THEORETICAL CONFORMATIONAL ANALYSIS OF TRIPEPTIDES WITH A GLYCINE RESIDUE

V. Z. Pletnev, F. A. Kadymova, and E. M. Popov

UDC 547.962

The conformations of amino-acid residues realized in proteins correspond to the low-energy forms of the corresponding free monopeptides. This shows that the conformational states of each residue are determined to a considerable degree by the closest interactions which, however, do not come into contradiction with the interactions between the neighboring residues in the sequence and residues remote along the chain but spatially close in the globule. At the present time, the geometric and thermodynamic parameters of the stable forms have been found for all 20 natural amino-acid residues. However, this extremely valuable information cannot be used directly in the conformational analysis of even comparatively short oligopeptides in the overwhelming majority of cases, since many residues have a large number of approximately equivalent low-energy forms. Even a tripeptide fragment is a fairly complex subject for investigation from the aspects of the monopeptide, since it requires the consideration of an average of 10^3 variants of the structures.

The results of experimental [1-6] and theoretical [7-9] investigations show that a given (n) residue in a polypeptide chain may experience a considerable influence of the three or four preceding (n − 1 to n − 4) and following (n + 1 to n + 4) residues, interactions with the closest neighbors (n − 1 and n + 1) being the strongest, as a rule. However, interactions with more remote residues (n − 5, n + 5, etc.) are small only if the peptide chain does not form a loop or an inflection. The interactions arising in this case between residues we assign to the long-range interactions and have not so far taken them into consideration. Thus, to account for the interactions of one amino-acid residue with its neighbors (let us call them medium-range interactions) it is possible to limit ourselves to a nonapeptide fragment.

It is desirable to begin the investigation of the conformational possibilities of a nonapeptide with the central tripeptide fragment and then gradually to complicate the problem by extending the amino-acid residues from both ends of the chain, taking into account in this process the results of the calculation of the preceding fragment. The large number of low-energy forms that, as already mentioned, exist for each residue practically excludes the solution of such a problem on the basis of a direct use of information on monopeptides. In view of this, it appeared to interest to investigate the possibility of the a priori prediction of a comparatively small set of preferred forms of the tripeptide (n − 1, n, n + 1) with the completely known conformational states of the dipeptides (n − 1, n) and (n, n + 1). Such an approach may open up a route to the solution of the problem of medium-range interactions under two conditions. In the first place, where the interactions of the two neighboring residues (n − 1, n and n, n + 1) lead to a considerable differentiation of the optimum forms of each of them and, in the second place, where the most favorable conformations of the tripeptide consist of combinations of the most favorable forms of the corresponding dipeptides. The first promising results in relation to the fulfillment of the conditions have already been obtained [10-14].

The solution of the conformational problem for tripeptides may apparently be additionally simplified if in the formulation of the initial approximations we take into account not

M. M. Shemyakin Institute of Bioorganic Chemistry, Academy of Sciences of the USSR.
Translated from Khimiya Prirodnykh Soedinenii, No. 5, pp. 631-637, September-October, 1975. Original article submitted June 17, 1974.

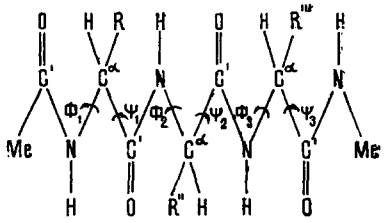

Fig. 1. Calculation model of the molecules: I. Ac-Gly-L-Ala-L-Ala-NHMe (R' = H, R" = R"' = Me), II. Ac-L-Ala-Gly-L-Ala-NHMe (R" = H, R' = R"' = Me), III. Ac-L-Ala-L-Ala-Gly-NHMe (R"' = H, R' = R" = Me).

only the complementarity of the central residue (n) with respect to the residues (n − 1) and (n + 1) but also the influence exerted by skeletal interactions between (n − 1) and (n + 1). This leads to a reduction in the number of conformations of the tripeptide fragment subjected to analysis provided that the conformational states realized in proteins are close to the optimum both with respect to the interactions of the main chains and of the side chains.

In tripeptide protein fragments only four types of main chains are found. The first of them, the most numerous, consists of a peptide skeleton and C^{β} atoms — this is the alanine type of main chain. It is formed by all the natural amino-acid residues with the exception of Gly and Pro. The association of glycine or proline with any other residues comprises the following two types of main chain. Finally, the fourth type simultaneously includes residues of the alanine type and Gly and Pro. From the point of view of skeletal interactions, each type has a completely determined conformational specific nature. We have previously investigated the conformational possibilities of the molecules Ac-L-Ala-L-Ala-L-Ala-NHMe [15] and Ac-L-Ala-L-Pro-L-Ala-NHMe [14], modelling two types of peptide skeleton. A comparison of the calculated optimum forms of these molecules with the conformational states of the corresponding main chains of the tripeptide protein fragments led to the conclusion that in proteins a stabilizing interaction of the side chains is exerted with the low-energy forms of the main chains.

The present paper illustrates the results of an investigation, analogous in its aim, of the conformational possibilities of a peptide skeleton including a glycine residue. We have calculated all the optimum forms of the following tripeptide molecules with a Gly residue in the first, second, and third positions:

Ac-Gly-L-Ala-L-Ala-NHMe (I)

Ac-L-Ala-Gly-L-Ala-NHMe (II)

Ac-L-Ala-L-Ala-Gly-NHMe (III)

The conformational characteristics of these molecules are determined mainly by the Gly residue. Glycine, possessing the highest lability among amino-acid residues, can fulfill diverse and, at first sight, mutually exclusive functions in the spatial organization of the protein molecule. On the one hand, it can ensure the relative conformational freedom of local sections of the peptide chain from the point of view of close-range and medium-range interactions and thereby create favorable conditions for the realization of long-range interactions. On the other hand, glycine under the action of only close- and medium-range interactions, can also promote the formation of autonomous conformationally rigid sections of the peptide chain. In this case, because of the labile Gly link in the protein fragments, favorable stabilizing interactions between the residues may come into being. The calculation of the molecules (I-III) was performed with reference to the conditions of a polar medium and without taking hydrogen bonds into account. Hydrogen bonds of types 5→1 and 4→1 which are found in proteins and are possible in the molecules (I-III) have a very low energy in aqueous solution (0-1.5 kcal/mole [16, 17]). They are realized in conformations of the chain suitable with respect to nonvalent interactions, and therefore the fact that intramolecular hydrogen bonds are not taken into account does not affect the definitive conclusions of the work.

A model of the (I-III) molecules and the symbols adopted are given in Fig. 1. The lengths of the bonds and the valence angles were taken as being the same as in the calculation of the methylamide of N-acetyl-α-methylalanine [18]. The calculation of the optimum forms was performed by searching for minima in the potential energy with variations in the six angles or rotation φ and ψ. As the zero approximations for each tripeptide we selected the 64 structures represented by all the combinations of the four forms of each residue corresponding to the optimum conformations R, B, L, and P* of the Ac-L-Ala-NHMe and Ac-Gly-NHMe molecules

*On the φ —ψ conformational map, the R form is present in the first quadrant ($\varphi = \psi = 180\text{-}0°$), B is in the second quadrant ($\varphi = -180\text{-}0°$, $\psi = 0\text{-}180°$), L in the third quadrant ($\varphi = \psi = 0\text{-}180°$), and P in the fourth quadrant ($\varphi = 0\text{-}180°$, $\psi = -180\text{-}0°$).

TABLE 1. Geometric and Energy Parameters of the Preferred Conformations of the Molecules Ac-Gly-L-Ala-L-Ala-NHMe and Ac-L-Ala-L-Ala-Gly-NHMe

Identifier	Dihedral angles of rotation, deg						Energy, kcal/mole	
	φ_1	ψ_1	φ_2	ψ_2	φ_3	ψ_3	E_{nonv}	E_{tot}
Ac-Gly-L-Ala-L-Ala-NHMe								
R—R—R	—70	— 65	—63	— 42	—63	— 37	0	0
L—R—R	65	68	—68	— 50	—68	— 49	2,0	—0,4
B—R—R	—72	102	—67	— 50	—69	— 48	2,0	—0,3
P R—B	58	—106	—66	— 47	—82	149	0,6	0,2
L—R—B	65	66	—66	— 49	—66	155	2,6	0,4
P—R—R	74	—103	—68	— 50	—68	— 50	0,4	0,6
R—B—R	—64	— 66	—65	156	—69	— 46	2,7	0,6
L—B—R	65	66	—69	153	—71	— 41	3,0	0,9
B—R—B	—72	102	—67	— 50	—65	156	2,7	1,0
L—B—B	65	65	—67	154	—68	152	3,5	1,0
Ac-L-Ala-L-Ala-Gly-NHMe								
R—R—R	—74	— 51	—65	— 45	—63	— 53	0	0
R—R—L	—63	— 49	—67	— 49	65	62	1,3	—0,9
R—B—L	—67	— 50	—64	156	64	67	2,0	—0,1
B—R—R	—65	155	—66	— 49	—65	— 65	1,9	0,1
B—R—B	—64	150	—60	— 45	—72	88	2,1	0,3
R—B—R	—64	— 49	—67	155	—65	— 63	2,2	0,3
R—R—P	—67	— 50	—68	— 50	74	—103	1,6	0,4
R—R—B	—70	— 50	—69	— 48	—69	96	0,7	0,5
R—L—L	—67	— 49	52	59	66	66	2,4	0,5
B—B—L	—63	157	—63	157	65	67	3,0	0,7
B—B—P	—64	156	—64	155	71	—101	3,1	0,7
B—B—R	—66	155	—66	155	—65	— 66	3,2	0,8
R—B—B	—67	— 50	— 63	157	—73	102	2,4	0,9
B—R—L	—69	153	—67	— 47	66	66	2,5	0,9
L—R—R	55	59	—66	— 49	—66	— 66	2,7	0,9

[19, 20]. For the free Ala residue the most favorable, and approximately equiprobable, conformations are R and B, and then come L and P. For the Gly residue, all four conformations are approximately equiprobable, R being equivalent to L and B to P.

The calculation was performed with account of nonvalent and electrostatic interactions and also of torsional energy. The nonvalent interactions were calculated from the Lennard-Jones potential with Scott and Scheraga's parameters [21]. The methyl groups were approximated by spheres with a diameter of 3.7 Å, and their interaction with one another and with other atoms was evaluated from the Lennard-Jones potential with parameters found in our laboratory. The electrostatic energy was calculated by means of the Coulomb law with the charges on the atoms suggested by Momany et al., [22] and a value of the dielectric constant $\varepsilon = 10$, corresponding to a polar medium [20]. The torsional potentials and the barriers to rotation around the C^{α}—N and C^{α}—C' bonds were taken as in previous work [23]. The reckoning of the angles of rotation was in accordance with the IUPAC-IUB nomenclature [24].

The results of the calculation of the molecules (I-III) are shown in Fig. 2 in the form of histograms and in Tables 1 and 2. To denote the forms of the main chain, a system of identifiers has been adopted. An identifier consists of three letters, each of which characterizes one of the four possible conformational states of a residue (R, B, L, and P). Examples of the correspondence of the identifiers to the angles of internal rotation are given in Tables 1 and 2. The heights of a column on the histogram as compared with the optimum form of the given tripeptide corresponds to the magnitude of its relative energy. For the least preferred forms the energies on the histograms are not given, and it is indicated only that they exceed 5 kcal/mole. The figure below the column corresponds to the frequency with which the given conformation of the main chain is found in the corresponding tripeptide fragments in proteins*; and the figures on the histogram of the Ac-Gly-L-Ala-L-Ala-NHMe molecule (Fig. 2a) relate to protein fragments of the type of Gly-X-Y, those for the Ac-L-Ala-Gly-L-

*X-Ray structural information for myoglobin [25], lysozyme [26], α-chymotrypsin [27], carboxypeptidase A [28], cytochrome C [29], and insulin [30] was used.

TABLE 2. Geometric and Energy Parameters of the Preferred Conformations of the Molecule Ac-L-Ala-Gly-L-Ala-NHMe

Identifier	Dihedral angles of rotation, deg						Energy, kcal/mole		l_{term} Å
	φ_1	ψ_1	φ_2	ψ_2	φ_3	ψ_3	E_{nonv}	E_{tot}	
R—R—R	—74	— 46	—61	—58	—59	— 49	0	0	5,7
R—P—R	—66	— 45	73	—79	—69	— 41	1,2	—0,8	8,7
L—L—R	51	57	64	62	—69	— 49	2,3	0	8,5
B—R—R	—62	158	—67	—66	—68	— 50	1,9	0,1	7,3
R—L—R	—67	— 49	65	64	—68	— 49	2,1	0,3	11,2
R—B—R	—64	— 48	—66	97	—69	— 48	1,9	0,6	9,9
R—L—L	—67	— 48	62	64	57	56	2,5	0,6	8,9
B—L—R	—62	157	65	67	—68	— 50	2,5	0,6	10,5
R—B—B	—66	— 48	—68	96	—70	153	2,4	0,7	9,4
R—L—B	—67	— 49	65	64	—68	153	2,6	0,8	12,2
L—R—R	54	60	—65	—68	—68	— 50	2,7	1,0	8,8
B—R—B	—70	151	—62	—62	—71	152	2,5	1,0	9,1

TABLE 3. Distribution with Respect to Energy of the Conformational States of the Main Chains of Tripeptide Fragments with a Gly Residue in Proteins

Type of fragment	Number of fragments	Energy range, kcal/mole					
		less than 1	1–2	2–3	3–4	4–5	greater than 5
Gly-X-Y	80	45	32	2	1	—	—
X-Gly-Y	76	34	39	3	—	—	—
X-Y-Gly	70	44	16	6	1	3	—

Note: X, Y ≠ Gly and Pro. X-ray structural information on myoglobin, lysozyme, α-chymotrypsin, carboxypeptidase A, cytochrome C, and insulin was used.

Ala-NHMe molecule (Fig. 2b) to X-Gly-Y, and those for Ac-L-Ala-L-Ala-Gly-NHMe (Fig. 2c) to X-Y-Gly (X, Y ≠ Gly, Pro).

The tables show, in addition to the total energy (E_{tot}), the energies of the nonvalent interactions (E_{nonv}), which makes it possible to judge the stabilizing role of other energy components (in all cases, the torsional contribution is insignificant). In Table 2 the distances between the carbon atoms of the terminal methyl groups (l_{term}) are also given.

As was to be expected, the incorporation of glycine into the composition of a tripeptide increases its conformational possibilities. In actual fact, calculation has shown that with the replacement of even one Ala residue by Gly in a tripeptide the number of low-energy conformations rises considerably. While in the Ac-L-Ala-L-Ala-L-Ala-NHMe molecule the number of forms with energies less than 2 kcal/mole is 16 [15], in the molecule of (I) it is 25, in (II) 29, and in (III) 30. The increase in conformational freedom takes place because the interactions within the limits of the main chain do not prevent the realization in the Gly residue of conformations with an angle $\varphi > 0°$ (L and P forms). As can be seen from Fig. 2 and Tables 1 and 2, the state of the chain with Gly in the L and P forms is found as frequently as with it in the R and B forms. Moreover, the L—R—R (I), R—P—R (II), and R—R—L (III) conformations possess the lowest energies. Their preferential nature is due to the comparatively low energy of the nonvalent interactions and the mainly stabilizing effect of the electrostatic interactions. From the value of E_{nonv}, the global conformations in all cases are the α-helical forms R—R—R, which is due to the large negative contribution of the dispersion interactions. The possibility of the realization of a considerable number of low-energy

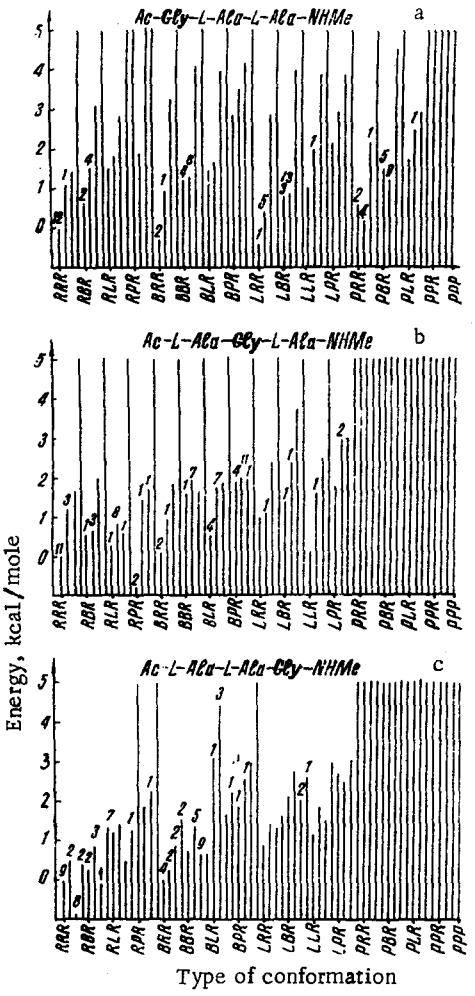

Fig. 2. Distribution of the conformation of Ac-Gly-L-Ala-L-Ala-NHMe in the Gly-X-Y fragments (a), Ac-L-Ala-Gly-L-Ala-NHMe in the X-Gly-Y fragment (b) and of Ac-L-Ala-L-Ala-Gly-NHMe in the X-Y-Gly fragments (X, Y ≠ Gly, Pro) (c) in six proteins [25-30]. The identifiers of the forms not shown in the histogram have a running third index. For example, after the R–R–R forms the R–R–B, R–R–L, and R–R–P forms follow immediately.

conformations with Gly in the L and P forms is due to the presence of this residue in proteins in the majority of those sections of the chain which form inflections and loops [31]. As is well known, at angles $\varphi > 0°$ the main chain sharply changes its direction.

A comparison of the results of the calculation of the molecules (I-III) with the structure of protein tripeptide fragments of types Gly-X-Y, X-Gly-Y, and X-Y-Gly shows that the skeletal conformations observed in a protein correspond to the low-energy forms of the corresponding isolated main chains. Table 3 gives the distribution with respect to energy of the conformational states of the tripeptide fragments of six proteins including the Gly residue. In 93% of the protein fragments considered (210 out of 226), the state of the main chains correspond to the conformations of the molecule (I-III) possessing energies of less than 2 kcal/mole. Nevertheless, the structure of the protein chain is affected by a large number of other factors beside the skeletal interactions that are absent in (I-III). Among them may be mentioned the interaction between voluminous and charge-bearing side chains and of the side chains with the main chain, and interactions with residues remote along the chain. Each of the factors mentioned makes a considerable energy contribution. For example, the energy of the interaction of the side chains of two neighboring phenylalanine residues may amount to −3 to −4 kcal/mole [12]. The fact that, in spite of this, in all the cases considered a good correspondence is observed between the conformations of the tripeptide fragments in the proteins and the low-energy forms of the molecules (I-III) shows that other forms of interaction do not contradict one another and exist with the favorable forms of the main chain. Thus (see also [32, 33, 13-15]), we come to the conclusion of the agreement in proteins of the whole multiplicity of close-, medium-, and long-range interactions. This property is characteristic only of evolution-selected amino-acid sequences. It is just because of the agreement between all the types of interactions and with the condition of their maximum degree of saturation that protein chains, unlike synthetic, artificial, polypeptides, possess a unique tertiary structure.

SUMMARY

1. The conformational states of glycine residues — R, B, L, P — in the low-energy forms of the tripeptides (I-III) are found with equal frequency.

2. The conformations of the main chain in sections of protein with glycine residues correspond to the preferred forms of the tripeptide fragments (I-III).

3. The close-, medium-, and long-range interactions in proteins do not contradict one another and exist at the favorable forms of the main chain.

LITERATURE CITED

1. M. Goodman, M. Langsam, and J. G. Rosen, Biopolymers, 4, 305 (1966).
2. M. Goodman, A. S. Verdini, C. Toniolo, W. D. Phillips, and F. A. Bovey, Proc. Nat. Acad. Sci. U.S., 64, 444 (1969).
3. M. Goodman, F. Naider, and C. Toniolo, Biopolymers, 10, 1719 (1971).
4. M. Goodman, F. Naider, and R. Rupp, Bioorg. Chem., 1, 310 (1971).

5. A. Yoron, E. Katchalski, A. Berger, G. D. Fasman, and H. A. Sober, Biopolymers, 10, 1107 (1971).
6. S. Kawanishi and S. Sano, Biopolymers, 12, 3166 (1973).
7. P. K. Ponnuswamy, P. K. Warme, and H. A. Scheraga, Proc. Nat. Acad. Sci., US, 70, 830 (1973).
8. D. A. D. Parry and E. Suzuki, Biopolymers, 7, 189 (1969).
9. G. M. Lipkind and E. M. Popov, Int. J. Pept. Prot. Res., 5, 371 (1973).
10. G. M. Lipkind, S. F. Arkhipova, and E. M. Popov, Molekul. Biol., 4, 331 (1970).
11. G. M. Lipkind, S. F. Arkhipova, and E. M. Popov, Molekul. Biol., 4, 509 (1970).
12. M. A. Kreissler, N. A. Ahmedov, G. M. Lipkind, S. F. Arkhipova, and E. M. Popov, J. Chim. Phys., 71, 913 (1974).
13. V. Z. Pletnev, E. P. Gromov, and E. M. Popov, Khim. Prirodn. Soedin., 618 (1975).
14. E. P. Gromov, V. Z. Pletnev, and E. M. Popov, Khim. Prirodn. Soedin., 626 (1975).
15. V. Z. Pletnev, F. A. Kadymova, and E. M. Popov, Molekul. Biol., 9, No. 2, 213 (1975).
16. I. A. Schellman, Compt. Rend. Trav. Lab. Carlsberg, Ser. Chim., 29, 223 (1955).
17. J. M. Klotz and J. S. Fransen, J. Amer. Chem. Soc., 84, 3461 (1962).
18. V. Z. Pletnev, E. P. Gromov, and E. M. Popov, Khim. Prirodn. Soedin., 224 (1973).
19. E. M. Popov, G. M. Lipkind, S. F. Arkhipova, and V. G. Dashevskii, Molekul Biol., 622 (1968).
20. G. M. Lipkind, S. F. Arkhipova, and E. M. Popov, Zh. Strukt. Khim., 11, 121 (1970).
21. R. A. Scott and H. A. Scheraga, J. Chem. Phys., 45, 2091 (1966).
22. F. A. Momany, R. F. McQuire, J. F. Yan, and H. A. Scheraga, J. Phys. Chem., 75, 2286 (1971).
23. E. M. Popov, V. G. Dashevskii, G. M. Lipkind, and S. F. Arkhipova, Molekul. Biol., 2, 612 (1968).
24. IUPAC-IUB Commission on Biochemical Nomenclature, Biochemistry, 9, 3471 (1971).
25. H. C. Watson, Progr. Stereochem., 4, 229 (1969).
26. D. C. Phillips, Proc. Nat. Acad. Sci. U.S., 57, 484 (1967); G. N. Ramachandran and V. Sasisekharan, Advan. Protein Chem., 283 (1968).
27. J. J. Birktoft and D. M. Blow, J. Mol. Biol., 68, 187 (1972).
28. W. N. Lipscomb, G. M. Reeke, J. Jean, A. Hartsuck, F. A. Quiocho, and P. H. Bethpage, Phil. Trans. Roy. Soc., B 257, 177 (1970).
29. R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper, and T. Margoliash, J. Biol. Chem., 246, 1511 (1971); T. T. Wu and E. A. Kabat, J. Mol. Biol., 75, 13 (1973).
30. T. L. Blundell, J. F. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola, and M. Vilayan, Nature (Lond), 231, 506 (1971).
31. P. N. Lewis, F. A. Momany, and H. A. Scheraga, Biochim. Biophys. Acta, 303, 211 (1973).
32. E. M. Popov and G. M. Lipkind, Molekul. Biol., 5, 624 (1971).
33. G. M. Lipkind and E. M. Popov, Molekul. Biol., 5, 667 (1971).