Vol. 102, No. 1, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
September 16, 1981 Pages 335-340

MASS SPECTROMETRIC DIFFERENTIATION OF LEUCINE AND ISOLEUCINE IN PROTEINS DERIVED FROM BACTERIA OR CELL CULTURE

Walter C. Herlihy, David Kidwell, Bea Meeusen, and K. Biemann
Department of Chemistry, Massachusetts Institute of Technology
Cambridge, MA 02139

Received August 4,1981

Summary

The differentiation of leucine and isoleucine is a well known difficulty in mass spectrometric peptide sequencing. A technique has been developed which allows these two amino acids to be distinguished by growing a bacterial or cell culture in a medium containing γ,δ-*dl*-dideuteroleucine. The isotopically labelled residue is incorporated into the cell's proteins, and the resulting mass spectra of leucine containing peptides exhibit sequence ions 2 amu higher than the corresponding isoleucine peptides.

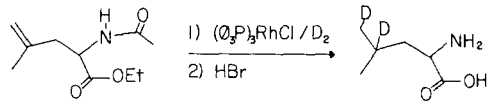
Introduction

In the past ten years the application of mass spectrometry to the determination of peptide and protein structure has rapidly increased (1,2,3). Two strategies have been highly developed and applied to a variety of unknown structures. The methodology developed in this laboratory (1,2) consists of hydrolysis of a polypeptide or protein to a complex mixture of di- to pentapeptides which are converted, without prior separation, to the corresponding O-trimethylsilyl polyamino alcohol derivatives (4). The mixture is separated by gas chromatography, and the polyamino alcohols are identified from their mass spectra. The GCMS technique has been applied to numerous peptides and proteins (5,6,7). and has been shown to effectively complement both the Edman degradation (8,9,10) and the more recent approach of deriving protein sequences from the nucleotide sequence of the corresponding gene (11).

Another approach (3) consists of hydrolysis of a polypeptide or protein to a complex mixture which is fractionated into simple (2-4 component) mixtures by ion exchange chromatography. After conversion to the N-acetyl, N-O-permethyl

Abbreviations: xLeu, Leucine or Isoleucine; GCMS, gas chromatographic mass spectrometry; TMS, trimethylsilyl

0006-291X/81/170335-06$01.00/0



1 dideuteroleucine

Scheme 1. Synthesis of dideuteroleucine.

derivatives, the peptides are introduced directly into the ion source of the mass spectrometer. Morris has recently reviewed applications of this technique (3).

An inherent problem with both mass spectrometric techniques is that it is difficult to distinguish the isomeric amino acids leucine and isoleucine. They can be differentiated from the spectra of polyamino alcohols only when they occur in the C-terminal position of a dipeptide (12). However, larger peptides carry most of the sequence information, and in these leucine and isoleucine are indistinguishable. We have developed a technique to resolve this ambiguity for proteins which can be derived from bacteria or cell culture and have demonstrated its utility with peptides derived from *E. coli* alanine tRNA synthetase (AlaRS).

Materials and Methods

Synthesis of γ,δ-*dl*-dideuteroleucine. Ethyl-2-acetamido-2-carbethoxy-4-methyl-4-pentenoate *1* was prepared by the method of Albertson and Archer (13). Tris-(triphenylphosphine) rhodium (I) chloride (1g) was weighed into a reaction flask and dissolved in 50 ml of dry benzene. The apparatus was connected to a manifold, and the solution was frozen with liquid nitrogen. The system was evacuated and then flushed with deuterium. The flask was allowed to warm to room temperature and stirred for 20 min to saturate the catalyst with deuterium. Then, 67.5 g of *1* (.25 m) in 250 ml of dry benzene was added. The solution was refrozen, the flask was evacuated and refilled with deuterium. The flask was again allowed to warm to room temperature and stirred overnight under a deuterium atmosphere. The benzene was removed under reduced pressure, and the residue was refluxed for 3 hr with 300 ml of concentrated HBr. The HBr solution was concentrated to ca, 100 ml, poured into 1.5 l of water, and neutralized with NaOH to pH 6. After chilling overnight, the product was collected by filtration, washed and dried to yield 20 g (60%) of γ,δ-*dl*-dideuteroleucine, m.p. 269-270. NMR and mass spectral analysis showed two deuteriums incorporated to a level of 95%.

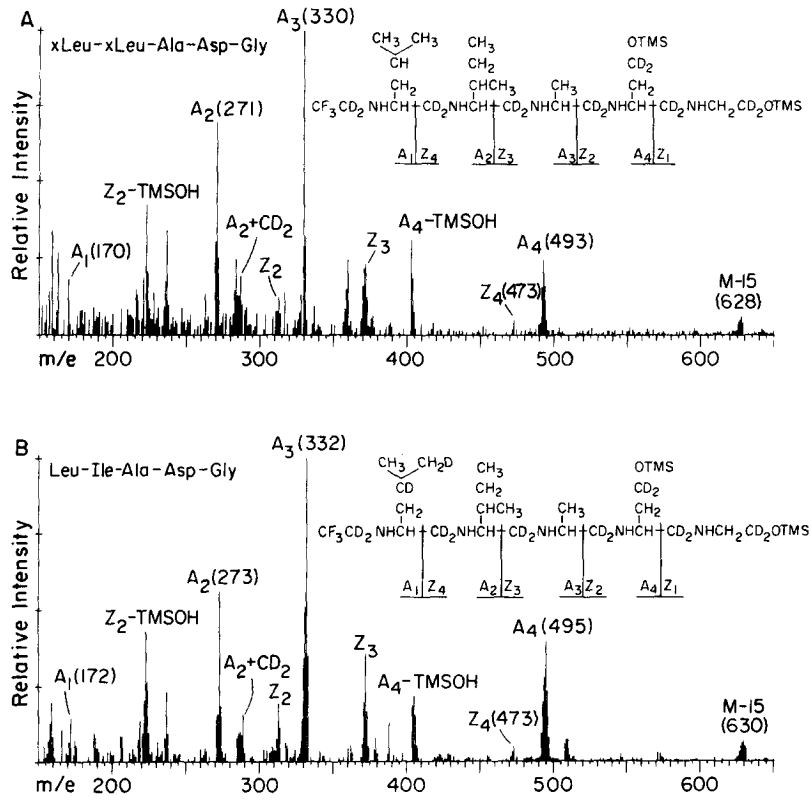
E. coli growth medium. *E. coli* containing a recombinant plasmid (14) carrying the AlaRs gene were grown in Zubay's medium (15) modified as follows: the yeast extract concentration was reduced to 4 gm/l and 2 gm of γ,δ-*dl*-dideuteroleucine was added to each liter of medium. Fifty grams of cells were isolated from eight liters of culture.

Isolation and Mass Spectrometry. AlaRS was isolated by the method of Putney *et al.* (16). A 365 amino acid fragment (T-1) corresponding to the N-terminal segment of AlaRS was produced by limited tryptic digestion, isolated and hydrolyzed with thermolysin as previously described (11). Details of conversion of the peptides to polyamino alcohols and the conditions employed for the GCMS experiment have been described elsewhere (4).

Results and Discussion

In order to make it possible to differentiate leucine from isoleucine by mass spectrometry, the protein is prepared in such a way that all leucines are primarily the dideutero analog. This is achieved by growing the bacteria or cell culture in a medium containing an excess of dideuteroleucine. Large quantities of *dl*-dideuteroleucine can be prepared easily and inexpensively, and the racemate need not be separated into the enantiomers since the cells utilize only the L form. The mass spectra of the derivatives of peptides derived from these proteins exhibit the expected sequence ions for isoleucine containing peptides; however, if the peptide contains a leucine, all of the corresponding sequence ions are shifted upward by 2 amu because the side chain of this amino acid now contains two deuteriums. Thus leucine and isoleucine are readily differentiated regardless of the peptide's length or the position of the leucine/isoleucine. The utility of this technique is demonstrated with our work on the sequence of alanine tRNA synthetase, a protein 875 amino acids in length.

The amino acid sequence of AlaRS has recently been determined (17) from both the DNA sequence of its gene and peptide sequences determined from GCMS experiments on partial hydrolyzates of large (>350 amino acids) segments of the protein. The tetra- and pentapeptides identified by GCMS were used to insure that the DNA was translated in the correct reading frame throughout its length. Tetra- and pentapeptides are sufficiently long that they normally occur only once in the correct amino acid sequence and not in either of the amino acid sequences derived from translation of the two incorrect reading frames. Occasionally, however, such an ambiguity will occur. For example, in a thermolytic digest of the N-terminal segment of the protein (T-1), the peptide xLeu-xLeu-Ala-Asp-Gly was identified from the spectrum shown in Figure 1A. This peptide matched in the apparently correct reading frame (established by other peptides) at the sequence Leu-Ile-Ala-Asp-Gly ^{297}, but also corresponded to the sequence Leu-Leu-Ala-Asp-Gly ^{172} in a second, apparently incorrect reading frame. It was therefore possible that the phasing of the DNA was incorrect in one of these regions.



Legend: Figure 1 A) Mass spectrum of the derivative of xLeu-xLeu-Ala-Asp-Gly
B) Mass spectrum of Leu-Ile-Ala-Asp-Gly obtained from cells grown in medium containing dideuteroleucine. The deuterium in the peptide backbone and the side chain of aspartic acid result from reduction of the peptide with hexadeuterodiborane(4). The deuterated reducing agent allows aspartic acid to be distinguished from threonine which would otherwise be isomeric(18).

Since these two sequences could be resolved only if leucine/isoleucine could be distinguished, a culture of *E. coli* was grown on a medium to which dideuteroleucine was added. Following thermolytic digestion of T-1, a GCMS experiment was carried out, and the spectrum of xLeu-xLeu-Ala-Asp-Gly shown in Figure 1B was recorded. Comparison of the spectra shown in Figure 1A and 1B shows that the A_1 ion is shifted from m/z 170 to m/z 172 confirming that the N-terminal amino acid is leucine. If the second amino acid were also leucine, one would expect the A_2 ion to be shifted upward by 4 amu to m/z 275 since this fragment ion would then contain 2 dideuteroleucines. Inspection of Figure 1B, however,

reveals that the A_2 ion is at m/z 273 and the second amino acid must therefore be isoleucine. All other "A" ions in Figure 1B are also shifted upward by 2 amu because they all contain the N-terminal dideuteroleucine. In contrast, the Z_2 and Z_3 ions are not shifted since these fragments do not contain the N-termina amino acid. A weak ion at m/z 473, the Z_4 ion, is observed in both spectra indicating that the fourth amino acid from the C-terminus is isoleucine. A dideuteroleucine in that position would have shifted the Z_4 ion to m/z 475.

Clearly the peptide observed in the original thermolytic digest was Leu-Ile-Ala-Asp-Gly corresponding to the residues 293-297 in the reading frame of the DNA confirmed with other peptides, and the possibility of a reading frame error at positions 168-172 can therefore be eliminated. Ten other leucine and iosleucine containing peptides were identified in this experiment and in all cases the sequence assigned by mass spectrometry was in agreement with that predicted from the DNA sequence. We believe that the approach outlined here for the mass spectrometric differentiation of leucine and isoleucine is unambiguous and applicable to all cases where the protein can be produced by growing cells in media doped with labelled leucine or labelled isoleucine.

Acknowledgements

The authors thank P.R. Schimmel and S.D. Putney for many useful discussions and the National Institutes of Health (GM 05472) for financial support.

References

1. Biemann, K. (1978) Pure and Appl. Chem. 50, 149-158.
2. Biemann, K. (1980) in Biochemical Applications of Mass Spectrometry, First Supplementary Volume, ed. by G.R. Waller and O.C. Dermer, pp. 169-525, Wiley, New York.
3. Morris, H.R. (1980) Nature 286, 447-452.
4. Carr, S.A., Herlihy, W.C. and Biemann, K. (1981) Biomed. Mass Spectrom. 8, 51-61.
5. Anderegg, R.J., Biemann, K., Buchi, G., and Cushman, M. (1976) J. Am. Chem. Soc. 98, 3365-3370.
6. Anderegg, R.J., Biemann, K., Manmade, A., and Ghosh, A.C. (1979) Biomed. Mass Spectrom. 6, 129-134.
7. Hudson, G. and Biemann, K. (1976) Biochem. Biophys. Res. Comm. 71, 212-220.
8. Gerber, G.E., Anderegg, R.J., Herlihy, W.C., Gray, C.P., Biemann, K., and Khorana, H.G. (1979) Proc. Natl. Acad. Sci. (USA) 76, 227-231.
9. Khorana, H.G., Gerber, G.E., Herlihy, W.C., Gray, C.P., Anderegg, R.J., Nihei, K., and Biemann, K. (1979) Proc. Amer. Acad. Sci. (USA) 76, 5046-5050.

10. Herlihy, W.C., Anderegg, R.J., and Biemann, K. (1981) Biomed. Mass Spectrom. 8, 62-69.
11. Herlihy, W.C., Royal, N.J., Biemann, K., Putney, S.D., and Schimmel, P.R. (1980) Proc. Natl. Acad. Sci. (USA) 77, 6531-6535.
12. Nau, H. and Biemann, K. (1976) Anal. Biochem. 73, 154-174.
13. Albertson, N.F. and Archer, J. (1945) J. Am. Chem. Soc. 67, 308-310.
14. Putney, S.D., Melendez, D.L., and Schimmel, P.R. (1981) J. Biol. Chem. 256, 205-211.
15. Zubay, G., Chambers, D.A., and Cheong, L.C. (1970) in The Lactose Operon, ed. by J.R. Beckwith and D. Zipser, pp. 375-391, Cold Spring Harbor, New York.
16. Putney, S.D., Sauer, R.T., and Schimmel, P.R. (1981) J. Biol. Chem. 256, 198-204.
17. Putney, S.D., Royal, N.J., Neuman deVegvar, H., Herlihy, W.C., Biemann, K., and Schimmel, P.R. (1981) submitted to Science.
18. Biemann, K., Gapp, F. and Seibl, J. (1959) J. Am. Chem. Soc. 81, 2274-2275.