDGEMENTS

The authors thank the Medical Research Council for financial support.

REFERENCES

1. Morris, H.R., Panico, M., Barber, M., Bordoli, R.S., Sedgwick, R.D. and Tyler, A.N. (1981) Biochem. Biophys. Res. Commun. 101, 623-631.
2. Morris, H.R., Panico, M. and Etienne, A.T. (1983) Proc. 31st Ann. Conf. on Mass Spec. and Allied Topics, Boston, 683-685.
3. Haskins, N., M-SCAN Ltd, Silwood Park, Ascot, Berks, U.K.
4. Lemaire, S., Chouinard, L., Denis, D., Morris, H.R. and Panico, M. (1982) Biochem. Biophys. Res. Commun. 108, 51-58.
5. Morris, H.R., Dell, A. and McDowell, R.A. (1981) Biomed. Mass Spectrom. 8, 463-473.
6. Morris, H.R., Dell, A., Banner, A.E., Evans, S., McDowell, R.A., Hazelby, D. (1977) Proc. 25th Ann. Conf. on Mass Spec. and Allied Topics, 73-75.
7. Dell, A. and Morris, H.R. (1982) Biochem. Biophys. Res. Commun. 106, 1456-1462.
8. Morris, H.R., Etienne, A.T., Dell, A. and Albuquerque, R. (1980) J. Neurochem. 34. 574-582.