Conclusions

X-ray diffraction photographs of t-Boc-(L-X)$_n$OMe (X = Ala, Val, Leu, and n = 5, 6, 7) show the characteristic reflections of a cross-β structure. In addition, the good orientation of suitably prepared specimens has enabled a fairly complete determination of the unit cell of the pentapeptides to be made.

As regards chain orientation, an appreciable proportion of the Leu pentapeptide is in the antiparallel arrangement, while from the available data of the Ala and Val pentapeptides it has not been possible to establish whether the arrangement of the chains within the sheets is parallel or antiparallel.

References and Notes

(1) This work is part 50 of the series; for part 49 see C. Toniolo, *Macromolecules*, 11, 437 (1978)
(2) M. Palumbo, S. Da Rin, G. M. Bonora, and C. Toniolo, *Makromol. Chem.*, **177**, 1477 (1976).
(3) C. Toniolo and M. Palumbo, *Biopolymers*, **16**, 219 (1977).
(4) J. S. Balcerski, E. S. Pysh, G. M. Bonora, and C. Toniolo, *J. Am. Chem. Soc.*, **98**, 3470 (1976).
(5) M. M. Kelly, E. S. Pysh, G. M. Bonora, and C. Toniolo, *J. Am. Chem. Soc.*, **99**, 3264 (1977).
(6) The following abbreviations are used: t-Boc, *tert*-butyloxycarbonyl; OMe, methoxy; Ala, alanine; Leu, leucine; Val, valine.
(7) G. M. Bonora, D. Nisato, and C. Toniolo, *Makromol. Chem.*, **176**, 2535 (1975).
(8) G. M. Bonora, A. Maglione, and C. Toniolo, *Polymer*, **15**, 767 (1974).
(9) C. Toniolo, G. M. Bonora, and A. Fontana, *Int. J. Pept. Protein Res.*, **6**, 371 (1974).
(10) R. D. B. Fraser and T. P. MacRae in "Conformation in Fibrous Proteins and Related Synthetic Polypeptides", Academic Press, New York, N.Y., 1973.
(11) S. Arnott, S. D. Dover, and A. Elliott, *J. Mol. Biol.*, **30**, 201 (1967).
(12) J. O. Warwicker, *Acta Crystallogr.*, **7**, 565 (1954).
(13) R. E. Marsh, R. B. Corey, and L. Pauling, *Acta Crystallogr.*, **8**, 710 (1955).
(14) T. Komoto, K. Y. Kim, M. Oya, and T. Kawai, *Makromol. Chem.*, **175**, 283 (1974).
(15) E. M. Bradbury, L. Brown, A. R. Downie, A. Elliott, R. D. B. Fraser, W. B. Hanby, and T. R. R. McDonald, *J. Mol. Biol.*, **2**, 276 (1960).
(16) A. Elliott, R. D. B. Fraser, T. P. MacRae, I. W. Stepleton, and E. Suzuki, *J. Mol. Biol.*, **9**, 10 (1964).

Conformational Energy Calculations of the Effects of Sequence Variations on the Conformations of Two Tetrapeptides[1]

István Simon,[2] George Némethy, and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14853.

Received December 8, 1977

ABSTRACT: Conformational energy calculations were carried out on the two terminally blocked tetrapeptides *N*-acetyl-Thr-Asp-Gly-Lys-*N'*-methylamide and *N*-acetyl-Ala-Asp-Gly-Lys-*N'*-methylamide. The first peptide is a sequence variant of tetrapeptides studied earlier in this laboratory. The second peptide occurs in a bend at residues 94–97 in staphylococcal nuclease. A selection strategy is described which helps to accelerate the search of starting conformations used for energy minimization. The strategy involves exhaustive searches for conformations of fragments of the molecule which are stabilized by specific interactions and subsequent combination of fragments, prior to minimization. Several groups of low-energy conformations were found. They are compactly folded structures, but they differ from the "standard" chain reversals. One group, which is of low energy in both peptides, is stabilized by Asp···Asp and Asp···Lys backbone–side chain hydrogen bonds. Another group, of low energy in the Thr-containing peptides, is stabilized by a network of hydrogen bonds involving polar atoms of both backbone and side chains of the Thr, Asp, and Lys residues. The conformation corresponding to the sequence fragment in staphylococcal nuclease has relatively high energy, indicating that the bend observed in the protein is stabilized by interactions involving parts of the protein outside the tetrapeptide sequence.

I. Introduction

The balance between the roles of short-range and of medium- or long-range interactions in determining the existence and location of chain reversals is one of the important questions in the analysis of protein folding.[3,4] One can approach this problem by analyzing the tendency of oligopeptides to take up compactly folded conformations. This study is a continuation of earlier theoretical and experimental work from this laboratory[5] on specific tetrapeptide sequences.

In this paper, two tetrapeptides composed of L-amino acids are investigated. One of them, Ala-Asp-Gly-Lys, to be referred to as peptide A, appears as a bend conformation at residues 94–97 of staphylococcal nuclease.[6] The other, Thr-Asp-Gly-Lys, to be referred to as peptide T, is a permutation of four sequences (one of which appears as a bend at residues 35–38 of α-chymotrypsin) studied in earlier work[5] and should provide further information about the role of sequence variations when the amino acid composition is constant. In the two tetrapeptide sequences studied theoretically in this paper, only the first residue is different. Their comparison should, therefore, furnish some clues about the role of side-chain interactions in short-range conformational stability. The threonyl side chain in peptide T, for example, may participate in interactions which must be absent in peptide A, where alanine replaces threonine. An experimental study of these two tetrapeptides is in progress.[7]

Similar theoretical conformational studies on two biologically active tetrapeptides are reported in the following two papers.[8,9] The peptides are sequence variants of each other and show the large effect of variations in the sequence upon preferred conformations.

II. Computational Methods

Conformational Energy Calculations. The tetrapeptides were considered in the form of *N*-acetyl-*N'*-methylamides. The use of the CH_3CO and $NHCH_3$ blocking groups[10] eliminates charges at the terminal groups and in effect simulates an oligopeptide inside a protein sequence. The aspartyl and lysyl side chains were taken to be uncharged. The justification for this choice has been discussed elsewhere.[5,11]

0024-9297/78/2211-0797$01.00/0

Table I. Minimum-energy Conformations of Ac-Ala-Asp-Gly-Lys-NMe

Classification[a]		Short-hand notation for dihedral angles[b]													Energy, ΔE kcal/mol
		Ala	Asp				Gly	Lys							
		(ϕ,ψ)	(ϕ,ψ)	χ^1	χ^2	$\chi^{3,2}$	(ϕ,ψ)	(ϕ,ψ)	χ^1	χ^2	χ^3	χ^4	χ^5		
(i)	(a)	A	A	g^+	p	t	D*	F	t	t	t	t	g^+		0.0[c]
(i)L	(b)	A	A	g^-	p	t	D*	F	t	t	t	g^-	g^+		0.9
(ii)	(c)	A	A	g^-	n	t	C*	C	t	t	t	t	g^+		0.9
(i)	(d)	C	A	g^-	p	t	D*	C	t	t	t	t	g^+		1.4
(ii)		C	A	g^-	n	t	C*	C	t	t	t	t	g^+		1.7
(ii)L		A	A	g^+	n	t	C*	C	t	t	t	g^-	g^+		1.8
(i)		E	A	g^+	p	t	D*	C	t	t	t	t	g^+		1.9
(ii)		E	A	g^-	n	t	C*	C	t	t	t	t	g^+		2.1
(i)		D	A	g^-	p	t	D*	C	t	t	t	t	g^+		2.4
(i)		G	A	g^-	p	t	D*	C	t	t	t	t	g^+		2.5
(ii)		D	A	g^-	n	t	C*	C	t	t	t	t	g^+		2.6
(ii)		G	A	g^+	n	t	C*	C	t	t	t	t	g^+		3.0
X-ray structure[d]	(e)	D	A*	g^-	p	--[e]	D*	F	(see Table II)						>10^3
Minimized[f]	(f)	C	A*	g^-	p	t	C*	C	g^-	t	g^+	t	t		6.9
"Standard" bends															
Type II with H-bond[g]		E	C	g^+	g^+	c	D*	E	g^+	t	t	t	g^+		6.7
Type I'		E	A*	t	g^+	c	D*	E	g^+	t	t	t	g^+		7.1
γ		D	D	g^-	g^+	t	C*	E	g^+	t	t	t	g^+		8.0
Type II		E	C	g^-	p	t	A*	E	g^+	t	t	t	g^+		8.1
Type I		E	A	g^-	p	t	B	E	g^+	t	t	t	g^+		11.5

Footnotes to Table I

[a] The classification is explained in the text. The letters (a) to (f) in parentheses refer to the columns in Table II.

[b] Backbone and side-chain dihedral angles are described in terms of the short-hand notation[16] described in Section II. Symbols not explained there are as follows: p: χ^2_{Asp} near 115°; n: χ^2_{Asp} near -50°; c: $\chi^{3,2}_{Asp}$ near 0° (cis).

[c] E_0 = -19.1 kcal/mol.

[d] Ref 6.

[e] Not defined.

[f] This conformation was obtained by starting from the conformation of residues 94 to 97 in the X-ray crystallographic structure of staphylococcal nuclease, (e) in the Table, and minimizing the energy.

[g] See footnote 28 and step (2) of the selection strategy.

A description of the computational method, of the parameters describing geometry and interaction energies, and of minimization methods is given elsewhere.[12,13] The total energy, E, of any conformation was taken as the sum of the contributions from intramolecular nonbonded, electrostatic, hydrogen bonding, and torsional interactions. No solvent effects were included in the calculation. The effective dielectric constant in the electrostatic term was taken as two.[12,14] Energies are reported in this paper as $\Delta E = E - E_0$. The conformation with the lowest energy, E_0, is used as the reference state for each peptide.

Bond lengths and bond angles were maintained fixed. All dihedral angles were allowed to vary, except ω_0 (fixed at 180°), the dihedral angle of the peptide bond linking the N-terminal CH_3CO group to the first residue, and the two dihedral angles defining the orientation of the two terminal methyl groups (fixed at 60°).

Nomenclature. The nomenclature and conventions adopted by an IUPAC-IUB Commission were used.[15] A short-hand notation, introduced in a recent paper from this laboratory,[16] is used to denote regions around the low-energy minima on the (ϕ,ψ) map for blocked single residues. Such a notation is useful for approximate comparison of related conformations or in the description of selection strategies (section III), whenever the exact value of the dihedral angles is not important. For the convenience of the reader, a brief summary of the notation is given here: A = right-handed α-helical region, B = bridge region around $(\phi,\psi) = (-100°,0°)$, C = region around the seven-membered C_7^{eq} ring conformation, D = region around (−150°,70°), E = extended region, F = region around (−80°,160°), and G = region around (−160°,−60°). The same letters with asterisks denote the corresponding regions in the right-hand half of the conformational map, with $\phi > 0°$, i.e., A* = left-handed α-helical region, etc. In side chains, ranges of the χ's near the staggered conformations are denoted by t, g^+, and g^-, corresponding to values within ±30° of χ = 180°, 60°, and −60°, respectively.[17] A few other short-hand symbols, used occasionally in this paper, are explained in the footnotes to Table I.

III. Selection Strategy for Starting Conformations

Already in tetrapeptides, it is not feasible to carry out a complete search of conformational space, as was done in searching for the energy minima of single amino acid residues[16] or of dipeptides.[18–21] Were one to start with all combinations of the low-energy potential energy minima reported for the constituent amino acids,[16] the theoretical total number of conformations to be tested for a tetrapeptide would be of the order of 10^6. A selection strategy has to be devised which allows the progressive but systematic exclusion of starting conformations which are not likely to yield low-energy conformations upon minimization. In fact, designing the selection strategy constitutes the main conceptual task in conformational energy computations of the type described here. By contrast, the subsequent computation of the potential energies and the energy minimization require large computing times, but they use well-established methods and computer programs.[12,13] The best method of selection may differ from one peptide to the other. Therefore, we describe in detail the selection strategy used in this work. It differs somewhat from the ones used previously in this laboratory.[5,22] In a study of gramicidin S, the initial selections were restricted by symmetry, by the necessity for ring closure, and by the use of available NMR data.[22] No such restrictions existed in this study. Howard et al.[5] selected some of the optimal conformations for the single amino acids[23] and then minimized the energy of the entire tetrapeptide. For the sequences studied here, it was more efficient to investigate di- and tripeptide fragments at first and, thereby, to restrict the number of starting conformations of the tetrapeptide.

The present selection strategy consisted of three steps, to be described below in detail: (1) Favorable low-energy conformations of terminally blocked dipeptides and of a terminally blocked tripeptide were found, and starting conformations of the tetrapeptides were "built up" by combining such conformations with each other and with low-energy conformations of single residues. (2) In an independent selection of starting points for the tetrapeptides, several chain-reversal conformations were chosen. Some of the side-chain dihedral angles were kept fixed in steps (1) and (2). (3) In a final step, those side-chain dihedral angles, which had been kept constant in the earlier steps, were allowed to vary. The conformational space defined by these dihedral angles was searched completely (for the backbone conformations found in the earlier steps) in order to find all low-energy side-chain conformations. Finally, the energy was minimized, allowing variation of all dihedral angles.

Selection strategies of a similar type were adopted in the following two papers.[8,9] However, details of the strategies differ in each case, because of differences in conformational properties of the constituent amino acids and dipeptides, and in order to take into account specific features of each molecule.

(1) Reduction of the Number of Starting Conformations by a Buildup from Shorter Sequences. In this step of the strategy, terminally blocked dipeptide and tripeptide sequences were examined, in the hope that certain conformations could be singled out because they might be much more stable than the others, due to particularly strong interresidue interactions, usually side chain–side chain or side chain–backbone hydrogen bonds.[24] Such stabilization is indicated by a large lowering of the energy of the dipeptide or tripeptide from the value expected for a structure in which the energy is essentially the sum of the energies of the component residues in the proper backbone conformations.[25] If such conformations can be found, they serve as preferential starting points for energy minimization on longer sequences. This leads then to a reduction of the number of starting conformations.

The use of this argument will be illustrated below for terminally blocked Thr-Asp.

It was assumed, as before,[5,18–21] that the starting conformations for energy minimization can be chosen among the combinations of the local energy minima for terminally blocked amino acids.[16] A recently computed refined set of minima for the individual amino acids[16] was used, instead of the more selective set[23] used earlier.[5] Eight local minima had been found[16] for alanine, six of them of low relative energy (ΔE < 5 kcal/mol), and seven for glycine. All of them were included as possible starting points for Ala and Gly, respectively. However, for the other amino acids, very many local minima occur with closely spaced energies, because of the occurrence of several side-chain positions for each backbone conformation. We considered only conformations with ΔE < 3 kcal/mol for each blocked amino acid.[26] There are 32 such conformations for aspartic acid, 174 for lysine, and 15 for threonine.[16] Except as noted otherwise, these numbers were reduced further by selecting one or more representative side-chain conformations in each of the seven (ϕ,ψ) regions[16] (viz., C, E, D, A, F, G, A*) in which low-energy minima are found.

Dipeptides. The strategy just outlined did not give useful results for three of the four dipeptides. For terminally blocked Gly-Lys, combination of the seven minima for glycine with the 56 lowest energy conformations of lysine, and minimization of the energy of the 392 resultant dipeptides, gave a continuous sequence of conformations with energies increasing in small steps. No dipeptide conformations could be singled out as being particularly favorable, as compared with the conformation of blocked single residues. The same conclusion was reached from the results of another study in this laboratory[19,20] for the 162 conformations of blocked Asp-Gly and the 130 conformations of blocked Ala-Asp with ΔE < 3 kcal/mol. Therefore, no reduction of the number of starting conformations could be achieved by considering the dipeptides Gly-Lys, Asp-Gly, and Ala-Asp by themselves.

However, the technique was useful for terminally blocked Thr-Asp. The energy was minimized for 220 dipeptide conformations, chosen in the following way. Out of the combinations of the 15 threonine and 32 aspartic acid minima[16] with ΔE_{Thr} < 3 kcal/mol and ΔE_{Asp} < 3 kcal/mol, only those were selected for which $\Delta E_{Thr} + \Delta E_{Asp}$ < 3 kcal/mol as well.[25] Upon minimization, four dipeptide conformations were found whose energies are much lower than those of all others, because of a hydrogen bond between the threonyl hydroxyl group and the $O^{\delta 1}$ of the aspartyl side chain. The threonyl residue has the same conformation in all of them, A g^+g^+, with dihedral angles varying less than 10°. This occurs in combination with four different aspartic acid conformations, all of which have similar values of ϕ (Figure 1A). The relative energies of the four conformations are ΔE = 0.0, 1.4, 1.5, and 1.7 kcal/mol. All other conformations have ΔE > 2.9 kcal/mol. There are 15 conformations in the energy range 2.9 < ΔE < 4.0 kcal/mol. Fourteen of them combine the most stable threonyl conformation, C g^+g^+, with various aspartyl conformations, and one has threonine in the conformation A g^+g^+. Many conformations occur with ΔE > 4.0 kcal/mol. The four conformations shown in Figure 1A, of much lower energy than the others, were used preferentially as starting conformations in the tetrapeptide studies. The preference of threonine for the C and A conformations was observed in the earlier study of Gly-Thr-Asp-Lys as well.[5] One of the low-energy conformations found[5] for that tetrapeptide contains the Thr-Asp side-chain hydrogen bond described here.

Tripeptide. The sequence Asp-Gly-Lys, common to the two tetrapeptides, was examined next, using 140 starting conformations. These were chosen as combinations of representative conformations from the low-energy (ϕ,ψ) regions of each of three amino acids, obeying the condition $\Delta E_{Asp} + \Delta E_{Gly} + \Delta E_{Lys}$ < 5 kcal/mol. In a first cycle of minimization, only ϕ and ψ of the three residues were varied.[27] The aspartyl side chain was fixed in some of the most favorable orientations for each of the starting (ϕ,ψ) values. The lysyl side chain was fully extended. After minimization, four different low-energy conformations were found, all with ΔE < 1.5 kcal/mol. All other conformations had ΔE > 3.7 kcal/mol and were rejected. A second cycle of minimization was carried out on the four favorable conformations. All dihedral angles (including the ω's and χ's) were allowed to vary. The energies stayed within a 1.4-kcal/mol interval. The final dihedral angles of the glycyl residue occupy the same low-energy region of the (ϕ,ψ) map in all four conformations. The final dihedral angles for aspartic acid and for lysine occur in two distinct regions each. The four conformations represent all combinations of these regions (Figure 1B). As a further test, minimization was carried out for six more conformations. These were combinations of some low-energy Asp-Gly and Gly-Lys conformations which contained the same conformation of glycine. All six conformations reached ΔE > 3.5 kcal/mol after minimization. They were not retained for further use.

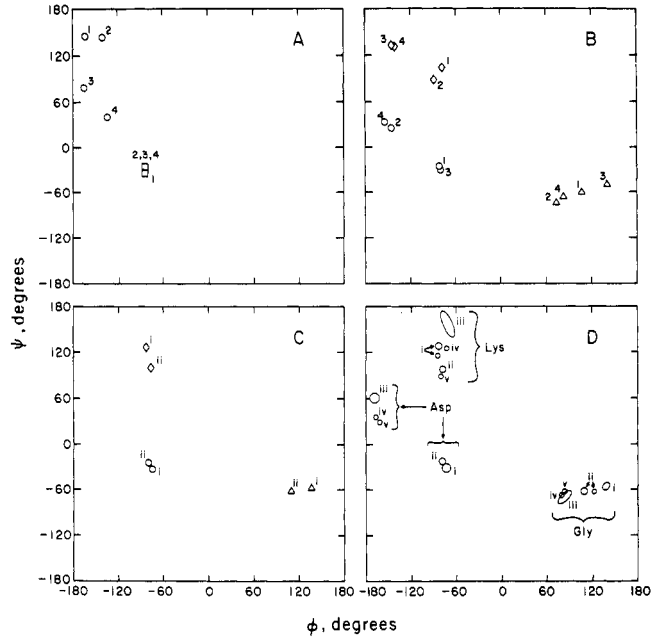

Figure 1. (A) Location of the four lowest energy conformations of blocked Thr-Asp on the (ϕ,ψ) map: □ = threonine, ○ = aspartic acid. Numbers refer to the four different conformations, with aspartic acid position and total energy as follows: (1) E g^+g^-t, ΔE = 0.0 kcal/mol; (2) E g^-g^+t, ΔE = 1.4 kcal/mol; (3) D g^+g^-c, ΔE = 1.5 kcal/mol; (4) D g^+g^+t, ΔE = 1.7 kcal/mol. (c denotes a dihedral angle which is near cis for $\chi_{Asp}^{3,2}$). Threonine occurs as A g^+g^+ in all cases. In the subsequent computation, conformation 3 led to the conformations of groups (iii) and (iv) of peptide T. (B) Location of the four lowest energy conformations of blocked Asp-Gly-Lys on the (ϕ,ψ) map: ○ = aspartic acid, △ = glycine, ◇ = lysine. The lysyl side chain is fully stretched, and the aspartyl side chain has the conformation g^+g^-t. Numbers refer to the four conformations, with the following total energies: (1) ΔE = 0.0 kcal/mol; (2) ΔE = 0.4 kcal/mol; (3) ΔE = 0.8 kcal/mol; (4) ΔE = 1.5 kcal/mol. In the subsequent computation, conformation 1 led to the conformations of groups (i) and (ii) respectively of both peptides A and T. (C) Location of the backbone dihedral angles of the aspartyl (○), glycyl (△), and lysyl (◇) residues in the low-energy conformations of blocked Ala-Asp-Gly-Lys (peptide A), in groups (i) and (ii) of Table I, as marked in the figure. The dihedral angles for different conformations within each group differ from each other by amounts less than the areas covered by the symbols in the figure. (D) The locations of the backbone dihedral angles of the aspartyl, glycyl, and lysyl residues in the low-energy conformations [groups (i) to (v) of Table III] of blocked Thr-Asp-Gly-Lys. The areas marked by roman numerals enclose all conformations falling into the various groups for each of the amino acid residues indicated.

Peptide A. In a first cycle of minimization, in which only ϕ and ψ were varied, 24 starting conformations were tested, viz., all combinations of the four low-energy Asp-Gly-Lys conformations (Figure 1B) with the six low-energy minima[16] for alanine. Low-energy minima were found for only one of the tripeptide conformations, viz., AC*C (marked 1 in Figure 1B), combined with the six alanine conformations. All six minima of peptide A occurred in an energy range of $0 \leq \Delta E < 3$ kcal/mol. Of the other starting conformations, most of those having aspartic acid in the D region (marked 2 and 4 in Figure 1B) converged upon minimization to the minima just mentioned. Thus, it appears that the molecule is sufficiently flexible to allow movements of the aspartyl and lysyl residues from one low-energy region of conformational space into another during minimization. All other conformations minimized to $\Delta E > 7$ kcal/mol and, therefore, could be discarded. Changes in side-chain orientations did not give new conformations with low energies, except as discussed below in the Results section. A second cycle of minimization, during which all variable dihedral angles were allowed to adjust, lowered the energies somewhat more.

Peptide T. Two sets of starting conformations were tested using the same two cycles of minimization as for peptide A. One set was obtained by combining the four low-energy Asp-Gly-Lys conformations with the 15 low-energy minima of blocked threonine.[16] Just as for peptide A, only one of the tripeptide conformations, AC*C, minimized to low energies when combined with several threonine conformations. Another set of starting conformations was obtained by combining the four low-energy Thr-Asp conformations (Figure 1A) with various low-energy Gly-Lys conformations. The latter were chosen because they gave low-energy Asp-Gly-Lys tripeptides upon combination with the Asp conformations in the four Thr-Asp conformations. The conformation of lowest energy was obtained from this set by using step (3). A few other combinations in this set gave conformations with $\Delta E < 5$ kcal/mol. Several others with $8 < \Delta E < 11$ kcal/mol were of high energy and were, therefore, not considered further.

(2) The Use of Bends as Starting Conformations. A number of starting conformations corresponding to "standard" bend (chain reversal) structures[24] were selected. The Asp and Gly dihedral angles were chosen in two ways. First, the assigned approximate values[24] for bends of types I, I', and II were taken. Second, these bends were modified to the form in which a nearly straight Lys···Ala N–H···O=C hydrogen bond occurs.[28] The initial dihedral angles of alanine and lysine were chosen to bring the two ends of the molecule close together in order to favor their interactions. In addition, the γ turn[32] was tested; in this conformation, the glycyl residue forms a C_7^{ax} ring and there is an additional Asp-NH···O=C-Lys hydrogen bond.

The dihedral angles reported in the x-ray crystallographic analysis of staphylococcal nuclease[6] for residues 94–97 were also used as a starting conformation.

The energies of all "standard" bend structures of peptide A stayed relatively high after minimization ($\Delta E > 6.7$ kcal/mol). Model studies indicated that the substitution of threonine for alanine would not result in strong new interactions so that these bends very likely are high-energy conformations in peptide T as well.

(3) Variation of the Lysyl Side-Chain Conformation. During the minimizations in steps (1) and (2), only the fully stretched conformation of the lysyl side chain ($\chi^i = 180°$, i = 1 to 5) was used as a starting conformation. Even though the χ^i values could change by a few degrees during minimization, they could not move out of the local minimum designated as t. In step (3) of the strategy, side-chain conformations were explored in which one or more dihedral angles occur in regions g^+ or g^-. Although rotation around the lysyl C–C bonds from a t to a g^+ or g^- position increases the torsional and nonbonded energies, favorable interactions involving the terminal NH_2 group *may* counteract this energy change. Because of the large number of possible lysyl conformations, viz., 243, testing was done in two cycles, starting from the low-energy conformations obtained in step (1). First the side-chain dihedral angles of lysine were varied systematically to cover all staggered positions ($\chi^i = \pm 60°$ and 180°, for i = 1, 2, 3, 4), while all other dihedral angles were held fixed. Each side-chain orientation was tested for potential hydrogen bonds between the H (or N) atoms of the terminal $N^\zeta H_2$ group and hydrogen bond acceptors (or donors) elsewhere in the molecule by computing interatomic distances.[33] Whenever a distance below the cutoff limit[34] was found, the second cycle of testing was applied, i.e., the energy of the molecule was minimized, allowing variation of *all* dihedral angles. Usually, the presence of a hydrogen bond was confirmed by energetic criteria. Occasionally, the energy of the resultant conformation was lower than that of the parent conformation with the stretched lysyl side chain. For example, this was the case for the global minimum of peptide T. However, in most cases, the final energy was higher in spite of the newly formed hydrogen bond.

Step 1 of this selection strategy, in the search for preferences based on the conformational behavior of fragments of the oligopeptide, is applicable, in the form presented here, primarily to sequences with many polar side chains, because such peptides are likely to have many backbone–side chain interactions involving hydrogen bonds.[16,18–21,23,26] It is less useful for oligopeptides with few or no polar side chains, such as enkephalin.[37] However, different approaches, using this step of the selection strategy, can apply to oligopeptides with fewer side-chain interactions.[8,9] Step 3, the systematic variation of side-chain dihedral angles, is an important part of various selection strategies. Its use resulted in many significant results in the study of enkephalin,[37] and it was useful in another tetrapeptide.[9]

IV. Results

Peptide A. The results are summarized in Table I, using the short-hand notation.[16] Conformations are listed in order of increasing energy. The dihedral angles for the conformations of lowest energy and for the one derived from the x-ray structure are listed in Table II.

All conformations with $\Delta E < 3.1$ kcal/mol are closely related (Figure 1C). As seen in Table I, they contain essentially a single conformation[38] of the Asp-Gly-Lys tripeptide, viz., AD*C, in combination with all low-energy conformations of alanine. They are marked (i) in Table I. The compactly folded tripeptide structure is stabilized by a network of hydrogen bonds involving the aspartyl carboxyl group: Asp-NH···$O^{\delta 2}$CO-Asp with an H···O distance $d_{H\cdots O}$ = 2.2 Å, Asp-$CO^{\delta 2}$H···O=C'-Lys with $d_{H\cdots O}$ = 1.7 Å, and Lys-NH···$O^{\delta 1}$=C-Asp with $d_{H\cdots O}$ = 2.2 Å. The conformation with the lowest energy [(a) in Tables I and II] is shown in Figure 2. It does not satisfy the criteria for a bend structure, because the $C^{\alpha}_i \cdots C^{\alpha}_{i+3}$ (Ala···Lys) distance of 7.5 Å exceeds the 7 Å limit defined for bends.[24] However, compactness of the folding is indicated by the small $C^{\alpha}_{i-1} \cdots C^{\alpha}_{i+4}$ distance, 4.1 Å, between the α carbons of the two terminal groups. The other conformations in group (i) are also compact, and they are stabilized by the same hydrogen-bonded network. However, both $C^{\alpha} \cdots C^{\alpha}$ distances just mentioned usually are somewhat larger than in (a) because of differences in the orientation of the alanyl residue.

For each conformation in group (i), there exists a variant with slightly higher energy, marked (ii) in Table I. Group (ii) differs from group (i) essentially by a shift of the carboxyl hydrogen of Asp from one oxygen to the other, or equivalently, by a 180° rotation of the carboxyl group[39] around the C^{β}–C^{γ}

Table II. Dihedral Angles and Energies for Some Low Energy Conformations of Ac-Ala-Asp-Gly-Lys-NMe

Dihedral Angle	Conformation[a] (a)	(b)	(c)	(c)	(e) X-ray[b]	(f) X-ray minimized[c]
ϕ_i	-70	-70	-73	-83	-127	-81
ψ_i	-38	-39	-40	77	106	83
ω_i	179	179	180	179	179	179
ϕ_{i+1}	-73	-73	-78	-78	65	54
ψ_{i+1}	-33	-32	-24	-31	33	47
ω_{i+1}	-173	-174	-179	-174	180	178
χ^1_{i+1}	62	62	63	66	-62	-53
χ^2_{i+1}	124	125	-50	112	129	96
$\chi^{3,2}_{i+1}$	174	174	-179	176	-	-180
ϕ_{i+2}	139	137	109	134	119	108
ψ_{i+2}	-57	-57	-61	-59	-29	-39
ω_{i+2}	175	174	178	177	-165	-174
ϕ_{i+3}	-83	-81	-76	-83	-97	-79
ψ_{i+3}	131	130	100	128	138	93
χ^1_{i+3}	-174	-173	-174	-174	+68	-66
χ^2_{i+3}	176	177	178	177	144	175
χ^3_{i+3}	179	-176	179	179	9	76
χ^4_{i+3}	179	-69	179	179	-135	175
χ^5_{i+3}	62	75	62	62	-	177
Energy, kcal/mol	0.0	0.9	0.9	1.4	$> 10^3$	6.9
C^α distances, Å						
$C^\alpha_i \cdots C^\alpha_{i+3}$	7.5	7.5	8.1	7.6	-	4.8
$C^\alpha_{i-1} \cdots C^\alpha_{i+4}$	4.1	4.1	5.3	3.9	-	4.1

[a] The letters in parentheses refer to the classification in Table I. Dihedral angles are in degrees. χ for Ala was always within 1° of 60°. [b] Ref. 6. [c] See footnote f of Table I.

Table III. Minimum-energy Conformations of Ac-Thr-Asp-Gly-Lys-NMe

Classification[a]	Short-hand notation for dihedral angles[b] Thr (ϕ,ψ)	Thr χ^1	Thr $\chi^{2,1}$	Asp (ϕ,ψ)	Asp χ^1	Asp χ^2	Asp $\chi^{3,2}$	Gly (ϕ,ψ)	Lys (ϕ,ψ)	Lys χ^1	Lys χ^2	Lys χ^3	Lys χ^4	Lys χ^5	Energy, ΔE kcal/mol
(iii) (a)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	g^-	g^+	0.0[c]
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	g^+	g^+	1.5
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	g^+	t	1.6
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	g^-	g^+	1.7
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	t	t	1.8
(v) (e)	A	g^+	g^+	D	g^+	p	t	C*	C	t	t	t	g^-	g^+	1.9
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^+	t	t	t	g^-	2.0
(iii)	A	g^+	g^+	D	g^-	g^-	c	C*	F	g^+	t	t	g^+	g^+	2.0
(iv)	A	g^+	p	D	g^+	p	t	C*	C	t	t	t	t	g^+	2.3
(iii)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^-	t	t	g^-	t	2.5
(iii) (b)	A	g^+	g^+	D	g^+	g^-	c	C*	F	t	t	t	t	g^-	2.7
(iii)	A	g^+	g^+	D	g^-	g^-	c	C*	F	g^+	t	t	g^+	g^-	2.7
(i) (c)	A	g^+	t	A	g^+	p	t	D*	F	t	t	t	g^+	g^+	2.9
(i)	A	g^+	t	A	g^+	p	t	D*	F	t	t	t	t	g^+	3.0
(i)	A	g^+	g^+	A	g^+	p	t	D*	C	t	t	t	t	g^+	3.2
(i)	E	t	t	A	g^+	p	t	D*	C	t	t	t	t	g^+	3.2
(i)	C	g^+	g^+	A	g^-	p	t	D*	C	t	t	t	t	g^+	3.3
(v)	A	g^+	g^+	D	g^+	p	t	C*	C	g^-	t	t	g^+	g^+	3.3
(i)	E	t	g^+	A	g^+	p	t	D*	C	t	t	t	t	g^+	3.5
(ii) (d)	C	g^-	g^-	A	g^+	n	t	C*	C	t	t	t	t	g^+	3.5
(ii)	A	g^+	t	A	g^+	n	t	D*	C	t	t	t	t	g^+	3.6
(i)	A	g^-	g^+	A	g^+	p	t	D*	F	t	t	t	t	g^+	3.7
(ii)	E	t	t	A	g^+	n	t	C*	C	t	t	t	t	g^+	3.7
(ii)	E	t	g^+	A	g^+	n	t	C*	C	t	t	t	t	g^+	3.8
(vi)	A	g^+	g^+	D	g^+	g^-	c	C*	F	g^-	g^-	t	g^-	t	3.8
(ii)	A	g^+	g^+	A	g^+	n	t	C*	C	t	t	t	t	g^+	3.9
(ii)	A	g^-	t	A	g^+	n	t	C*	C	t	t	t	t	g^+	3.9
(vi)	A	g^+	g^+	D	g^+	g^-	c	C*	E	t	t	t	t	g^+	4.3
(vi) (f)	A	g^+	g^+	E	g^+	g^-	t	C	G	t	t	t	t	g^+	5.0

[a] The classification is explained in the text. The letters (a) to (f) in parentheses refer to columns in Table IV. [b] See footnote b of Table I. [c] E_0 = -25.0 kcal/mol.

Table IV. Dihedral angles and Energies for Some Low-Energy Conformations of Ac-Thr-Asp-Gly-Lys-NMe

Dihedral Angle	Conformation[a] (a)	(b)	(c)	(d)	(e)	(f)
ϕ_i	-87	-82	-76	-88	-76	-80
ψ_i	-32	-32	-39	83	-35	-35
ω_i	-173	-171	176	180	-172	174
χ^1_i	52	50	44	57	47	47
χ^2_i	90	90	165	65	79	87
ϕ_{i+1}	-166	-166	-74	-81	-153	-163
ψ_{i+1}	59	55	-32	-24	30	133
ω_{i+1}	-168	-168	-175	-179	170	179
χ^1_{i+1}	51	52	63	64	59	50
χ^2_{i+1}	-85	-86	112	-47	116	-69
$\chi^{3,2}_{i+1}$	-25	-24	174	179	168	175
ϕ_{i+2}	83	76	137	109	82	-83
ψ_{i+2}	-71	-69	-54	-63	-64	70
ω_{i+2}	171	169	174	179	171	-177
ϕ_{i+3}	-81	-67	-85	-77	-80	-159
ψ_{i+3}	168	142	130	99	91	-57
χ^1_{i+3}	76	-173	-173	-174	-173	-178
χ^2_{i+3}	-176	173	176	177	178	176
χ^3_{i+3}	-177	179	179	179	179	179
χ^4_{i+3}	-73	179	179	179	179	180
χ^5_{i+3}	72	62	62	62	62	63
Energy, kcal/mol	0.0	2.7	2.9	3.5	1.9	5.0
C^α distances, Å						
$C^\alpha_i \cdots C^\alpha_{i+3}$	6.2	6.3	7.4	8.1	6.9	8.9
$C^\alpha_{i-1} \cdots C^\alpha_{i+4}$	4.2	4.1	4.1	9.7	4.7	9.2

[a] The letters in parentheses refer to the classification in Table III. Dihedral angles are in degrees.

bond, and by an accompanying decrease[38] of about 30° in ϕ_{Gly}. The latter change helps to maintain the Asp-CO$^{\delta 2}$H···O=C′-Lys hydrogen bond.

The fully stretched lysyl side chain points away from the molecule (Figure 2). Its displacement into other stable (staggered) positions generally increases the energy, as is seen for the two conformations marked (i)L and (ii)L in Table I. The same lysyl side-chain orientation can occur for all other conformations of groups (i) and (ii) in Table I but with a computed energy $\Delta E > 3$ kcal/mol.

The various "standard" chain reversal (bend) conformations all have energies $\Delta E > 6.7$ kcal/mol. They are not favored for this tetrapeptide because such conformations lack the specific side-chain interactions described above.

The conformation obtained by energy minimization starting with the dihedral angles reported for residues 94 to 97 in the crystal structure of staphylococcal nuclease[6] has an energy $\Delta E = 6.9$ kcal/mol. Neither the starting nor the energy-minimized conformation are similar to the low-energy conformations in groups (i) and (ii). The backbone dihedral angles of the Asp-Gly fragment of the x-ray structure differ by 10 to 40° from the values defined[24] for a type I′ chain reversal. However, the x-ray conformation is not a local minimum of the free tetrapeptide, as seen from the large changes in ϕ and ψ, especially of Ala and Lys, upon minimization. The lysyl side chain is forced into a very unfavorable position in the protein; the reported value for χ^3 is 9°, which is near an eclipsed high-energy position; χ^2 and χ^4, too, differ by about 35 and 50°, respectively, from a staggered value. Minimization in the free peptide A brings these dihedral angles much nearer the staggered positions. Apparently, the conformation of residues 94 to 97 in staphylococcal nuclease is stabilized by medium-range interactions with neighboring residues in the sequence, by long-range interactions in the protein, by solvent effects, or by any combination of these, since these four residues are near the surface of the molecule.[6] The present calculations show merely that the observed structure is not stabilized sufficiently by interactions occurring *within* the tetrapeptide. A similar conclusion was reached by Hurwitz and Hopfinger for the bend occurring at residues 35–38 of chymotrypsin.[36]

Peptide T. The results are summarized in Table III, using the short-hand notation. Conformations are listed in order of increasing energy. The low-energy conformations of Table III can be classed into several related groups which will be discussed separately (Figure 1D). Except for group (vi), only

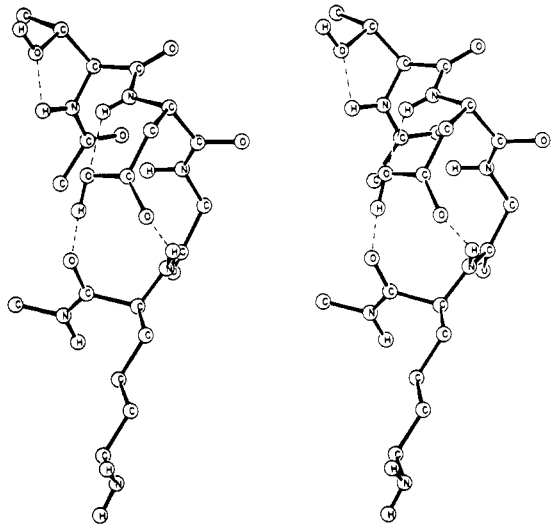

Figure 2. Stereoscopic illustration of the lowest energy conformation, (a), of peptide A and conformation (c) of peptide T. The positions of backbone and side-chain atoms are identical (within 0.1 Å) in both peptides, except for the C$^{\gamma 2}$, O$^{\gamma}$, and H$^{\gamma}$ atoms of threonine in peptide T which do not occur in peptide A. Hydrogen bonds are indicated by dashed lines. Backbone and side-chain aliphatic hydrogen atoms have been omitted for clarity.

conformations with $\Delta E < 4$ kcal/mol are listed in Table III. The dihedral angles for the important distinct conformations in various groups are given in Table IV.

Groups (i) and (ii) are combinations of a single conformation[38] of the Asp-Gly-Lys tripeptide, AD*C (or AD*F) or AC*C, respectively, and of various low-energy conformations of threonine. The threonyl side chain points away from the rest of the molecule in most cases and it does not participate in strong interactions. The conformations of peptide T in group (i) are almost identical with the corresponding conformations of peptide A. Group (ii) arises from group (i) by the 180° rotation of the aspartic acid carboxyl group,[38,39] as described for group (ii) of peptide A. The lowest energy conformation of peptide T in group (i) is identical (except for the added atoms of the threonyl side chain) with the lowest energy conformation for peptide A [it is marked (c) in Tables III and IV and it is shown in Figure 2]. Of the many conformations in groups (i) and (ii), with energies in the range $2.9 \leq \Delta E \leq 6.0$

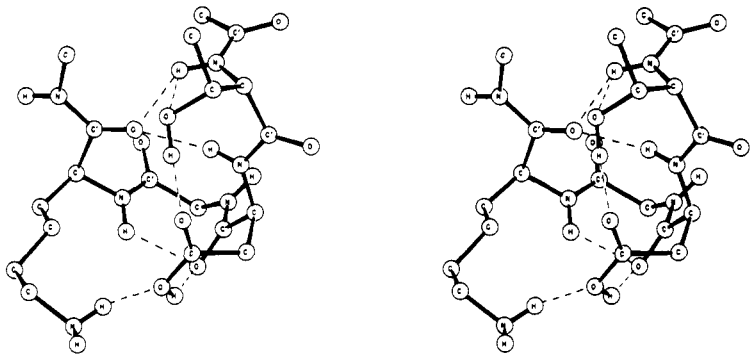

Figure 3. Stereoscopic illustration of the lowest energy conformation, (a), of peptide T. Except for the lysyl side chain, the figure also very nearly represents conformation (b) of peptide T. Hydrogen bonds are indicated by dashed lines. Backbone and side-chain aliphatic hydrogen atoms have been omitted for clarity.

kcal/mol, only the most significant ones, with $\Delta E < 4$ kcal/mol above the global minimum, are listed in Table III. Placing the lysyl side chain into rotamer positions other than the fully extended one in these groups (step 3 of the selection strategy) always raises the energy to $\Delta E > 4$ kcal/mol; thus, no new conformations which make significant contributions are generated in this manner.

The lowest energy conformation obtained by combining the low-energy Thr-Asp dipeptides with low-energy Gly-Lys dipeptides (step 1 of the strategy) is the parent conformation of group (iii) in Table III. The subsequent application of step (3) of the selection strategy to this conformation gave several conformations of even lower energy, as discussed below. In the parent conformation, with a fully stretched lysine side chain, the energy is $\Delta E = 2.7$ kcal/mol [(b) of Tables III and IV]. This is a very compact structure, stabilized by a network of hydrogen bonds (Figure 3). The most important one is the Thr-$O^{\gamma}H\cdots O^{\delta 1}$=C-Asp bond with $d_{H\cdots O} = 1.7$ Å. This is the bond which stabilizes the Thr-Asp dipeptides of Figure 1A. The presence of this bond allows the formation of the Asp-$CO^{\delta 2}H\cdots O$=C′-Asp bond with $d_{H\cdots O} = 1.7$ Å and the weaker intra-threonyl $NH\cdots O^{\gamma}$ bond with $d_{H\cdots O} = 2.1$ Å. A weak backbone Lys-NH···O=C′-Asp hydrogen bond with $d_{H\cdots O} = 2.2$ Å occurs because glycine is in the seven-membered C_7^{ax} ring conformation. The tetrapeptide forms a bend according to the distance criterion:[24] the $C^{\alpha}_i\cdots C^{\alpha}_{i+3}$ (Thr···Lys) distance is 6.3 Å. Compactness is indicated also by the terminal group $C^{\alpha}_{i-1}\cdots C^{\alpha}_{i+4}$ distance of 4.1 Å.

Further stabilization of this conformation is possible in some cases when the lysyl side chain is bent so that it interacts with the rest of the peptide. This may happen, for example, when the ϵ-NH_2 group comes within hydrogen-bonding distance of the aspartic acid carboxyl group. Nine such conformations were found. They are shown in group (iii) in Table III. The energy of eight of them is below that of the parent conformation (b), and four have the Lys···Asp hydrogen bond. One of them is the global minimum of peptide T found in this study. It is marked (a) in Tables III and IV, and its structure is shown in Figure 3. The Lys-$N^{\delta}H\cdots O^{\delta 2}$-C-Asp hydrogen bond has $d_{H\cdots O} = 1.9$ Å. The other hydrogen bond lengths and the $C^{\alpha}\cdots C^{\alpha}$ distances are identical within ±0.1 Å with those of conformation (b), discussed above. Most dihedral angles, too, are the same within a few degrees, except for ϕ and ψ of Lys (Table IV). In some, though not all, of the conformations in group (iii), the carbonyl oxygen of lysine moves closer to the amide hydrogens of threonine and aspartic acid, resulting in weak hydrogen-bonding interactions. In conformation (a), $d_{H\cdots O} = 2.2$ Å for Thr-NH···O=C′-Lys and $d_{H\cdots O} = 2.1$ Å for the Asp-NH···O=C-Lys interaction. Apparently, the backbone conformation in (b) is so stable that it is perturbed very little by the added lysyl side-chain hydrogen bond.

A 180° rotation of the carboxyl group of aspartic acid in the conformations of group (iii) gives rise to group (iv). All conformations of this class have $\Delta E > 3.9$ kcal/mol, except for one with $\Delta E = 2.3$ kcal/mol. The latter conformation is derived from the parent conformation (b) of group (ii). It is listed in Table III.

Starting with other combinations of low-energy Thr-Asp and Gly-Lys dipeptides, several more minima are reached for peptide T. They are placed in groups (v) or (vi), depending on the orientation of the carboxyl group of the aspartic acid. Only one of them [(e) in Tables III and IV] has low energy, $\Delta E = 1.9$ kcal/mol. All those with $\Delta E < 6$ kcal/mol are listed in Table III. The first four conformations listed in these groups differ from those in groups (iii) and (iv) mainly by a change in the orientation of the entire lysyl residue. The hydrogen-bonded network described for conformation (b) is not changed much in them. The last conformation listed in group (vi) [(f) of Tables III and IV] differs very much from the others. The strong Thr···Asp side-chain interaction is maintained. In the rest of the molecule, a seven-membered ring with a Lys-NH···O=C′-Asp hydrogen bond, due to the C_7^{eq} conformation of glycine, and a weak hydrogen bond between the C-terminal NH group and the $O^{\delta 2}$-C-Asp contribute to the stability.

The energies of the various "standard" bend conformations are relatively high, because the threonyl side chain cannot participate in strong interactions.

V. Discussion

The results indicate that these tetrapeptides are flexible and may exist as an equilibrium mixture of many conformations. However, the calculations suggest that a few groups of related conformations are energetically favored in both tetrapeptides studied here. The reason for this is that some part of the molecule can be held in a rigid position by a network of strong hydrogen bonds. Flexibility of the remaining part of the molecule, e.g., of the first residue of both tetrapeptides in group (i), gives rise to a related group of conformations with energies occurring within a small interval. In peptide A, the group of stable conformations, listed in Table I and having $\Delta E \leq 3$ kcal/mol, is separated by an energy gap of 2 kcal/mol from a large group of medium-energy conformations. The latter group includes the "standard" bend conformations as well as several others, not shown in Table I, which were eliminated by the selection strategy. In peptide T, no large energy gaps occur; the distribution of conformations over a wide energy range is more or less continuous. However, only a few groups of conformations were found with $\Delta E < 5$ kcal/mol.

Solvent effects, omitted in the present calculations, may alter the relative stability of various conformations. In aqueous solution, two opposing factors may operate. On the one hand, the net strength of intramolecular hydrogen bonds is weak-

ened. This would reduce the stability of the favored conformations found here. On the other hand, hydrophobic interactions between the terminal methyl groups would be strengthened in water. Presumably, the first effect is more significant. Thus, the compact conformations considered here may make a sizable contribution in aqueous solution as well, although they are expected to be less favored than in a nonpolar solvent. Ionization of the carboxyl group of Asp would disrupt part of the hydrogen-bond networks shown in Figures 2 and 3, destabilizing some conformations to some extent. However, at neutral pH in aqueous solution, both the lysyl ϵ-amino group and the aspartyl carboxyl group are ionized. The favorable $-NH_3^+\cdots{}^-OOC-$ interaction would still stabilize conformations in which these two groups are close to each other, i.e., most of the group (iii) low-energy conformations of Table III, including the conformation of Figure 3.

The main interactions mentioned in the Results section depend strongly on the amino acid sequence. The very strong Thr···Asp side-chain interaction, stabilizing many conformations of peptide T, occurs because these two residues are neighbors in the sequence. The hydrogen bond closes a ten-membered ring. The two side chains occasionally can interact in other peptides as well, in which Thr and Asp are not nearest neighbors.[5] However, their effect is less pronounced because the hydrogen bond confers less rigidity on the peptide chain. The networks of hydrogen bonds linking backbone or side-chain atoms of aspartic acid with the lysyl backbone in all low-energy conformations of both peptide T and A can form presumably because the intervening glycyl residue permits the chain to turn sharply, utilizing the C_7 conformations or related forms of glycine. The lysyl–aspartic acid side-chain interaction may contribute to stabilization as in group (iii) of peptide T, but it seems to be of secondary importance in most cases, because of the unfavorable energy requirement for the bending of the long lysyl side chain.

Some of the backbone–side chain hydrogen bonds occurring in conformations of groups (i), (ii), and (iii) have been found in computed low-energy conformations of the blocked single amino acid residues[16,23] and of dipeptides.[18–21] However, the simultaneous presence of intra- and interresidue hydrogen bonds in the networks found may provide some additional stabilization.[23] In the course of minimization, ϕ and ψ values occasionally moved from one low-energy region of the blocked single amino acid maps into another region. Such movement occurred for lysine between regions C and F and between regions E and F, all of which are separated by very shallow energy barriers.[16] Aspartic acid sometimes moved from region D to region A, across the bridge in region B. This shows that starting the minimization from the blocked single-residue minima does not restrict the computation by biasing the selection of starting points.

Most low-energy conformations found here are compactly folded. Some, though not all, can be classified as bends in terms of distance criteria[24] for α carbons. However, they do not correspond to the "standard" types of chain reversals which are defined in terms of a particular hydrogen-bond geometry and/or a specified set of dihedral angles.[24,28]

The theoretical relative probability of occurrence of bends[40] in peptide T is 11, relative to peptide L (Gly-Thr-Asp-Lys) of the earlier study.[5] This is rather low (cf. ref 5). However, some of the low-energy conformations found in this study are compact but do not satisfy the criteria for bends.[24,40]

The computations can be used to predict values for the NH–C^αH coupling constants. The values of ϕ occur within one or two narrowly defined ranges for all low-energy conformations for every residue except the N-terminal ones (Figures 1C and 1D). Predicted average values of $^3J_{NHC^\alpha H}$ and $^3J_{C^\alpha HC^\beta H}$ can be calculated from Karplus-like relationships.[41,42] They are shown in Table V.

Table V
Predicted Vicinal Coupling Constants[a] for Peptides A and T

Peptide	Residue	$^3J_{NHC^\alpha H}$, Hz	$^3J_{C^\alpha HC^\beta H}$,[b] Hz
A	Ala	5.9	
	Asp	6.2	2.5, 3.1
	Gly	5.7[c]	
	Lys	7.5	4.0, 10.5
T	Thr	8.9	4.2
	Asp	5.9	1.6, 4.0
	Gly	6.3[c]	
	Lys	7.0	1.2, 5.5

[a] Based on the J vs. ϕ relations derived in ref 41 and 42. Calculated as the weighted average of the coupling constants for the conformations listed in Tables I and III. Estimated uncertainties are ±0.4 Hz for $^3J_{NHC^\alpha H}$ and ±1.0 Hz for $^3J_{C^\alpha HC^\beta H}$. [b] Where two values of J are listed, they correspond to the two different protons on the β carbon. [c] Mean value.

The terminal methyl groups are separated from each other by 4 to 5 Å in the two lowest energy conformations of peptide A [(a) and (b) in Table I] and in all conformations of groups (i), (iii), and (v) of peptide T with energies $\Delta E < 3.2$ kcal/mol. Therefore, it is predicted that a Nuclear Overhauser Effect might be observed between the terminal methyl groups of peptide T in a solvent in which the distribution of conformations computed here (Table III) is valid. Very small Nuclear Overhauser effects have been observed in related tetrapeptides.[5]

Recently, we showed[43] that observation of a Nuclear Overhauser effect between the C^αH of a given residue and the amide NH of the residue following it in the sequence may give information about the value of the dihedral angle ψ. Using the calibration curve that we presented,[43] and averaging over the low-energy conformations listed in Tables I and III, the following predictions can be made of the maximally observable Nuclear Overhauser Effect under favorable conditions.[44] In peptide A, a barely observable effect (below 6%) is predicted for the C^αH of Ala when the NH of Asp is irradiated and none for Asp with irradiation of the NH of Gly. In peptide T, no effect is predicted for the C^αH of Thr, but a sizable effect (up to 20%) is predicted for the C^αH of Asp. We also showed that further information may be obtained for threonine residues by observation of Nuclear Overhauser Effects involving the side-chain C^βH proton.[44] However, no effect is predicted for those threonyl conformations predominating in peptide T.

Work is in progress to test the various NMR predictions presented here.[7]

Acknowledgment. We wish to thank M. S. Pottle and Dr. S. S. Zimmerman for stimulating discussions, as well as for their help in setting up the computer programs and in supplying computed data on small peptides. We thank Dr. S. J. Leach for critical comments on the manuscript and S. Rumsey and R. W. Tuttle for help with graphic display programs.

Miniprint Material Available: Full size copies of Tables I–IV containing minimum-energy conformations for Ac-Ala-Asp-Gly-Lys-NMe (Table I) and Ac-Thr-Asp-Gly-Lys-NMe (Table III) and dihedral angles and energies for some low-energy conformations of Ac-Ala-Asp-Gly-Lys-NMe (Table II) and Ac-Thr-Asp-Gly-Lys-NMe (Table IV) (5 pages). See ordering information for supplementary material on any current masthead page.

References and Notes

(1) (a) This work was supported by research grants from the National Institute of Aging and the National Institute of Arthritis and Metabolic Diseases of the National Institutes of Health, U.S. Public Health Service (AG-00322 and AM-13743), and from the National Science Foundation (PCM75-08691). (b) Presented, in part, at the 11th F.E.B.S. Meeting, Copenhagen, August, 1977.

(2) Institute of Biochemistry of the Hungarian Academy of Sciences, Budapest, Hungary. Exchange fellow of the International Research and Exchange Board.

(3) C. B. Anfinsen and H. A. Scheraga, *Adv. Protein Chem.*, **29**, 205 (1975).

(4) G. Némethy and H. A. Scheraga, *Q. Rev. Biophys.*, **10**, 239 (1977).

(5) J. C. Howard, A. Ali, H. A. Scheraga, and F. A. Momany, *Macromolecules*, **8**, 607 (1975).

(6) A. Arnone, C. J. Bier, F. A. Cotton, V. M. Day, E. E. Hazen, Jr., D. C. Richardson, J. S. Richardson, and A. Yonath, *J. Biol. Chem.*, **246**, 2302 (1971).

(7) I. D. Rae, S. J. Leach, and H. A. Scheraga, work in progress.

(8) S. Fitzwater, Z. I. Hodes, and H. A. Scheraga, *Macromolecules*, following paper in this issue.

(9) J. S. Anderson and H. A. Scheraga, *Macromolecules*, this issue.

(10) Amino acids and oligopeptides in this derivative form will be referred to as "terminally blocked" throughout this paper.

(11) J. C. Howard, F. A. Momany, R. H. Andreatta, and H. A. Scheraga, *Macromolecules*, **6**, 535 (1973).

(12) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).

(13) A computer program named ECEPP (Empirical Conformational Energy Program for Peptides) was used.[12] The Fortran computer program for ECEPP, its description, and the associated geometric and energy parameters are available on magnetic tape from the Quantum Chemistry Program Exchange, as program No. QCPE 286. See footnote 60 of ref 12 for the procedure to obtain this material.

(14) R. F. McGuire, G. Vanderkooi, F. A. Momany, R. T. Ingwall, G. M. Crippen, N. Lotan, R. W. Tuttle, K. L. Kashuba, and H. A. Scheraga, *Macromolecules*, **4**, 112 (1971).

(15) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).

(16) S. S. Zimmerman, M. S. Pottle, G. Némethy, and H. A. Scheraga, *Macromolecules*, **10**, 1 (1977).

(17) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969, p 56.

(18) S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, **16**, 811 (1977).

(19) S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, in press.

(20) S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, in press.

(21) S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, in press.

(22) M. Dygert, N. Gō, and H. A. Scheraga, *Macromolecules*, **8**, 750 (1975).

(23) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Isr. J. Chem.*, **11**, 121 (1973).

(24) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Biochim. Biophys. Acta*, **303**, 211 (1973).

(25) K. Nishikawa, F. A. Momany, and H. A. Scheraga, *Macromolecules*, **7**, 797 (1974).

(26) The use of this criterion as a reasonable cutoff throughout this study was justified by preliminary investigations, from which it appeared that favorable interactions between any pair of amino acid side chains (or a side chain and the backbone) rarely contribute as much as 3 kcal/mol to the total energy in small peptides. Thus, it is very unlikely that inclusion of higher energy amino acid minima could lead to conformations with energies below those found here. This argument may not hold true in a protein where multiple interactions and small adjustments in many degrees of freedom can compensate for the occurrence of some local high-energy conformations. See also ref 18 to 21 and 25.

(27) Keeping ω fixed at 180° in this step reduced the time required for the computations. In addition, it allowed the spanning of a larger representative region of conformational space. Several times during the minimization, (ϕ,ψ) passed from one low-energy region into another one, especially for aspartic acid. This result suggests that false results are not likely to occur because of trapping in a higher energy region in early steps of selection, which might have prevented convergence to the true minimum.

(28) The dihedral angles usually quoted[24,29] for the "standard" bends correspond to conformations in which the distance between the C^{α}_i and the C^{α}_{i+3} atoms is low (<7 Å).[24] However, there exists no $NH_{i+3}\cdots O{=}C_i$ hydrogen bond for these dihedral angles or it is strongly bent. In the form of β bends usually drawn, with a nearly straight $NH_{i+3}\cdots O{=}C_i$ hydrogen bond (e.g., ref 24, 29, and 30), the dihedral angles take different values.[31]

(29) C. M. Venkatachalam, *Biopolymers*, **6**, 1425 (1968).

(30) J. L. Crawford, W. N. Lipscomb, and C. G. Schellman, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 538 (1973).

(31) A. J. Geddes, K. D. Parker, E. D. T. Atkins, and E. Beighton, *J. Mol. Biol.*, **32**, 343 (1968).

(32) G. Némethy and M. P. Printz, *Macromolecules*, **5**, 755 (1972).

(33) A computer program written in this laboratory by S. S. Zimmerman was used in this testing.[16]

(34) A cutoff distance of 2.7 Å was used. This was chosen as a safe criterion with which it is unlikely that one excludes possible hydrogen bonds; most hydrogen bonds which could be formed here have optimal H···acceptor atom distances[12,35] of 1.6 to 1.9 Å, and, in a less likely case, 2.2 Å.

(35) G. Némethy and H. A. Scheraga, *J. Phys. Chem.*, **81**, 928 (1977).

(36) F. I. Hurwitz and A. J. Hopfinger, *Int. J. Pept. Protein Res.*, **8**, 543 (1976).

(37) Y. Isogai, G. Némethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 414 (1977).

(38) The dihedral angle ϕ of glycine has values around 137° and 109° in classes (i) and (ii), respectively. The difference is not significant; the conformations in both classes occur in the *same* rather flat low-energy region for glycine.[16] The change in letter code from D* to C* is an artifact in this particular case. The boundaries of the letter-code regions were designed[16] to separate low-energy regions of the other amino acids and are not always significant for glycine.

(39) As pointed out by A. W. Burgess (private communication), after a 180° rotation around the C^{β}–C^{γ} bond of aspartic acid the carboxyl hydrogen moves to the other oxygen of the group, if one of the orientations of the carboxyl group is strongly favored energetically, e.g., by a hydrogen bond to the carbonyl oxygen. The energy difference between corresponding conformations in classes (i) and (ii) (of Tables I and III) in this case is 0.5 to 1.0 kcal/mol, so that a sizable fraction (0.15 to 0.3) of the molecules in each conformation can exist in form (ii). See also ref 23.

(40) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2293 (1971).

(41) M. T. Cung, M. Marraud, and J. Néel, *Macromolecules*, **7**, 606 (1974).

(42) K. G. R. Pachler, *Spectrochim. Acta*, **20**, 581 (1964).

(43) S. J. Leach, G. Némethy, and H. A. Scheraga, *Biochem. Biophys. Res. Commun.*, **75**, 207 (1977).

(44) P. H. Von Dreele, I. D. Rae, and H. A. Scheraga, *Biochemistry*, **17**, 956 (1978).