Calculation of the Structures of Collagen Models. Role of Interchain Interactions in Determining the Triple-Helical Coiled-Coil Conformation. 4. Poly(glycylalanylprolyl)[1]

George Némethy, Martha H. Miller, and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.

Received December 3, 1979

ABSTRACT: The conformational space of regularly repeating structures of poly(glycylalanylprolyl), $(GAP)_n$, was examined for stable triple-stranded complexes. The three strands were assumed to be equivalent. The structures generated included (a) coiled coils with either screw or rotational symmetry and (b) parallel-chain complexes with either screw or rotational symmetry. The dihedral angles for rotation about the single bonds of the three residues in the GAP unit were treated as the independent variables which repeated in each tripeptide unit. In parallel-chain structures, the interchain distance and orientation also were independent variables. In contrast to the results found in earlier papers of this series on $(Gly\text{-}Pro\text{-}Pro)_n$, $(Gly\text{-}Pro\text{-}Hyp)_n$, and $(Gly\text{-}Pro\text{-}Ala)_n$, coiled-coil and parallel-chain triple-stranded complexes had comparable energies. The lowest energy conformations with both of these kinds of symmetries differed from solid structures of $(GAP)_n$ polymers observed under various conditions. This disagreement can be rationalized, however, in terms of solvent effects, not included thus far in our calculations. The fact that the energies of structures with different kinds of symmetry are not very different suggests that minor solvent effects may lead to different structural preferences in various solvents, in agreement with observations. A collagen-like triple-helical structure is not of low energy for $(GAP)_n$ unless the flexibility of the prolyl ring is taken into account. This result points out the importance of including this flexibility in conformational energy computations on tightly packed structures.

I. Introduction

This paper is part of a series of studies[2-4] on the conformational properties of collagen-like regular-sequence poly(tripeptide)s which have been used as models of collagen.[5,6] The first three parts of the series described work on $(GPP)_n$, $(GPH)_n$, and $(GPA)_n$, respectively, where G = glycyl, P = prolyl, H = hydroxyprolyl, and A = alanyl. In this paper, we describe studies on poly(glycylalanylprolyl), abbreviated as $(GAP)_n$. The results will be compared with those[4] on $(GPA)_n$.

Alanine occurs frequently in position X of the GXY repeating unit of collagen, and Gly-Ala-Hyp is one of the frequently occurring triplets,[7,8] as summarized in the preceding paper.[4] Even though Pro occurs rarely in position Y in collagen, compared with Hyp, we studied $(GAP)_n$ rather than $(GAH)_n$ in the present work. This was done in order to facilitate comparisons with $(GPP)_n$ and $(GPA)_n$, considered in the earlier papers of this series.[2,4] The choice of this sequence was justified further by previous work[3] which indicated that the hydroxyprolyl OH group interacts very weakly with neighboring residues in the collagen-like triple-stranded structure in the absence of water. In certain other structures, some lowering of the total energy may occur because of the presence of interchain hydrogen bonds involving the OH group. A comparison of the results obtained for $(GPP)_n$ and $(GPH)_n$, however, indicated that such hydrogen bonds do not occur in low-energy conformations.[2,3]

The structures of $(GPA)_n$ and $(GAP)_n$ are considerably different in both the solid state and in solution. Evidence was summarized earlier[4] on the occurrence of the collagen-like triple-helical conformation in $(GPA)_n$ and in several other polymers of the sequence $(GPY)_n$. In contrast, $(GAP)_n$ can exist in several forms in the solid state, depending on the solvent used for the preparation of the solid. Its solution conformation also is solvent-dependent. Upon drying from aqueous solution, a polyproline II-like structure is obtained, with hydrogen-bonded sheets formed by antiparallel polypeptide chains.[9,10] A similar structure is obtained by drying from dioxane, acetone,[9] or an ethylene glycol–hexafluoroisopropyl alcohol mixture,[10] but a collagen-like triple-helical structure is obtained by drying from trifluoroethanol, as indicated by the X-ray diffraction pattern.[10] The solution conformation also differs in these solvents, as indicated by infrared and circular dichroic measurements.[10] The spectra of a polymer sample with $n \approx 60$, taken in trifluoroethanol or in the ethylene glycol–hexafluoroisopropyl alcohol mixture, are similar to those of collagen. The spectrum of a sample with $n \approx 20$, taken in water, corresponds to that of a disordered polypeptide chain. Because of the use of chains of different lengths in the two sets of measurements, it is not certain whether the difference in the spectra is due entirely to a solvent effect or whether it is also influenced by the difference in the degree of polymerization. In any case, however, the data cited indicate that the conformational behavior of $(GPA)_n$ and $(GAP)_n$ differ considerably. The differences extend to other poly(tripeptide)s with the sequence $(GXP)_n$. Poly(Gly-Leu-Pro) and poly(Gly-Phe-Pro) are less ordered than poly(Gly-Pro-Leu) and poly(Gly-Pro-Phe) in solution.[11,12]

In this study, we compare the computed conformational properties of $(GAP)_n$ with those of $(GPA)_n$ and $(GPP)_n$ in various kinds of triple-stranded arrangements. The work described here was carried out in the same manner reported in the three earlier papers of this series.[2-4] Conformational space was explored for both the single-stranded polymer and for coiled-coil and parallel-chain triple-stranded complexes. Minimum-energy conformations were computed and were compared with those reported earlier[2,4] for $(GPP)_n$ and $(GPA)_n$.

II. Computational Procedure

The assumptions and the computational procedures used were those reported in the first three papers of the series.[2-4] The same selection of geometrical and energy parameters was made,[2-4,13] with one exception. A C^α–H bond length of 1.09 Å was used here, in contrast to earlier work from this laboratory[13] and to the first papers of this series,[2-4] where the bond length of 1.00 Å was used. The shorter value was introduced earlier to define the bond length in the computer program ECEPP, developed in this laboratory.[13] Neutron diffraction data[14] published since then indicate that the C^α–H distance is 1.09 Å, like the

0024-9297/80/2213-0914$01.00/0

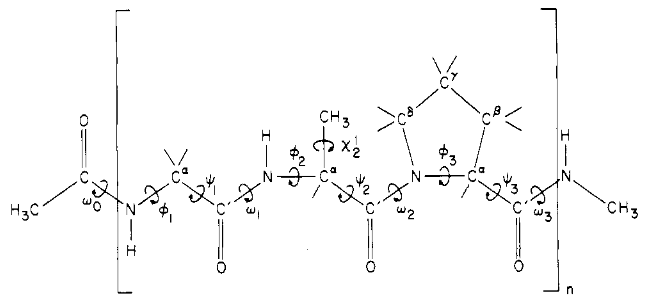


Figure 1. Structure of $(GAP)_n$, indicating the dihedral angles. The end groups used in the computation are shown.

other C–H distances in amino acid residues. Dihedral angles and energies of most conformations of $(GPP)_4$, $(GPH)_4$, and $(GPA)_4$ do not change noticeably as a result of this modification. Therefore, the results reported in the earlier papers[2-4] are unaffected by the change in bond length. Several of the conformations of $(GAP)_4$ of interest, however, are affected by the change. Therefore, the revised value of 1.09 Å was used in this paper.

The procedures described earlier[2] were used for the generation of chain conformations and for function minimization. Six variable dihedral angles were considered, as in the work[4] on $(GPA)_n$. Trans and cis conformations (with ω fixed at 180 or 0°) were considered only for the peptide bond preceding Pro. In the other two peptide bonds, preceding Gly and Ala, ω was fixed only at 180°. The repeating unit GAP is shown in Figure 1.

Selection of Starting Conformations and Energy Minimization. Essentially the same procedures were followed in the study of $(GAP)_n$ as those already described[4] for $(GPA)_n$.

Coiled-Coil Structures with Screw Symmetry. As the first step, minimum-energy conformations of single-stranded $(GPA)_4$ chains were determined, in the manner described before.[4] All conformational energy minima obtained earlier[15] for *N*-acetyl-*N'*-methyl-L-alanyl-L-prolinamide and *N*-acetyl-*N'*-methyl-L-prolylglycinamide were combined and used as starting points for the minimization of the energy of single-stranded $(GAP)_4$ with the *N*-acetyl and *N'*-methyl terminal blocking groups. The minima of the terminally blocked Gly-L-Ala dipeptide did not have to be considered separately because they are all included in the set selected above, except for those which are incompatible with a Pro-Gly sequence. This computation gave 90 minimum-energy conformations for single-stranded $(GAP)_4$, with an energy spread of 30.4 kcal/mol of $(GAP)_4$. These single-stranded minimum-energy structures were packed into triple-stranded coiled-coil complexes with screw symmetry and examined by a grid search as described[4] for $(GPA)_n$. This was followed by energy minimization, using all six dihedral angles (including $\chi_2{}^1$) as variables. In addition, the minimum-energy coiled-coil structures found earlier for $(GPP)_4$ (Tables V and VI of ref 2) also were used as starting points in this minimization.

Parallel-Chain Structures with Screw Symmetry. As in the three previous papers,[2-4] the only structures with this symmetry that were tested were those having a polypeptide conformation with $\Theta = 0°$. A total of 243 conformations were generated by means of the gridding procedure described in papers 2 and 3.[3,4] Energy minimization was carried out on these conformations with respect to chain separation and orientation in the triple-stranded complexes, with fixed dihedral angles. This is in contrast to paper 3 where no energy minimization was necessary.

Coiled-Coil and Parallel-Chain Structures with Rotational Symmetry. The procedure described in paper 3[4] was followed, using the same 50 starting conformations for energy minimization as in paper 3.

Computations with Altered Puckering of the Prolyl Ring. It was necessary (in the case of $(GAP)_n$), as an additional procedure, to test the effect of changing the geometry of the prolyl ring. The "down" puckering, with $\phi = -75.0°$ and $\chi^1 = 18.7°$, was used in all previous studies.[2-4] In the present work on $(GAP)_n$, both puckerings of the prolyl ring were tested, "down", as defined above, and "up", i.e., the conformation[13,16] with $\phi = -67.6°$ and $\chi^1 = -6.1°$, also denoted as exo.[17] Preliminary calculations indicated that atomic overlaps resulting in high energies in *some* of the low-energy coiled-coil $(GAP)_n$ structures could be relieved by allowing some flexibility in the proline ring geometry. This was not the case for the low-energy structures treated in the first three papers.[2-4] The geometry of the prolyl ring was changed to "up" puckering in all minima obtained with "down" puckering for both coiled-coil and parallel-chain structures, as described above. These conformations were then used as starting points for new energy minimizations. The energy minimizations were carried out with respect to the variables described above for each kind of symmetry.

III. Results

A. Triple-Stranded Structures. The energy minimization of the coiled-coil structures with screw symmetry and with "down" puckering resulted in 48 minimum-energy structures with counterclockwise disposition of the chains and 48 structures with clockwise disposition of the chains. The energies ranged from −51.0 to +5.6 kcal/mol of $(GAP)_4$. Minimization of all these structures with the puckering changed to "up" showed that there are only six counterclockwise and three clockwise conformations in which the energy is lower for the "up" than for the "down" puckering. The dihedral angles for the minima with $E < -26.0$ kcal/mol of $(GAP)_4$, i.e., for the conformations which are less than 25.0 kcal/mol of $(GAP)_4$ above the global minimum, are given in Tables I and II, for the two kinds of chain dispositions, respectively. The conformations are listed in order of increasing energy. Tables III and IV, respectively, list the helical parameters and the inter- and intramolecular energy components for these conformations.

Twenty-three minimum-energy triple-stranded coiled-coil and parallel-chain structures with rotational symmetry were found, with energies ranging from −29.6 to −5.4 kcal/mol of $(GAP)_4$. They are all of high energy and were not considered further. A total of 243 minimum-energy parallel-chain structures with screw symmetry were found, all of them with counterclockwise disposition of the chains. The energies ranged from −43.1 to 0.0 kcal/mol of $(GAP)_4$.

There are very few similarities between the low-energy conformations of $(GAP)_n$ and of $(GPP)_n$, as seen from a comparison of Tables I and II in this paper with Tables V and VI of the first paper.[2] It was pointed out earlier[4] in the case of $(GPA)_n$ that many coiled-coil structures of this polymer have no $(GPP)_n$ counterparts because their ϕ_{Ala} differs from the fixed value of ϕ_{Pro} in $(GPP)_n$. The same is true for $(GAP)_n$ as well. Even in those conformations in which ϕ_{Ala} is near −75°, the other dihedral angles usually differ widely from any of the combinations found for $(GPP)_n$, just as in the case of $(GPA)_n$. Consequently, D and Θ also differ from the collagen-like structure. The reasons for this difference were discussed earlier.[4] The energy difference between the lowest energy conformation with clockwise and the one with counterclockwise disposition ($\Delta E = 1.9$ kcal/mol of $(GAP)_4$) is even less than in the case of $(GPA)_n$. The few conforma-

Table I. Dihedral angles and energies of triple-stranded coiled-coil structures of $(GAP)_4$, with screw symmetry and counterclockwise disposition of the chains

Dihedral angle (degrees)[a]							Energy [kcal/mol of $(GAP)_4$]	Prolyl ring puckering[b]
ϕ_1	ψ_1	ϕ_2	ψ_2	χ_2^1	ω_2	ψ_3		
136	88	-137	82	-62	180	79	-51.0	
-61	-75	-150	75	58	180	82	-47.1	
-62	-66	-145	99	51	180	164	-45.4	U
-78	66	-149	77	57	180	-39	-42.6	U
88	-72	52	66	60	180	144	-40.4	
66	-73	53	76	58	180	142	-40.1	U
149	-79	-144	95	58	180	-58	-37.4	U
83	-75	-79	150	55	0	-45	-35.5	
82	-79	-73	148	57	0	-47	-35.3	
-164	175	-151	151	57	180	174	-34.4	
65	51	-113	146	63	0	-23	-34.1	
76	55	-101	148	57	0	-21	-33.8	
-142	-71	-156	73	58	180	169	-33.0	
134	169	-158	80	54	180	73	-32.9	
178	-163	55	61	65	180	146	-31.6	U
-77	169	-76	161	-48	180	155	-31.0[c]	U
-166	119	-64	139	61	180	159	-29.8	
-151	81	-110	139	63	0	166	-28.7	
-110	103	-140	141	60	0	141	-27.7	
-160	-64	53	78	63	180	-23	-26.9	
-155	123	-72	155	61	180	166	-26.4	
-137	166	-101	156	52	180	97	-26.3	

[a] Only ϕ_1, ψ_1, ϕ_2, ψ_2, χ_2^1, and ψ_3 were varied in energy minimization; ω_2 was kept fixed at the values indicated; $\omega_1 = \omega_3 = 180°$ for all conformations.

[b] The prolyl ring has the "up" puckering in the conformations marked U, and the "down" puckering in all other conformations.

[c] This conformation is the closest one to the collagen II-like structure.

Table II. Dihedral angles and energies of triple-stranded coiled-coil structures of $(GAP)_4$, with screw symmetry and clockwise disposition of the chains.

Dihedral angle (degrees)[a]							Energy [kcal/mol of $(GAP)_4$]	Prolyl ring puckering[b]
ϕ_1	ψ_1	ϕ_2	ψ_2	χ_2^1	ω_2	ψ_3		
-59	-72	-158	72	58	180	81	-49.1	
61	-67	-159	78	57	180	-50	-48.6	U
-160	-170	-156	155	49	180	144	-46.1	
-171	-61	-151	74	57	180	-180	-39.5	
-81	106	-155	148	53	0	149	-39.0	
153	-172	-154	73	57	180	-49	-38.3	
149	69	-149	82	58	180	78	-34.8	
106	52	-47	148	56	0	-24	-34.6	
77	-159	-160	157	40	180	77	-31.4	
-70	44	-73	155	66	180	121	-28.1	
61	-77	52	76	64	180	155	-28.0	
65	-69	-77	148	59	0	-37	-27.3	
86	-66	-85	150	52	0	-44	-27.3	
-82	72	-67	152	61	180	171	-27.1	
-63	-77	-155	110	58	180	155	-26.6	

[a] Only ϕ_1, ψ_1, ϕ_2, ψ_2, χ_2^1 and ψ_3 were varied in energy minimization; ω_2 was kept fixed at the values indicated; $\omega_1 = \omega_3 = 180°$ for all conformations.

[b] The prolyl ring has the "up" puckering in the conformations marked U, and the "down" puckering in all other conformations.

Table III. Helical Parameters, Interchain Energy, and Intrachain Energy of Triple-Stranded Coiled-Coil Structures of $(GAP)_4$, with Screw Symmetry and Counterclockwise Disposition of the Chains[a]

Helical Parameters			Energy[b] [kcal/mol of $(GAP)_4$]				
D	Θ[c]	ρ_1[d]	Interchain		Intrachain		Total
(Å)	(degrees)	(Å)	E_{ES}	E_{NB}	E_{ES}	E_{NB}	E_{tot}
3.2	75.9	6.6	-4.5	-33.7	-6.9	-5.9	-51.0
3.9	-73.9	5.7	-0.6	-27.0	-3.9	-15.7	-47.1
3.4	75.8	6.5	-10.6	-36.0	0.2	1.4	-45.4
4.0	74.6	6.5	1.3	-20.5	-6.0	-10.6	-42.6
4.4	75.0	6.5	1.0	-28.7	-4.1	-6.6	-40.4
4.1	75.3	6.5	-0.0	-29.5	-5.2	-5.4	-40.1
3.7	75.4	5.9	-0.4	-33.0	-3.6	-0.4	-37.4
5.5	-66.3	3.0	-2.4	-23.5	-7.0	-2.7	-35.5
5.5	-66.5	3.1	-2.4	-23.1	-6.9	-2.9	-35.3
9.5	-88.3	2.9	-3.5	-25.8	-6.0	0.7	-34.4
2.0	63.1	5.5	-5.2	-31.0	-1.2	3.2	-34.1
3.6	70.5	4.5	-3.9	-30.8	-1.4	2.2	-33.8
4.0	-82.5	5.6	4.0	-24.9	-1.2	-10.8	-33.0
5.4	77.1	5.7	-0.6	-20.3	-6.9	-5.3	-32.9
7.4	91.1	2.9	[illegible]	-20.3	-2.0	-1.5	-31.6
6.9	43.1	1.7	-6.6	-25.1	-3.5	7.2	-31.0[e]
9.2	-31.3	2.0	-4.6	-31.9	-5.6	12.4	-29.8
4.6	91.3	3.5	-0.4	-34.8	0.5	5.9	-28.7
4.7	66.9	3.4	-0.5	-33.6	-3.1	9.0	-27.7
4.7	-77.9	5.5	-5.6	-21.9	3.2	-2.6	-26.9
9.2	-76.7	2.6	1.2	-27.5	-4.7	4.7	-26.4
6.0	-110.4	3.4	-2.3	-26.7	-6.6	8.8	-26.3

[a] The conformations are listed in the same order as in Table I.

[b] The subscript ES refers to the electrostatic energy term, NB refers to the nonbonded term. Both of these terms include the contributions to the hydrogen bond energy. In addition to the four energy terms shown here, E_{tot} also contains a torsional energy term E_{TOR}, contributed by the side-chain rotation about the $C^{\alpha}-C^{\beta}$ bond in Ala and a prolyl ring energy which depends on whether ω = 0° or 180° preceding Pro.

[c] The sign of Θ determines the handedness of the major helix.[3]

[d] The distance of the glycyl N atom from the z-axis.

[e] This conformation is the closest one to the collagen II-like structure.

Table IV. Helical Parameters, Interchain Energy, and Intrachain Energy of Triple-Stranded Coiled-Coil Structures of $(GAP)_4$, with Screw Symmetry and Clockwise Disposition of the Chains[a]

Helical Parameters			Energy[b] [kcal/mol of $(GAP)_4$]				
D	Θ[c]	ρ_1[d]	Interchain		Intrachain		Total
(Å)	(degrees)	(Å)	E_{ES}	E_{NB}	E_{ES}	E_{NB}	E_{tot}
4.8	-77.2	5.1	-2.2	-30.0	-3.5	-15.3	-49.1
4.4	-72.5	5.5	-2.3	-24.0	-6.9	-15.6	-48.6
8.4	-122.2	3.0	-2.2	-40.0	-6.6	0.0	-46.1
4.9	-88.8	4.9	1.9	-29.4	-1.7	-10.3	-39.5
6.0	83.1	3.1	-1.3	-33.9	-2.7	-0.8	-39.0
7.0	-73.6	3.9	1.6	-29.0	-4.3	-6.6	-38.3
3.7	76.9	6.3	0.2	-19.9	-6.6	-5.4	-34.8
4.5	79.8	3.6	-5.3	-32.9	-2.1	5.6	-34.6
6.2	78.2	5.0	-1.0	-25.5	-8.1	-1.5	-31.4
7.5	-67.6	3.7	-1.5	-20.0	-7.9	1.4	-28.1
4.4	70.4	6.9	3.7	-18.0	-4.2	-4.6	-28.0
4.6	-62.3	4.1	-0.1	-17.6	-7.7	-2.0	-27.3
5.3	-63.1	3.4	1.6	-21.0	-7.1	-1.2	-27.3
5.4	-70.1	6.2	-0.6	-17.4	-6.6	-2.5	-27.1
3.2	66.7	7.2	-4.3	-23.2	-0.4	1.0	-26.6

[a] The conformations are listed in the same order as in Table II.

[b] The subscript ES refers to the electrostatic energy term, NB refers to the nonbonded term. Both of these terms include the contributions to the hydrogen bond energy. In addition to the four energy terms shown here, E_{tot} also contains a torsional energy term E_{TOR}, contributed by the side-chain rotation about the $C^{\alpha}-C^{\beta}$ bond in Ala and a prolyl ring energy which depends on whether ω = 0° or 180° preceding Pro.

[c] The sign of Θ determines the handedness of the major helix.[3]

[d] The distance of the glycyl N atom from the z-axis.

Table V. Cartesian Coordinates[a] of One Tripeptide Unit[b] of the Lowest-Energy Triple-Stranded Coiled-Coil Minimum-Energy Structure with Screw Symmetry[c] of $(GAP)_n$ with "Down" Puckering of the Prolyl Ring

Residue	Atom	X	Y (Angstrom units)	Z
GLYCINE	N	6.62	0.00	0.00
	HN	5.78	-0.32	0.44
	CA	6.93	1.44	0.06
	HA	7.19	1.71	1.05
	HA	7.60	1.73	-0.65
	C'	5.53	2.19	0.14
	O	5.27	2.54	-1.39
ALANINE	N	4.76	2.42	0.91
	HN	4.98	2.14	1.74
	CA	3.50	3.13	0.66
	HA	3.53	3.68	-0.28
	CB	2.35	2.13	0.62
	HB	1.41	2.66	0.48
	HB	2.50	1.44	-0.21
	HB	2.32	1.57	1.55
	C'	3.35	4.14	1.81
	O	2.68	3.86	2.80
PROLINE-DOWN	NP	4.01	5.31	1.63
	CA	3.95	6.36	2.64
	HA	4.06	5.93	3.63
	CB	5.11	7.29	2.31
	HB	4.82	8.33	2.43
	HB	5.96	7.11	2.97
	CG	5.48	6.99	0.86
	HG	5.16	7.80	0.21
	HG	6.57	6.90	0.76
	CD	4.81	5.68	0.47
	HD	4.19	5.81	-0.42
	HD	5.54	4.91	0.24
	C'	2.60	7.07	2.62
	O	2.47	8.16	2.07

[a] Expressed in a right-handed coordinate system whose z-axis coincides with the major helical axis (the horizontal axis of Fig. 2).

[b] The coordinates of the other tripeptide units in the same and in neighboring strands can be obtained by the use of symmetry operations as described in ref. 2.

[c] This is the lowest-energy conformation computed for $(GAP)_4$. It is the first conformation listed in Tables I and III. It is shown in Fig. 2.

Table VI. Cartesian Coordinates[a] of One Tripeptide Unit[b] of the Lowest-Energy Triple-Stranded Parallel-Chain Structure with Screw Symmetry[c] of $(GAP)_n$ with "Down" Puckering of the Prolyl Ring

Residue	Atom	X	Y (Angstrom units)	Z
GLYCINE	N	[illegible]	[illegible]	[illegible]
	HN	[illegible]	[illegible]	[illegible]
	CA	2.25	[illegible]	1.34
	HA	[illegible]	1.51	0.26
	HA	2.40	0.00	1.45
	C'	1.69	1.42	2.31
	O	0.48	-1.94	2.53
ALANINE	N	2.59	2.56	3.05
	HN	3.57	2.53	2.86
	CA	2.20	-3.36	4.19
	HA	1.32	3.93	3.91
	CB	3.33	-4.32	4.56
	HB	3.03	-4.92	5.42
	HB	3.53	-4.98	3.71
	HB	4.37	3.75	4.80
	C'	1.83	-2.43	5.35
	O	2.43	1.37	5.52
PROLINE	NP	0.82	-2.87	6.15
	CA	0.36	-2.69	7.28
	HA	0.30	-1.04	7.02
	CB	1.01	-2.66	7.53
	HB	-1.14	-2.74	8.71
	HB	1.00	2.01	7.25
	CG	-1.07	-4.03	6.97
	HG	-0.59	4.82	7.72
	HG	2.01	-4.17	6.46
	CD	0.08	4.12	5.96
	HD	0.71	-4.90	6.18
	HD	-0.29	-4.23	4.96
	C'	1.35	-2.19	8.45
	O	2.10	3.17	8.55

[a] Expressed in a right-handed coordinate system whose z-axis coincides with the major helical axis (the horizontal axis of Fig. 3).

[b] The coordinates of the other tripeptide units in the same and in neighboring strands can be obtained by the use of symmetry operations as described in ref. 2.

[c] This conformation is shown in Fig. 3. It is described in the text.

Table VII. Cartesian Coordinates[a] of One Tripeptide Unit[b] of the Triple-Stranded Coiled-Coil Conformation[c] of $(GAP)_n$ Which Is Closest to the Collagen-like Conformation of Other Poly(tripeptide)s, with "Up" Puckering of the Prolyl Ring

Residue	Atom	X	Y (Angstrom units)	Z
GLYCINE	N	1.67	0.00	0.00
	HN	0.99	-0.73	0.11
	CA	1.44	1.19	0.81
	HA	1.59	2.08	0.20
	HA	0.40	1.20	1.15
	C'	2.38	1.22	2.01
	O	3.04	0.23	2.32
ALANINE	N	2.40	2.37	2.67
	HN	1.86	3.18	2.41
	CA	3.25	2.55	3.83
	HA	4.22	2.09	3.62
	CB	3.46	4.04	4.09
	HB	3.92	4.18	5.07
	HB	4.12	4.46	3.32
	HB	2.50	4.56	4.06
	C'	2.62	1.84	5.03
	O	1.43	1.54	5.02
PROLINE-UP	NP	3.47	1.58	6.06
	CA	3.01	0.91	7.26
	HA	2.46	0.00	7.00
	CB	4.27	0.58	8.04
	HB	4.30	1.14	8.98
	HB	4.31	-0.48	8.30
	CG	5.44	0.96	7.15
	HG	6.23	1.44	7.73
	HG	5.88	0.08	6.68
	CD	4.89	1.92	6.11
	HD	5.04	2.96	6.39
	HD	5.37	1.78	5.14
	C'	2.03	1.79	8.04
	C	2.04	3.01	7.90

[a] Expressed in a right-handed coordinate system whose z-axis coincides with the major helical axis.

[b] The coordinates of the other tripeptide units in the same and in neighboring strands can be obtained by the use of symmetry operations as described in ref. 2.

[c] This is the conformation marked c in Table I.

tions of coiled-coil $(GAP)_n$ which are similar to the low-energy conformations of $(GPP)_n$ have much higher relative energies. In other words, many non-collagen-like triple-stranded coiled-coil and parallel-chain complex conformations of $(GAP)_n$ are low in energy, as compared with the collagen-like structure. This makes the *relative* energy of the collagen-like structure high.

The coordinates of the lowest energy conformation of $(GAP)_n$ are given in Table V. This is a coiled-coil structure with right-handed major helical sense and a counterclockwise disposition of the chains, with an energy of −51.0 kcal/mol of $(GAP)_4$. Its dihedral angles are $(\phi_1, \psi_1, \phi_2, \psi_2, \chi_2^1, \omega_2, \psi_3)$ = (136°, 88°, −137°, 82°, −62°, 180°, 79°) (first line of Tables I and III). Its structure is shown in Figure 2. It is very different from the lowest energy conformations found for the other poly(tripeptide)s.[2-4] It is a helix with much lower pitch than collagen. Its helical parameters are D = 3.2 Å and Θ = 75.9°.

The lowest energy parallel-chain triple-stranded structure with screw symmetry is shown in Figure 3. Its energy

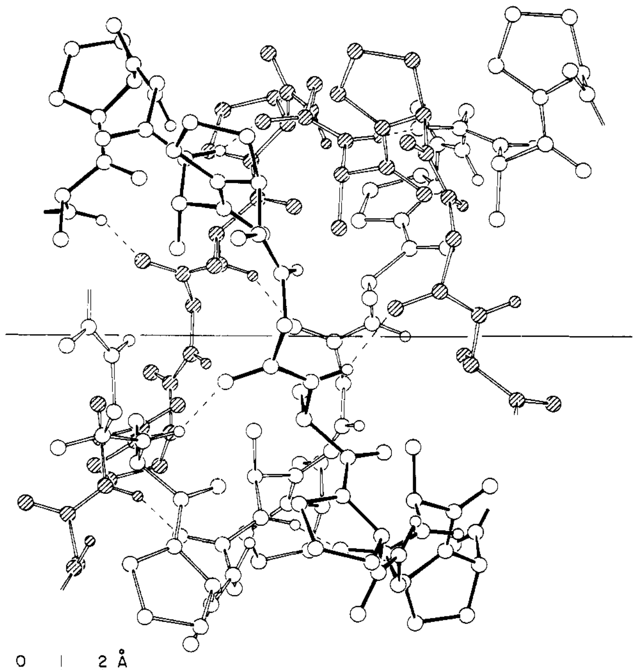


Figure 2. Triple-stranded coiled-coil structure with screw symmetry of $(GAP)_n$. The lowest energy coiled-coil structure with "down" puckering of the prolyl ring is shown (cf. first line of Table I). The chains form right-handed helices around the z axis ($\Theta > 0$), and the three strands are arranged in a counterclockwise disposition. Strand 1 is shown with dark bonds, strand 2 is shown with shaded atoms, and strand 3 is shown with unshaded bonds and atoms. The axis of symmetry is the z axis, shown as a horizontal line, with coordinates increasing to the right. The x axis is in the plane of the figure, pointing toward the top, and the y axis points up from the plane of the figure. The N–H···O hydrogen bonds connecting the strands are shown with dashed lines.

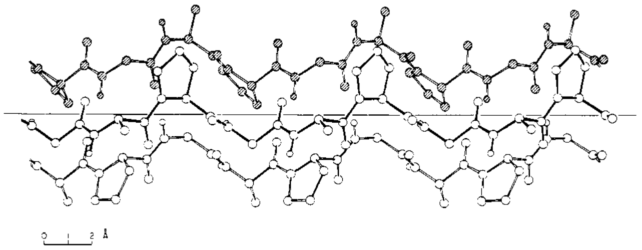


Figure 3. Triple-stranded parallel-chain structure with screw symmetry of $(GAP)_n$. This is the parallel-chain structure of lowest energy found for the "down" puckering of the prolyl ring. The three strands are arranged in a counterclockwise disposition. Strand 1 is shown with open bonds and atoms, strand 2 is shown with shaded atoms, and strand 3 is shown with dark bonds. The axis of symmetry is the z axis, shown as a horizontal line, with coordinates increasing toward the right. The x axis is perpendicular to the plane of the figure and points downward, and the y axis is in the plane, pointing toward the top.

is −43.1 kcal/mol of $(GAP)_4$. Its coordinates are listed in Table VI. The disposition of the three chains is counterclockwise, and the translational repeat is $D = 9.3$ Å. The interchain orientation angle is 156.8°. It is the angle in the xy plane between the vector from the axis of symmetry to the major axis of one of the strands and the vector from the major axis to the projection of a glycyl N atom onto this plane. The distance between the strands is 3.1 Å, measured between the projections of glycyl N atoms onto the xy plane. The dihedral angles are $(\phi_1, \psi_1, \phi_2, \psi_2, \chi_2^1, \omega_2, \psi_3) = (-83°, 144°, -75°, 145°, 60°, 180°, 155°)$.

The coiled-coil structure which is closest in conformation to the "collagen II"-like[18] structure (most stable[2] for $(GPP)_n$), is relatively high in energy. With the "down" puckering of proline, the computed energy is −23.4 kcal/mol of $(GAP)_4$. Energy minimization with the "up" puckering of the prolyl ring decreases the energy to −31.0 kcal/mol of $(GAP)_4$, even though the internal energy of each prolyl residue is 0.215 kcal/mol higher for the "up" puckering.[13,16] The dihedral angles of the conformation with "up" puckering are $(\phi_1, \psi_1, \phi_2, \psi_2, \chi_2^1, \omega_2, \psi_3) = (-77°, 169°, -76°, 161°, -48°, 180°, 158°)$. They are close to those of the most stable $(GPP)_4$ conformation. Agreement of the dihedral angles is worse in the conformation with "down" puckering. The coordinates of this structure are given in Table VII. Its helical parameters are $D = 8.9$ Å and $\Theta = 43.1°$ (line 16 of Tables I and III).

The energy of the collagen II-like conformation is relatively high with both proline puckerings, as compared to some other coiled-coil and parallel-chain structures, because a few close atomic approaches cause an increase in the nonbonded energy. The analysis of the interactions indicated that these atomic overlaps are due mainly to the change in the backbone geometry upon substitution of alanine for proline, and in particular, to the change in the $NC^{\alpha}C'$ bond angle. The ECEPP computer program used in this laboratory for peptide computations employs fixed bond lengths and bond angles.[13] These geometrical parameters were determined as average values over a variety of crystal structures of small molecules.[19] This standard data set uses different values of $\tau(NC^{\alpha}C')$ for Pro and Ala, viz., $\tau = 111.0$ and 109.3°, respectively. The change of 1.7° in this bond angle, when Ala is substituted for Pro in the X position of $(GPP)_4$, is sufficient to result in repulsive interactions involving some atoms of the prolyl ring and several nearby atoms in the "collagen II"-like triple helix.[20] Some of these interactions become less unfavorable when the puckering of the prolyl ring is changed. This result points out the importance of allowing for flexibility in the proline geometry[17,21] in computations involving sterically closely packed structures. If the three ω's are allowed to vary (instead of being fixed at 180°), there is no change in the dihedral angles and in the energy of the collagen II-like conformation. Thus, the increased flexibility of the backbone cannot relieve the unfavorable interactions in this conformation. The energy of the lowest energy conformation, however, is lowered by 1.5 kcal/mol of $(GAP)_4$. This change does not alter the relationships discussed in this paper. A more detailed discussion of the effect of various substitutions in positions X and Y on interactions in the collagen-like triple helix will be presented elsewhere.[22]

B. Comparison with Experiment. As discussed in the Introduction, two different crystalline forms of $(GAP)_n$ have been observed, depending on the solvent used to prepare the solid. One of them, prepared from trifluoroethanol,[10] has collagen-like reflections, with $D = 8.55$ Å. (No value of Θ was reported.) We find a corresponding structure in our computations, with "up" puckering of the prolyl ring, $D = 8.9$ Å and $\Theta = 43.1°$. It is very similar to the computed[2-4] collagen-like structures of $(GPP)_n$, $(GPH)_n$, and $(GPA)_n$. The computed and observed values of D agree very well. The energy of our computed structure, however, is high, as compared with other coiled-coil and parallel-strand complexes.

Two very similar solid structures are obtained from water[9] and from acetone,[9] dioxane,[9] or an ethylene glycol–hexafluoroisopropyl alcohol mixture.[10] In both of them, the polypeptide chains have an extended conformation, similar to polyproline II, and they associate into hydrogen-bonded sheets, with chains running antiparallel.

The sheets are packed together in the solid. Both structures contain two tripeptides per unit cell. The two structures differ slightly in the placement of the strands relative to each other and hence in the spacing of equivalent chains in the sheets and in the spacing of the sheets. There is a 9.40-Å repeat distance, however, in both structures. This is very close to the 9.36-Å repeat distance in the direction of the polypeptide strands in polyproline II.[23] The computed lowest energy parallel-chain triple-stranded complex also has a repeat distance $D = 9.3$ Å. All three strands, however, run parallel, and the packing of the chains in this complex is very different from that observed experimentally, there being three tripeptides per unit cell. This is not necessarily a contradiction, because we restricted our computations to triple-stranded complexes with an axis of symmetry. Therefore, the computation did not generate any sheetlike structures. The only conclusion that can be reached is that computed polyproline II-like parallel-chain conformations are stable, in agreement with experiment.

IV. Discussion

The energies of the computed lowest energy triple-stranded structures with different kinds of symmetry were as follows: −51.0, −43.1, and −29.6 kcal/mol of $(GAP)_4$, for coiled-coil structures with screw symmetry, for parallel chains with screw symmetry, and for structures with rotational symmetry. The coiled-coil structure which is closest to collagen II in its conformation has a higher energy, viz., −31.0 kcal/mol of $(GAP)_4$.

The results obtained here for $(GAP)_n$ differ in two significant respects from those obtained earlier[2-4] for $(GPP)_n$, $(GPH)_n$, and $(GPA)_n$. First, in the other three polymers, all of the parallel-chain, i.e., not supercoiled, triple-stranded complexes were much higher in energy than the most stable coiled-coil structures. The energy difference between the lowest energy parallel-chain and coiled-coil structures amounted to 14.3 kcal/mol for $(GAP)_4$ and 23.8 kcal/mol for $(GPP)_4$. In contrast, the energy difference of the lowest energy coiled-coil and parallel-chain complexes of $(GAP)_n$ is much smaller, viz., 7.9 kcal/mol of $(GAP)_4$. Second, the lowest energy structure found for all three other polymers was a "collagen II"-like triple helix,[17] while this structure was not the lowest energy one in $(GAP)_n$.

Even though the lowest energy triple-stranded structure computed here for $(GAP)_n$ does not agree with structures determined experimentally for this poly(tripeptide),[9,10] the general trend of the calculations for the various sequences agrees with experiment. For the other three poly(tripeptide)s, parallel-chain complexes without supercoiling were computed to have much higher energies than the coiled-coil triple helices. Experimentally, no structures with parallel packing of chains have been observed for them, either. On the other hand, energy differences between the two kinds of triple-strand complexes of $(GAP)_n$ are much smaller. Both kinds of complexes have been observed[9,10] for $(GAP)_n$ under various solvent conditions.

The increased number of low-energy triple-stranded structures for $(GAP)_n$, as compared with $(GPA)_n$, suggests that the flexibility of the former chain is larger. There are 11 coiled-coil complexes of $(GAP)_n$ within a 12.0 kcal/mol of $(GAP)_4$ interval (i.e., 3.0 kcal/mol in terms of the GAP repeat unit) above the lowest energy structure (Tables I and II), but only four coiled-coil complexes of $(GPA)_n$ within the same energy range. If this reflects the flexibility of the two sequences in randomly coiled chains, too, it would help to explain why poly(tripeptide)s with the sequences $(GPY)_n$ are generally more ordered in solution[10-12] than those with sequences $(GXP)_n$. A recent study[24] attributed differences in the stability of triple-helical structures of $(GPY)_n$ and $(GXP)_n$ poly(tripeptide)s to differences in the conformational preferences of Gly-Pro and Pro-Gly dipeptides. The higher tendency of Pro-Gly to form β bends,[15] as compared with that of Gly-Pro, makes extended structures (necessary to form collagen-like triple helices) relatively less favorable[24] in $(GXP)_n$. The decreased tendency for Pro-Gly dipeptides to form extended structures (as evidenced by the higher bend probability[15]) may be reflected in the computed preference of $(GAP)_4$ to form compactly coiled triple-stranded complexes instead of the more extended collagen-like triple helix (see Tables III and IV).

The disagreement between the most stable structures computed and observed for $(GAP)_n$ can be rationalized in terms of solvent effects. It is seen experimentally that the conformation of $(GAP)_n$ is sensitive to changes in solvent, both in solution and in the solid state. In our calculations, solvent effects have not been considered thus far. The fact that the computed energies for supercoiled and nonsupercoiled structures are not very different suggests that solvent interaction energies might alter the relative magnitudes of the potential energies and that different kinds of solvation may cause one or the other of the low-energy structures to be preferred. Energy changes of 2–3 kcal/mol of the GAP repeat unit are sufficient to alter the relative stability of various coiled-coil and parallel-chain complexes. Interactions with the solvent could result in such energy differences.

The "collagen II"-like triple helix[17] was computed to be very stable for $(GPP)_n$, $(GPH)_n$, and $(GPA)_n$, as compared to other triple-stranded conformations, but not so stable for $(GAP)_n$. Destabilization of this structure for $(GAP)_n$ results from small differences of the residue geometries of Pro and Ala, especially of bond angles. This finding agrees with experimental observations on the relative ease of triple-helix formation by poly(tripeptide)s[9-12] with sequences $(GPY)_n$, compared to those with sequences $(GXP)_n$. An earlier, approximate theoretical computation resulted in similar conclusions.[25]

The effects of altering residue geometry (e.g., by changing from "down" to "up" puckering of the prolyl ring) are more pronounced in $(GAP)_n$ than in $(GPA)_n$ or $(GPP)_n$, as seen from the comparison of dihedral angles and energies of stable conformations. This is a consequence of small differences in the local environment of residues in position X or Y along the sequence in tightly packed triple-stranded complexes. A more detailed analysis of conformational effects accompanying amino acid substitutions in positions X and Y in the collagen II structure will be published later.[22]

The studies presented here and in the first three papers of the series[2-4] indicate that stable coiled-coil triple helices can be formed by poly(tripeptide)s of the sequence $(GXY)_n$, where G = Gly and X or Y or both can be either Pro or Ala (or Y can be Hyp). In most cases considered, the most stable triple-stranded conformation is very close to that observed experimentally for collagen and for collagen-like $(GXY)_n$ poly(tripeptide)s. Only in the case of $(GAP)_n$ is this not the lowest energy computed conformation, but its energy is not very high. The results indicate that substitution of either amino or imino acid residues in positions X and Y of collagen do not result in conformational destabilization of the collagen triple helix. The collagen-like conformation of the poly(tripeptide) chain is not stable in an isolated strand for any of the sequences. Thus, interchain energy accounts for a large measure of the stability of the collagen-like triple-helical structure.

Acknowledgment. We thank Dr. S. S. Zimmerman for helpful discussions and M. S. Pottle and S. Rumsey for extensive help with the computations. This work was supported by research grants from the National Science Foundation (PCM75-08691) and from the National Institute of General Medical Sciences (GM-14312) and the National Institute on Aging (AG-00322) of the National Institutes of Health.

Miniprint Material Available: Full-size copies of Tables I-VII (7 pages). Ordering information is given on any current masthead page.

References and Notes

(1) Presented before the Division of Cellulose, Paper, and Textile at the 178th National Meeting of the American Chemical Society, Washington, D.C., Sept 1979, Abstract No. CELL 13.
(2) Miller, M. H.; Scheraga, H. A. *J. Polym. Sci., Polym. Symp.* **1976**, *54*, 171.
(3) Miller, M. H.; Nĕmethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 470.
(4) Miller, M. H.; Nĕmethy, G.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 910.
(5) Traub, W.; Piez, K. A. *Adv. Protein Chem.* **1971**, *25*, 243.
(6) Bhatnagar, R. S.; Rapaka, R. S. In "Biochemistry of Collagen"; Ramachandran, G. N., Reddi, A. K., Eds.; Plenum Press: New York and London, 1976; p 479.
(7) Hulmes, D. J. S.; Miller, A.; Parry, D. A. D.; Piez, K. A.; Woodhead-Galloway, J. *J. Mol. Biol.* **1973**, *79*, 137.
(8) Fietzek, P. P.; Glanville, R. W., personal communication.
(9) Segal, D. M.; Traub, W. *J. Mol. Biol.* **1969**, *43*, 487.
(10) Doyle, B. B.; Traub, W.; Lorenzi, G. P.; Blout, E. R. *Biochemistry* **1971**, *10*, 3052.
(11) Scatturin, A.; Tamburro, A. M.; Del Pra, A.; Bordignon, E. *Int. J. Pept. Protein Res.* **1975**, *7*, 425.
(12) Tamburro, A. M.; Scatturin, A.; Del Pra, A. *Int. J. Pept. Protein Res.* **1977**, *9*, 310.
(13) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
(14) (a) Al-Karaghouli, A. P.; Koetzle, T. F. *Acta Crystallogr., Sect. B* **1975**, *31*, 2461. (b) Power, L. F.; Turner, K. E.; Moore, F. H. *Ibid.* **1976**, *32*, 11. (c) Kvick, Å.; Al-Karaghouli, A. R.; Koetzle, T. F. *Ibid.* **1977**, *33*, 3796.
(15) Zimmerman, S. S.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 811.
(16) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 698.
(17) Ramachandran, G. N.; Bansal, M. *Curr. Sci.* **1976**, *45*, 647.
(18) Rich, A.; Crick, F. H. C. *Nature (London)* **1955**, *176*, 915.
(19) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.
(20) The role of this change of bond angle was tested by generating a triple-stranded $(GAP)_4$ complex in the conformation corresponding to the lowest energy structure[2] of $(GPP)_4$, using an artificial Ala residue, with $\tau(NC^{\alpha}C') = 111.0°$, as in proline. The repulsive interactions were reduced by the change in geometry.
(21) Madison, V. *Biopolymers* **1977**, *16*, 2671; **1979**, *18*, 1563.
(22) Nĕmethy, G.; Scheraga, H. A., to be submitted for publication.
(23) Cowan, P. M.; McGavin, S. *Nature (London)* **1955**, *176*, 501.
(24) Brahmachari, S. K.; Ananthanarayanan, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 5119.
(25) Hopfinger, A. J.; Walton, A. G. *Biopolymers* **1970**, *9*, 29.

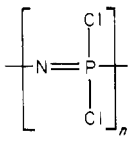
X-ray Diffraction Analysis of the Structure of Poly(dichlorophosphazene)[1]

H. R. Allcock,* R. A. Arcus, and E. G. Stroh

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802. *Received October 2, 1979*

ABSTRACT: Improved X-ray diffraction photographs, containing 63 unique reflections, have been obtained for highly purified poly(dichlorophosphazene) with the use of inclined-beam X-ray techniques. The results have been analyzed by a combination of unit cell indexing, model building, conformational analysis, and helical and optical transform methods. The chain-repeat distance is 4.92 Å. An exact structural solution was precluded by an unusual combination of factors that lead to disorder of the system. However, the data are consistent with a cis–trans planar chain conformation and with a chain-packing arrangement that allows appreciable "rotational" disorder between adjacent chains and additional disorder with respect to the directional "sense" of different chains. These factors help to explain the unusual physical properties of this polymer.

In this paper we explore the conformational and chain-packing structure of poly(dichlorophosphazene), $(NPCl_2)_n$ (I), by means of an analysis of X-ray diffraction data.



$$\left[-N{=}\overset{Cl}{\underset{Cl}{P}}-\right]_n$$

I

The objectives of this study were to obtain improved molecular structural data for this polymer and to correlate the unusual physical properties of this material with the molecular and chain-packing features. Poly(dichlorophosphazene) is one of the simplest polyphosphazenes known. As such, it provides a vital starting point for an understanding of the structural characteristics of the more complex poly(organophosphazenes).[2-12]

Two other high-polymeric phosphazenes have been investigated by X-ray diffraction techniques. The structure of the low-temperature form of poly(difluorophosphazene), $(NPF_2)_n$, was solved by Allcock, Kugel, and Stroh.[13] The organophosphazene poly[bis(*p*-chlorophenoxy)phosphazene] studied by Bishop and Hall,[14] but the structure proposed had a high discrepancy factor (45%). Chain repeat distance information was obtained by Allcock, Kugel, Valan, Siegel, and Stroh,[3,15] for the polymers $[NP(OCH_2CF_3)_2]_n$ and $[NP(OC_6H_5)_2]_n$, but difficulties were encountered in the complete structure solutions of these materials.

Poly(dichlorophosphazene) itself has been studied by earlier investigators. One of the earliest attempts to solve the structure of any polymer (inorganic or organic) was carried out in 1936 by Meyer, Lotmar, and Pankow[16] with a cross-linked form of $(NPCl_2)_n$. On the basis of an analysis

0024-9297/80/2213-0919$01.00/0