THE AMINO ACID SEQUENCE AND CONFIGURATION OF TENTOXIN

W. L. Meyer,[1] L. F. Kuyper,[1] R. B. Lewis, G. E. Templeton[2]
and S. H. Woodhead[2]

Departments of [1]Chemistry and [2]Plant Pathology, University of Arkansas,
Fayetteville, Arkansas 72701

Received November 6, 1973

Summary. - The full structure cyclo[L-leucyl-N-methyl-trans-dehydrophenylalanyl-glycyl-N-methyl-L-alanyl] (IV) is assigned to tentoxin, the chlorosis-inducing phytotoxin from Alternaria tenuis Nees., on the basis of ^{1}H nmr, mass spectrometric, and degradative data from tentoxin and its N,N-dimethyl, dihydro, and N,N-dimethyldihydro derivatives.

Earlier we reported evidence to show that tentoxin, the phytotoxic peptide from Alternaria tenuis Nees. (1-5), is a cyclic tetrapeptide containing one unit each of glycine, leucine, N-methylalanine, and N-methyldehydrophenylalanine (6). The recent report (7) that tentoxin corresponds to the amino acid sequence cyclo-[N-methyldehydrophenylalanyl-L-leucyl-glycyl-N-methyl-L-alanyl] prompts us to present in abbreviated form some of our own results on this subject, for they are not in complete accord with those of the Wisconsin group (7).

Materials and Methods. - Crystalline tentoxin (T,* mp 172-175° (3-6)) was hydrogenated at 1 atm in 95% ethanol over 5% palladium on carbon to produce dihydrotentoxin (TH_2, mp 298-300°; found: C, 63.23; H, 7.56; N, 13.31. $C_{22}H_{32}N_4O_4$ requires C, 63.44; H, 7.74; N, 13.45), in which the olefinic bond has been saturated. N-Methylation of T and TH_2 using methyl iodide and sodium hydride in dimethylformamide (8) afforded respectively N,N-dimethyltentoxin (TMe_2, mp 104-108°; m/e of molecular ion 444.2577; $C_{24}H_{34}N_4O_4$ requires 442.2578) and N,N-dimethyldihydrotentoxin (TMe_2H_2, mp 130-133°; found; C, 64.68; H, 8.11; N, 12.60. $C_{24}H_{36}N_4O_4$ requires C, 64.84; H, 8.16; N, 12.60).

*Compound abbreviations are used as follows: T = tentoxin; TH_2 = dihydrotentoxin; TMe_2 = dimethyltentoxin; TMe_2H_2 = N,N-dimethyldihydrotentoxin.

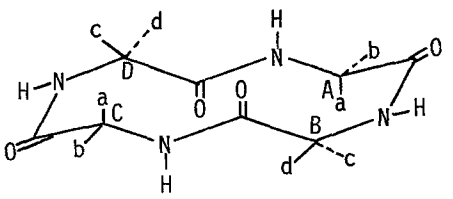
^{1}H Nmr spectra show each of these substances to be single diastereomers. Analogous N-methylation of T and TH_2 using methyl iodide-d_3 provided the hexadeutero derivatives of TMe_2 and TMe_2H_2.

Hydrolysis of T and TH_2 with 6N hydrochloric acid at 100° for 24 hours afforded amino acid mixtures which were treated with L-leucine N-carboxy anhydride to produce mixtures of the corresponding L-leucyl dipeptides (9). These were compared by ion exchange chromatography (9) with analogous L-leucylation products of authentic glycine, sarcosine, D- and L-leucine, D- and L-N-methylalanine (10), and D- and L-N-methylphenylalanine (10).

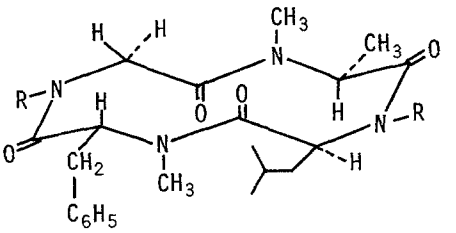
^{1}H Nmr spectra were obtained at 60 and 90 MHz from chloroform-d solutions using Varian A-60 and Bruker HFX-90 spectrometers respectively. High resolution mass spectra were obtained from the Florida State University Mass Spectrometry Center; an independent determination of the masses of ions at m/e 214, 215, and 216 in the T spectrum was kindly provided by Professor K. L. Rinehart of the University of Illinois.

Results and Discussion. - Analysis (9) of the L-leucyl dipeptide mixture from hydrolysis of TH_2 showed the N-methylphenylalanine to be of the D-configuration. N-Methylalanine and leucine from both TH_2 and T proved to be L, in accord with the Wisconsin result with T (7) and our own earlier report concerning the leucyl unit of T (11).

^{1}H Nmr spectra of T, TH_2, TMe_2, and TMe_2H_2 are well resolved, and all chemical shifts and coupling constants except those associated with leucine β- and γ-protons can be assigned with the assistance of spectra of the two d_6 derivatives and appropriate spin-decoupling and INDOR experiments. The nearly uniform similarity of these properties of tentoxin and its derivatives with those reported for a series of simpler cyclic tetrapeptides (12) leads to the deduction that the tentoxin compounds also occupy the typical (12) cyclotetra-

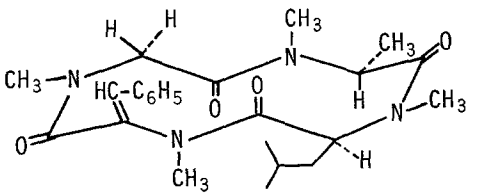


I

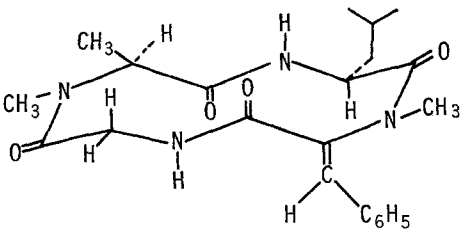


II; TH_2: R = H

TMe_2H_2: R = CH_3



III



IV

peptide conformation, I, which is based on an X-ray diffraction study of cyclotetrasarcosyl (13). Thus these spectra can be analyzed in terms of chemical shifts and coupling constants which are characteristic (12) of the various proton types in this system (indicated a - d in I). Sufficient elements of this analysis are presented here both to illustrate its method and potential as a technique of peptide sequencing and to allow assignment of a unique structure to tentoxin.

In TH_2 and TMe_2H_2 respectively the α-proton chemical shifts of the alanyl (τ_α = 4.79; 4.73) and phenylalanyl (τ_α = 4.35; 4.50) residues are at too low field to correspond to any but type a protons of conformer I (12). Accordingly these two residues must be non-adjacent, and since alanine is L and phenylalanine D, the former may be assigned to location A and the latter to location C in structure I (the a-proton in I is that of an L-α-amino acid at A but of a D-α-amino acid at C; locations B and D are similarly related). In accord with this deduction are the observations that in both compounds the glycine α-proton shifts and geminal couplings correspond (12) to a B or D location (TH_2:

τ_α = 5.79 and 6.43, J_{gem} = 18.0 Hz; $\underline{TMe_2H_2}$: τ_α = 5.75 and 6.55, J_{gem} = 18.0 Hz), as do the vicinal NH:CH couplings in $\underline{TH_2}$ ($J_{N\alpha}$ = 9.2 and 6.0 Hz) (14). Inasmuch as the shifts and vicinal NH:CH coupling of leucine's α-proton in this series ($\underline{TH_2}$: τ_α = 5.95, $J_{N\alpha}$ = 8 Hz; $\underline{TMe_2H_2}$: τ_α = 5.78) are characteristic (12, 14) of a type c proton, and since the leucine is L, it may be assigned to location B. Therefore glycine is placed at D, and the sequence derived for $\underline{TH_2}$ and $\underline{TMe_2H_2}$ is cyclo[L-MeAla-L-Leu-D-MePhe-Gly], with the conformation II. None of the reactions leading to these compounds from I could have altered the sequence, and accordingly this sequence (but not necessarily conformation) may also be assigned to I itself.

Analogous analysis of spectra of I and $\underline{TMe_2}$, with appropriate consideration of the shielding influence of the C=C in these molecules, leads to similar sequence deductions. Thus in $\underline{TMe_2}$ leucine may be assigned to location B (being L with a type c τ_α of 5.60), glycine to D (type cd J_{gem} of 17.5 Hz and type cd τ_α's of 5.76 and 6.53), and alanine to A (being L with a type a τ_α of 4.79), leaving the dehydrophenylalanyl unit at C (III). Interestingly I itself clearly differs in conformation from these three derivatives. Its spectrum remains in good accord with the basic system I, but leucine must be assigned to location A (being L with a type a α-proton shielded by the C=C, τ_α 5.97, and a type a $J_{N\alpha}$ of 9.0 Hz), dehydrophenylalanine to B (so its C=C can shield leucine's α-proton), and glycine to C (type ab J_{gem} of 15.0 Hz, type ab τ_α's of 4.93 and 6.52, and type ab $J_{N\alpha}$'s of 10 and < 2 Hz), leaving alanine at D (IV). This is, of course, the amino acid sequence which was deduced independently in the other three cases.

The configuration of CO and C_6H_5 about the C=C bond of I and $\underline{TMe_2}$ may be assigned as trans on the basis of the remarkable difference in vinyl proton chemical shifts between I (τ 2.45) and $\underline{TMe_2}$ (τ 3.24), among other evidence. The conformational change of I (IV) to $\underline{TMe_2}$ (III) clearly moves that proton out of a very deshielding environment, which would be provided by the nodal plane of the DehydroPhe-Gly peptide bond if the vinyl proton is cis, but not trans, to it in IV.

High resolution mass spectra of tentoxin and these derivatives independently lead us to the deduction that the sequence of T is cyclo[Leu-MeDehydroPhe-Gly-MeAla]. Structures for each of the six possible sequences of each of the four compounds were dissected (by computer so as to avoid inadvertent omission of any possibility) to derive molecular and hypothetical structural formulas for all fragments which could result from cleavage of every possible pair of ring bonds, with or without rearrangement of one hydrogen to or from the ionic fragment. All ions observed with significant intensity (>2% of base peak) were compared with each of these six lists of potential fragments which could derive from the six sequences without major rearrangements or side-chain cleavages, taking into account also in each case the isotopic shifts expected for deuteromethylation. Of 81 observed fragments from T, TMe_2, TH_2, and TMe_2H_2 which can be correlated on this basis with at least one, but not all six sequences, all but 8 are accomodated by cyclo[Leu-MeDehydroPhe-Gly-MeAla] and its derivatives, whereas the five alternatives leave 29-43 ions unexplained in these terms. Furthermore, these 8 residual ions can be readily derived from the assigned sequence by appropriate side-chain fragmentation processes.

Although space does not permit detailed discussion of the mass spectrometric data at this point, brief comment is in order concerning cyclo[Gly-MeAla-MeDehydroPhe-Leu], the sequence assigned by the Wisconsin group from data which included mass spectrometry (7). First, the ion $C_{13}H_{12}NO_2$ in the spectrum of T, which was formulated as CH_2=CHCONMeC(=CHPh)CO^+ and used as evidence for the segment MeAla-MeDehydroPhe (7) can derive as well from Leu-MeDehydroPhe by loss of an isopropyl radical. The only other two ions reported (7) to require a MeAla-MeDehydroPhe segment were m/e 215 and 216, described as $C_{13}H_{13}NO_2$ and $C_{13}H_{14}NO_2$. Our data indicate these fragments to be minor components of m/e 215 and 216 multiplets, the major components of which are $C_{12}H_{11}N_2O_2$ and $C_{12}H_{12}N_2O_2$ respectively. The latter compositions correspond to a MeDehydroPhe-Gly segment rather than MeAla-MeDehydroPhe, and since they include the glycine nitrogen should shift to $C_{13}H_{13}N_2O_2$, $C_{13}H_{10}D_3N_2O_2$, $C_{13}H_{14}N_2O_2$, and $C_{13}H_{11}D_3N_2O_2$

in spectra of $\underline{TMe_2}$ and its $\underline{d_6}$ derivative, as observed. Furthermore, $C_{11}H_{12}N_2O$, one of the ten most intense peaks from $\underline{T}$, and the homologous $C_{12}H_{14}N_2O$ and $C_{12}H_{11}D_3N_2O$, which are base peaks from $\underline{TMe_2}$ and its $\underline{d_6}$ derivative, are directly derivable from a MeDehydroPhe-Gly segment, and $C_{16}H_{20}N_2O_3$ and $C_{16}H_{17}D_3N_2O_3$, which are base peaks from $\underline{TMe_2H_2}$ and its $\underline{d_6}$ derivative, correspond readily to MeDehydroPhe-Gly-MeAla; these ions are among the 43 for which rearrangement or side-chain fragmentation interpretations must be sought for the Wisconsin sequence (7).

Accordingly, we propose for tentoxin the complete structure cyclo[L-leucyl-N-methyl-trans-dehydrophenylalanyl-glycyl-N-methyl-L-alanyl], with the conformation IV in chloroform solution. An X-ray crystallographic structure determination to confirm this assignment is in progress.

ACKNOWLEDGMENTS

We are grateful to Professor K. L. Rinehart of the University of Illinois for an independent determination of the masses of the m/e 214, 215, and 216 ions in the tentoxin spectrum. This research was supported in part by research grant CSRS-916-15-24 to the Arkansas Agricultural Experiment Station. The Bruker NMR spectrometer was obtained with the partial assistance of National Science Foundation Grant GP-18291 to the Department of Chemistry.

REFERENCES

1. N. D. Fulton, K. Bollenbacher, and G. E. Templeton, Phytopathology, 55, 49 (1965).

2. G. E. Templeton, C. I. Grable, N. D. Fulton, and K. Bollenbacher, Phytopathology, 55, 1079 (1965); 57, 516 (1967).

3. C. I. Grable, G. E. Templeton, and W. L. Meyer, Phytopathology, 56, 897 (1966).

4. G. E. Templeton, C. I. Grable, N. D. Fulton, and W. L. Meyer, Proceedings of the Mycotoxin Research Seminar, U. S. Dept. of Agriculture, Washington, D. C., June 8-9, 1967, p. 27.

5. G. E. Templeton, W. L. Meyer, C. I. Grable, and C. W. Sigel, Phytopathology, 57, 833 (1967).

6. W. L. Meyer, G. E. Templeton, C. I. Grable, C. W. Sigel, R. Jones, S. H. Woodhead, and C. Sauer, Tetrahedron Lett., 2357 (1971).

7. M. Koncewicz, P. Mathiaparanam, T. F. Uchytil, L. Sparapano, J. Tam, D. H. Rich, and R. D. Durbin, Biochem. Biophys. Res. Comm., 53, 653 (1973).

8. G. Marino, L. Valente, R. A. W. Johnstone, and F. Mohammedi-Tabrizi, Chem. Comm., 357 (1972).

9. J. M. Manning and S. Moore, J. Biol. Chem., 243, 5591 (1968).

10. P. Quitt, J. Hellerbach, and K. Bolger, Helv. Chim. Acta, 46, 327 (1963).

11. J. T. Daniel and G. E. Templeton, Proc. Arkansas Acad. Sci., 24, 47 (1970).

12. J. Dale and K. Titlestad, Chem. Comm., 656 (1969); 1403 (1970); 255 (1972).

13. P. Groth, Acta Chem. Scand., 24, 780 (1970).

14. G. N. Ramachandran, R. Chandrasekaran, and K. D. Kopple, Biopolymers, 10, 2113 (1971).