A THEORETICAL ANALYSIS OF THE CONFORMATION OF Ac-L-Ala-L-Pro-L-Ala-NHMe WITH INTRAMOLECULAR HYDROGEN BONDS

É. P. Gromov and V. Z. Pletnev

UDC 547.962.541

We have previously [1] made a detailed study of the spatial structure of the tripeptide compound Ac-L-Ala-L-Pro-L-Ala-NHMe without hydrogen bonds and have compared the results with the calculated conformations of the L-Ala-L-Pro and L-Pro-L-Ala fragments [2]. Analysis has shown that the most favorable forms of the tripeptide are combinations of the low-energy complementary forms of the dipeptide fragments. The aim of the present investigation was to study the conformations of the tripeptide compound Ac-L-Ala-L-Pro-L-Ala-NHMe with intramolecular hydrogen bonds. The interest in this investigation is due not only to the possibility of predicting the optimum conformations of a given concrete compound in the case of nonpolar media but also of finding canonical forms with H bonds realized in sections of peptide–protein systems with proline residues. In addition, a comparison of the forms of Ac-L-Ala-L-Pro-L-Ala-NHMe found with the conformational states of the dipeptide fragments -Ala-Pro and -Pro-Ala [2] will make it possible to evaluate the conformational role of the interaction of the terminal residues in the structure of the tripeptide investigated.

Figure 1 shows a model of the Ac-L-Ala-L-Pro-L-Ala-NHMe molecule with the potentially possible intramolecular hydrogen bonds (I-V). Table 1 gives the letter symbols of the conformations considered with eight probable systems of H bonds (S1-S8). The letters R, B, L, and P denote the regions of low potential energies on the (φ, ψ) conformational map. The R form is found in the first quadrant of the conformational map ($\varphi = \psi = -180$–$0°$), B in the second quadrant ($\varphi = -180$–$0°$; $\psi = 0$–$180°$), L in the third quadrant ($\varphi = \psi = 0$–$180°$), and P in the fourth quadrant ($\varphi = 0$–$180°$, $\psi = -180$–$0°$). The letters M and H denote the convoluted forms of the residue with a hydrogen bond of the 3–1 type and corresponds to the same low-energy regions as B and P. The brace connecting the first and third letters or two neighboring letters shows the presence of a 5–1 or 4–1 H bond forming a 13-membered or a 10-membered ring, respectively. Primes attached to the letter symbols denote the cis configuration of the tertiary amide group. Thus, three letters and a brace above them show the conformational type of the molecule with the corresponding system of H bonds.

The optimum conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe were found by minimizing the total potential energy over the 11 angles of rotation — ω_0, φ_1, ψ_1, ω_1, χ_1, ψ_2, ω_2, φ_3, ψ_3, ω_3, χ_3

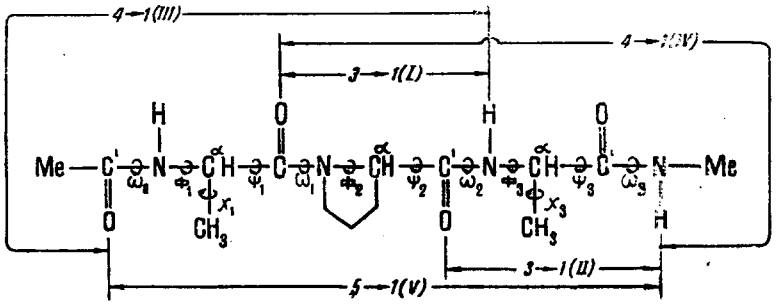

Fig. 1. Model for calculation of the methylamide of N-acetyl-L-alanyl-L-prolyl-L-alanine with intermolecular hydrogen bonds.

M. M. Shemyakin Institute of Bioorganic Chemistry, Moscow. Translated from Khimiya Prirodnykh Soedinenii, No. 2, pp. 233-238, March-April, 1976. Original article submitted June 13, 1975.

TABLE 1. Probable Forms of Ac-L-Ala-L-Pro-L-Ala-NHMe with Various Systems of Intramolecular Hydrogen Bonds

System	Number of H bonds	Type of H bonds	Probable conformations with H bonds*
S1	1	5—1 (V)	$\overline{R{-}R{-}R^{1}}$
S2	1	4—1 (IV)	$B{-}\overline{R{-}R^{2}}$, $B{-}\overline{R{-}B^{3}}$, $L{-}\overline{R{-}R^{4}}$, $L{-}\overline{B{-}L^{5}}$, $B{-}\overline{B{-}L^{6}}$, $L{-}\overline{R{-}B^{7}}$, $R{-}\overline{R{-}R^{8}}$, $R{-}\overline{R{-}B^{9}}$, $R{-}\overline{B{-}L^{10}}$
S3	1	3—1 (II)	$B'{-}R{-}M^{11}$, $B{-}R{-}M^{12}$, $L{-}R{-}M^{13}$, $B{-}R{-}H^{14}$, $L{-}R{-}H^{15}$, $B{-}B{-}H^{16}$, $L{-}B{-}H^{17}$, $R{-}R{-}M^{18}$, $L{-}B{-}M^{19}$, $B{-}B{-}M^{20}$, $B'{-}R{-}H^{21}$, $L'{-}B{-}M^{22}$, $B'{-}B{-}M^{23}$, $B'{-}B{-}H^{24}$, $R{-}R{-}H^{25}$, $L'{-}R{-}M^{26}$, $R{-}B{-}H^{27}$, $L'{-}B{-}H^{28}$, $R{-}B{-}M^{29}$, $L'{-}R{-}H^{30}$, $R'{-}B{-}H^{31}$, $R'{-}R{-}M^{32}$, $R{-}R{-}H^{33}$, $R'{-}B{-}M^{34}$
S4	2	4—1 (III, IV)	$\overline{R{-}R{-}R^{35}}$, $\overline{R{-}R{-}B^{36}}$, $\overline{P{-}R{-}B^{37}}$, $\overline{P{-}R{-}R^{38}}$
S5	1	4—1 (III)	$\overline{R{-}R{-}R^{39}}$, $\overline{R{-}R{-}B^{40}}$, $\overline{B'{-}R{-}L^{41}}$, $\overline{R{-}R{-}L^{42}}$, $\overline{B'{-}R{-}R^{43}}$, $\overline{B'{-}R{-}B^{44}}$, $\overline{B'{-}B{-}R^{45}}$, $\overline{B'{-}B{-}B^{46}}$, $\overline{B'{-}B{-}L^{47}}$, $\overline{L'{-}R{-}B^{48}}$, $\overline{L'{-}R{-}L^{49}}$, $\overline{L{-}B{-}B^{50}}$, $\overline{L{-}B{-}L^{51}}$, $\overline{L'{-}R{-}R^{52}}$, $\overline{L{-}B{-}R^{53}}$
S6	1	3—1 (I)	$B{-}M{-}L^{54}$, $L{-}M{-}L^{55}$, $L{-}M{-}R^{56}$, $B{-}M{-}R^{57}$, $L{-}M{-}B^{58}$, $B{-}M{-}B^{59}$, $R{-}M{-}L^{60}$, $R{-}M{-}R^{61}$, $R{-}M{-}B^{62}$
S7	2	3—1 (I) 3—1 (II)	$L{-}M{-}M^{63}$, $B{-}M{-}M^{64}$, $R{-}M{-}M^{65}$
S8	2	4—1 (III) 3—1 (II)	$B'{-}\overline{R{-}M^{66}}$, $\overline{R{-}R{-}H^{67}}$, $\overline{R{-}R{-}M^{68}}$, $B'{-}\overline{R{-}H^{69}}$, $\overline{B'{-}B{-}M^{70}}$, $B'{-}\overline{B{-}H^{71}}$, $\overline{L'{-}R{-}M^{72}}$, $L'{-}\overline{R{-}H^{73}}$, $\overline{R{-}B{-}H^{74}}$, $\overline{R{-}B{-}M^{75}}$

*The numbers of the conformations are shown as superscripts to the right of the letter symbols.

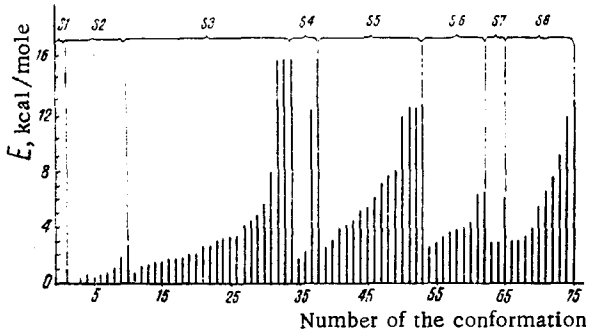

Fig. 2. Energy distribution of the potentially probable conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe with the S_1-S_8 systems of hydrogen bonds.

(see Fig. 1). In the calculation of the total potential energy, we took into account the energy contributions from the nonvalent and electrostatic interactions of the atoms, inhibited rotation around the valence bonds, and hydrogen bonding. The parameters and functions used in the calculation were taken from the literature [1-5]. As the zero approximations we took the optimized values of the angular parameters of the complementary conformations of the hydrogen-bound dipeptides Ac-L-Ala-L-Pro-NHMe and Ac-L-Pro-L-Ala-NHMe [2]. The

TABLE 2. Geometric Parameters (degrees)* and Energies (kcal/mole) of the Optimum Conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe with Intramolecular Hydrogen Bonds

Type of conformations	ω_0	φ_1	ψ_1	ω_1	χ_1	φ_2	ψ_2	ω_2	φ_3	ψ_3	ω_3	χ_3	$L_{term.}$, Å	E_{tot} ($\varepsilon = 4$)
B—R—R	179,7	—66,8	152,3	174,5	—59,8	—55	—28,3	174,5	—55,0	—33,3	—179,2	59,4	6,2	0
B—R—B	179,7	—67,3	153,2	175,2	59,6	—55	—36,6	173,1	—103,2	32,9	—179,7	—58,3	5,5	0,2
L—R—R	179,9	52,5	65,5	—178,7	59,9	—60	—27,4	174,1	—57,2	—30,5	—179,2	59,9	8,2	0,4
L—B—L	179,9	52,4	65,8	178,3	60,0	—55	109,5	—176,0	54,3	32,9	—179,4	—56,8	8,1	0,4
B—B—L	179,8	—65,4	151,8	177,8	59,2	—55	110,7	—176,3	54,4	32,0	—179,7	—56,8	6,4	0,5
B′—R—M	—178,2	—50,5	149,2	—3,7	58,6	—70	—26,0	179,2	—73,4	54,8	—177,7	—59,3	3,9	0,6
L—R—B	179,9	52,7	65,9	—179,6	—59,1	—55	—39,3	176,0	—98,9	31,2	179,2	—58,2	7,8	0,8
R—R—R	178,8	—54,5	—53,0	177,9	22,4	—65	—39,7	179,7	—59,0	—39,2	178,9	—58,3	5,9	0,9
B—R—M	179,7	—67,6	152,0	176,6	59,4	—70	—41,5	177,1	—70,7	62,0	179,9	59,9	7,7	1,0
R—R—R	—178,5	—60,1	—48,8	—169,2	—49,3	—60	—30,8	176,7	—58,7	—32,0	—179,1	—58,8	7,3	1,0
L—R—M	179,9	52,4	64,4	179,9	—59,7	—65	—49,5	177,2	—71,3	60,5	—179,7	—59,9	9,0	1,1
B—R—H	179,7	—67,5	151,7	177,5	59,4	—70	—38,5	179,1	69,2	—65,8	—179,3	—50,2	11,6	1,5
L—R—H	179,9	52,4	64,1	—179,8	—59,6	—65	—43,1	178,9	69,3	—65,9	—179,3	—50,2	11,2	1,6
B—B—H	179,7	—66,7	151,4	177,2	59,8	—60	110,6	—176,0	69,1	—61,9	179,1	—47,6	8,1	1,7
L—B—H	179,9	52,1	64,5	179,9	—59,8	—55	113,3	—177,4	69,5	—62,6	175,7	—47,9	9,0	1,8
R—R—B	—178,1	—59,1	—48,6	—169,9	—48,7	—65	—31,5	173,6	—102,1	25,9	—179,5	—58,4	5,8	1,8
R—R—M	—178,9	—57,2	—45,5	—176,7	56,2	—65	—49,6	177,5	—71,6	60,7	—179,9	—59,9	6,7	1,9
R—R—R	—174,2	—48,8	—48,4	—175,4	—38,4	—55	—26,9	—179,4	—65,1	—29,8	—179,3	—51,5	7,4	1,9

*The angles of rotation are given in accordance with the IUPAC—IUB nomenclature [7].

TABLE 3. Energy Structure of the Optimum Conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe with Hydrogen Bonds

Type of conformations	E_{tot} (ε = 4), kcal/mole	Sum of the energies of Ala-Pro and of Pro-Ala, kcal/mole	Type of conformations	E_{tot} (ε = 4), kcal/mole	Sum of the energies of Ala-Pro and of Pro-Ala, kcal/mole
B-$\overline{\text{R-R}}$	0	−0,6	R-$\overline{\text{R-R}}$	1,0	1,6
B-$\overline{\text{R-B}}$	0,2	−0,1	L-R-M	1,1	−0,2
L-$\overline{\text{R-R}}$	0,4	−0,5	B-R-H	1,5	0,1
L-$\overline{\text{B-L}}$	0,4	0	L-R-H	1,6	0,2
B-$\overline{\text{B-L}}$	0,5	0,3	L-B-H	1,8	0,8
B′-R-M	0,6	−0,3	B-B-H	1,8	1,1
L-$\overline{\text{R-B}}$	0,8	0	R-$\overline{\text{R-B}}$	1,8	2,1
$\overline{\text{R-R-R}}$	0,9	—	$\overline{\text{R-R-R}}$	1,9	1,6
B-R-M	1,0	−0,3	R-R-M	1,9	1,9

optimum forms were calculated with the trans and cis conformations of the tertiary amide group and with the four values of the angle φ_2 of the proline residue −55, −60, −65, and −70°. Thus, we considered a total of 75 structures which comprise all the potentially possible conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe with intramolecular hydrogen bonds.

The conformational calculations were performed on a BÉSM-6 computer by a universal program [6] written in ALGOL-60 language.

The histogram of Fig. 2 gives the calculated values of the energies of the optimum conformations with all the systems of hydrogen bonds considered (the levels of the relative energies are bounded by a value of 16 kcal/mole). The most preferred with respect to energy are the conformations of types S1-S4. Table 2 gives the calculated geometric values of the angles φ, ψ, ω, and χ of the optimum conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe, the energies of which do not exceed 2.0 kcal/mole. The geometric parameters of the optimum forms are given at the most favorable values of the angle φ_2 of the proline residue. Also shown for all the calculated forms are the distances between the C^{α} atoms of the terminal methyl groups (L_{term}), which form useful information in the selection of the zero approximations for fragments of cyclic structures.

The deviation of the angles of rotation (φ, ψ, ω) in the calculated conformations of the tripeptide from the optimum parameters of the corresponding forms of the dipeptide fragments Ala-Pro and Pro-Ala [2] are mostly less than 5°. The results obtained show a definite role of short-range interaction in the structures of fragments containing a Pro residue. As can be seen from Table 2, the most probable conformations in nonpolar media (ε = 4) are those with the trans configuration of the tertiary amide group and with one hydrogen bond of the 4—1 type forming a 10-membered ring with the Pro-Ala fragment (IV, Fig. 1). In Table 3, the values of E_{tot} of the most favorable forms of Ac-L-Ala-L-Pro-L-Ala-NHMe with H bonds are compared with the additive sums of the energies of the dipeptide fragments Ala-Pro and Pro-Ala.

It follows from the facts given that for many spatial forms of the tripeptide the probability of their realization can be evaluated to a good approximation by the sum of the energies of the component dipeptide forms. Nevertheless, the energies of a number of conformations show an appreciable deviation from additivity, which is due mainly to the steric interaction of the terminal residues. On the whole, however, it must be observed that the most favorable structures of Ac-L-Ala-L-Pro-L-Ala-NHMe are determined mainly by the interaction of the adjacent units. The interactions of the terminal residues have a correlating nature.

Let us compare the calculated figures obtained with the results of experimental investigations of the conformations of peptide materials containing proline residues. An x-ray structural study of the crystalline linear peptides Gly-Pro-Leu-Gly and Gly-Pro-Leu-Gly-Pro [8, 9] has shown the presence of a 4—1 H bond ($\overline{\text{R-B}}$ type) in the Pro-Leu section. The realization of a similar bond is predicted by calculation for one of the most suitable forms of Ala-Pro-Ala — B—$\overline{\text{R-B}}$. The tertiary amide groups are present in the trans configurations in all the compounds. It follows from the NMR results that the formation of a similar H bond is quite probable in the cycloheptapeptide compound evolidin in the Pro-Val section [10]. In addition, analysis has shown that the most suitable structures with H bonds for Ac-L-Ala-L-Pro-L-Ala-NHMe are fairly frequently found in the β-inflections of protein chains at the

X-Pro-Y sections (where X and Y are any amino-acid residues). As an example we can give the fragments Ser^{159}-Pro^{160}-Cys^{161} in carboxypeptidase A [11], and Trp^{27}-Pro^{28}-Trp^{29} in α-chymotrypsin [12], assuming the B—R—R form, and also the Ile^{213}-Pro^{214}-Asp^{215} sections in carboxypeptidase A, realized in the form B—R—B.

SUMMARY

1. The optimum conformations of Ac-L-Ala-L-Pro-L-Ala-NHMe with eight possible systems of intramolecular hydrogen bonds have been calculated by the method of theoretical conformational analysis.

2. The structures found serve as canonical forms for the study of the conformational states of sections of peptide—protein systems with proline residues.

3. It has been shown that the interaction of adjacent residues in the structure of the compound mentioned play the dominating role.

LITERATURE CITED

1. É. P. Gromov, V. Z. Pletnev, and E. M. Popov, Khim. Prirodn. Soedin., 626 (1975).
2. V. Z. Pletnev, É. P. Gromov, and E. M. Popov, Khim. Prirodn. Soedin., 618 (1975).
3. V. Z. Pletnev, É. P. Gromov, and E. M. Popov, Bioorgan. Khim., 1, 328 (1975).
4. H. A. Scheraga, Advan. Phys. Org. Chem., 6, 103 (1968).
5. V. Z. Pletnev, E. M. Popov, and F. A. Kadymova, Theoret. Chim. Acta, 35, 93 (1974).
6. V. Z. Pletnev, E. P. Gromov, and F. A. Kadymova, Zh. Strukt. Khim., 16, 165 (1975).
7. IUPAC—IUB Commission on Biochemical Nomenclature, Biochemistry, 9, 3471 (1970).
8. T. Ueki, T. Ashida, M. Kakudo, Y. Sasada, and Y. Katsube, Acta Cryst., B25, 1840 (1969).
9. T. Ueki, S. Bando, T. Ashida, and M. Kakudo, Acta Cryst., B27, 2219 (1971).
10. K. D. Kopple, Biopolymers, 10, 1139 (1971).
11. W. N. Lipscomb, G. N. Reeke, J. A. Hartsuck, F. A. Quiocho, and P. H. Bethge, Phil. Trans. Roy. Soc. Lond., B257, 177 (1970).
12. J. J. Birktoft and D. M. Blow, J. Mol. Biol., 68, 187 (1972).

STRUCTURE OF GOSSYVERTIN — A NEW PHYTOALEXIN OF THE COTTON PLANT

A. K. Karimdzhanov, A. I. Ismailov, Z. S. Abdullaev, Sh. Yu. Islambekov, F. G. Kamaev, and A. S. Sadykov

UDC 547.972

The phytoalexins, which arise in plant tissues damaged by phytopathogenic microorganisms, must be assigned to a special class of antibiotics [1]. Investigations of the chemical nature of the phytoalexins of the cotton plant infected by verticilliaceous wilt have only just begun, and there is fairly contradictory information in the literature. Thus, Bell [2, 3] regarded gossypol to be the main fungitoxic substance for the fungus. Zaki and Erwin [4] consider as phytoalexins of the cotton plant two compounds of phenolic nature which they isolated from an infected plant and which they called vergosin and hemigossypol. In investigations by Sadykov [5], it was shown that isohemigossypol — 8-formyl-1,2,7-trihydroxy-5-isopropyl-3-methylnaphthalene — isolated from plants infected with *Verticillium dahliae* Kleb. is also a phytoalexin of the cotton plant.

In addition to isohemigossypol, we have found in the stems of infected plants another series of compounds with the nature of phytoalexins, and in the present paper we give the results of a chemical study of one of these compounds.

Division of Bioorganic Chemistry, Academy of Sciences of the Uzbek SSR, Tashkent.
Translated from Khimiya Prirodnykh Soedinenii, No. 2, pp. 238-242, March-April, 1976. Original article submitted July 14, 1975.