Theoret. Chim. Acta (Berl.) 35, 113—128 (1974)

Molecular Orbital Calculations on the Conformation of Polypeptides and Proteins

XI. Conformational Studies on "Tripeptide" Models

Bernard Maigret and Bernard Pullman

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S. Paris

Received March 8, 1974

Quantum mechanical calculations using the PCILO method have been performed on the tripeptide model CH_3CO–X–Y–$NHCH_3$. Competition between C_5, C_7, C_{10} rings and open structures has been investigated through mapping of the whole $\{\Phi, \Psi\}$ conformational space and energy minimization. From these results, it appears that the C_{10} ring simulating the folding named *U*-turn, involving a hydrogen bond between the $i...i+3$ residues, is the most probable structure although not the most stable in energy. The results are used for predicting the frequency of *U*-turns in proteins. α-chymotrypsin is given as an example.

Key words: Polypeptides, conformation of ~ – Proteins, conformation of ~ – α-chymotrypsin

A series of recent publications [1–7] has shown the importance of the so-called "turns" in polypeptides and proteins as a key factor in their three-dimensional structure. Most of the turns present in proteins or in cyclic or linear oligo- and polypeptides are stabilized by an intra-molecular N–H...O=C hydrogen bond which involves the amino acid residues $i...i+3$. We shall call this type of a turn a "*U*-turn" (Fig. 1). Although the importance of the "*U*-turn" is well recognized, the knowledge of the factors leading to its stability is limited: essentially geometrical H-bond criteria are used in the literature for its definition (Table 1). It seems obvious that deeper studies on the electronic properties of this hydrogen-bonded ring system (which we shall denote by the symbol C_{10}, in extension to the symbols C_5 and C_7 used in the study of dipeptides [9]) are necessary.

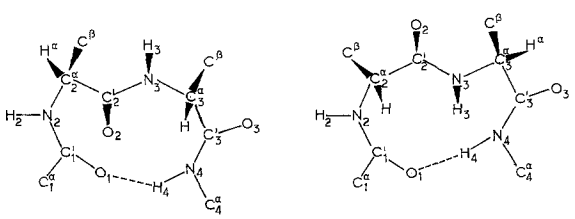

Fig. 1. Type I and II *U*-turns in an LL sequence having the hydrogen bond $i...i+3$, corresponding to the conformation denoted C_{10}

Table 1. Geometrical criteria found in the literature

Reference	[12]	[6]	[2]	[3]
H...O	1.6 Å $\leqq \leqq$ 2.5 Å	1.8 Å $<<$ 2.1 Å	—	$\leqq$ 2.5 Å
$\widehat{ONH}$	$\leqq 35°$	$9° \leqq \leqq 28°$	—	$<30°$
$C^{\alpha}_i \ldots C^{\alpha}_{i+3}$	—	<7 Å	$\leqq 6.5$ Å	<5.7 Å
$(\vec{C}^{\alpha}_i, \vec{C}^{\alpha}_{i+1}, C^{\alpha}_{i+2}, C^{\alpha}_{i+3})$	—	—	$\geqq 90°$	—
Others	—	$i+1, i+2$ non helical	—	within Venkatachalam area [6] $\pm 15°$

Investigations in this field are performed on suitable model compounds and the first study is due to Venkatachalam [10] who used a simple stereochemical yes-or-no criterium on a "tripeptide" model. This first essay, essentially qualitative, indicated only that the existence of the "*U*-turn" was favored by satisfactory Van der Waals contacts. More sophisticated model studies have been performed through "empirical" potential-energy evaluations by the teams of Popov [11], Ramachandran [12], and Scheraga [7, 13]. The essential results obtained by these authors are summarized in Table 2 and Figs. 2a, b.

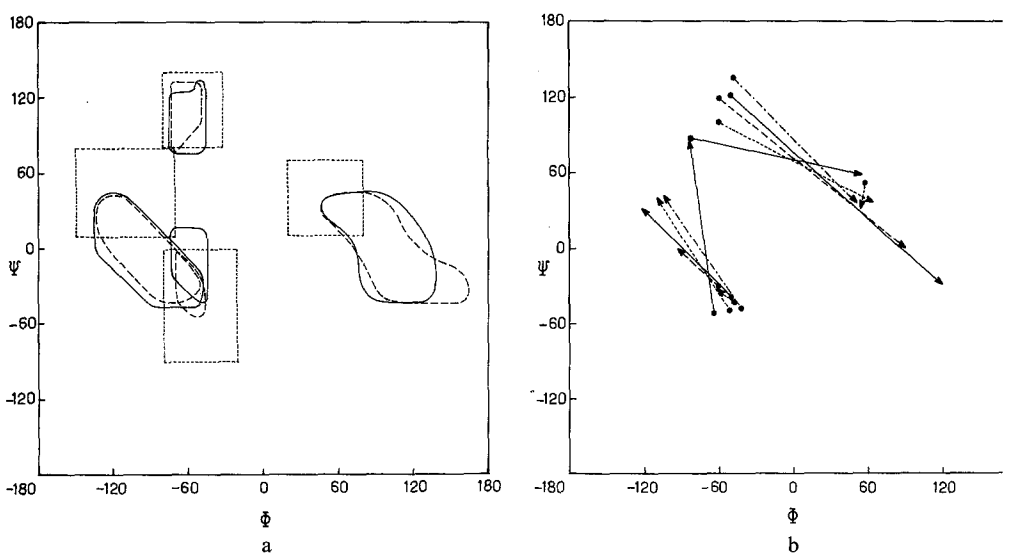

Fig. 2. Representation of the contours of stability of the *U*-turn on the ($\Phi\Psi$) map. ——— following Lewis *et al.* [7], — — — following Venkatachalam [10], ... following Chandrasekaran *et al.* [12]

Fig. 2a and b. Representation of the *U*-vector (running from Φ_{i+1}, Ψ_{i+1} to Φ_{i+2}, Ψ_{i+2}) on the (Φ, Ψ) map. ——— following Lewis *et al.* [7], — — — following Venkatachalam [10], ... following Chandrasekaran *et al.* [12], —·— following Lipkind *et al.* [11]

for the definition of "*U*-turns" in proteins

[4]	[5]	[7]	Our results for LL	Our results for LD	C_5 ring	C_7 ring
1.8 Å < < 2.1 Å	—	—	1.6 Å ≦ ≦ 2.2 Å	1.5 Å ≦ ≦ 2.2 Å	2.14 Å	1.73 Å
9° ≦ ≦ 28°	—	—	5° ≦ ≦ 41°	4° ≦ ≦ 50°	50°	21°
< 7 Å	—	< 7 Å	4.0 Å ≦ ≦ 6.4 Å	4.0 Å ≦ ≦ 5.8 Å	—	—
—	—	—	82° ≦ ≦ 173°	102° ≦ ≦ 173°	—	—
$i+1, i+2$ non helical	within Venkatachalam [6] or Mathews [8] areas	$i \ldots i+4$ non helical	—	—	—	—

Table 2. Conformational characteristics of the *U*-turn used or obtained by previous calculations

Authors	Residue $i+1$		Residue $i+2$		Type of C_{10}	Symbol used	Sequence used	Compounds on which calculations are performed
	Φ_2	Ψ_2	Φ_3	Ψ_3				
Venkatachalam [10]	−60,	−30	− 90,	0	I		LL	tripeptide model
	60,	30	90,	0	I′		LG	
	−60,	120	90,	0	II			
	60,	120	− 90,	0	II′			
	−60,	−30	− 60,	−30	III			
	60,	30	60,	30	III′			
Ramachandran *et al.* [12]	−50,	−50	−110,	40	I	Ia	L	tripeptide model
	−60,	100	60,	40	II	III	L, D	
Popov *et al.* [11]	−42,	−51	−104,	40	I	$\overbrace{R\ B}$	L	tripeptide model
	−48,	114	50,	36	II	$\overbrace{B\ L}$		
	−50,	−42	− 60,	−34	III	$\overbrace{R\ R}$		
	58,	50	55,	35	III′	$\overbrace{L\ L}$		
Kotelchuck *et al.* [13]	−50,	−40	−120,	−30	I		L	oxytocin
	−70,	−20	− 90,	0	I			
	−70,	−10	− 90,	0	I			
	−60,	−10	− 60,	−30	I			
	−70,	0	−120,	0	I			
	−70,	100	120,	0	II			
	−60,	100	60,	30	II			
	−70,	110	90,	0	II			
	−70,	120	90,	0	II			
	−50,	120	120,	−30	II			
Lewis *et al.* [7]	−64,	−52	− 83,	85	I		L	pentapeptide model
	54,	66	75,	−67	I′			
	−87,	82	52,	58	II			
	71,	−72	− 69,	−54	II′			
	−72,	−60	− 76,	−57	III			
	54,	46	56,	47	III′			

The present paper proposes a quantum-mechanical investigation on the energy conditions governing the stability of the "*U*-turn", considered as a C_{10} hydrogen-bonded ring.

1. The Procedure

The method utilized is the PCILO (Perturbative Configuration Interaction using Localized Orbitals) method applied previously to an extensive study of the conformation of the individual amino-acid residues of proteins within the "dipeptide" model. (For general summary see [14].) The program may be obtained from QCPE.

The model compounds on which the calculations on the "*U*-turn" are performed consist of three linked peptides units (Fig. 3):

$$CH_3CO\text{–}X\text{–}Y\text{–}NHCH_3$$

in which X is always a L-Alanyl residue, and Y is either a L-Alanyl, a D-Alanyl, or a Glycyl residue. Standard geometries [15] are used as input data. These model compounds allow the study of the stability of the C_{10} hydrogen-bonded ring compared with those of the C_5 and C_7 rings, as well as of the open stable structures. They make also possible the evaluation of the influence of middle range interactions upon the general contour of stability and the positions of energy minima of an individual residue. Such influences are neglected in the "dipeptide" model.

The following abreviations will be used: the "tripeptide" with Y = L-ALA will be denoted LL, the one with Y = D-ALA will be denoted LD and the one with Y = GLY will be denoted LG.

We choose as variables the angles of rotation around the four bonds of the peptide backbone adjacent to the Cα atoms: Φ_2, Ψ_2, Φ_3, Ψ_3 (Fig. 3) and we use the conventions recently adopted by the IUPAC-IUB Commission [16].

These four torsion angles are then the basis of a 4-dimensional conformational space in which each conformational state of any of the LL, LD, and LG models will be defined by a set of values $(\Phi_2^i, \Psi_2^j, \Phi_3^k, \Psi_3^l)$. The corresponding energy will be written $e_{ij,kl}^{XY}$. We have investigated this conformational space with a grid of 30° for each variable and have visualized the potential energy surface

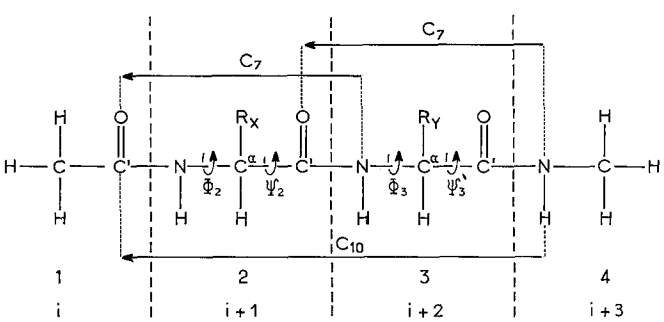

Fig. 3. The "tripeptide" model

by examination of the family of $(\Phi\,\Psi)$ maps. To this end we plotted for each LL, LD, and LG sequences the conformational (Φ_2, Ψ_2) maps at each Φ_3^k, Ψ_3^l values of the grid. We have thus obtained 144 (Φ_2, Ψ_2) submaps labelled $(\Phi_2, \Psi_2)_{kl}^{XY}$. Similarly 144 $(\Phi_3, \Psi_3)_{ij}^{XY}$ submaps are built. In this way all the regions corresponding to energy minima are determined. Already at this stage of the study, the examination of the sets of $(\Phi_3, \Psi_3)_{ij}^{XY}$ and $(\Phi_2, \Psi_2)_{kl}^{XY}$ submaps shows that the conformational energy corresponding to a couple of Φ, Ψ angles depends somewhat on the values of the Φ, Ψ couple at the adjacent residue. Consequently the rotations around the bonds of two adjacent C^α atoms are not completely independent, especially in regions lying near the energy minima. For this reason the conformational map obtained for the Glycyl residue looses the symmetry which appears on the "dipeptide" map. At the same time, these interactions do not cause however significant changes in the general location of the energy minima: except for new minima corresponding to the C_{10}-like structures, all *stable* conformations appear as combination of the "dipeptide" minima slightly shifted. This result is important because it emphasizes the possibility of the existence of a stable conformational code in polypeptide systems.

Moreover from the set of the $(\Phi_2, \Psi_2)_{kl}^{XY}$ submaps we are able to construct a global $(\Phi_2, \Psi_2)^{XY}$ conformational energy map: each E_{ij}^{XY} global state for the second residue (X) will consists of the lowest energy $e_{ij,kl}^{XY}$ conformational state found by varying kl:

$$E_{ij}^{XY} = \text{minimum} \left\{ (e_{ij,kl}^{XY})_{\substack{k=1,12 \\ l=1,12}} \right\}.$$

Working similarly for the global $(\Phi_3, \Psi_3)^{XY}$ map of the third residue (Y) we shall be able to draw for both residues X and Y individual conformational energy maps analogous to those presented in the "dipeptide" studies but which now take account of more remote interactions.

2. Results and Discussion

2.1. The Global Conformational Energy Maps

The global maps for the LL, LD, and LG sequences are presented in Figs. 4–9. The comparison of these maps with those obtained for the same individual residues in the "dipeptide" model [9, 14], shows that the general contour of stability of each residue (within, say, the usual limit of 5 kcal/mole above the global minimum) remains essentially unchanged. This proves that only short-range interactions between neighbouring peptide units are accountable for these general conformational limitations. This contour (which may be considered as a constrain in the conformational space) is a fundamental characteristic of the individual residue and is respected whatever its surroundings may be.

But if the general aspect of these global energy maps is similar for the "dipeptide" and the "tripeptide" models, there appear in these last models new large stable areas corresponding to C_{10} structures indicating the major importance taken by the "*U*-turn" in the new model system. We observed essentially two areas of conformational stability for the C_{10} ring defined roughly for

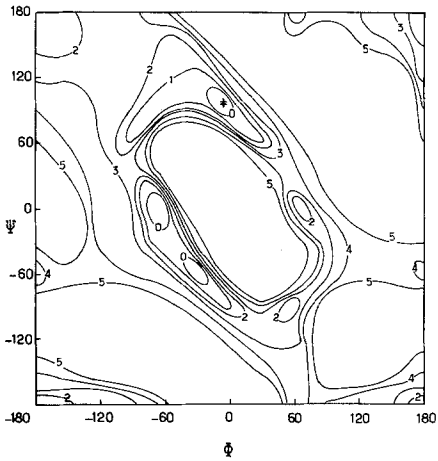

Fig. 4

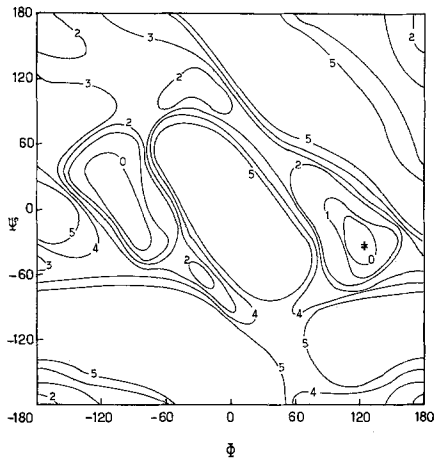

Fig. 5

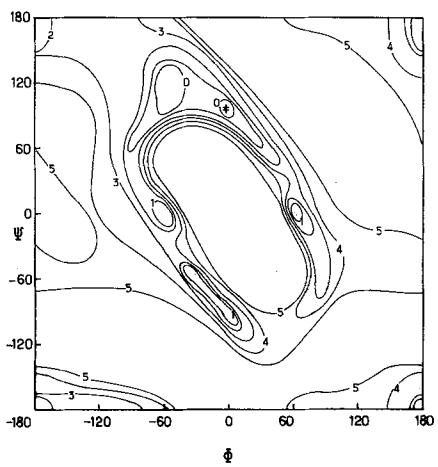

Fig. 6

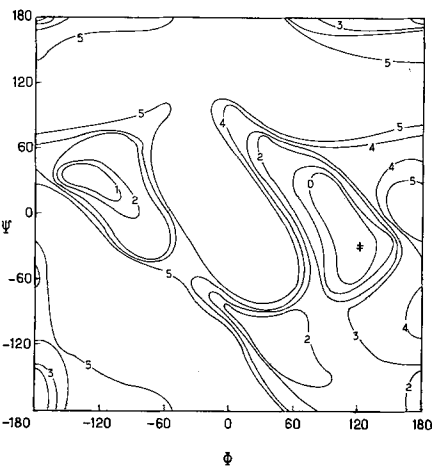

Fig. 7

Fig. 4. Global conformational energy map for the $(L\text{–}ALA)_2$ residue obtained by the "tripeptide" model for the LL sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

Fig. 5. Global conformational energy map for the $(L\text{–}ALA)_3$ residue obtained by the "tripeptide" model for the LL sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

Fig. 6. Global conformational energy map for the $(L\text{–}ALA)_2$ residue obtained by the "tripeptide" model for the LD sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

Fig. 7. Global conformational energy map for the $(D\text{–}ALA)_3$ residue obtained by the "tripeptide" model for the LD sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

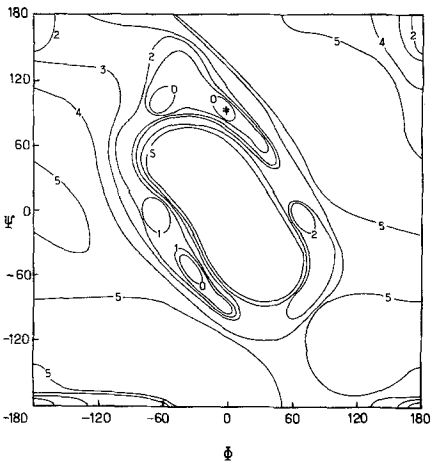

Fig. 8

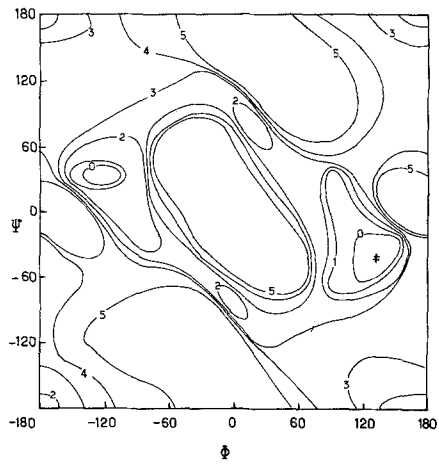

Fig. 9

Fig. 8. Global conformational energy map for the $(L\text{-}ALA)_2$ residue obtained by the "tripeptide" model for the LG sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

Fig. 9. Global conformational energy map for the $(GLY)_3$ residue obtained by the "tripeptide" model for the LG sequence. Isoenergy curves (kcal/mole) with respect to the global energy minimum (∓) taken as energy zero

LL as:

$\Phi_2 \Psi_2 = -60$ to $-30°$, -60 to $0°$ with $\Phi_3 \Psi_3 = -150$ to $-60°$, -30 to $60°$

$\Phi_2 \Psi_2 = -60$ to $0°$, 90 to $120°$ with $\Phi_3 \Psi_3 = 90$ to $150°$, -60 to $0°$,

for LD as:

$\Phi_2 \Psi_2 = -60$ to $-30°$, -60 to $0°$ with $\Phi_3 \Psi_3 = -150$ to $-90°$, 0 to $30°$

$\Phi_2 \Psi_2 = -60$ to $0°$, 90 to $150°$ with $\Phi_3 \Psi_3 = 30$ to $150°$, -60 to $30°$,

and for LS as:

$\Phi_2 \Psi_2 = -60$ to $-30°$, -60 to $0°$ with $\Phi_3 \Psi_3 = -150$ to $-90°$, 0 to $30°$

$\Phi_2 \Psi_2 = -60$ to $0°$, 90 to $150°$. with $\Phi_3 \Psi_3 = 90$ to $150°$, -60 to $30°$.

These two types of structures have been labelled by Venkatachalam [10] as C^{I}_{10} and C^{II}_{10} forms of the "*U*-turn". We show in Tybles 3–5 the geometrical characteristics (particularly those of the $i \ldots i+3$ H-bond) as well as the energies of the most stable $e^{XY}_{ij,kl}$ points of our grid.

2.2 Optimum Conformations in the Tripeptide Model

In order to refine the accuracy in the determination of the local energy minima of conformational interest we have performed energy minimizations using the Simplex method of Nelder and Mead [17]. To this end, we start the calculations from the $e^{XY}_{ij,kl}$ states recognized as belonging to low energy areas of our

Table 3. Geometrical and energetical characteristics of ($\Phi_2\Psi_2$, $\Phi_3\Psi_3$) points whose energy is less than 2.5 kcal/mole above the global minimum in an LL sequence

Conformation	Φ_2	Ψ_2	Φ_3	Ψ_3	Energy	$O_1 \ldots H_4$ Å	$\widehat{O_1N_4H_4}$ °	$C^\alpha_1 \ldots C^\alpha_4$ Å	$(\overrightarrow{C^\alpha_1C^\alpha_2}, \overrightarrow{C^\alpha_3C^\alpha_4})$ °
Type I	−30,	−60	−120,	30	−1.84	1.65	8.9	4.03	173.2
C_{10}	−60,	0	− 90,	−30	−1.73	1.75	18.1	5.39	112.3
	−60,	0	− 90,	0	−1.38	1.82	22.9	5.62	101.6
	−60,	0	−150,	30	−0.93	1.76	14.7	4.32	150.3
	−60,	−30	− 90,	0	−0.91	1.79	21.3	4.58	130.8
	−60,	0	−120,	0	−0.57	1.49	10.4	4.64	131.5
	−60,	0	− 60,	−30	−0.35	2.12	5.4	6.36	82.3
	−60,	−30	− 90,	30	−0.29	2.21	41.2	5.07	119.7
	−30,	−60	−120,	60	−0.24	1.99	31.8	4.49	162.0
	−60,	−30	− 60,	−30	−0.04	1.99	23.2	5.33	111.5
	−60,	0	−120,	30	0.04	1.99	39.1	5.12	120.5
	−60,	−30	− 60,	0	0.22	2.08	28.1	5.58	100.9
	−60,	0	− 90,	30	0.23	2.70	57.4	6.10	90.5
Type II	0,	90	120,	−30	−2.24	1.60	13.4	4.69	160.9
C_{10}	−60,	90	150,	−30	−1.11	1.77	13.5	5.38	143.0
	−30,	120	120,	−30	−1.04	1.91	12.2	5.67	137.6
	0,	90	120,	−60	−0.93	1.77	25.6	4.85	163.2
	30,	60	120,	−30	−0.36	1.65	8.9	4.03	173.2
	−60,	90	90,	0	−0.26	1.79	22.3	4.22	164.5
	−60,	120	90,	0	−0.16	1.81	21.9	4.91	159.7
	−60,	120	120,	−30	0.04	2.04	15.6	5.65	141.9
	−60,	90	90,	30	0.04	1.75	19.5	4.46	169.1
	−60,	120	90,	−30	0.10	2.20	40.4	4.85	163.9
	−30,	120	90,	−60	0.12	2.23	55.6	5.06	160.8
	−30,	120	120,	−60	0.18	2.10	26.4	5.78	142.4
	−30,	90	150,	−30	0.21	1.68	12.2	5.41	138.2
	−60,	90	120,	−30	0.25	1.97	38.0	4.57	168.0
Type I′	60,	0	150,	−30	0.01	1.76	14.7	4.32	150.3
C_{10}									
Type II′	0,	−90	−120,	30	−1.04	1.60	13.4	4.60	160.9
C_{10}	0,	−90	−120,	60	−0.15	1.77	25.6	4.85	163.2
	60,	−90	− 90,	−30	0.40	1.75	19.5	4.46	169.1
C_7C_7	−90,	60	− 90,	60	−0.57	5.43	79.5	8.50	26.6
	−90,	60	− 90,	30	−0.37	4.62	57.9	8.17	34.8
	−90,	30	− 90,	60	−0.11	4.93	82.3	7.77	54.7
	−90,	30	− 90,	30	−0.02	3.96	56.6	7.31	64.2
MC_7	−30,	120	− 90,	60	0.03	5.28	70.7	8.79	6.5
	−60,	90	− 90,	60	0.19	4.79	50.9	8.64	40.8
	0,	90	− 90,	60	0.21	4.54	52.9	8.67	18.2
C_5C_7	180,	180	− 90,	60	−0.12	6.32	46.3	8.00	99.0
	180,	150	− 90,	60	0.10	6.45	46.5	8.66	74.3
RC_7	−30,	−60	− 90,	60	−0.45	2.28	61.7	4.88	135.0
C_5C_5	180,	180	180,	180	0.00	9.18	108.6	10.93	0.0
C_7C_5	−90,	30	180,	150	0.16	6.11	118.5	6.94	124.4
RR	150,	120	150,	120	0.25	1.99	40.4	5.51	114.1
MM	−30,	120	0,	90	0.26	3.03	54.0	6.07	127.0

The zero of energy is obtained for Φ_2, $\Psi_2 \ldots 180°$, 180°, Φ_3, $\Psi_3 \ldots 180°$, 180° and the global minimum for Φ_2, $\Psi_2 = 0°$, 90°, Φ_3, $\Psi_3 \ldots 120°$, −30°.

Table 4. Geometrical and energetical caracteristics of (Φ_2, Ψ_2, Φ_3, Ψ_3) points whose energy is less than 2.5 kcal/mole above the global minimum in an LD sequence. The zero of energy is obtained for Φ_2, $\Psi_2 = 180°, 180°$, Φ_3, $\Psi_3 = 180°, 180°$, and the global minimum for Φ_2, $\Psi_2 = 0°, 90°$, Φ_3, $\Psi_3 = 120°, -30°$

Conformation	Φ_2	Ψ_2	Φ_3	Ψ_3	Energy kcal/mol	$H_2 \ldots O_1$ Å	$\widehat{O_1N_4H_4}$ °	$C^\alpha_1 \ldots C^\alpha_4$ Å	$(\overrightarrow{C^\alpha_1C^\alpha_2}, \overrightarrow{C^\alpha_3C^\alpha_4})$ °
Type I	−30,	−60	−120,	30	−1.47	1.65	8.9	4.03	173.2
C_{10}	−80,	0	−150,	30	−0.90	1.76	14.7	4.32	150.3
Type II	0,	90	120,	−30	−2.63	1.60	13.4	4.60	160.9
C_{10}	−60,	90	90,	30	−2.27	1.75	19.5	4.46	169.1
	−60,	90	90,	0	−1.92	1.79	22.3	4.22	164.5
	−60,	120	90,	0	−1.79	1.81	21.9	4.91	159.7
	0,	90	120,	−60	−1.69	1.77	25.6	4.85	162.2
	−30,	120	120,	−30	−1.44	1.91	12.2	5.67	137.6
	−30,	120	90,	−60	−1.32	2.23	55.6	5.06	160.8
	−60,	90	150,	−30	−1.24	1.77	13.5	5.38	143.0
	−60,	90	60,	30	−1.08	2.09	4.3	4.57	145.9
	−60,	120	60,	30	−1.02	2.00	24.5	4.75	164.5
	−60,	90	120,	0	−0.88	1.50	11.2	4.64	159.8
	−60,	120	60,	0	−0.62	2.07	27.9	4.53	161.5
	−30,	120	120,	−60	−0.58	2.10	26.4	5.78	142.4
	−60,	120	120,	−30	−0.45	2.04	15.6	5.65	141.9
	−30,	90	30,	60	−0.34	1.95	27.6	5.18	129.5
	−30,	120	30,	60	−0.31	2.10	46.7	4.99	153.4
	0,	90	90,	−60	−0.23	1.93	56.1	4.64	135.6
	−60,	90	120,	−30	−0.21	1.97	38.0	4.57	168.0
	−60,	150	60,	0	−0.15	2.19	29.6	5.09	156.9
Type I′	60,	0	90,	30	−0.88	1.75	18.1	5.39	112.5
C_{10}	30,	60	120,	−30	−0.76	1.65	8.9	4.03	173.2
	60,	0	90,	0	−0.34	1.82	22.9	5.62	101.6
	60,	0	150,	−30	−0.04	1.76	14.7	4.32	150.3
Type II′	0,	−90	−120,	30	−0.68	1.60	13.4	4.60	160.9
C_{10}									
$C_7C'_7$	−90,	60	90,	−60	−0.70	5.25	81.3	5.95	112.3
	−90,	60	90,	−30	−0.18	4.37	58.1	5.30	122.4
	−90,	30	90,	−60	−0.37	5.69	75.0	6.80	84.9
MC'_7	−60,	120	90,	−30	−0.91	2.20	40.4	4.85	163.9
	−60,	150	90,	−30	−0.60	2.37	23.5	5.81	140.1
C_7M'	−90,	60	30,	−120	−0.34	5.54	95.1	8.43	151.1

grid of points or to combinations of the "dipeptide" minima (notations recalled in Table 6). The rotational Φ, Ψ parameters of the most stable conformations (limited to 5 kcal/mole above the global minimum) and their energies are listed in Tables 7 and 8 for the LL and LD sequences respectively. (Those for the LG sequence have not been calculated.)

In contrast to the direct results of the grid, the lowest energy is now obtained for a zig-zag structure stabilized by two $1 \rightarrow 2$ hydrogen bonds. Because there are several possibilities of constructing such a conformation (with the peptide

Table 5. Geometrical and energetical caracteristics of (Φ_2, Ψ_2, Φ_3, Ψ_3) points whose energy is less than 2.5 kcal/mole above the global minimum in an LG sequence. The energy of conformation Φ_2, $\Psi_2 = 180°$, 180°; Φ_3, $\Psi_3 = 180°$, 180° is taken as zero. The global minimum of the grid corresponds to conformation Φ_2, $\Psi_2 = 0°$, 90°, Φ_3, $\Psi_3 = 120°$, −30°

Conformation	Φ_2	Ψ_2	Φ_3	Ψ_3	Energy kcal/mole	$H_4 \ldots O_1$ Å	$\widehat{O_1N_4H_4}$ °	$C^\alpha_1 \ldots C^\alpha_4$ Å	$(\overrightarrow{C^\alpha_1 C^\alpha_2}, \overrightarrow{C_3 C_4})$ °
Type I	−30,	−60	−120,	30	−1.85	1.65	8.9	4.03	173.2
C_{10}	−60,	0	−150,	30	−1.35	1.76	14.7	4.32	150.3
	−30,	−60	−120,	60	−0.63	1.99	31.8	4.49	162.0
	−60,	0	− 90,	−30	−0.57	1.75	18.1	5.39	112.3
	−60,	0	− 90,	0	−0.19	1.82	22.9	5.62	101.6
Type II	0,	90	120,	−30	−2.64	1.60	13.4	4.60	160.9
	0,	90	120,	−60	−2.08	1.77	25.6	4.85	163.2
	−60,	90	150,	−30	−1.65	1.77	13.5	5.38	143.0
	−30,	120	120,	−30	−1.44	1.91	12.2	5.67	137.6
	−60,	90	90,	30	−1.05	1.75	19.5	4.46	169.1
	−30,	120	120,	−60	−0.95	2.10	26.4	5.78	142.4
	−60,	90	90,	0	−0.78	1.79	22.3	4.22	164.5
	−60,	120	90,	0	−0.65	1.81	21.9	4.91	159.7
	−60,	120	120,	−30	−0.44	2.04	15.6	5.65	141.9
	−60,	90	120,	0	−0.37	1.50	11.2	4.64	159.8
	−60,	120	90,	−30	−0.31	2.20	40.4	4.85	163.9
	−30,	90	150,	−30	−0.30	1.68	12.2	5.41	138.2
	−60,	90	120,	−30	−0.28	1.97	38.0	4.57	168.0
Type I′	30,	60	120,	−30	−0.76	1.65	8.9	4.03	173.2
C_{10}	60,	0	150,	−30	−0.46	1.76	14.7	4.32	150.3
Type II′	0,	−90	−120,	30	−1.08	1.60	13.4	4.60	160.9
C_{10}	0,	−90	−120,	60	−0.58	1.77	25.6	4.85	163.2
C_7C_7	−90,	60	− 90,	60	−0.30	5.43	79.5	8.50	26.6
MC'_7	−30,	120	90,	−60	−1.10	2.23	55.6	5.06	160.8
$C_7C'_7$	−90,	60	90,	−60	−0.40	5.25	81.3	5.95	112.3
C_7M'	−90,	60	0,	−90	−0.34	5.98	75.1	8.65	20.8
RC_7	−30,	−60	− 90,	60	−0.25	2.28	61.7	4.88	135.0

Table 6. Optimum "dipeptide" conformations

Φ, Ψ values °			Symbol used
−180,	−180	G, L, and D residues	C_5
− 90,	60	G, L, and D residues	C_7
90,	− 30	G, L, and D residues	C'_7
− 30,	− 60	G, L, and D residues	R
30,	60	G, L, and D residues	L
− 30,	120	G, L residues only	M
30,	−120	G, D residues only	M′

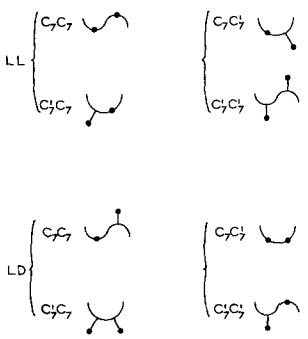

Fig. 10. Schematic representation of the zig-zag conformations stabilized by two hydrogen bonds of the 1→2 type. (●= C^{β} methyl groups)

backbone extended or folded, see Fig. 10), it was verified that the best results are obtained when the C^{β}-methyl groups of residues 2 and 3 are both equatorial with respect to the mean-plane of the molecules.

As concerns the C_{10} hydrogen-bonded ring, the minimization procedure leads to several solutions depending of the starting (Φ, Ψ) values: it seems that the areas of stability of the different types of the C_{10} ring in the $\{\Phi_2, \Psi_2, \Phi_3, \Psi_3\}$ conformational space consist of very flat zones in which small conformational holes of little more stability appear. Nevertheless if we use the Venkatachalam's notation [10] of the bending "U-vector" (which is a visualization of the C_{10} ring on the usual (Φ, Ψ) map by a vector joining the point at $\Phi, \Psi = \Phi_2 \Psi_2$ to the point $\Phi, \Psi = \Phi_3 \Psi_3$ corresponding to the most stable conformation of that ring) we can distinguish roughly two principal orientations of the vector among these minima:

1. for the C_{10}^{II} type of C_{10}, one of these orientations is a deformed ML combination (labelled $\overline{ML}$ in Popov's notations [11]), while the other is a modification of the MC_7' combination ($\overline{MC}_7'$).

2. for the C_{10}^{I} type of C_{10}, one of the orientations results from a deformation of the RC_7 combination ($\overline{RC}_7$), while the other is due to Φ, Ψ values being generally forbidden in most of the (Φ, Ψ) "dipeptide" maps: for this "U-vector" $\Phi \cong -100°$ to $-60°$ with Ψ near 0°.

Moreover one of the C_{10}-like structures leads to conformations close to the 3_{10} helix and is itself of a special type: we have labelled it C_{10}^{III} following Venkatachalam's notations [10].

We obtain also $C_{10}^{I'}$ and $C_{10}^{II'}$ structures for LL sequence, although hard-sphere calculations propose these type of structures only for GG and DD sequences (for type I') or for GL and DL sequences (for type II'). In the LD sequence the $C_{10}^{II'}$ type disappears due to the destabilisation of the R region for Φ_3, Ψ_3. It may be worth stressing also that we obtain an energy equivalence between C_{10}^{I} and C_{10}^{II} in the LL combination, whereas in LD C_{10}^{II} is about 1 kcal/mole more stable than C_{10}^{I}. In both compounds conformations without any hydrogen-bond are destabilized by at least 4 kcal/mole with respect to global minimum.

Table 7. Minimum energy conformations in the range of 5 kcal/mole above the deepest minimum

Conformation	C_7C_7	$C'_7C'_7$	$C'_7C'_7$	C^{I}_{10}	C^{II}_{10}	C^{III}_{10}	C^{I}_{10}	C_5C_7	MC_7	$C_7C'_7$
Φ_2	−74	75	75	− 29	− 12	−51	−61	−180	− 19	−76
Ψ_2	58	−51	−53	− 60	103	−28	− 6	169	109	59
Φ_3	−75	−75	74	−122	121	−54	−86	− 76	− 75	75
Ψ_3	57	55	−50	43	− 41	−30	−16	58	57	−27
Energy	− 3.6	− 3.0	− 1.8	− 1.8	− 1.8	− 1.7	− 1.6	− 1.6	− 1.5	− 1.5

Table 8. Minimum energy conformations in the range of 5 kcal/mole above the deepest minimum

Conformation	$C_7C'_7$	$C'_7C'_7$	C_7C_7	C^{II}_{10}	C^{II}_{10}	C^{II}_{10}	C'_7C_7	MC'_7	C_7M'	$C_5C'_7$
Φ_2	−75	74	−76	− 10	−61	− 46	75	− 40	− 74	179
Ψ_2	55	−52	53	102	89	104	−52	137	57	169
Φ_3	74	74	−76	121	90	48	−75	83	20	75
Ψ_3	−58	−56	54	− 42	19	36	54	− 37	−109	− 58
Energy	− 3.8	− 2.8	− 2.6	− 2.4	− 2.3	− 2.2	− 2.1	− 2.1	− 1.7	− 1.5

2.3 Probability Computations for Optimum Energy Conformations

Starting from the set of $e^{XY}_{ij,kl}$ conformational points, we are able to associate to each of them a statistical weight $W^{XY}_{ij,kl} = \exp(-\text{energy}(e^{XY}_{ij,kl})/RT)$ and a probability $P^{XY}_{ij,kl} = W^{XY}_{ij,kl}/Z^{XY}$ where $Z^{XY} = \sum_{ij,kl} \exp(-\text{energy}(e^{XY}_{ij,kl})/RT)$. From these weights and probabilities associated with conformational points we can calculate roughly the probability associated with each of the areas of energy minimum: $P^{XY}_S = \sum_{ijkl \in S} P^{XY}_{ij,kl}$ where the ij,kl points belong to the area S.

Although this simplified method for the introduction of librational entropy is far from perfect from a rigorous statistical point of view, it has been shown to give satisfactory results on the "dipeptide" model, compared to experimental conclusions [18].

We present in Table 9 the evaluation of the *most probable* conformations in the "tripeptide" models. It is not surprising to find that the C_{10} ring represents the most probable structure because we have already noticed the large conformational area occupied by this structure on the global maps as compared with those of other conformations (especially of the association of C_7 or C'_7 rings).

The results of Table 9 compare very satisfactorily with experimental infrared work carried out on the same compounds by Marraud *et al.* [19].

Table 9. Probabilities of occurence of the principal types of conformations in the XY sequence

Sequence	C^{II}_{10}	C^{I}_{10}	C'_7C_7	C_5	Others
LL	35%	30%	12%	8%	15%
LD	70%	8%	9%	5%	8%
LG	56%	15%	9%	6%	14%

in the LL sequence. The zero in energy is taken as that of the C_5C_5 conformation

C_7M	C_{10}^{I}	RC_7	C_7R	C_7C_5	C'_7M	C_{10}^{II}	$C_5C'_7$	$C_{10}^{II'}$	RC'_7	C_{10}^{II}
− 75	−43	−40	−74	− 76	76	− 55	176	65	−24	−10
59	−53	−56	57	47	− 51	102	173	− 73	−60	75
− 17	−96	−80	−27	−145	− 25	63	75	−106	74	15
107	32	41	−58	160	119	25	− 52	− 20	−53	66
− 1.4	− 1.3	− 1.1	− 1.0	− 0.7	− 0.7	− 0.7	− 0.6	− 0.3	− 0.1	− 0.1

in the LD sequence. The zero in energy is taken as that of the C_5C_5 conformation

C_{10}^{I}	RC'_7	MC'_7	$C_{10}^{I'}$	C'_7M'	C_5C_7	C_7C_5	C_{10}^{III}	C'_7L	C'_7C_5	C_5C_5
− 32	−29	− 20	64	75	−180	− 73	−50	75′	77	−180
− 51	−55	107	− 2	− 50	170	54	−34	−49	− 40	170
−129	74	− 75	96	18	− 76	−180	−49	26	−171	−180
41	−59	54	13	−106	53	−166	−32	60	−160	−171
− 1.2	− 0.8	− 0.7	− 0.7	− 0.5	− 0.5	− 0.2	− 0.2	− 0.1	0.1	0.0

2.4 The Tripeptide as a Model to Select U-turns in Proteins

In the paragraph concerned with the global energy conformational maps, we have defined areas which correspond to stable C_{10} structures. If we limit these areas to points which are more than 1% probable (following conclusions drawn for the "dipeptide" model [18]) we can define with a good approximation contours in which the selection of Φ, Ψ values for residues $i+1$ and $i+2$ should lead to "*U*-turn"-like foldings. This criterium leads to geometrical characteristics for the *U*-turns in peptide chains which are compared with those proposed by others in Table 1. We observed a general agreement although a number of local differences.

We can now apply our conformational criteria to proteins for which sets of Φ, Ψ are given or can be calculated from published coordinates in order to find the "*U*-turns" in their tertiary structures.

The method is the following:

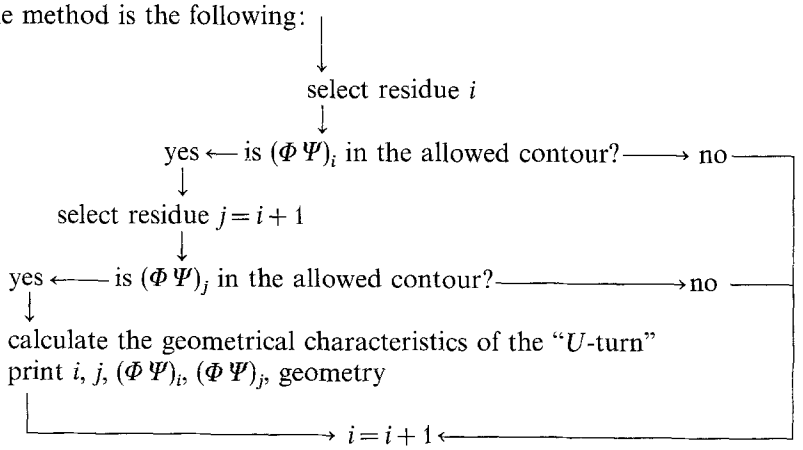

Table 10. Predicted turns in α-chymotrypsin

Ref.	[5]	[20]	[2]	[3]	[6]	[12]	[7]	Our results	From cristallographic analysis
Numbering of residues $i...i+3$	23 26 27 30 48 51 55 58 56 59 61 64 72 75 91 94 95 98 115 118 125 128 131 134 167 170 171 174 174 177 177 180 185 188 191 194 194 197 203 206 217 220 221 224 230 233 231 234	34 40[a] 45 52[a] 57 64[a] 70 79[a] 94 101[a] 111 119[a] 144 150[a]	4 7 16 19 23 26 25 28 27 30 34 40[a] 35 38 48 51 55 58 61 64 67 70 72 75 75 78 91 94 95 98 96 99 99 102 108 111 114 118[a] 115 118 125 128 131 134 169 173[a] 172 175 173 176 177 180 184 188[a] 190 194[a] 191 194 194 197 202 205 202 207[a] 203 207 217 220 221 224 230 233 231 234 232 235	4 7 23 26 27 30 35 38 48 51 56 59 61 64 72 75 91 94 96 99 99 102 108 111 115 118 125 128 131 134 168 171 173 176 177 180 191 194 194 197 203 206 217 220 221 224 231 234	23 26 27 30 35 38 48 51 56 59 61 64 72 75 91 94 96 99 99 102 108 111 115 118 125 128 131 134 152 155 175 180 177 180 185 188 191 194 203 206 217 220 221 224	23 26 27 30 72 75 91 94 99 102 115 118 125 128 131 134 171 175 177 180 191 194 194 197 217 220	23 26 25 28 27 30 35 38 48 51 55 58 56 59 61 64 72 75 91 94 95 98 99 102 108 111 115 118 125 128 131 134 152 155 173 176 177 180 185 189[a] 191 194 194 197 203 206 217 220 221 224 231 234	23 26 27 30 48 51 55 58 56 59 61 64 72 75 91 94 95 98 115 118 125 128 131 134 165 168 166 169 168 171 172 175 173 176 177 180 191 194 194 197 203 206 217 220 221 224 231 234 234 237 235 238 238 241	23 26 27 30 48 51 55 58 56 59 61 64 72 75 91 94 95 98 115 118 125 128 131 134 167 170 168 171 172 175 173 176 177 180 185 188 191 194 194 197 203 206 217 220 221 224 230 233 231 234 234 237
Total of	24	7	38	22	23	13	29	27	26
% of turns predicted with respect to crystallographic analysis [21]	84	—	73	76	69	46	80	88	—

[a] Sequences designed as loops and containing more than $i...i+3$ residues.

After examination of the $\Phi\Psi$ values in 10 proteins (lysozyme, myoglobin, erythrocruorin, carboxypeptidase, chymotrypsin, subtilisin, oxyhemoglobin, ribonuclease *S*, rubredoxin, α lactalbumin) we have obtained 199 "*U*-turns". The mean values for the Φ, Ψ of the $i+1$ and $i+2$ residues are the following:

type II $\quad \Phi_{i+1}, \Psi_{i+1} = 51°, 107° \quad \Phi_{i+2}, \Psi_{i+2} = 98°, -4°$,

type I + III $\quad \Phi_{i+1}, \Psi_{i+1} = -53°, -40° \quad \Phi_{i+2}, \Psi_{i+2} = -74°, -21°$.

As an example we give in Table 10 our predictions compared with others for α-chymotrypsin for which detailed cristallographic data are available [21].

3. Conclusions

Thus a comparison of the backbone dihedral angles of residues $i+1$ and $i+2$ of the stable conformations of the "tripeptide" with the minimum energy backbone conformations of the "dipeptide" model indicates that to a good approximation the former are simply combinations of "dipeptide" minima with the exception of C_{10} rings (which cannot appear in a model consisting of only two peptide bonds). This situation points to a remarquable stability of the "residual" conformational code in peptide systems. Calculations performed on the "tetrapeptide" LLL and "pentapeptide" LLLL models [22] strengthen this idea, as do also recent calculations by Popov [11] and Scheraga [7]. This makes conformational analysis on large peptide systems more reliable. Second, most of the calculations performed recently on LL [10] and LLLL [7] models find the zig-zag model stabilized by two $1 \rightarrow 2$ hydrogen bonds as the most stable structure from an energy point of view, in good agreement with our results. There are differences, however, as concerns the formation of bends in such models. Thus e.g. Scheraga [10] finds that minimum energy bent conformations do not posses an i to $i+3$ backbone hydrogen bond and are the result of RC_7 and C_7L combinations, while we find rather C_{10}-like structures. We do not think that Scheraga's results are due to end effects in the "pentapeptide" used (instead of a "tripeptide" in our calculations) but rather to this author's minimisation procedure which jumps over the C_{10} energy minimum regions to fall into the best ones (here RC_7 and C_7L conformations).

Finally, it seems that our calculations reflect satisfactorily the short- and middle range inter-residual interactions in polypeptides, as visible from the very good predictions obtained for the formation of "*U*-turns" in proteins, compared with other computations. Also, because of the different stabilities obtained for the C_{10} rings in the LL, LD and LG sequences, we may say that the existence of the bends of the "*U*" type is essentially a problem involving the nature of side-chains at residues $i+1$ and $i+2$, even if a potentiality of forming such turns exists at the backbone level.

References

1. Geddes, A.J., Parker, K.D., Atkins, E.D.T., Beighton, E.: J. Mol. Biol. **32**, 343 (1968)
2. Kuntz, I.D.: J. Am. Chem. Soc. **94**, 4009, 8568 (1972)
3. Crawford, J.L., Lipscomb, W.N., Schellman, C.G.: Proc. Natl. Acad. Sci. USA **70**, 538 (1973)
4. Bunting, J.R., Athey, T.W., Cathou, R.E.: Biochim. Biophys. Acta **285**, 60 (1972)

5. Krigbaum, W.R., Knutton, S.P.: Proc. Natl. Acad. Sci. USA **70**, 2809 (1973)
6. Lewis, P.N., Momany, F.A., Scheraga, H.A.: Proc. Natl. Acad. Sci. USA **68**, 2293 (1971)
7. Lewis, P.N., Momany, F.A., Scheraga, H.A.: Biochim. Biophys. Acta **303**, 211 (1973)
8. Mathews, B.W.: Macromol. **5**, 818 (1972)
9. Pullman, B., Maigret, B.: In: Conformation of biological molecules and polymers (E. D. Bergmann and B. Pullman, Eds.), p. 13. New York: Academic Press 1973
10. Venkatachalam, C.M.: Biopolymers **6**, 1425 (1968)
11. Lipkind, G.M., Arkipova, S.F., Popov, E.M.: Molecular Biologiya **4**, 331, 509 (1970)
12. Chandrasekaran, R., Lakshminarayanan, A.V., Pandya, U.V., Ramachandran, G.N.: Biochim. Biophys. Acta **303**, 14 (1973)
13. Kotelchuck, D., Scheraga, H.A., Walter, R.: Proc. Natl. Acad. Sci. USA **69**, 3629 (1972)
14. Pullman, B., Pullman, A.: Advan. Protein Res. **16**, 387 (1974)
15. Scheraga, H.A.: In: Advan. Phys. Org. Chem. **6** (1968)
16. IUPAC. IUPAB Comm. Biochem. Nomencl., Biochemistry **9**, 3471 (1970)
17. Nelder, J., Mead, R.: Comp. J. **8**, 42 (1965)
18. Maigret, B.: Thesis, Paris (1971)
19. Marraud, M., Boussard, G., Neel, J.: Manuscript in preparation
20. Nagano, K.: J. Mol. Biol. **75**, 401 (1973)
21. Birktoft, J.J., Blow, D.M.: J. Mol. Biol. **68**, 187 (1972)
22. Maigret, B., Pullman, B.: Manuscript in preparation

Prof. Dr. B. Pullman
Institut de Biologie Physico-Chimique
Fondation Edmond de Rothschild
13, Rue Pierre et Marie Curie
F-75005 Paris, France