for a scholarship. Thanks are also due to Mr. J. Zoeller for helpful discussions.

A. H. Cowley, M. H. Hnoosh
Department of Chemistry, University of Texas
Austin, Texas 78712

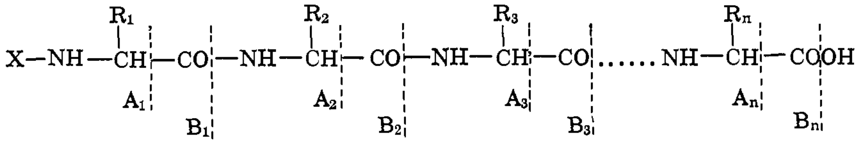
Received March 25, 1966

Computer-Aided Interpretation of High-Resolution Mass Spectra. II.[1] Amino Acid Sequence of Peptides[2]

Sir:

In a continuation of our efforts to make use of computers in the interpretation of complex mass spectrometric data[1] we have selected the determination of the amino acid sequence in peptides as a specific example. This problem seemed to lend itself particularly well to a complete solution even at this early stage because the structures of oligopeptides follow a few strict requirements which can be simply expressed in computer language.

In amino acids and peptides the CH–CO bonds as well as the CO–N bonds cleave relatively easily[3] with retention of the positive charge at the CH and CO groups, respectively, and the mass spectrum of an oligopeptide (I) thus always contains peaks due to "amine fragments" A_1, A_2, A_3, etc., and the molecular ion. The mass of ion A_1 must correspond to that of X–NH–CH plus the substituent at C_α of any one of the amino acids present in the peptide (H for glycine, CH_3 for alanine, etc.). There should also be an ion of mass A_1 plus CONHCH plus another α substituent, etc., and the molecular ion must be of mass A_n + COOH. Cleavage of the CO–NH bonds gives rise to a series of "amino-acyl fragments" B_1, $B_2 \ldots B_n$, differing from the above only by the addition of the mass of CO.



$$X{-}NH{-}\overset{R_1}{\underset{A_1}{CH}}{-}\underset{B_1}{CO}{-}NH{-}\overset{R_2}{\underset{A_2}{CH}}{-}\underset{B_2}{CO}{-}NH{-}\overset{R_3}{\underset{A_3}{CH}}{-}\underset{B_3}{CO}\ldots\ldots NH{-}\overset{R_n}{\underset{A_n}{CH}}{-}\underset{B_n}{CO}OH$$

I

While those peaks may be quite intense, they can easily be overshadowed by others arising from fragments that have lost the N-terminal.[3] It was suggested[3c] that one attaches a homologous pair of long acyl chains to the N-terminus (*e.g.*, N-heptadecanoyl and N-octadecanoyl) to shift the ions containing the N-terminal acid to high mass and relies on the expected high abundance of the amino-acyl ions. This holds for peptides containing uncomplicated, aliphatic amino acids,[4] but the tendency to side-chain fragmentation of polyfunctional amino acids leads sometimes to very intense peaks not fitting this simple scheme. N-Trifluoroacetyl peptide esters[3a,d] often have the tendency to eliminate the trifluoroacetyl group and require a detailed consideration of the mass spectrometric behavior of N–TFA peptides, of metastable ions, etc.[3b] The correct interpretation of conventional mass spectra of peptide derivatives of unknown sequence is thus quite tedious and time consuming. If the mass spectrometric technique is to be applied to the many small peptides produced in the course of the determination of the primary structure of polypeptides of biological interest, a much faster and objective approach is desirable.

A complete high-resolution mass spectrum[5] of a small peptide lends itself particularly well to automatic determination of the amino acid sequence, because the data represent the accurate mass (in the order of millimass units) of *all ions regardless of their relative abundance and significance. Thus, the elemental composition, a structurally unique parameter, becomes the criterion, rather than the more subjective abundances.* Given the exact mass of the N-terminal substituent X plus -NHCH<[6] (*e.g.*, 159.0320 for phthaloyl, 192.9931 for chlorophthaloyl, 125.0088 for trifluoroacetyl, 140.9793 for chlorodifluoroacetyl, 163.0633 for carbobenzoxy, etc.), the computer then adds the masses of the "side chain" of all possible amino acids (*e.g.*, 1.0078 for Gly, 15.0234 for Ala, *etc.*), each time comparing the sum (a possible A_1 fragment) with the masses found. By addition of the mass of CO (27.9949), ion B_1 is searched for. The next amino acid is identified by adding 56.0136 (the "backbone-unit" -CONHCH<) to the mass of ion A_1 plus the accurate masses of all possible "side chains," testing the data for a fit each time, etc., etc. For each fragment of type A found, 44.9977 (CO_2H) is added to test whether the C terminal has been reached (59.0133 for methyl esters). Typical fragments (M − CO_2, M − H_2O, and M − CH_3OH) are tested to recognize peptides exhibiting no molecular ion.

Since secondary fragmentation processes may lead to ions of the same elemental composition as A or B ions (*e.g.*, loss of C_4H_8 or C_3H_6 from leucine leads to ions corresponding to the presence of glycine and alanine, respectively), for this reason the sequence representing the largest fraction (in terms of intensity) of the entire spectrum (*i.e.*, utilizing the more intense peaks of a given type) is considered the best result. Various sequence-specific fragment ions are then searched for as well (see Figure 1).

---

(1) Part I: K. Biemann and W. J. McMurray, *Tetrahedron Letters*, 647 (1965).

(2) This work was supported by grants from the National Institutes of Health (GM 05472 and GM 09352). The use of the facilities of the MIT Computer Center is gratefully acknowledged. We are indebted to Mrs. V. Beecher for writing the program.

(3) (a) E. Stenhagen, *Z. Anal. Chem.*, **181**, 462 (1961); (b) K. Biemann, "Mass Spectrometry," McGraw-Hill Book Co., Inc., New York, N. Y., 1962, p 294; (c) E. Bricas, *et al.*, *Biochemistry*, **4**, 2254 (1965); (d) F. Weygand, A. Prox, H. H. Fessel, and K. K. Sun, *Z. Naturforsch.*, **20B**, 1169 (1965).

(4) (a) M. Barber, P. Jolles, E. Vilkas, and E. Lederer, *Biochem. Biophys. Res. Commun.*, **18**, 469 (1965); M. Barber, *et al.*, *Tetrahedron Letters*, 1331 (1965); (c) G. Laneelle, J. Asselineau, W. A. Wolstenholme, and E. Lederer, *Bull Soc. Chim. France*, (1965) 2133.

(5) Techniques for the automatic recording of accurate mass data in digital computer compatible form have been described previously (P. Bommer, D. Desiderio, W. J. McMurray, and K. Biemann, Twelfth Annual Symposium on Mass Spectrometry, Montreal, Ontario, Canada, June 1964).

(6) NCH for phthaloyl derivatives. An aromatic and/or halogenated substituent X results in a unique elemental composition of the A and B ions, assuring their unambiguous identification.

Oligopeptides containing two to five amino acids, such as Gly, Val, Ala, Leu, Ser, Met, S-Benzyl-Cys, Asp(NH_2), Pro, and Phe, have as yet been tested. Figure 1 shows a typical result produced by the computer, which had been provided with an automatically recorded high-resolution mass spectrum (density profile of a photographic plate), the type of N-terminal substituent, and the accurate masses of the "side chains" of the nineteen amino acids (HO-Pro, Orn, Asp, Glu, Glu(NH_2), Lys, His, Tyr, and Try, in addition to those listed above) to be considered for the sequence.

```
 SEQUENCE FOUND FOR SAMPLE 372-09-4

    CARBOGENZOXY-VAL -APG[a]-LEU -OME -

BASED ON FOLLOWING DATA

AMINE FRAGMENTS
1NTENS1TY[b]          666    29    49    25
ERROR[c]              -2.0   2.9   2.8   1.7
AMINO ACYL FRAGMENTS
1NTENS1TY[b]           66    70     3
ERROR[c]               2.2   1.0   1.4
SUM OF 1NTENS1T1ES IS-   908    TOTAL INTENSITY =  9733[d]
AVERAGE ERROR 1S-  1.986

   INT        FRAGMENT LOST FROM R-VAL-
    19        VAL-H[e]

   1NT        FRAGMENT LOST FROM R-VAL-APG-
     2        NH3
     2        CONH2

   1NT        FRAGMENT LOST FROM R-VAL-APG-LEU-
     1        LEU-H[e]

   1NT        FRAGMENT LOST FROM R-VAL-APG-CO -
    36        CONH2

   INT        FRAGMENT LOST FROM R-VAL-APG-LEU-CO -
     1        C3H6

   INT        FRAGMENT LOST FROM MOLECULAR ION
     3        H2O
     2        CBM[f]
     3        VAL-H
    27        NH3
     1        CONH2
    14        LEU-H
     3        C3H6

SUM OF 1NTENS1T1ES WITH FRAGMENTS LOST   1022
```

Figure 1. Reproduction of the result of a sequence analysis provided by the computer (explanatory footnotes added manually). (a) APG = asparaginyl; (b) in the order of A_1 through A_n, and B_1 through B_n, respectively; (c) difference between calculated and found mass, in millimass units; (d) above mass of R (163 for carbobenzoxy); (e) side chain of valine and leucine, respectively, minus one hydrogen transferred to peptide chain;[3] (f) benzylcarbamate.

In view of the small amount of material required (microgram quantities) and of the extreme speed (1-3 min of computer time) with which this objective and exhaustive interpretation is achieved, this approach shows considerable promise for the routine determination of the amino acid sequence of small peptides obtained upon partial hydrolysis of the oligopeptides resulting from the enzymatic cleavage of large polypeptides. It should also be useful in synthetic work, since the principle is independent of the end groups and additional protecting groups, the mass of which can be read in with the data.

K. Biemann, C. Cone, B. R. Webster

Department of Chemistry, Massachusetts Institute of Technology
Cambridge, Massachusetts

Received March 25, 1966

Enthalpies of Transfer from Water to Dimethyl Sulfoxide for Some Ions and Molecules[1,2]

Sir:

Large changes in the rates of certain reactions have been noted as a result of changing the medium from a hydroxylic to a dipolar aprotic solvent.[3-7] Although the other activation parameters need not parallel the free energy of activation (see below), the relative enthalpies of solvation for typical molecules and ions should be important to an eventual understanding of these medium effects.[8] We present here some preliminary results which furnish strong support for the frequently cited notion of a large difference between cation and anion solvation in the two kinds of systems.[7]

In Table I are presented calorimetric partial molal heats of solution, $\Delta\bar{H}_s$ (defined below), for a few selected nonelectrolytes and salts in water and dimethyl sulfoxide (DMSO) and the derived enthalpy of transfer ($\delta\Delta\bar{H}_s$) for each compound from water to DMSO. The nonelectrolytes give (by this criterion) remarkably perfect solutions in DMSO, the values for all twelve compounds lying between -1.32 and $+1.30$ kcal/mole. The hydrogen-bonding solutes chloroform, water, and methanol show an exothermic interaction whereas the steady endothermic trend for the larger alcohols probably reflects the increasing heat of vaporization which must be supplied to separate their molecules. In the alcohol series, there is a gradual increase in the enthalpies of transfer. This seems to be evidence for a progressive increase in exothermic "structure making"[9] around the larger molecules in water.

The results for the salts are better interpreted by calculating single ion enthalpies of transfer. It is customary to compare cations with each other through salts of a common anion and to construct a similar, but separate, scale for anions. If we make the special assumption that the tetraphenylarsonium cation and the tetraphenylboride anion have equal enthalpies of transfer, it is possible to put anions and cations on a common scale. This is a reasonable assumption since the main difference between the two ions is the sign of their respective charges which lie buried within similar large organic envelopes.[10] These single ion enthalpies of transfer, listed in Table II, bring out the difference in the solvating abilities of water and DMSO. The large negative heats of transfer for the small metal cations indicate that DMSO is the better solvating medium for these ions. Wu and Friedman have recently arrived at an opposite conclusion for another dipolar

---

[1] Solvent Effects in Organic Chemistry. IX. Previous paper is E. M. Arnett and G. Mach, *J. Am. Chem. Soc.*, **88**, 1177 (1966).

[2] Supported by National Science Foundation Grant GP-2014.

[3] C. A. Kingsbury, *J. Am. Chem. Soc.*, **87**, 5409 (1965).

[4] D. J. Cram, "Fundamentals of Carbanion Chemistry," Academic Press Inc., New York, N. Y., 1965 (see index for references to DMSO).

[5] H. E. Zaugg and A. D. Schaefer, *J. Am. Chem. Soc.*, **87**, 1857 (1965); **83**, 837 (1961).

[6] R. Gompper, *Angew. Chem. Intern. Ed. Engl.*, **3**, 560 (1964).

[7] A. J. Parker, *Quart. Rev.*, (London), 163 (1962).

[8] I. P. Evans and A. J. Parker, *Tetrahedron Letters*, 163 (1966).

[9] For key references see previous papers in this series, especially (a) E. M. Arnett and D. R. McKelvey, *Record Chem. Progr.*, **26**, 185 (1965); (b) E. M. Arnett, W. G. Bentrude, J. J. Burke, and P. McC. Duggleby, *J. Am. Chem. Soc.*, **87**, 1541 (1965).

[10] E. Grunwald, G. Baughman, and G. Kohnstam, *ibid.*, **82**, 5801 (1960). Tables I and II indicate that the comparison of $\delta\Delta\bar{H}_s$ for these ions with tetraphenylmethane would be neither grossly in error nor completely exact.