

CONFIGURATIONS OF POLYPEPTIDE CHAINS WITH EQUIVALENT CIS AMIDE GROUPS

BY LINUS PAULING AND ROBERT B. COREY

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,* CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIF.

Communicated November 29, 1951

In recent papers¹⁻³ we have pointed out that stable configurations of polypeptide chains must show the bond distances and bond angles found in simple peptides and related simple substances, that each amide group must be planar, with either the cis-configuration or the trans-configuration about the C'—N bond, and that the carbonyl and imino groups (except for proline or hydroxyproline residues) must be involved in the formation of hydrogen bonds, with N—H···O distance approximately 2.8 Å, and with the oxygen atom nearly on the N—H axis. We have also presented evidence indicating that certain configurations around the two single bonds from the amide groups to the α carbon atom are energetically favored, and have discussed the 36 polypeptide-chain structures involving equivalent trans-amide groups with one or another of six favored azimuthal orientations about the N—C single bond and also one of six favored orientations about the C—C' single bond.⁴ In the present paper we discuss the 36 configurations of equivalent cis-amide groups corresponding to these favored orientations, and also a few other configurations of cis-amide groups.

In the study of these structures we have been greatly aided by the use of a special set of molecular models, constructed of wood on the scale 2.5 cm. = 1 Å. The wooden pieces representing the atoms are joined together by the use of steel pins to form a model of a polypeptide chain, and the models are fitted with screw clamps on the steel pins which permit the orientations around the bonds to be fixed. The model representing the hydrogen atom of the imino group is cupped in such a way that the model representing the oxygen atom may fit into it, at a distance between N and O corresponding to 2.8 Å. The position of the cup is adjustable, to allow a bend in the hydrogen bond of as great as 20°. The models conform to the accepted van der Waals radii of the atoms, permitting the easy recognition of serious steric hindrance for a proposed configuration. A detailed description of these molecular models will be published later.

The favored orientations about the N—C bond (the bond between the nitrogen atom of an amide group and the adjacent α carbon atom) are those in which one of the three other bonds formed by the α carbon is either cis or trans to the N—H bond. These orientations can be described by an azimuthal angle φ , as follows: $\varphi = 0^\circ$, CC' cis to NH; $\varphi = 60^\circ$, CH trans to NH; $\varphi = 120^\circ$, CR cis to NH; $\varphi = 180^\circ$, CC' trans to NH:

$\varphi = 240^\circ$, CH cis to NH; $\varphi = 300^\circ$, CR trans to NH. The similar favored orientations about the C—C' bond can be described by an azimuthal angle χ , as follows: $\chi = 0^\circ$, CN cis to CO; $\chi = 60^\circ$, CH trans to CO; $\chi = 120^\circ$, CR cis to CO; $\chi = 180^\circ$, CN trans to CO; $\chi = 240^\circ$, CH cis to CO; $\chi = 300^\circ$, CR trans to CO.

The 36 configurations are described in table 1. In general they are helical structures; in a few cases the helix has degenerated to a ring. Many of the structures are eliminated by serious steric hindrance. Of those that are not thus eliminated, some cannot enter into hydrogen-bond formation with adjacent similar polypeptide chains, and are eliminated for this reason. No configurations with intramolecular hydrogen bonds are contained in this set. For each of the configurations that is not eliminated by steric hindrance, the pitch of the helix (axial length per turn), the number of residues per turn, the approximate orientation of the CO and NH groups relative to the helical axis, and the suitability of the structure to form hydrogen bonds are given in the table. A helical molecule can form a crystal in which it is linked to equivalent molecules by equivalent hydrogen bonds only if it has a twofold, threefold, fourfold, or sixfold axis of symmetry.

Discussion of the Structures.—No helical structures suitable for proteins or polypeptides in general have been discovered by the foregoing considerations. The most interesting of the helical structures described in table 1 is number 16, with three residues per turn. These helical molecules could form hydrogen bonds with similar helices arranged in a hexagonal lattice. The dimensions of the unit are such as to make the structure a possible one for polyalanine, but not for other polypeptides or for proteins. Structures 4 and 5, with four residues per turn, could form hydrogen-bonded tetragonal crystals, but are probably eliminated by steric hindrance. Similarly, structures 28 and 29, with three residues per turn, could form hexagonal crystals, but also are probably eliminated by steric hindrance. The helices 24 and 35, with 4.5 residues per turn, are interesting as configurations for the polypeptide chain, but are unable to form equivalent hydrogen bonds with adjacent chains because of the absence of a crystallographic symmetry element.

We might ask whether satisfactory helical configurations of polypeptide chains with equivalent cis-amide groups could be formed by permitting orientations around the single bonds to the α carbon atom other than the postulated favored orientations. No configurations of this sort involving intramolecular hydrogen bonds exist. In order to form intermolecular hydrogen bonds between equivalent amide groups the helix would have to have a twofold, threefold, fourfold, or sixfold screw axis of symmetry. Configurations with threefold and fourfold screw axes are described in table 1, and have been discussed in the preceding paragraph. The geometry of the

TABLE 1

CONFIGURATIONS OF EQUIVALENT *cis*-AMIDE GROUPS WITH FAVORED ORIENTATIONS ABOUT C—N AND C'—C BONDS^a

A. *Trans-trans* Configurations:

1. CH *trans* to NH, CH *trans* to CO; $\varphi = 60^\circ$, $\chi = 60^\circ$. A four-residue puckered ring is formed, but the hydrogen atoms attached to the α carbon atom are at the corners of a square only about 1.4 Å on edge. The consequent steric hindrance is probably so great as to eliminate this structure. The carbonyl and imino groups point out and up, and hydrogen bonded tetragonal layers could be formed, with $a = 9.6$ Å (the unit containing two cyclic tetrapeptide molecules), and with c determined by the nature of the side chains.

2. CR *trans* to NH, CR *trans* to CO; $\varphi = 300^\circ$, $\chi = 300^\circ$. A four-residue puckered ring, resembling 1. The four R groups are crowded together, and strong steric hindrance eliminates this structure.

3. CC' *trans* to NH, CN *trans* to CO; $\varphi = 180^\circ$, $\chi = 180^\circ$. A planar cyclic dipeptide. This is the diketopiperazine structure. The molecules can form long strings held together by hydrogen bonds, as in the diketopiperazine crystal.⁵

4. CH *trans* to NH, CR *trans* to CO; $\varphi = 60^\circ$, $\chi = 300^\circ$. Helix with pitch 11.8 Å, four residues per turn. Amide groups in vertical plane, with NH and CO projecting out diagonally. Each amide group could form two hydrogen bonds with a corresponding amide group of a similar chain, with reversed orientation, giving a tetragonal crystal, with $a = 7.5$ Å, $c = 11.8$ Å, eight residues per unit. However, an H atom and an R group on α carbon atoms adjacent to each amide group are directed toward one another, and steric hindrance eliminates this structure for all R groups except H. The structure probably could hence be assumed only by polyglycine; but for polyglycine the density with this structure would be 1.15 g. cm.⁻³, which is too small. The structure would hence not permit maximum stabilization for polyglycine by electronic van der Waals attraction, and it is accordingly eliminated.

5. CR *trans* to NH, CH *trans* to CO; $\varphi = 300^\circ$, $\chi = 60^\circ$. Four-residue helix with pitch 11.8 Å, resembling 4, and probably eliminated for the same reasons.

6. CH *trans* to NH, CN *trans* to CO; $\varphi = 60^\circ$, $\chi = 180^\circ$. Ruled out by steric hindrance of C and H of NH.

7. CR *trans* to NH, CN *trans* to CO; $\varphi = 300^\circ$, $\chi = 180^\circ$. Ruled out by steric hindrance of N and C.

8. CC' *trans* to NH, CH *trans* to CO; $\varphi = 180^\circ$, $\chi = 60^\circ$. Ruled out by steric hindrance of H (of CH) and C'O.

9. CC' *trans* to NH, CR *trans* to CO; $\varphi = 180^\circ$, $\chi = 300^\circ$. Ruled out by steric hindrance of α C and O.

B. *Cis-cis* Configurations:

10. CH *cis* to NH, CH *cis* to CO; $\varphi = 240^\circ$, $\chi = 240^\circ$. Ruled out by steric hindrance of C and C.

11. CR *cis* to NH, CR *cis* to CO; $\varphi = 120^\circ$, $\chi = 120^\circ$. Ruled out by great steric hindrance.

12. CC' *cis* to NH, CN *cis* to CO; $\varphi = 0^\circ$, $\chi = 0^\circ$. A nearly strainless planar ring containing seven residues. Some steric hindrance may occur between R groups, which are all on one side of the plane, with the H atoms on the other side. The CO and NH groups could form intermolecular hydrogen bonds with equivalent molecules, except that the sevenfold axis prevents the formation of a crystal containing only equivalent amide groups.

TABLE 1—Continued

13. CH cis to NH, CR cis to CO; $\varphi = 240^\circ$, $\chi = 120^\circ$. Ruled out by steric hindrance of R and NH.

14. CR cis to NH, CH cis to CO; $\varphi = 120^\circ$, $\chi = 240^\circ$. Ruled out by steric hindrance of H of CH and H of NH.

15. CH cis to NH, CN cis to CO; $\varphi = 240^\circ$, $\chi = 0^\circ$. Tight helix with pitch 6.6 Å, three residues per turn. Ruled out by steric hindrance of R and CO of next turn of the helix.

16. CR cis to NH, CN cis to CO; $\varphi = 120^\circ$, $\chi = 0^\circ$. A tight helix with pitch 6.6 Å and three residues per turn, resembling 15. Close contact of CH and CO of next turn, but probably not enough to eliminate the structure. NH points out laterally, CO out and down by 60° . A hexagonal lattice could be formed, each chain being surrounded by six parallel chains, with three of which it forms hydrogen bonds using its NH groups, and with the other three hydrogen bonds using its CO groups. The hexagonal unit has $a = 6.7$ Å, $c = 6.6$ Å, three residues per unit. The assumed density 1.40 g. cm.⁻³ corresponds to residue weight 72. That of alanine is 71; accordingly only polyalanine might be expected to have this structure.

17. CC' cis to NH, CH cis to CO; $\varphi = 0^\circ$, $\chi = 240^\circ$. Helix with pitch 7.1 Å, three residues per turn. Ruled out by steric hindrance of R and H of NH, and by inhibition of hydrogen-bond formation between chains through steric hindrance between H of CH and O of adjacent chain (see 18).

18. CC' cis to NH, CR cis to CO; $\varphi = 0^\circ$, $\chi = 120^\circ$. Helix with pitch 7.1 Å, three residues per turn, resembling 17. H of CH in contact with H of NH. The NH groups point closely toward R of CHR separated from it by two residues: R shields NH so that no satisfactory hydrogen bond could be formed; hence the structure is ruled out.

C. Trans-cis Configurations:

19. CH trans to NH, CH cis to CO; $\varphi = 60^\circ$, $\chi = 240^\circ$. Ruled out by steric hindrance of H of CH and H of NH.

20. CR trans to NH, CR cis to CO; $\varphi = 300^\circ$, $\chi = 120^\circ$. Helix with pitch 7.8 Å, three residues per turn. CO points out, NH points out and up. NH is close to axis of helix and is in steric interference with adjacent H and R groups, serious enough to eliminate the structure.

21. CC' trans to NH, CN cis to CO; $\varphi = 180^\circ$, $\chi = 0^\circ$. Ruled out by great steric hindrance.

22. CH trans to NH, CR cis to CO; $\varphi = 60^\circ$, $\chi = 120^\circ$. Ruled out by great steric hindrance.

23. CR trans to NH, CH cis to CO; $\varphi = 200^\circ$, $\chi = 240^\circ$. Ruled out by steric hindrance of R and NH.

24. CH trans to NH, CN cis to CO; $\varphi = 60^\circ$, $\chi = 0^\circ$. Open helix with pitch 8.1 Å, 4.5 residues per turn. H of CH in tight helix, in contact with equivalent adjacent atoms. NH points out laterally, CO out and down at 45° . Not suited to formation of intermolecular hydrogen bonds with equivalent molecules.

25. CR trans to NH, CN cis to CO; $\varphi = 300^\circ$, $\chi = 0^\circ$. Open helix with pitch 8.1 Å, 4.5 residues per turn, resembling 24, but with H and R interchanged. Ruled out by steric hindrance of R and R (R—R = 1.8 Å).

26. CC' trans to NH, CH cis to CO; $\varphi = 180^\circ$, $\chi = 240^\circ$. Ruled out by great steric hindrance.

TABLE 1—Continued

27. CC' trans to NH, CR cis to CO; $\varphi = 180^\circ$, $\chi = 120^\circ$. Ruled out by great steric hindrance.

D. Cis-trans Configurations:

28. CH cis to NH, CH trans to CO; $\varphi = 240^\circ$, $\chi = 60^\circ$. Helix with pitch 8.5 A, three residues per turn. Intermolecular hydrogen bonds can be formed with three adjacent chains, oppositely directed. Hexagonal crystal with $a = 9.7$ A, $c = 8.5$ A, containing six residues, with average residue weight about 97; calculated density then 1.35 g. cm.⁻³. Probably eliminated by steric hindrance of H of CH, O, and R.

29. CR cis to NH, CR trans to CO; $\varphi = 120^\circ$, $\chi = 300^\circ$. Helix with pitch 8.5 A, three residues per turn, resembling 28. Probably eliminated by steric hindrance between R, O, and H of CH.

30. CC' cis to NH, CN trans to CO; $\varphi = 0^\circ$, $\chi = 180^\circ$. Ruled out by great steric hindrance of α C and H of NH.

31. CH cis to NH, CR trans to CO; $\varphi = 240^\circ$, $\chi = 300^\circ$. Tight helix with 2.6 residues per turn. Ruled out by steric hindrance of R, α C, and N.

32. CR cis to NH, CH trans to CO; $\varphi = 120^\circ$, $\chi = 60^\circ$. Ruled out by steric hindrance of CH and CO.

33. CH cis to NH, CN trans to CO; $\varphi = 240^\circ$, $\chi = 180^\circ$. Ruled out by steric hindrance of C and C'.

34. CR cis to NH, CN trans to CO; $\varphi = 120^\circ$, $\chi = 180^\circ$. Ruled out by great steric hindrance of α C and NC'.

35. CC' cis to NH, CH trans to CO; $\varphi = 0^\circ$, $\chi = 60^\circ$. Open helix with pitch 8.5 A, 4.5 residues per turn, resembling 24, but with CO and NH interchanged. Ruled out because of inability to form a hydrogen-bonded structure.

36. CC' cis to NH, CR trans to CO; $\varphi = 0^\circ$, $\chi = 300^\circ$. Open helix with pitch 8.1 A, 4.7 residues per turn, resembling 24. CO points out laterally, NH diagonally. R groups are on the axis; hence the configuration is ruled out by steric hindrance (R—R = 1.7 A).

^a The values of φ and χ for structures 1 to 36 in this table apply also to structures 1 to 36 in the corresponding table for trans-amide groups.⁴

cis-amide group is such that no acceptable configuration of equivalent residues with a twofold screw axis can be constructed. Helical arrangements of polypeptide chains of equivalent cis-amide groups with a sixfold screw axis can be built in many closely similar ways. Thus the orientations $\varphi = 30^\circ$ and $\chi = 355^\circ$ lead to a helix with six residues per turn and pitch 8.6 A. This helix can form hydrogen bonds with equivalent adjacent helices to give a hexagonal crystal with $a = 10.8$ A and $c = 8.6$ A. The CHR groups are near the helical axis of each molecule, the hydrogen atoms being in contact with their centers at the distance 1.5 A from the axis. The β carbon atoms of the side chains are at radius 2.6 A. Each chain is linked to each of the six adjacent chains by a pair of hydrogen bonds formed by an HNCHRCO group. With assumed density 1.35 g. cm.⁻³, the mean residue weight is 11.8. This structure is accordingly a possible one, though perhaps unlikely, for polypeptides or proteins with average residue weight

about 118. It is interesting to note that two helices of this sort could be intertwined, without evident steric hindrance, in such a way as to lead to a hexagonal crystal with $a = 10.8$ Å and $c = 4.3$ Å, the latter identity distance corresponding to the translation from an equivalent point on one to a similar point on the other helix intertwined with it. The residue weight for this structure is calculated to be about 60, and accordingly it is a possible structure for polyglycine.

Helices with six residues per turn are also obtained by other pairs of orientations around the C—N and C—C bonds. For example, the orientations $\varphi = 22^\circ$, $\chi = 338^\circ$ lead to a somewhat similar six-residue helix, with pitch 11.7 Å. It is possible that three helices of this sort could be intertwined, giving molecules that could form crystals with a hexagonal unit with $a = 10.8$ Å and $c = 3.9$ Å, containing six residues, with residue weight about 54. This three-chain structure is accordingly also a possible one for polyglycine. We consider it to be unlikely, however, because there is probable instability connected with the *cis* configurations of the amide groups, rather than the *trans* configuration, and also instability connected with the orientations around the N—C and C—C' bonds.

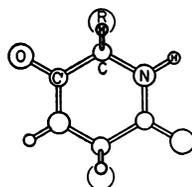


FIGURE 1

A drawing of the planar cyclic dipeptide diketopiperazine, structure 3 of table 1.

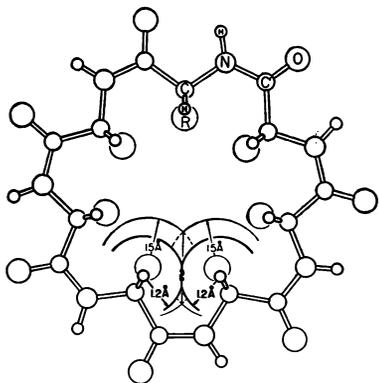


FIGURE 2

A drawing of a possible structure of a cyclic heptapeptide, structure 12 of table 1.

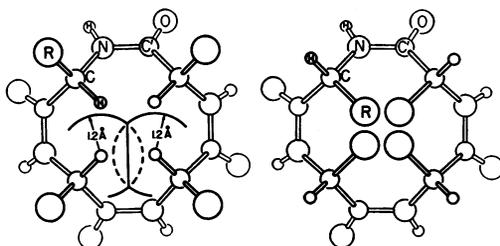


FIGURE 3

A drawing of two structures for cyclic tetrapeptides, structures 1 and 2 of table 1. Structure 2 is ruled out, except for tetraglycyl, by steric hindrance between R groups.

There are four cyclic structures described in table 1. Of these the only one that is known to be represented in nature is the dipeptide ring, structure 3, which is the diketopiperazine structure (Fig. 1). The planar seven-residue ring, structure 12, may well be shown by cyclic heptapeptides,

despite its inability to form hydrogen bonds with similar molecules (Fig. 2). Presumably heptapeptides crystallized from a solvent would tend to include molecules of the solvent in the crystals. The other two cyclic peptides described in table 1 are the cyclic tetrapeptides, structures 1 and 2 (Fig. 3). Structure 2 is ruled out, except for tetraglycyl, by steric hindrance between R groups. Structure 1 would presumably be assumed by cyclic tetrapeptides; some instability is conferred on it by the close approach of hydrogen atoms. Except for this instability, the tetragonal crystals involving staggered sheets of the cyclic tetrapeptide molecule connected by hydrogen bonds would be predicted for all cyclic tetrapeptides, the space

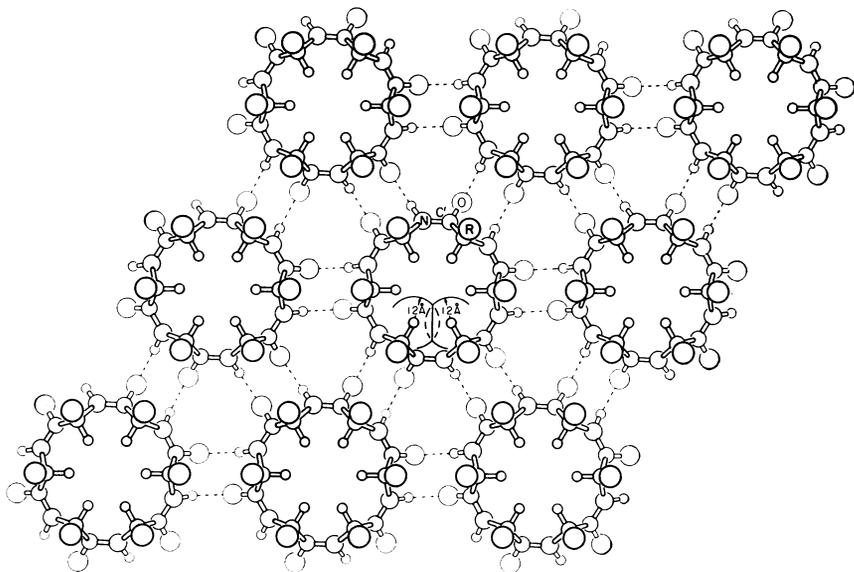


FIGURE 4

A drawing of a possible structure of a cyclic hexapeptide with unfavorable orientations about the bonds to the α carbon atom. Possible hydrogen bonding is shown between adjacent molecules, forming a two-dimensional crystalline sheet.

occupied by side chains being provided by suitable selection of the value of the length of the c axis of the tetragonal unit.

In addition, there are six cyclic structures involving equivalent cis-amide groups with unfavorable orientations about the bonds to the α carbon atom. These are two tripeptide structures, two pentapeptide structures, and two hexapeptide structures, all similar to the tetrapeptide structures 1 and 2. The three structures analogous to structure 2, with the R groups pointing toward the symmetry axis, are eliminated by steric hindrance. The other three structures involve a tight fit of the hydrogen atoms, which would confer some instability upon them. The pentapep-

tide cannot form hydrogen-bonded crystals with all residues equivalent, but the tripeptide can form crystals of this sort. In the cyclic tripeptide molecules the orientation about the N—C axis is $\varphi = 280^\circ$, and that about the C—C' axis is $\chi = 283^\circ$. Hexagonal crystals involving staggered layers can be formed by the cyclic tripeptide molecules, with two slightly different kinds of molecules alternating. Those of one kind form pairs of hydrogen bonds with their three neighbors with use of CONH groups in each pair, whereas the pairs of hydrogen bonds are formed by the molecules of the other kind with use of NHCHRCO groups. The lateral edge of the hexagonal unit is $a = 10.4$ A, the value of c being determined by the side chains. For cyclic triglycyl, for which mirror-image three-residue rings may be formed, the staggered sheet could be built up of equivalent residues, the pairs of hydrogen bonds being formed by the groups CONH of each molecule, or perhaps better by the groups NHCHRCO.

The cyclic hexapeptide with CH groups turned toward the hexagonal axis of the ring has orientations $\varphi = 337^\circ$, $\chi = 337^\circ$. The hydrogen atoms are about 1.7 A apart, introducing a slight instability due to steric hindrance. The molecules could form a hexagonal polar sheet with similar molecules, the hydrogen bonds being bent about 20° from the N—H axis. The value of a is 10.8 A, that of c being determined by the nature of the side chains. The hydrogen bonds are probably of the NHCHRCO type, a pair of bonds formed by this group in one molecule connecting it with each of the six surrounding molecules in the sheet, as shown in figure 4. No cyclic hexapeptide has yet been subjected to X-ray investigation.

This investigation was aided by grants from The Rockefeller Foundation, The National Foundation for Infantile Paralysis, and The National Institutes of Health, Public Health Service.

* Contribution No. 1646.

¹ Pauling, L., and Corey, R. B., *J. Am. Chem. Soc.*, **72**, 5439 (1950).

² Pauling, L., Corey, R. B., and Branson, H. R., these PROCEEDINGS, **37**, 205 (1951)

³ Pauling, L., and Corey, R. B., *Ibid.*, **37**, 235, 241, 251, 256, 261, 272, 282 (1951).

⁴ Pauling, L., and Corey, R. B., *Ibid.*, **37**, 729 (1951).

⁵ Corey, R. B., *J. Am. Chem. Soc.*, **60**, 1598 (1938).