Applications of Artificial Intelligence for Chemical Inference. IX. Analysis of Mixtures without Prior Separation as Illustrated for Estrogens[1]

Dennis H. Smith, Bruce G. Buchanan, Robert S. Engelmore, Herman Adlercreutz, and Carl Djerassi*

Contribution from the Departments of Chemistry and Computer Science, Stanford University, Stanford, California 94305, and Department of Clinical Chemistry, University of Helsinki, Meilahti Hospital, 00290 Helsinki 29, Finland. *Received November 17, 1972*

Abstract: Determination of the structure of each contributor to a complex mixture without prior separation is illustrated. Successful analysis of eight naturally occurring mixtures of estrogenic steroids has been accomplished *via* computerized interpretation of a variety of mass spectral data acquired on underivatized, unseparated mixtures. These data include high resolution mass spectra, low ionizing voltage spectra, and metastable ion spectra. Judicious selection of these data on metastable ions sufficient to distinguish among candidate molecular structures is shown to yield improved performance with concomitant reduction in sample consumption and analysis time. This approach complements conventional methods of analysis of such mixtures as it offers greater structural specificity whereas the technique of combined gas chromatography–low resolution mass spectrometry applied to the analysis of derivatized samples is better suited for quantitation and detection of trace components.

The identities and distributions of estrogens in biological fluids and tissues are important problems, for example, in the determination of fetal well-being and the relationship between estrogen metabolism and disease. Classical methods for clinical determination of estrogens include colorimetric and fluorometric procedures. More recently, gas chromatographic methods[2] (gc) and combined gas chromatography–low resolution mass spectrometry (gc–lrms)[3,4] have been applied to this problem. The gc–lrms technique is a powerful tool for routine analyses and has facilitated the identification of new estrogen metabolites.[3–5] Gas chromatography in connection with a high resolution mass spectrometer (gc–hrms) has been proposed for general analysis of mixtures of steroids.[6]

With the exception of gc–hrms, the above techniques suffer to a greater or lesser degree from an inability to provide structure-specific information, particularly when confronted with hitherto unknown structures. The gc–lrms technique is more specific than gc analysis alone, but cannot avoid ambiguities of elemental composition. However, optimal use of gc-lrms coupled with appropriate reference compounds makes it possible to arrive at a definite structure of an unknown estrogen detected in biological material.[4b] Gc–ms at either high or low resolution requires derivatization of samples to increase volatility for most classes of steroids, including the estrogens. The parent compounds, however, frequently exhibit mass spectra which are more characteristic of molecular structure than suitable gc derivatives. For this reason, a discussion of gc–hrms techniques for steroids has recommended gc analysis of underivatized compounds where possible.[6]

Estrogens are a class of steroids whose fragmentation patterns are intimately related to molecular structure;[7] furthermore, their mass spectra have been subjected to extensive analysis by computer.[8,9] This relationship is sometimes diminished through derivative formation[10] (*e.g.*, trimethylsilyl ethers and especially acetates) as competing fragmentations involving the new functionality may predominate. Unless sufficient reference spectra of standard compounds are available, identification of an unknown derivatized compound is difficult. These considerations suggest that the ability to study unseparated, underivatized mixtures of estrogens (or other classes of compounds) would provide an extra dimension of structure-specific information which would complement gc and gc–ms data. This ability would also make derivatization steps unnecessary thus minimizing sample loss and contamination, simultaneously saving time. Ideally, this analysis could be carried out in the presence of significant amounts of impurities.

A computer program has recently been described which reasons about mass spectral data from first principles[8] (originally referred to as a "planner"). It is immaterial to the planner whether the mass spectral data (high resolution (hrms), metastable ion (MI), and low ionizing voltage (LEV)) and other chemical

(1) For part VIII, see D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, *J. Amer. Chem. Soc.*, **94**, 5962 (1972). This work was supported by a grant (No. RR-612-03) from the National Institutes of Health and a grant from the Ford Foundation, New York (to H. A.).

(2) H. Adlercreutz and T. Luukkainen in "Gas Phase Chromatography of Steroids," K. Eik-Ness and E. C. Horning, Ed., Springer-Verlag, West Berlin, 1968.

(3) H. Adlercreutz and T. Luukkainen, *Advan. Biosci.*, **3**, 53 (1969).

(4) (a) H. Adlercreutz and T. Luukkainen, *Ann. Clin. Res.*, **2**, 365 (1970); (b) H. Adlercreutz and T. Luukkainen, *Z. Klin. Chem. Klin. Biochem.*, **9**, 421 (1971).

(5) T. Luukkainen and H. Adlercreutz, *Biochim. Biophys. Acta*, **107**, 579 (1965).

(6) K. Kaiser, H. Obermann, G. Remberg, M. Spiteller-Friedmann, and G. Spiteller, *Monatsh. Chem.*, **101**, 240 (1970).

(7) C. Djerassi, J. M. Wilson, H. Budzikiewicz and J. W. Chamberlin, *J. Amer. Chem. Soc.*, **84**, 4544 (1962).

(8) D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg and C. Djerassi, *J. Amer. Chem. Soc.*, **94**, 5962 (1972).

(9) D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, J. Lederberg, and C. Djerassi, *Tetrahedron*, in press.

(10) R. A. Okerholm, S. J. Clark, and H. H. Wotiz, *Anal. Biochem.*, **44**, 1 (1971).

data brought to bear on the analysis result from a single compound or mixture of compounds. The planner attempts a detailed structural analysis of each molecular ion (component). This approach may be contrasted with previous applications of high resolution mass spectra to complex mixtures where molecular ions of certain compositions were presumed to represent specific structures or structural types.[11,12] There are, to our knowledge, only two examples in the literature describing approaches toward rationalization of the molecular structure of each component in a complex mixture based on computerized analysis of the high resolution mass spectrum of the mixture.[13-15] One example,[13,14] in the area of sequence determination of each member of a mixture of peptides, points out the limitations of the high resolution mass spectrum alone, and utilizes other mass spectral data acquired on such mixtures (MI data and fractional volatilization curves) to aid analysis.

The approach described herein, in addition to being more systematic and thorough than a manual interpreter (whose basic limitation is patience), also has the advantage of generality.[8] It is not restricted to estrogenic steroids and can potentially be extended to mixtures of other types of compounds.

Results

The philosophy and operation of the analysis program ("estrogen planner") have been discussed in considerable detail previously.[8] The planner's capability for analysis of mixtures was also illustrated for spectra of inadvertent mixtures.[8] For the results described in this report, the planner has been made cognizant of LEV data, which it uses as an aid to the process of inference of molecular ions. This is a valuable supplement to or replacement of metastable ion information in the determination of molecular ions, especially in cases where sample quantities did not allow a significant amount of MI data to be collected.

The general analytical procedure is carried out in a stepwise fashion as illustrated in Figure 1. Because manual examination of data from one step is time consuming relative to the residence time of the sample in the mass spectrometer, separate aliquots of sample are used. The high resolution spectra are recorded as the sample is volatilized slowly from the direct insertion probe of the instrument. Several spectra representative of the qualitative content (containing the same ions, but with varying relative abundances) of the mixture can usually be obtained, one of which is processed further. An LEV spectrum is then recorded at sufficiently low ionizing voltages so that only molecular ions are observed. Based on the results of hrms, attention can be focused on the molecular weight range of interest;[16] scanning the entire spectrum yields no additional information as far as the molecular ions are concerned. This is one point where sample may be conserved.

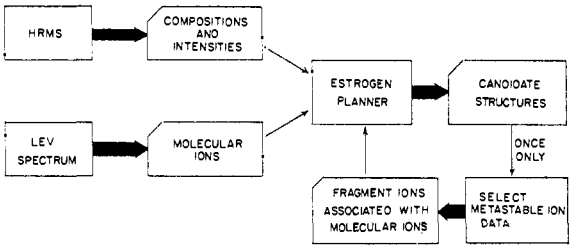


Figure 1. General scheme for analysis of estrogen mixtures. Manual intervention (light arrows) is required in running the samples, setting up the estrogen planner for a run and evaluation of candidate structures to select MI data. Heavy arrows indicate steps performed by a computer program.

The hrms and LEV data, which allow inference of molecular ions, are sufficient to permit a first run through the estrogen planner. This run is usually made with all thresholds and special constraints removed[8] to allow association of all possible fragment ions with each molecular ion. There is no way of associating specific fragments with a particular molecular ion from a single high resolution mass spectrum alone. Thus, it is likely that several of the possible fragment ions associated with a molecular ion do not arise from this molecular ion. The procedure results in varying numbers of candidate structures, including the correct one, unless key fragment ions from a minor component are absent from the spectrum. The structures can be differentiated if necessary on the basis of data on defocused metastable ions (*e.g.*, Barber–Elliott technique[17]) which associates specific fragment ions unambiguously with their respective molecular ions.[18] Note that this differentiation may require only a small number of MI determinations in contrast to the unguided collection of all metastable ion data.[13,14a] This has the result of significant conservation of sample and analytical time while removing ambiguities of structure. Specific examples illustrate these points in more detail (*vide infra*).

The mixtures studied were obtained from pregnancy urine by the method of Adlercreutz and Luukkainen.[19] The procedure has been slightly modified and will be described in detail by Adlercreutz.[20] The most abundant estrogens were quantitated by gc in Helsinki. Gc–lrms analysis of a large number of pregnancy urine samples had revealed the presence of both known and unknown estrogens. Only those estrogens for which some structural information was available are reported in Table I (last column). Nonestrogenic

(11) T. Aczel, D. E. Allan, J. H. Harding, and E. A. Knipp, *Anal. Chem.*, 42, 341 (1970).

(12) W. K. Seifert and R. M. Teeter, *Anal. Chem.*, 42, 750 (1970).

(13) F. W. McLafferty, R. Venkataraghavan, and P. Irving, *Biochem. Biophys. Res. Commun.*, 39, 274 (1970).

(14) (a) F. W. McLafferty, *Advan. Mass Spectrom.*, 5, 589 (1971); (b) H.-K. Wipf, P. Irving, M. McCamish, R. Venkataraghavan, and F. W. McLafferty, *J. Amer. Chem. Soc.*, 95, 3369 (1973).

(15) W. J. Richter, B. R. Simoneit, D. H. Smith, and A. L. Burlingame, *Anal. Chem.*, 41, 1392 (1969).

(16) Estrogens usually display intense molecular ions, which may be picked out directly from the hrms data with relative certainty. This is not true in general, however, as other classes of compounds may show molecular ions of low abundance or fragment ions may masquerade as molecular ions. Either or both LEV and MI data should be obtained if possible as confirmatory evidence.

(17) M. Barber and R. M. Elliott, presented at the 14th Annual Conference on Mass Spectrometry and Allied Topics, Montreal, June 1964.

(18) Previous experiments have established that every key fragmentation of estrogens utilized by the analysis program[8] yields an observable metastable transition in the first field-free region.[17] See (a) D. H. Smith, A. M. Duffield, and C. Djerassi, *Org. Mass Spectrom.*, 7, 367 (1973); (b) D. H. Smith, unpublished results.

(19) H. Adlercreutz and T. Luukkainen in "Gas Chromatography of Hormonal Steroids," R. Scholler and M. F. Jayle, Ed., Gordon and Breach, New York, N. Y., 1968, p 499.

(20) H. Adlercreutz in "Methods of Hormone Analysis," H. Brewer and H. L. Krüskemper, Ed., Georg Thieme Verlag, Stuttgart, in press.

Table I. Comparison of Results Obtained from the Estrogen Planner and Conventional Analysis by Gc and Gc–Lrms

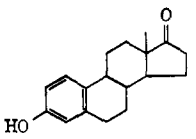
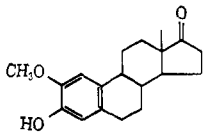
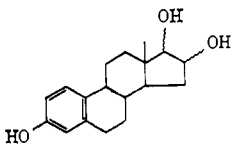
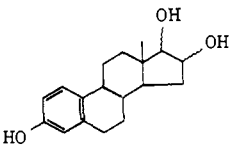
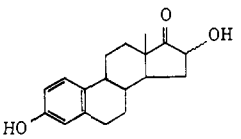
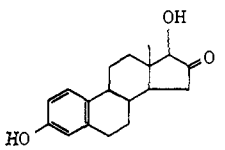
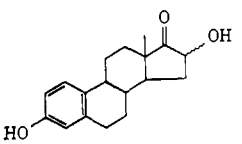
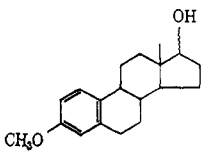
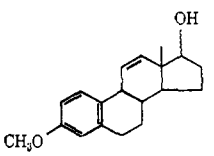
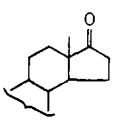
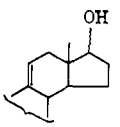
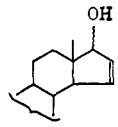
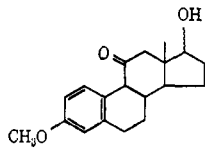
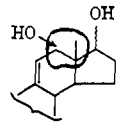
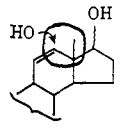
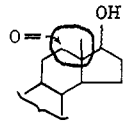
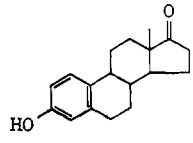
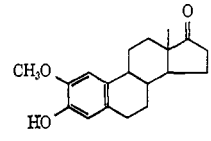
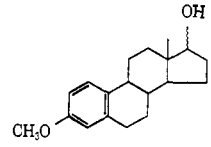
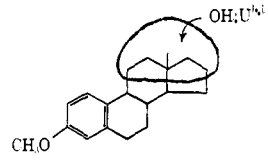
Mixture[a]	Amount,[c] μg	Molecular ions[b]	Results: Estrogen planner	Results: Conventional analysis
A	119	270 ($C_{18}H_{22}O_2$)[d]		80% estrone
		300 ($C_{19}H_{24}O_3$)[e]		20% 2-methoxyestrone
		286 ($C_{18}H_{22}O_3$)	*f*	Not reported
		298 ($C_{19}H_{22}O_3$)	*f*	Not reported
		284 ($C_{18}H_{20}O_3$)	*f*	Not reported
B	99	288 ($C_{18}H_{24}O_3$)[d]		Estriol plus trace amounts of others
C	76	288 ($C_{18}H_{24}O_3$)[e]		Estriol plus trace amounts of others
D	58	286 ($C_{18}H_{22}O_3$)		70% {16α-hydroxyestrone, 16β-hydroxyestrone}
				25% 16-oxoestradiol-17β
				5% 15α-hydroxyestrone
E	34	286 ($C_{18}H_{22}O_3$)		68% {16α-hydroxyestrone, 16β-hydroxyestrone}
				23% 16-oxoestradiol-17β
				9% 15α-hydroxyestrone
F[g]	26	286 ($C_{19}H_{26}O_2$)		~70% {estradiol-17α 3-methyl ether, estradiol-17β 3-methyl ether}
		284 ($C_{19}H_{24}O_2$)		~20% 11-dehydroestradiol-17α 3-methyl ether plus small amounts of several unknowns with one additional double bond
				
				

Table I (*Continued*)

Mixture[a]	Amount,[c] μg	Molecular ions[b]	Results: Estrogen planner	Results: Conventional analysis
				
		300 ($C_{19}H_{24}O_3$)		Unknown
				
				
				
G	24	270 ($C_{18}H_{22}O_2$)		~90% estrone
		300 ($C_{19}H_{24}O_3$)		~10% 2-methoxyestrone
		286 ($C_{18}H_{22}O_3$)	*f*	Not reported
H[g]	14	286 ($C_{19}H_{26}O_2$)		~80% {estradiol-17α 3-methyl ether / estradiol-17β 3-methyl ether}
		284 ($C_{19}H_{24}O_2$)		~20% 11-dehydroestradiol-17α 3-methyl ether plus several unknowns with one or two additional double bonds[3,4b]

[a] See text for more complete discussion. [b] Inferred by the planner from LEV and hrms data. The planner selects the best candidate if more than one ion is present at the nominal mass determined from an LEV experiment, using a set of plausibility criteria (see ref 8). [c] Weight determined by gc quantitation as trimethylsilyl ethers. Prior to computer analysis, the trimethylsilyl groups were removed by mild hydrolysis. [d] Extensive metastable ion data obtained. [e] Some metastable ion data obtained. [f] Molecular ion too weak for reliable analysis without extensive metastable ion data. [g] As 3-methyl ether derivatives. [h] See text for discussion. [i] A hydroxyl, presumably at C-17, and an unsaturation on rings C or D.

material was known to be present in all mixtures but was not reported and is not discussed in the subsequent paragraphs. This material included significant amounts of dioctyl phthalate and other, generally unknown materials. The constitution of the mixtures was unknown, however, to those involved in the computer-aided analyses; they were analyzed as true unknowns. The conclusions of the estrogen planner are compared to the results from conventional analyses in Table I.

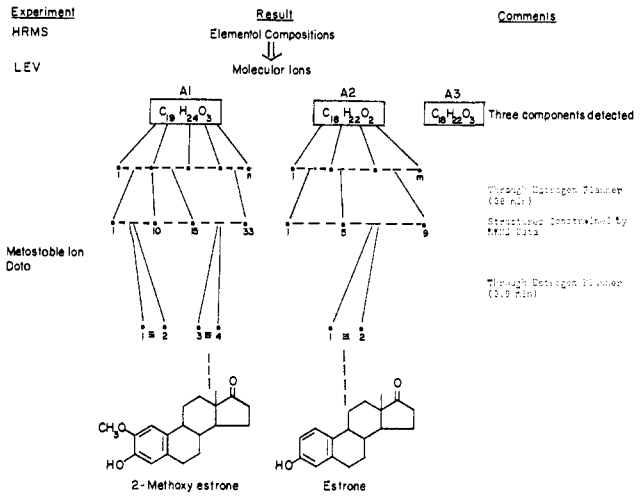


Figure 2. Results obtained for mixture A illustrating the increased specificity and decreased analysis time with input of additional information.

Discussion

General Comments. The mixtures analyzed in this study proved to consist of at most four major components, primarily as the result of extensive prior separations. Although they may not qualify as particularly complex mixtures, the performance of the estrogen planner is certainly encouraging. As noted previously,[8] mass spectra of underivatized estrogens may not permit differentiation between epimeric compounds, particularly if the spectrum is of a mixture of compounds.[21] For this reason, stereochemistry of substituents is not specified in the structures in Table I. As predicted previously,[8] structural analysis of compounds representing minor contributors to the mixture (*e.g.*, *m/e* 286 ($C_{18}H_{22}O_3$), mixtures A and G) is prone to errors resulting from consideration of spurious or ambiguous evidence. Without MI data structural analysis of minor constituents was not attempted, although a determination of the elemental composition is a valuable piece of information in itself.

Mixture A. These results, which agree with conventional analysis, are unambiguous because sample quantities permitted acquisition of extensive metastable data. *M/e* 286 ($C_{18}H_{22}O_3$) is a component not reported in the conventional analysis. LEV data indicate the existence of monounsaturated analogs of *m/e* 286 (*m/e* 284, $C_{18}H_{20}O_3$) and *m/e* 300 (*m/e* 298, $C_{19}H_{22}O_3$). These low abundance molecular ions represent hitherto unknown structures and will be the objects of further investigation.

Mixture B. The single structure which was determined corresponds to estriol, the result obtained from conventional analysis. Other estrogens may be present in this particular fraction. However, the high abundance of estriol relative to other possible contributors has precluded analysis by either technique.

Mixture C. Results are similar to those obtained for mixture B.

Mixture D. The planner found evidence for the two indicated structures, but not for the smallest constituent (15α-hydroxyestrone). Note that all components have the same molecular weight and formula, so that there is only a single molecular ion. The nonepimeric compounds can be differentiated by the planner when the spectra of pure compounds are analyzed. Evidence for both structures (Table I) indicates that both may be present. Confirmatory evidence could in principle be obtained from metastable ion data, but the diminished quantity of material precluded such experiments.

The compounds in mixtures D and E are extremely labile. The time lapse between isolation and final analysis was relatively long and may have resulted in significant destruction of the estrogens, resulting in problems of detection of the compounds present in lesser amounts, notably 15α-hydroxyestrone.

Mixture E. This is a similar fraction to mixture D. The planner determined the correct structure for the major component but had insufficient evidence to build structures for the other components.

Mixture F. A single, correct structure was determined for the major component. The second component (*m/e* 284, $C_{19}H_{24}O_2$) has four possible structures, varying in the placement of the unsaturation in rings C and D, one of which is correct. The four possibilities should be differentiated by the planner when spectra of pure compounds are analyzed (11-dehydroestradiol-17α has never been synthesized so it is not possible to guarantee that its spectrum is unique). In this instance, fragment ions from other molecular ions may masquerade as necessary data for these candidates. Metastable ion data are required to determine whether all or only some of the indicated structures are present. The second structure (estrone methyl ether) is not possible based on the chemistry of the isolation procedure which yields this nonketonic fraction. The unknown component (*m/e* 300, $C_{19}H_{24}O_3$) is unexpected. Candidate structures are indicated in Table I. Because the molecular ion is of low abundance it is possible that fragment ions are absent which would permit other possibilities. The identity of this component is under continuing investigation, but the presence of a ketonic function in the nonketonic fraction is unlikely. Other estrogens with additional double bonds, in agreement with previous work,[3,4b] are suggested by the low ionizing voltage and high resolution mass spectral data.

Mixture G. Results are similar to those obtained for mixture A. Again, there is a previously unreported contributor (*m/e* 286, $C_{18}H_{22}O_3$).

Mixture H. This is a fraction similar to mixture F. In this case, however, a spurious peak from an impurity masqueraded as a key fragment while the correct fragment was absent. This results in some ambiguity in placement of, particularly, the unsaturation for *m/e* 284 ($C_{19}H_{24}O_2$).

Metastable Ion Information. The results described have been obtained with the maximum available data from the complementary techniques of hrms and acquisition of MI and LEV data. Where sample quantities permitted, MI data were acquired.[17] As alluded to previously, manual selection of relevant MI data to be acquired allows easier distinction among candidate structures and more rapid analysis. This point is

(21) Some success has been achieved in differentiation of epimeric trimethylsilyl ether derivatives *via* careful comparison of normal and LEV mass spectra. See H. Adlercreutz and T. Luukkainen, in ref 19, p 93.

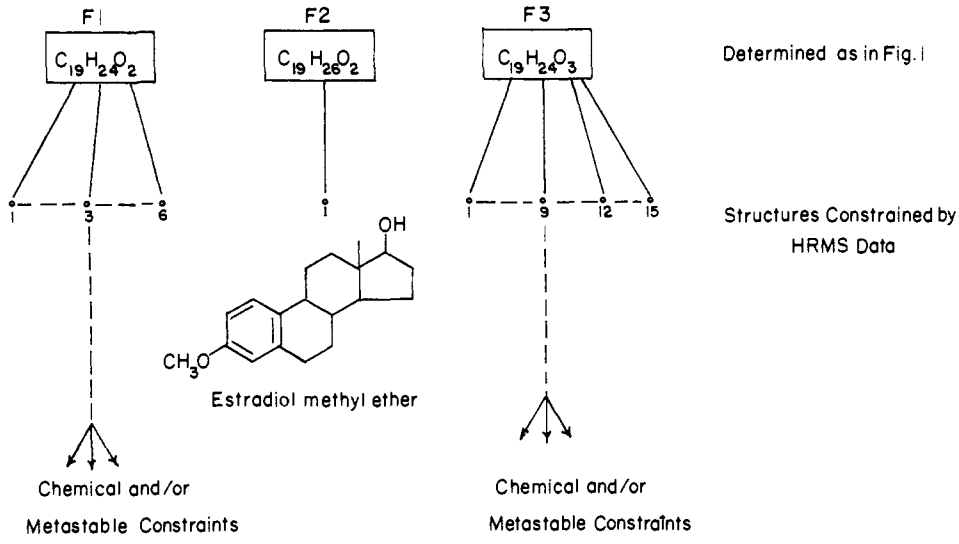


Figure 3. Intermediate results for mixture F indicating that a single, correct structure can sometimes be inferred without additional MI or chemical data.

amply demonstrated with the example of mixture A (Table I). The effects of constraining possible structures with available data are illustrated in Figure 2. Hrms and LEV data result in inference of the indicated molecular ions A1, A2, and A3. Molecular ion A3 was not processed further. There are many possible structures of estrogens (labeled $1 - n$ for A1, $1 - m$ for A2) which are allowed by the molecular formulas alone. Subsequent analysis of the hrms data by the estrogen planner results in 33 structures for A1 and nine structures for A2. Thus the constraints of the hrms data (compositions of all observed fragment ions) allow specification of a greatly reduced set of candidate structures. These results were examined and a set of MI experiments was specified to at least partially distinguish among these structures. These experiments resulted in a complete specification of each of the five key fragment ions[8] for A2, and two out of five key fragment ions for A1. A rerun of the estrogen planner (this time constrained by both hrms and MI data) resulted in a set of four structures for A1 (two pairs of equivalent structures) and two equivalent structures for A2. (The equivalent structures are the enol forms of the C-17 keto group.) "Natural" rules[8] remove one pair of structures for A1, resulting in the final conclusions indicated in Figure 2. The factor of 10 improvement in analysis time (38 min *vs.* 3.9 min) results from the fact that the MI data forbid certain branches from being followed in the tree (Figure 2) representing construction of structures. If a complete set of MI data can be obtained (five determinations), as could be accomplished for molecular ion A2, processing time may be further reduced to typically 10–20 sec per molecular ion. This time reduction becomes very important if automation and control of the experimental procedure are desired, particularly in view of limited sample quantities.

Another example is provided by the case of mixture F (see Figure 3). The constraints of the hrms and LEV data alone yielded six structures for molecular ion F1, one structure for F2, and 15 structures for F3. In this example, molecular ion F2 has a unique solution, and attention can thus be focused on molecular ions F1 and F3 for subsequent collection of MI data if required or possible.

It is notable that the structure of a compound may be inferred in the absence of MI data, as is the case for several examples in Table I. In the general case, however, elimination of all but one candidate structure on chemical grounds may not be possible. In every case, these conclusions can be placed on a much firmer footing by acquisition of MI data.[14b]

Conclusions

As evidenced by the data in Table I, the computer program is capable of excellent performance in analyzing the major constituents of microgram quantities of unknown mixtures of estrogens, even in the presence of significant impurities. It is clear from this study that the gc–lrms technique is much more suited for quantitation of these mixtures and for detection of trace constituents and separation of epimeric compounds. The specificity of the estrogen planner in terms of molecular structure provides important data which complement and extend results obtained from the more classical approaches.

The data provided by the estrogen planner suggest further mass fragmentographic[22] and mass chromatographic[23] studies on similar fractions. This makes it possible to localize, in the gc traces, previously undetected compounds, the presence of which is suggested by the estrogen planner, and to further elaborate their structures by monitoring specific ions. Subsequent purification and separation of the compounds in question will provide more material for detailed structural studies utilizing the estrogen planner.

It would be feasible to automate all steps indicated in Figure 1. Automated collection of metastable data has been described previously.[13,14] Control of a procedure of this type, directed by intermediate results of the estrogen planner, would provide a powerful analytical tool for structure elucidation based on

(22) (a) C. C. Sweeley, W. H. Elliott, J. Fries, and R. Ryhage, *Anal. Chem.*, **38**, 1549 (1966); (b) C.-G. Hammar, B. Holmstedt, and R. Ryhage, *Anal. Biochem.*, **25**, 532 (1968).

(23) (a) R. A. Hites and K. Biemann, *Anal. Chem.*, **42**, 855 (1970); (b) R. Reimendal and J. Sjövall, *ibid.*, **44**, 21 (1972).

mass spectral data. In addition, elimination of manual interpretation and manual instrument control would permit all necessary experiments to be performed on a single aliquot of sample. This would allow complete structural analysis of considerably smaller sample quantities (potentially <1 μg total mixture) thus conserving precious samples.

Experimental Section

High resolution mass spectra were obtained utilizing a Varian-MAT 711 mass spectrometer operated at an ionizing voltage of 70 eV, an ionizing current of 1.6 mA, and a scan rate of 22 sec/decade at a nominal resolving power of 10,000. LEV spectra were recorded at an ionizing voltage of 13.5 eV (uncorrected). This value was determined (utilizing standard samples of estrone, estrone methyl ether, and estradiol) to yield only molecular ion peaks in the molecular ion region of estrogens (above $C_{18}H_{24}$[8]). First field-free region metastable ions were analyzed on the Varian-MAT 711, with additional metastable ion data collected utilizing a Varian-MAT 311 (second field-free region) (see Acknowledgment).

The computer times mentioned above are those for an IBM 360/50 at the Stanford Medical School's ACME facility, a computer which is four to five times slower for this program than the computer utilized in a previous study (IBM 360/67).[8]

Acknowledgment. We are indebted to Dr. A. L. Burlingame, Space Sciences Laboratory, University of California, Berkeley, for making available to us time on a Varian-MAT 311 mass spectrometer for determination of some of the metastable ion data obtained for mixtures A–C.

Rotamer Stability of Histidine and Histidine Derivatives[1]

Robert J. Weinkam and Eugene C. Jorgensen*

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143.
Received August 22, 1972

Abstract: The nuclear magnetic resonance vicinal coupling constants for histidine and *im*-benzyl-, *N*-acyl-, and *O*-methylhistidines have been obtained in deuterium oxide solutions at acidic, neutral, and basic pH. Equations used in calculating relative rotamer populations have been derived which include terms for the functions of dihedral angles, substituent electronegativities, and orientation in vicinal coupling. A model is presented in which the rotamer populations of histidine and its derivatives are determined by an electrostatic interaction between the carboxylate anion and the imidazole ring, which stabilizes the conformation in which these groups are in close proximity.

As a result of our interest in the solution structure of histidine analogs and peptides which contain these analogs,[2] we have reexamined the proton magnetic resonance data concerning histidine and histidine derivatives. An understanding of histidine conformations and the relative importance of the interactions which determine these orientations were found to be important in problems of peptide secondary and tertiary structure.[2] Proton magnetic resonance techniques are particularly well suited to an investigation of these questions. Previous papers have described the relative populations of rotamer conformations[3] of histidine and factors which affect the energy differences.[4] Others have noted the variations of pmr[5] and ^{13}C nmr[6] parameters with pH and have attempted to correlate these with electronic structure.[7] Pmr studies attempting to describe the structure of the low energy histidine rotamer require assignment of the frequency of the β-methylene hydrogens. The chemical shift differences of these protons are small and the reasons given for the assignments have not been convincing.[3b,8] As a result of this difficulty, the conformation of the preferred histidine rotamer has not been established.

Solid-state conformations as determined by X-ray crystallography[9] indicate that energy differences separating conformers are comparable to the hydrogen bonding and packing energies in the crystal. A unit cell of L-*N*-acetylhistidine contains both an open (imidazole and carboxylate groups are trans) and closed (adjacent) conformation. As the identification of the preferred amino acid conformer is necessary for application to problems of peptide solution structure, we have examined this aspect in the pmr studies in this paper.

Work with histidine analogs[2a,b] in which the 4-imidazolyl ring was replaced by the stable free radical 1,3-dioxy-4,4,5,5-tetramethyldihydro-2-imidazoloyl showed electron spin resonance line broadening associated with low rotation rates[10] when the carboxyl group was in the anionic form. On protonation esr

(1) Supported in part by the Public Health Service Research Grant AM 08066 from the National Institute of Arthritis and Metabolic Disease.

(2) (a) R. J. Weinkam and E. C. Jorgensen, *J. Amer. Chem. Soc.*, **93**, 7028 (1971); (b) *ibid.*, **93**, 7033 (1971); (c) *ibid.*, **93**, 7038 (1971); (d) E. C. Jorgensen and R. J. Weinkam, *Peptides, Proc. Eur. Peptide Symp., 11th*, 311 (1973).

(3) (a) K. G. R. Pachler, *Spectrochim. Acta*, **20**, 581 (1964); (b) R. B. Martin and R. Mathur, *J. Amer. Chem. Soc.*, **87**, 1065 (1965); (c) J. J. M. Rowe, J. Hinton, and K. L. Rowe, *Chem. Rev.*, **70**, 1 (1970).

(4) J. R. Cavanaugh, *J. Amer. Chem. Soc.*, **92**, 1488 (1970).

(5) (a) F. Taddei and L. Pratt, *J. Chem. Soc.*, 1533 (1964); (b) K. G. R. Pachler, *Spectrochim. Acta*, **19**, 2085 (1963).

(6) W. Horsley, H. Sternlicht, and J. S. Cohen, *J. Amer. Chem. Soc.*, **92**, 680 (1970).

(7) G. Del Re, B. Pullman, and T. Yonezawa, *Biochim. Biophys. Acta*, **75**, 153 (1963).

(8) K. D. Kopple and M. Oshnishi, *J. Amer. Chem. Soc.*, **91**, 962 (1969).

(9) T. J. Kistenmacher and R. E. March, *Science*, **172**, 945 (1971).

(10) R. W. Kreilick, J. Becher, and E. F. Ullman, *J. Amer. Chem. Soc.*, **91**, 5122 (1969).