Conformational Energy Calculations on the Contraceptive Tetrapeptide H-Thr-Pro-Arg-Lys-OH[1]

Janet S. Anderson 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 contraceptive tetrapeptide H-Thr-Pro-Arg-Lys-OH (TPRK). Three selection strategies were used to pick starting conformations for energy minimization. In one strategy, combinations of single-residue and dipeptide minima were chosen; in the second, standard bend and regular repeating conformations were used; and in the third, Lys and Arg side chains were rotated in a search for new side chain–backbone or side chain–side chain hydrogen bonds. Many low-energy conformations were found and many of these were bends stabilized by interresidue hydrogen bonds. One bend was 1.1 kcal/mol lower in energy than all other conformations. It is not a "standard" chain reversal but has one hydrogen bond from the NH of the arginyl backbone to the C=O of the threonyl backbone and one from the threonyl side chain hydroxyl group to the C=O of the lysyl backbone. There were a number of low-energy conformations related to the global minimum one by variations of side-chain dihedral angles. Comparisons are made between the structures of TPRK and tuftsin, a sequence variant of TPRK.

I. Introduction

H-Thr-Pro-Arg-Lys-OH (TPRK)[2] is a contraceptive tetrapeptide found in the oviducal lumen of two-day pregravid hamsters.[3] Its amino acid sequence[4] indicates that it is a sequence variant of the phagocytosis-stimulating peptide tuftsin (TKPR), whose conformation was studied in the previous paper.[5] TPRK has two residues (Thr and Lys) in common with other tetrapeptides studied in this laboratory,[6,7] and it is of interest to compare the structure of TPRK with those of these other tetrapeptides to see how their conformations depend on the sequence and type of amino acid residues present. For example, the position of proline in tuftsin and in TPRK could have a different effect on the probability of formation of a bend because terminally blocked[8] Pro-X dipeptides have a higher probability of bend formation than do blocked X-Pro dipeptides,[9] where X is some other amino acid residue. Therefore, a conformational analysis of TPRK was carried out in conjunction with similar studies of other tetrapeptides.[5,7]

II. Computational Methods

TPRK has H and OH end groups on its N and C termini, respectively, and these groups (un-ionized) were used in the empirical energy calculations. The side-chain groups of Arg and Lys were taken to be uncharged to simulate the effects of ion shielding in water.[6] A description of the computational method, of the parameters describing the geometry and interaction energies, and of minimization methods is given elsewhere.[10,11] Bond lengths and bond angles were maintained fixed. The total energy, E, of any conformation is given as the sum of intramolecular nonbonded, electrostatic, hydrogen-bonded, torsional, and proline internal energies. No solvent effects were included. The effective dielectric constant in the electrostatic term was taken[10] as 2. Energies are reported as $\Delta E = E - E_0$, where E_0 is the energy at the global minimum. All dihedral angles were allowed to vary in the final minimizations except ω (the dihedral angle for rotation about the peptide bond) for Pro, Arg, and Lys, which was fixed at 180°, ϕ for Pro which was fixed at −75°, and $\chi^{6,1}$ and $\chi^{6,2}$ for Arg which were fixed at 180°. An exception was made in strategy 2 when minimizing the energies of standard bend conformations; i.e., all ω's were used as variables since Lewis et al.[12] found that this was necessary, when treating bend starting conformations, in order to maintain the bend while relieving steric overlaps. Starting values for ω for Thr were both 180 and 0° because of the higher probability that cis amide groups can precede proline.[13,14] ω for Thr was also allowed to vary in the minimization.

The minimizing algorithms used here were MINOP and POWELL. MINOP uses an unconstrained optimization algorithm with function and gradient values.[15] MINOP was faster than POWELL,[16] which was used in previous work from this laboratory, when tested[18] on molecules with greater than seven variables.

III. Selection Strategy for Starting Conformations

Because of the multiple-minimum problem in calculations of this type, it is desirable to have as complete a coverage of conformational space as possible in order to be able to minimize the energy to the global minimum. However, if one used all possible combinations of low-energy single-residue minima[17] (SRM) for TPRK, one would have to minimize the energies of the order of 2×10^6 starting conformations. To reduce these to an economically feasible number (less than 10^4), strategies are devised to select starting conformations which will probably lead to low-energy-minimized conformations. Implicit in these selection strategies is the assumption that low-energy conformations of TPRK can be found from combinations of SRM.

As suggested in ref 7, it was decided to examine a dipeptide fragment of TPRK to see if a particular dipeptide conformation had special stability and then to combine the low-energy conformations of the dipeptide with combinations of SRM of the other two residues (strategy 1). The dipeptide Pro-Arg was chosen because it is at the center of the only possible β-type bend in this tetrapeptide and its backbone dihedral angles would determine if a bend were able to form. A bend is a likely conformation of a tetrapeptide and would have to be included in any selection strategy. The use of SRM to form starting conformations of the Pro-Arg dipeptide is justified by empirical energy calculations on blocked Pro-X dipeptides[9] which indicate that most of the conformations of X in its SRM were present in the minimum-energy structures of Pro-X. Previous work on tuftsin[19] had led to a set of 23 Pro-Arg dipeptide minima. They were found by minimizing terminally blocked[8] combinations of earlier SRM[20] from this laboratory and then grouping together conformations with similar dihedral angles. Because they were already available, these dipeptide minima (listed in Table I) were used in a *preliminary* search in strategy 1 for low-energy conformations of TPRK, as described below.

There seemed to be no special advantage in searching for low-energy conformations of TPR or PRK tripeptides be-

0024-9297/78/2211-0812$01.00/0

Table I
Classes of Terminally Blocked Dipeptide Minima for Pro-Arg

Class[a]	Dihedral angles, deg								Conformational letter codes[b]
	ψ_P	ϕ_R	ψ_R	χ_R^1	χ_R^2	χ_R^3	χ_R^4	χ_R^5	
1	79	−69	129	−174	168	176	−91	90	CC
2	160	−69	130	−174	168	175	−94	87	FF
3	79	−158	133	−174	167	176	−91	91	CE
4	78	−70	−49	−174	166	173	−99	82	CA
5	159	−69	−48	−175	166	172	−102	75	FA
6	−47	−70	−48	−175	166	172	−102	76	AA
7	−45	53	48	−56	−164	−167	−58	−59	AA*
8	77	−65	154	71	−159	−177	84	75	CF
9	−47	−64	152	70	−156	177	−101	69	AF
10	−49	−71	127	−174	168	175	−93	89	AC
11	158	−158	136	−174	167	175	−92	90	FE
12	−46	−158	137	−174	166	175	−95	85	AE
13	−26	−158	137	−174	166	175	−95	86	AE
14	79	52	51	−58	−167	−173	−85	92	CA*
15	159	52	54	−59	−167	−172	−81	92	FA*
16	162	−64	154	70	−159	−177	−95	79	FF
17	78	−159	−44	−175	165	173	−99	84	CG
18	158	−159	−43	−174	165	173	−99	84	FG
19	−46	−158	−41	−175	164	172	−102	76	AG
20	−25	−158	−40	−175	164	172	−102	76	AB
21	79	55	74	−167	172	174	83	76	CA*
22	79	55	74	−166	−178	−177	−80	97	CA*
23	78	55	−100	−169	175	178	−68	92	CC*

[a] $\phi_P = -75°$, $\omega_P = 180°$, $\omega_R = 180°$, $\chi_R^{6,1} = 180°$, $\chi_R^{6,2} = 0°$ for all classes. [b] Defined in ref 17.

Table II
Classes of Single-Residue Minima for Terminally Blocked Threonine

Class[a]	Dihedral angles, deg					Conformational letter code
	ϕ	ψ	χ^1	$\chi^{2,1}$	$\chi^{2,2}$	
1	−91	83	58	64	69	C
2	−166	135	−93	65	52	E
3	−82	81	−56	165	59	C
4	−90	84	−61	71	57	C
5	−86	74	45	161	63	C
6	−155	157	−177	165	64	E
7	−81	−50	45	166	63	A
8	−166	−54	−92	67	48	G
9	−148	36	176	−64	52	D
10	−80	−38	45	89	62	A
11	−69	−44	−57	166	59	A
12	−86	−46	−64	69	56	A
13	−85	150	46	165	64	F
14	54	65	−54	74	58	A*

[a] $\omega = 180°$ for all classes.

cause, while that strategy was useful in one case,[5] it did not work in others.[6,7] The TPR tripeptide that would be selected from X-Pro and Pro-X minima,[9] where X is any amino acid, would still have to be combined with at least 13 Lys single-residue minima (see selection strategy 1 outlined below) to be sure that all possible combinations of TPR + K were checked. In addition, because arginine had not been treated in previous studies of dipeptides,[9] it would not have been possible to build the tripeptide from previously calculated dipeptide minima.

The number of (more recent[17]) SRM for Thr and Lys with ΔE less than 3 kcal/mol was reduced from 16 to 14 and from 176 to 13 respectively by grouping minima with similar values of ϕ, ψ, and χ^1. A few conformations which had low frequencies of occurrence and relatively high energies (close to $\Delta E \sim 3$ kcal/mol) were ignored. An attempt was made to include in these groups the largest possible representation of different backbone dihedral angles. In each class, χ^i's for i greater than 1 were assigned as the values which correspond to the lowest energy conformation in the class. These classes are listed in Tables II and III. Although these SRM were obtained from terminally blocked Thr and Lys, and Thr has an unblocked N-terminus and Lys an unblocked C-terminus in TPRK, any unnecessary restrictions thereby imposed on ϕ and ψ of Thr or Lys are relieved during the minimization procedure. For example, the second lowest energy conformation of TPRK found, conformation number ii in Table V, has $\phi_T = -37°$, which is a value not present in the SRM of Table II or in the other selection strategies.

In the *preliminary* search of 4186 possible combinations of single residue and dipeptide minima for TPRK (14 Thr × 23 Pro-Arg × 13 Lys), 1623 representative conformations were actually generated, and their energies were calculated with ECEPP.[10,11] Out of these 1623, 190 had energies within approximately 6 kcal/mol of the lowest energy conformation generated. The distance, R, between C^α of Thr and C^α of Lys was also calculated in order to identify bends (those conformations with $R \leq 7$ Å).[12] Of the 190 conformations 70 had R

Table III
Classes of Single-Residue Minima for Terminally Blocked Lysine

Class[a]	Dihedral angles, deg							Conformational letter code
	ϕ	ψ	χ^1	χ^2	χ^3	χ^4	χ^5	
1	−83	85	−172	178	180	179	62	C
2	−86	79	−66	−178	180	179	62	C
3	−74	−48	−175	177	180	−179	177	A
4	−155	158	63	180	180	179	62	E
5	−150	130	−175	176	180	−179	178	E
6	−148	92	−174	178	179	179	62	D
7	−155	−60	−177	176	180	−179	178	G
8	−78	−49	−67	−179	180	179	63	A
9	−129	148	−67	−176	−179	−179	177	E
10	−152	44	54	173	179	−179	178	D
11	55	58	−62	−176	−179	−179	178	A*
12	58	65	−165	180	180	179	62	A*
13	−70	150	75	−172	−179	180	63	F

[a] $\omega = 180°$ for all classes.

Table IV
Classes of Single-Residue Minima for Terminally Blocked Arginine

Class[a]	Dihedral angles, deg						Conformational letter code
	ϕ	ψ	χ^1	χ^2	χ^3	χ^4	
1	−82	84	−67	−179	180	82	C
2	−75	109	−173	175	179	82	C
3	−157	129	−170	176	−179	−81	E
4	−72	−44	−173	175	178	−83	A
5	−75	−42	−69	−179	−179	82	A
6	−162	159	59	−179	−179	82	E
7	−138	144	−69	−175	180	−83	E
8	−163	−54	−176	175	178	−82	G
9	−78	132	−67	−174	−67	103	F
10	52	60	−61	−175	−178	−81	A*

[a] $\omega = 180°$, $\chi^5 = 0°$, $\chi^{6,1} = 180°$, $\chi^{6,2} = 180°$ for all classes.

< 7.5 Å and, therefore, could be considered to be potential bend conformations. These *preliminary* calculations also showed which Thr and Lys SRM (when combined with low-energy conformations of Pro-Arg) formed low-energy conformations of TPRK and which combinations usually led to high-energy conformations. It was found that energy minimization did not change the relative order of the conformations; i.e., a high-energy starting conformation never minimized to one with an energy comparable to that attained by a low-energy starting conformation after minimization. This information enabled us to eliminate the rest of the 4186 possible combinations of SRM by concentrating our calculations only on combinations of single-residue and dipeptide minima which would have been expected to lead to low-energy conformations of TPRK.

Twenty of the lowest energy conformations (ten bends and ten which were not bends) were subjected to three iterations of energy minimization with POWELL. Two bend conformations led to the same lowest energy structure of TPRK obtained in this *preliminary* search. This minimum was about 3 kcal/mol lower in energy than all of the other conformations.

To be sure that the 23 Pro-Arg dipeptide minima used above were representative of all possible conformations, it was decided to generate another set of dipeptide minima for strategy 1 by minimizing combinations of Pro and Arg SRM. The 144 SRM[17] of Arg were reduced to 10 in the same manner as done above for Thr and Lys. The classes of Arg minima used are listed in Table IV. All five Pro SRM[17] were used. They included three trans and two cis potential peptide bonds between Thr and Pro which will appear later when TPRK is formed from these dipeptide minima. The energies of these 50 combinations were minimized as terminally blocked dipeptides with MINOP for a maximum of ten iterations, with all dihedral angles as variables except those for rotation of the terminal methyl groups, ω for Pro and Arg, and χ^5, $\chi^{6,1}$, and $\chi^{6,2}$ of Arg. The energies of the resulting dipeptide minima were approximately equal to the sum of the energies of their SRM, so that no dipeptide minimum had significant stabilizing interactions between the two residues. Seventeen of these minima were similar to seventeen of the 23 Pro-Arg minima of the *preliminary* search, but 33 of the new minima were not found in the *preliminary* Pro-Arg dipeptide set.

Because the first set of dipeptide minima used in the *preliminary* search did not include all Pro-Arg conformations present in the second set, a second search in strategy 1 was begun using selected Thr and Lys SRM with all 50 new Pro-Arg dipeptide minima. The Thr minima used were classes 1, 2, 13, and 14 in Table II, and the Lys minima used were classes 1 and 5 in Table III. These SRM were selected because they usually produced low-energy TPRK conformations in the *preliminary* search. Also, since Pro restricts the conformational space of the residue before it[9] (see Discussion), it was not necessary to examine all 14 Thr SRM in Table II. In general, in the second search, the energy of each TPRK conformation was minimized with respect to only the backbone dihedral angles. Specifically, the energies of the conformations formed from Thr 1, Pro-Arg 1–50, and Lys 1 and 5 were minimized with respect to the backbone dihedral angles using MINOP for ten iterations (or to $\Delta E < 0.01$ kcal/mol, whichever came first). The results indicated that the conformations with Lys in class 5 (region E) almost always had a lower energy than the same conformation with Lys in class 1, showing that minimum-energy TPRK backbone structures depend only

Table V
Minimum-Energy Conformations of TPRK from Strategies 1 and 2

Conformation No.	Short-hand notation for dihedral angles[a] Thr[b] (ϕ,ψ)	ω	χ^1	$\chi^{2,1}$	Pro[c] (ϕ,ψ)	Arg[d] (ϕ,ψ)	χ^1	χ^2	χ^3	χ^4	Lys[e] (ϕ,ψ)	R,[f] Å	ΔE,[g] kcal/mol
i	F	t	g^+	t	C	A*	g^-	t	t	g^-	E	5.5	0.0
ii	H	t	g^+	t	A	C	g^-	t	t	g^+	E	6.3	1.1
iii	C	t	g^+	g^+	A	A	g^-	t	t	g^+	E	5.5	1.5
iv	F	t	g^+	t	A	C	t	t	t	g^+	E	6.4	1.9
v	C	t	g^+	g^+	A	A	g^-	t	t	g^+	C	5.2	2.3
vi	C	c	g^+	g^+	A	A	g^-	t	t	g^+	E	4.8	3.2
vii	C	t	g^+	g^+	C	G	t	t	t	g^-	E	6.0	3.2
viii	D	t	g^+	g^+	C	C	g^-	t	t	g^+	E	8.9	3.3
ix	D	t	g^+	g^+	C	A	g^-	t	t	g^+	E	8.6	3.3
x	D	t	g^+	g^+	C	A	g^-	t	t	g^+	C	8.6	3.6
xi	D	t	g^+	g^+	C	C	g^-	t	t	g^+	C	8.9	3.7
xii	C	c	g^+	g^+	A	A	t	t	t	g^-	E	4.8	3.8
xiii	C	t	g^+	g^+	A	C	g^-	t	t	g^+	E	5.9	4.2
xiv	A*	t	g^-	g^+	C	G	t	t	t	g^-	E	6.3	4.4
xv	D	t	g^+	g^+	C	A*	g^-	t	t	g^-	E	5.3	4.7
xvi	C	t	g^+	g^+	C	E	g^+	t	t	g^+	E	8.3	5.0
C_7^{eq}	C	t	g^+	g^+	C	C	g^-	t	t	g^+	C	8.9	3.8
II′	D	t	g^+	g^+	A	A	g^-	t	t	g^+	C	5.0	4.6

[a] Backbone dihedral angles are described by their conformational letter code.[17] Values for both ω and χ are as follows: t, near 180°; g^+, near 60°; g^-, near −60°; c, near 0°; p, near 115°. [b] $\chi^{2,2}$ for Thr is g^+ for all conformations. [c] ω for Pro = 180° for all conformations (except, for small variations around 180°, as noted in section II for standard bends). [d] ω for Arg = 180° (except, for small variations around 180°, as noted in section II for standard bends), χ^5 = c, $\chi^{6,1}$ = 180°, and $\chi^{6,2}$ = 180° for all conformations. [e] ω for Lys = 180° for all conformations (except, for small variations around 180°, as noted in section II for standard bends). For E conformations: χ^1, χ^2, χ^3, χ^4, χ^5 = t, t, t, t, t. For C conformations: χ^1, χ^2, χ^3, χ^4, χ^5 = t, t, t, t, g^+. [f] R is the distance from C^α of Thr to C^α of Lys. [g] $\Delta E = E - E_0$, where E_0 = −37.31 kcal/mol.

weakly on the Lys backbone conformation. (See Discussion for some experimental evidence to support this observation.) Therefore, the rest of the second search (with Thr 2, 13, and 14 and Pro-Arg 1–50) was carried out only with Lys in class 5 to reduce the number of minimizations required. The energies of all conformations (within 5 kcal/mol of the minimum) then were minimized with respect to all dihedral angles (except those mentioned in section II as being held fixed) using MINOP until the change in energy was less than 0.05 kcal/mol.

The second search led to a nearly continuous energy spectrum except for one distinct low-energy conformation 1.1 kcal/mol lower than the others. This was the same minimum-energy conformation of TPRK found in the *preliminary* search, which lends support to the use of these selection strategies. This minimum-energy conformation is a bend with R = 5.49 Å and has two hydrogen bonds stabilizing the bend.

In a second selection strategy (strategy 2), standard bend conformations and certain regular conformations such as the α helix were considered. The backbone dihedral angles for the bend conformations[21] were taken from ref 12. The backbone dihedral angles used for the regular α helix were (ϕ,ψ) = (−72°,−54°), for the regular extended conformation (−154°,153°), and for the regular C_7^{eq} conformation (−80°,80°).[22] Two sets of side-chain dihedral angles were selected: first from the lowest energy SRM[17] and next from their values in the minimum-energy conformation of TRPK found in the selection strategy 1. After minimization with respect to all dihedral angles to a tolerance of 0.05 kcal/mol (with MINOP), only the C_7^{eq} and the type II′ bend conformation had energies within 5 kcal/mol of the minimum.

A third selection strategy (strategy 3) involved the rotation of lysyl and arginyl side chains which had been fixed in certain minimum-energy conformations during the first and second strategies. Their side-chain dihedral angles were varied systematically to see if new side chain–backbone hydrogen bonds would form. This was a successful strategy in the search for a minimum-energy conformation of enkephalin,[23] a pentapeptide, and of other tetrapeptides.[7] For example, all the dihedral angles of TPRK were assigned their values at the minimum of the selection strategy 1 except the side-chain dihedral angles of Lys. These were varied over the values −60, 60, and 180°. Each conformation was tested for the presence of a potential Lys side-chain hydrogen bond (potential H···N or H···O hydrogen bond distance of 2.7 Å or less). If present, the energy of the conformation was minimized to a tolerance of 0.05 kcal/mol with respect to all backbone dihedral angles and the side-chain dihedral angles of Lys. A similar procedure was used for Arg.

Out of the 3^5, or 243, possible side-chain conformations of Lys, 109 conformations of TPRK had new potential hydrogen bonds formed between the Lys side chain and the rest of the molecule. Nine of these conformations had energies within 30 kcal/mol of the minimum-energy conformation from selection strategy 1. After minimization with respect to all dihedral angles with MINOP, six of the nine were within 5 kcal/mol of the minimum, but none was closer than 1.4 kcal/mol to the minimum. For Arg, 33 out of the 81 possible conformations had potential hydrogen bonds, and eight of the 33 had low energies. Six of these eight were within 5 kcal of the minimum but none was closer than 2.7 kcal/mol.

No new conformation generated in strategy 3 had an energy less than the minimum that was found previously. Therefore, the search for new hydrogen bonds by variation of χ's was carried out only for the lowest minimum-energy backbone conformation.

IV. Results

Strategy 1 yielded the global minimum of TPRK and 12 other conformations (numbers i–iv, vi–ix, and xii–xvi) with energies less than 5 kcal/mol from the global minimum. These conformations have different backbone dihedral angles except that all have Lys in the E state for reasons specified in the selection strategy. Table V lists the dihedral angles of these

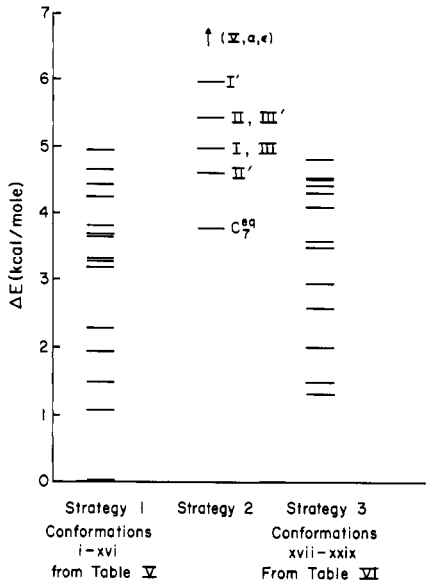

Figure 1. Low-energy conformations of TPRK in order of increasing energy. Column 1 is the results from strategy 1, column 2 from strategy 2, and column 3 from strategy 3. The ordinate is $\Delta E = E - E_0$, where E_0 is the energy of the global minimum, −37.31 kcal/mol. The arrow in column 2 means that the α, ϵ, and V conformations had energies greater than ΔE = 7 kcal/mol. Large Roman numerals represent the standard bend conformations. Small Roman numerals are conformation numbers in Tables V and VI.

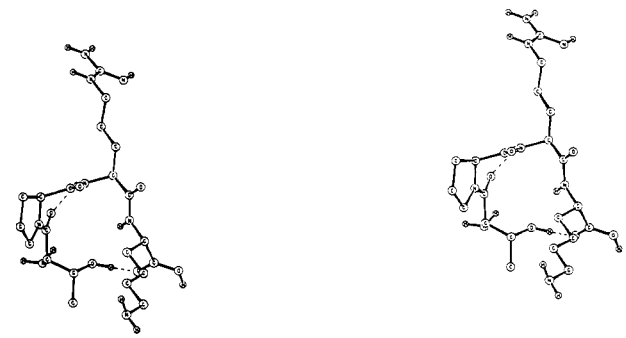

Figure 2. Stereoscopic illustration of the lowest energy conformation of TPRK. Hydrogen bonds are indicated by dashed lines. Backbone and side-chain aliphatic hydrogen atoms have been omitted for clarity.

conformations, and their relative energies are plotted in Figure 1.

A few of the low-energy conformations were reminimized with Lys in the C state. They always yielded higher energies and the TPR dihedral angles did not change. The conformations which minimized to a $\Delta E < 5$ kcal/mol are listed in Table V and shown in Figure 1. They are conformations v, x, and xi and were derived from conformations iii, ix, and viii, respectively, by changing Lys from the E to the C state.

Ten of the thirteen conformations (numbers i–iv, vi, vii, and xii–xv) are compact structures, or bends, characterized by $R < 7$ Å. Each one has between one and four hydrogen bonds stabilizing the structure. Two of the ten bend conformations (numbers vi and xii) have a cis Thr-Pro peptide bond and differ only in the arginyl side-chain dihedral angles. They have a very compact structure, with $R = 4.8$ Å, and are stabilized by at least three interresidue hydrogen bonds.

The other three (nonbend) conformations (numbers viii, ix, and xvi) have $R > 8.3$ Å, and their hydrogen bonds are mostly intraresidue. They are generally higher in energy than the bends.

The global minimum (conformation number i) is a bend with $R = 5.49$ Å and has two hydrogen bonds: one from the amino hydrogen of Arg to the carbonyl oxygen of Thr and one from the hydroxyl hydrogen of Thr to the carbonyl oxygen of Lys. The latter hydrogen bond brings the two ends of the molecule together while the former stabilizes the Pro-Arg bend. This is a type V bend according to the classification of Lewis et al.,[12] and there is a seven-membered ring present in the Pro-Arg bend formed by the hydrogen bond between Thr and Arg. This is shown in Figure 2, a stereodiagram of the global minimum of TPRK.

Selection strategy 2, the minimization of standard bend and regular repeating conformations, led to two low-energy conformations, the C_7^{eq} conformation and the type II′ standard bend. Their dihedral angles are listed in Table V. The C_7^{eq} conformation has an open structure, with $R = 8.9$ Å, and has two hydrogen bonds, both to the carbonyl oxygen of Thr: one from the hydroxyl hydrogen of Thr and one from the amino hydrogen of the arginyl backbone which makes a seven-membered ring. The type II′ standard bend conformation[21] minimized to a ΔE of 4.6 kcal/mol. It has an R value of 5.0 Å and an intraresidue hydrogen bond from the hydroxyl hydrogen to the carbonyl oxygen of Thr. The energies of the other minima derived from selection strategy 2 are shown in Figure 1.

Selection strategy 3, the variation of side-chain dihedral angles of Lys and Arg of the global minimum conformation, gave 13 new low-energy conformations (numbers xxii–xxix) but none with lower energy than that of the global minimum. Even though these conformations were selected to give potential hydrogen bonds between the rotatable side chain and the rest of the molecule, only one conformation (number xxvi) minimized to form such a hydrogen bond (from the ϵ-NH_2 hydrogen of the lysyl side chain to the carbonyl oxygen of Pro), and this particular conformation was 3.5 kcal/mol higher in energy than the minimum. The dihedral angles for these new low-energy conformations are listed in Table VI, and their energies are plotted in Figure 1. A few of these conformations possess the intraresidue threonyl hydrogen bond between the amino hydrogen and the hydroxyl oxygen that is present in some of the SRM of Thr. These hydrogen and oxygen atoms are 2.46 Å apart in the global minimum and so form a weak hydrogen bond. It is interesting that more hydrogen bonds do not necessarily increase the stability of the minimum-energy conformation.

Table VII lists all the dihedral angles of the low-energy conformations of TPRK whose stereodiagrams are included as Figures 2 through 5.

V. Discussion

The preferred conformation for TPRK is a bend stabilized by two hydrogen bonds. The arginyl and lysyl side chains could be flexible in solution as shown by the many low-energy conformations (numbers xvii–xxix) with the same backbone dihedral angles as the global minimum but with different side-chain dihedral angles. There are other low-energy conformations with different backbone and side-chain dihedral angles but, if the probability of occurrence of these conformations at 25 °C is computed by the method described previously,[13] no single conformation has a higher than 10% probability. The global minimum has a 60% probability of occurrence. In these calculations, the partition function Q was summed over all states shown on Figure 1 with energies less than 5 kcal/mol from the minimum.

Using selection strategies, it is possible that every low-energy conformation of TPRK may not be present in Figure 1, but it is probable that any omitted conformation will vary only in side-chain dihedral angles and that the main conclusions presented here are still applicable.

Table VI
Minimum-Energy Conformations of TPRK from Strategy 3

Conformation No.[c]	Short-hand notation for dihedral angles[a,b] Arg χ^1	Arg χ^2	Arg χ^3	Arg χ^4	Arg χ^5	Lys χ^1	Lys χ^2	Lys χ^3	Lys χ^4	Lys χ^5	ΔE, kcal/mol
xvii	g^-	t	g^-	g^+	c	t	t	t	t	t	2.6
xviii	g^-	t	g^+	g^-	c	t	t	t	t	t	2.9
xix	g^-	p	g^+	g^+	c	t	t	t	t	t	3.6
xx	g^-	g^-	g^-	g^+	c	t	t	t	t	t	4.4
xxi	g^-	g^-	g^+	g^+	c	t	t	t	t	t	4.5
xxii	g^-	g^-	g^+	t	c	t	t	t	t	t	4.8
xxiii	g^-	t	t	g^-	c	g^-	t	t	g^+	g^-	1.3
xxiv	g^-	t	t	g^-	c	g^-	t	t	g^+	t	1.5
xxv	g^-	t	t	g^-	c	g^-	t	t	g^+	g^+	2.0
xxvi	g^-	t	t	g^-	c	g^+	t	g^+	g^+	t	3.5
xxvii	g^-	t	t	g^-	c	g^-	g^-	t	g^-	g^-	4.1
xxviii	g^-	t	t	g^-	c	g^-	g^-	t	g^-	t	4.3
xxix	g^-	t	t	g^-	c	g^-	g^-	t	g^-	g^+	4.5

[a] See footnote *a* of Table V. [b] For all these conformations, the dihedral angles of Thr and Pro, ϕ, ψ, ω, $\chi^{6,1}$, and $\chi^{6,2}$ of Arg, and ϕ, ψ, ω of Lys are the same as for conformation i of Table V. [c] In the first group, the conformation of Lys is constant. In the second group, that of Arg is constant.

Table VII
Dihedral Angles and Energies for Some Low-Energy Conformations of TPRK

Dihedral angle[a]	Conformation[b] (dihedral angle, deg) i	ii	xxvi	C_7^{eq}
ϕ_T	−76	−37	−69	−61
ψ_T	156	166	160	90
ω_T	172	157	173	178
χ_T^1	48	52	49	61
$\chi_T^{2,1}$	165	158	165	64
$\chi_T^{2,2}$	63	63	63	67
ψ_P	74	−40	67	85
ϕ_R	51	−86	51	−80
ψ_R	62	113	59	85
χ_R^1	−61	−66	−61	−67
χ_R^2	−175	178	−175	−174
χ_R^3	−178	−178	−178	−177
χ_R^4	−82	80	−82	86
χ_R^5	0	0	0	2
ϕ_K	−151	−147	−168	−79
ψ_K	135	136	171	120
χ_K^1	−175	−174	60	−174
χ_K^2	176	176	161	178
χ_K^3	180	180	62	180
χ_K^4	−179	−179	58	179
χ_K^5	178	178	161	62
ΔE, kcal/mol	0.0	1.1	3.5	3.8

[a] $\phi_P = -75°$; ω_P, ω_R, $\omega_K = 180°$; $\chi_R^{6,1} = 180°$; $\chi_R^{6,2} = 180°$ for all conformations. [b] The numbers refer to column 1 in Tables V and VI.

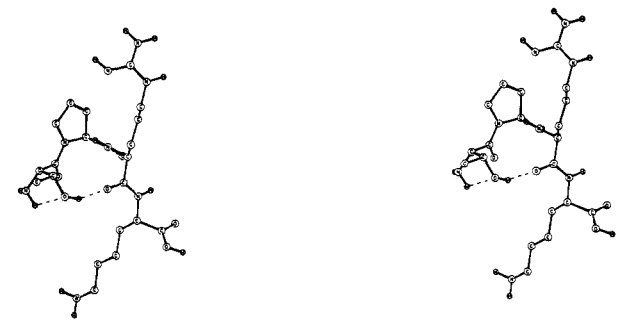

Figure 3. Stereoscopic illustration of the second lowest energy conformation of TPRK. (Conformation ii in Table V.) See Figure 2 for an explanation.

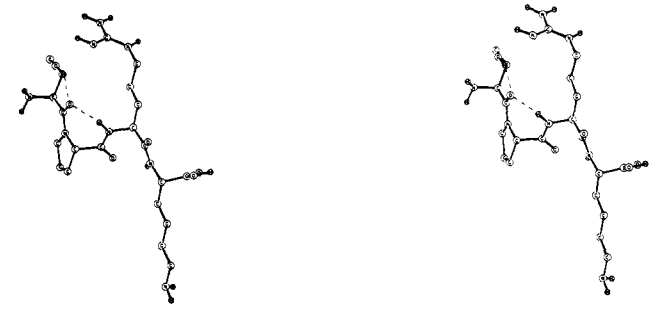

Figure 4. Stereoscopic illustration of the regular repeating C_7^{eq} conformation of TPRK. See Figure 2 for an explanation.

Figure 3 shows a stereodiagram of the second lowest energy conformation from strategy 1 (number ii). *R* for this conformation is 6.3 Å, larger than that for the global minimum but still characteristic of a bend. Even though it has very different backbone dihedral angles, its overall structure is very similar to the global minimum conformation because the arginyl and lysyl side chains are still extended out away from the rest of the molecule. This conformation has a threonyl intraresidue hydrogen bond from the amino hydrogen to the hydroxyl oxygen and an interresidue hydrogen bond from the hydroxyl hydrogen of threonine to the carbonyl oxygen of arginine.

Figure 4 shows the C_7^{eq} stereodiagram and the seven-membered ring formed by the hydrogen bond from the backbone amide hydrogen of Arg to the carbonyl oxygen of Thr. Figure 5 is the stereodiagram of the only conformation from strategy 3 (number xxvi) to have a new side chain–backbone hydrogen bond: from the ϵ-amino hydrogen of Lys to the carbonyl oxygen of Pro. There are three other hydrogen bonds in this conformation; two are the same as those present in the global minimum.

The presence of Pro in the second position of TPRK does limit the conformational space available to Thr but not to Arg, as predicted earlier.[9,14] There are seven different conformational letter codes in the classes of Thr SRM of Table II, C, A, E, G, D, F, and A*, but there are only five in the low-energy conformations of Table V, C, D, F, A*, and H (which is very close to the H–F boundary). In TPRK, Thr has an unblocked amino group while, in the SRM, Thr is blocked by an *N*-acetyl

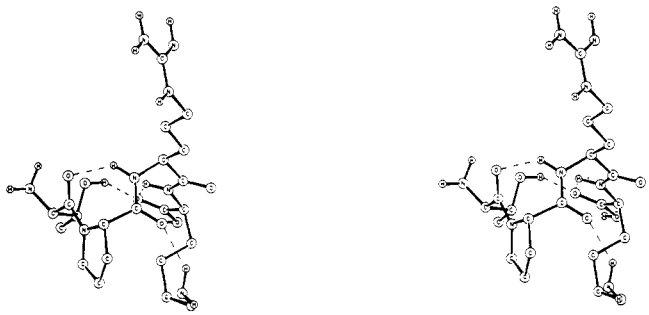

Figure 5. Stereoscopic illustration of the only conformation of TPRK from strategy 3 to have a lysyl side-chain hydrogen bond. (Conformation xxvi in Table VI.) See Figure 2 for an explanation.

group. This could explain why Thr is not more restricted by Pro in TPRK when comparing observed conformational letter codes. However, out of the six letter codes in the classes of Arg SRM in Table IV, five are still present in the low-energy conformations of TPRK.

The effect of solvent, which was not accounted for explicitly in these calculations, would be to destabilize the hydrogen bonds. It would probably have little effect on the polar arginyl and lysyl side chains since they are already extended in the minimum-energy conformation. Even though destabilization of bends could occur in a polar solvent, there are so many low-energy bend conformations that the molecule still may have a compact shape.

Lewis et al.[24] examined bend positions in proteins and computed a priori probabilities for a residue to occur in a definite position in a β bend. (A β bend was defined as a conformation with $R < 7$ Å and with the middle two residues not being in an α helix.) By multiplying these probabilities together for a given sequence of four residues, one can calculate the a priori probability for the sequence to be in a bend in a protein. For TPRK, the a priori probability is 1.5×10^{-4}, which is 2.7 times the probability that tuftsin will be in a bend. The energy calculations for tuftsin[5] showed that bend formation occurs but only for conformations with a cis Lys-Pro peptide bond. These conformations always had energies greater than 1.6 kcal/mol from the minimum energy calculated. The a priori probabilities of Lewis et al.[24] are not applicable to the prediction of bend conformations in cis peptides.

Lewis et al.[12] calculated the probability that *N*-acetyl *N'*-methyl Ala-Ala-Ala-Ala amide (A_4) would form a bend by calculating the energies of many possible conformations of A_4. They concluded that, if the a priori probability for bend formation for a tetrapeptide was greater than 10^{-4}, then bends will occur. The a priori probability for bend formation for A_4 is 5.9×10^{-6}. Therefore, TPRK has a definite tendency for bend formation, while tuftsin is marginal, and the results of this paper and those in ref 5 draw similar conclusions. It is important to remember that the a priori probabilities[24] were calculated for tetrapeptide sequences in proteins while the energy calculations for TPRK and tuftsin were carried out for the unblocked polypeptide. Also, A_4 conformations were examined using the blocking end groups which might have an effect on bend formation.

The presence of low-energy bends for TPRK but not for tuftsin is in agreement with studies of dipeptides.[9] In these studies, it was predicted that bends are favored for Pro-X but not for X-Pro because the conformational space for X in X-Pro is limited to conformations not usually found in bends because of the restrictions of the pyrrolidine ring. The preference of Pro to be in position 2 in a reverse turn of four residues was found to be quite strong in an examination of the x-ray coordinates of several proteins.[24,25] Of the residues in the second position of a bend 33% were Pro. This trend was also seen in NMR studies of tetrapeptides with Pro at position 2 and of tetrapeptides without Pro.[26] The smallest value of R found for tuftsin[5] with Lys in the lower energy trans conformation is 8.3 Å while R for the minimum energy conformation of TPRK is 5.5 Å. The stability of bends for most low-energy TPRK conformations is also due to two interresidue hydrogen bonds. Tuftsin appears to have no interresidue hydrogen bonds.

Table VIII
Predicted Vicinal Coupling Constants[a] for TPRK

Residue	$^3J_{NHC^\alpha H}$, Hz	$^3J_{C^\alpha HC^\beta H}$,[b] Hz
Thr	7.1	4.7
Pro	6.4	3.4
		10.5
Arg	5.5	13.7
		2.5
Lys	8.8	1.9
		13.6

[a] Based on the J vs. ϕ relations from ref 27 and 28 and calculated from the global minimum conformation. 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.

The NMR coupling constants, $^3J_{NHC^\alpha H}$ and $^3J_{C^\alpha HC^\beta H}$, were calculated for the minimum-energy conformation according to the methods of ref 27 and 28 and are listed in Table VIII. NMR experiments on TPRK are currently in progress,[29] and the preliminary results show a cis–trans isomeric ratio of 1 to 3 and that the Lys NH signals are shielded relative to those of Arg. This could be due in part to a reverse turn conformation like the global minimum shown in Figure 2.

Few physiological experiments have been performed with TPRK to determine the mechanism of its contraceptive action; hence, there is little experimental information with which to compare our calculated structure. However, Kent[30] showed that oral doses of TPRK were still effective in preventing conception. Since trypsin would be expected to cleave the Arg-Lys peptide bond, the tripeptide TPR by itself may be the active part of the peptide. Therefore, our energy calculations, which showed that the conformation of the lysyl side chain had little effect on the conformation of the rest of the molecule, may be supported by this experimental observation. It is impossible to say what the structure of the active agent is in vivo because there may be conformational changes on binding to the active site.

Acknowledgment. We thank Z. I. Hodes for use of his dipeptide minima from tuftsin, Dr. G. Némethy for helpful discussions, and S. Rumsey and R. W. Tuttle for help with graphic display programs.

References and Notes

(1) This work was supported by research grants from the National Institute of Arthritis and Metabolic Diseases of the National Institutes of Health, U. S. Public Health Service (AM-13743), and from the National Science Foundation (PCM75-08691). J.S.A. was an NIH Postdoctoral Fellow.
(2) T, P, R, and K are the one-letter abbreviations of these amino acids according to the *IUPAC-IUB Comm. Biochem. Nomencl., Arch. Biochem. Biophys.* **125**, 1–5 (1968).
(3) H. A. Kent, Jr., *Gen. Comp. Endocrinol.*, **22**, 397 (1974).
(4) H. A. Kent, Jr., *Biol. Reprod.*, **12**, 504 (1975).
(5) S. Fitzwater, Z. I. Hodes, and H. A. Scheraga, *Macromolecules*, 11, 805 (1978).
(6) J. C. Howard, A. Ali, H. A. Scheraga, and F. A. Momany, *Macromolecules*, **8**, 607 (1975).
(7) I. Simon, G. Némethy, and H. A. Scheraga, *Macromolecules*, **11**, 797 (1978).
(8) Terminally blocked dipeptides have *N*-acetyl and *N'*-methyl end groups in order to simulate dipeptides in the middle of a chain.
(9) S. S. Zimmerman and H. A. Scheraga, *Biopolymers*, **16**, 811 (1977).
(10) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys.*

Chem., 79, 2361 (1975).
(11) A computer program named ECEPP (Empirical Conformational Energy Program for Peptides) was used.[10] The Fortran computer program for ECEPP, its description, and the associated structural and energy parameters are available on magnetic tape from the Quantum Chemistry Program Exchange, as program No. QCPE 286. See footnote 60 of ref 10 for the procedure to obtain this material.
(12) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Biochim. Biophys. Acta*, **303**, 211 (1973).
(13) S. S. Zimmerman and H. A. Scheraga, *Macromolecules*, **9**, 408 (1976).
(14) A. E. Tonelli, *J. Mol. Biol.*, **86**, 627 (1974).
(15) J. E. Dennis and H. H. W. Mei, Technical Report No. 75-246 (1975), Department of Computer Science, Cornell University, Ithaca, N.Y. 14853.
(16) M. J. D. Powell, *Comput. J.*, **7**, 155 (1964).
(17) S. S. Zimmerman, M. S. Pottle, G. Némethy, and H. A. Scheraga, *Macromolecules*, **10**, 1 (1977).
(18) M. S. Pottle, unpublished results.
(19) Z. I. Hodes, unpublished results, prior to the work of ref 5.
(20) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Isr. J. Chem.*, **11**, 121 (1973).
(21) Actually the type I′, II′, and III′ bends could not be formed in TPRK because ϕ of Pro is restricted to −75°. These bend conformations were calculated with $\phi_{Pro} = -75°$, so that they are really type IV bends. Nevertheless, these bends are reported in the results as type I′, II′, and III′ bends since they were derived from the same dihedral angles for ψ_{Pro}, ϕ_{Arg}, and ψ_{Arg}.
(22) ϕ_{Pro} was again restricted to −75°.
(23) Y. Isogai, G. Némethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 414 (1977).
(24) P. N Lewis, F. A. Momany, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2293 (1971).
(25) J. L. Crawford, W. N. Lipscomb, and C. G. Schellman, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 538 (1973).
(26) K. D. Kopple and A. Go, *Biopolymers*, **15**, 1701 (1976).
(27) K. G. R. Pachler, *Spectrochim. Acta*, **20**, 581 (1964).
(28) M. T. Cung, M. Marraud, and J. Neel, *Macromolecules*, **7**, 606 (1974).
(29) I. D. Rae, S. J. Leach, and H. A. Scheraga, work in progress.
(30) H. A. Kent, *Gen. Comp. Endocrinol.*, **29**, 253 (1976).

A Method for Predicting Nucleation Sites for Protein Folding Based on Hydrophobic Contacts[1]

R. R. Matheson, Jr.,[2] and H. A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York, 14853.
Received December 27, 1977

ABSTRACT: A method, based on hydrophobic bonding, is introduced for predicting nucleation sites for protein folding. The amino acid sequence of the protein is searched for pockets of nonpolar residues whose (negative) free energy of interaction compensates for the increase in free energy that is required to bring them into contact to form hydrophobic pockets. The predicted nucleation sites (and their associated electrostatic properties) are used to rationalize some equilibrium and kinetic results on protein folding, including the relative amplitudes of absorption by transient species observed in kinetic studies. Finally, the method proposed here is compared to the proline-isomerization hypothesis for protein folding proposed by Brandts and co-workers.

I. Introduction

Several recent reviews on the nature of protein folding have been published.[3–5] There is convincing evidence from many systems that the process is not a two-state one[6–11] although it is highly, cooperative.[12,13] However, the number, order, and conformational characteristics of the stages through which a protein passes during folding are all topics of speculation. For the case of the thermally induced unfolding of ribonuclease A (i.e., passage from the native to more disordered conformations, with the disulfide bonds intact), a large body of experimental evidence is consistent with the hypothesis that a rather well-defined sequential pathway exists.[14] Yet, even for this well-studied system, it is not established whether or not this pathway is the exact reverse of the refolding process (i.e., from the disordered to the more native structure), whether it bears any relationship to the folding when disulfide bonds are broken, or whether it is applicable to the cooperative conformational changes induced by other denaturants.

A variety of computational procedures have been proposed for the determination of the structures of native proteins,[15–24] and one of these[18] has led to the hypothesis that the folding process can be considered to take place in three steps.[25] However, most of these consider only the energetic aspects of the problem without taking the conformational entropy into account explicitly. A number proportional to the conformational entropy can be obtained in principle by the procedure of Crippen,[22] but in practice it is practical to compute only an estimate.[22] The entropy does appear implicitly (in terms of a partition function) in statistical mechanical treatments of protein folding[23,24,26,27] and is treated indirectly by at least one computational method[18] which selects conformations from a distribution, and includes the effect of solvation in a contact free energy. In the treatment presented here, we shall account explicitly (albeit approximately) for the conformational entropy and for the free energy of formation of hydrophobic bonds.

In this paper, we shall consider the initial aspects of folding, viz., the nucleation process. More specifically, we shall examine the question of the existence of nucleation sites, and the molecular structural basis for their formation. By the term "nucleation site" we mean a specific conformation of a limited section of the polypeptide chain whose existence can significantly increase the rate of formation of the native structure of the protein from its unfolded state(s). The existence of the nucleation site is a necessary but not a sufficient condition for the formation of the native structure of a protein from the nascent polypeptide chain; i.e., it may be possible for the nucleating site to form and not produce proper folding if the additional, cooperative interactions which follow it can be prevented from occurring, e.g., by the presence of a denaturant or if the pH is unsuitable. Some general classes of nucleation sites have been described by Tsong et al.[28] and by Tanaka and Scheraga.[18,25] Brandts et al.[29] have proposed that the cis-trans isomerization of proline can act as a nucleation step for folding.

In considering the initial aspects of folding, we shall examine the possibility that nucleation can be accomplished by formation of a specific pocket in the polypeptide chain, stabilized by hydrophobic contacts. A method for identifying the residues involved in this specific pocket from a knowledge of the amino acid sequence of a protein will be discussed. We shall show that application of this method yields results in

0024-9297/78/2211-0819$01.00/0