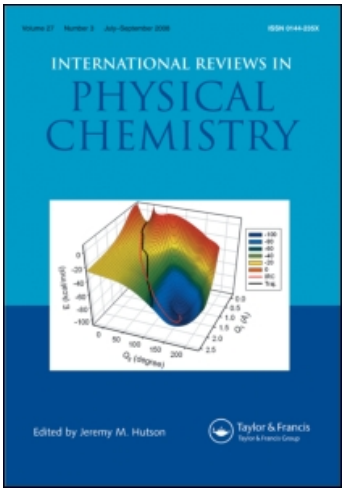


International Reviews in Physical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713724383

Searching for the simplest structural units to describe the three-dimensional structure of proteins

András Perczel[a]; Imre G. Csizmadia[b]
[a] Department of Organic Chemistry, Institute of Chemistry, Eötvös University, Budapest, Hungary [b] Department of Chemistry, University of Toronto, Toronto, Ontario, Canada

To cite this Article Perczel, András and Csizmadia, Imre G.(1995) 'Searching for the simplest structural units to describe the three-dimensional structure of proteins', International Reviews in Physical Chemistry, 14: 1, 127 — 168
To link to this Article: DOI: 10.1080/01442359509353307
URL: http://dx.doi.org/10.1080/01442359509353307

INTERNATIONAL REVIEWS IN PHYSICAL CHEMISTRY, 1995, VOL. 14, NO. 1, 127–168

Searching for the simplest structural units to describe the three-dimensional structure of proteins

by ANDRÁS PERCZEL

Department of Organic Chemistry, Institute of Chemistry, Eötvös University, Budapest, Hungary

and IMRE G. CSIZMADIA

Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A1

Ab initio computations have been carried out during the past several years on diamides of single amino acids (HCO–NHCH*R*–CONH$_2$ where R=H (glycine), –CH$_3$ (alanine), –CH(CH$_3$)$_2$ (valine) and –CH$_2$OH (serine)) exploring all possible backbone and side chain conformations. Selected conformations were studied in our laboratory on threonine (R=CH(CH$_3$)OH), cystein (R=CH$_2$–SH) and phenylalanine (R=CH$_2$–C$_5$H$_6$) diamides. Tri-, tetra-, penta-, hexa- and hepta-amide systems of poly-*L*-alanine (H–(CONH–CHCH$_3$–CONH)$_n$–H $2 \leq n \leq 6$) were also investigated at selected backbone conformations. All these studies confirmed the results of multidimensional conformation analyses: the *i*th amino acid residue in a polypeptide has a maximum of nine (9) discrete backbone conformations. These structures correspond to nine conformational centres on the 2D-Ramachandran map. On the basis of this finding, it can be shown that the folded secondary structure of any protein with known internal coordinates, can be described in terms of these nine discrete conformation types.

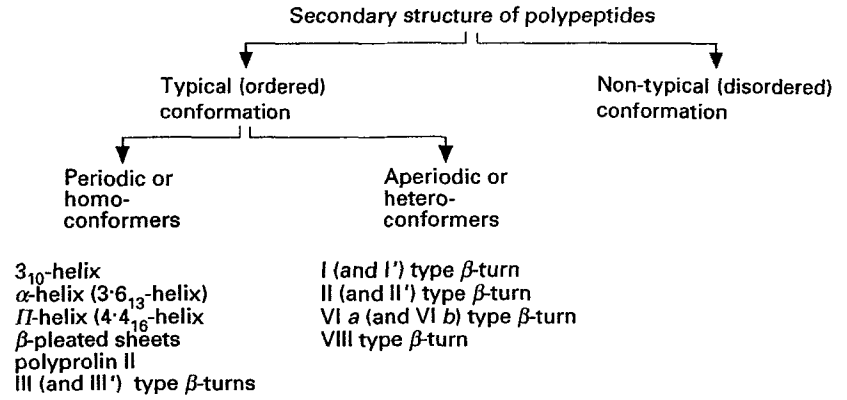
1. Introduction

To find better enzyme inhibitors or more selective drugs and biological ligand molecules, nowadays, gene technology provides an indispensable tool, called 'point mutation', for scientists. These mutation experiments involve the exchange of an amino acid residue with a more suitable candidate. Such a 'replacement' of the side chain functional group at the targeted sequential position, may affect not only the folding of the main chain, but also the conformation of the sequentially neighbouring amino acid residues [1–10]. Even if the individual structural data (bond lengths, angles and torsions), indispensable for quantitative structure determination, are known for the original protein (e.g. from X-ray diffraction experiments), the conformational consequences of such a point mutation are hardly predictable. Although the relative orientation of the backbone and the side-chain atoms determine the global structure of a peptide or a protein, we know very little about the conformers involved in a folding process [11–19].

Traditionally, those sections of a protein main-chain where any type of 'pattern' is observed, are usually called as typical or 'ordered' secondary structures. The remaining portions of a protein are characterized as non-typical or 'unordered' or 'disordered' structures. (For these 'unordered' sequential units the misleading 'random conformation' terminology was often used in the past.)

The ordered conformations themselves can be subdivided into 'periodic' and 'aperiodic' conformations. The most common conformations built up from periodic

0144–235X/95 $10·00



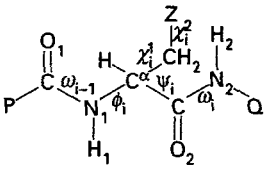
Scheme 1.

Table 1. The nine different β-turn types previously assigned in globular proteins.

β-turn type	Backbone torsional angle values				Conformation type(s)
	ϕ_1	ψ_1	ϕ_2	ψ_2	
I	−60	−30	−90	0	$\alpha_L\alpha_L$, $\alpha_L\gamma_L$, $\alpha_L\delta_L$
I′	60	30	90	0	$\alpha_D\alpha_D$, $\alpha_D\gamma_D$, $\alpha_D\delta_D$
II	−60	120	80	0	$\varepsilon_L\alpha_D$, $\varepsilon_L\gamma_D$, $\varepsilon_L\delta_D$
II′	60	−120	−80	0	$\varepsilon_D\alpha_L$, $\varepsilon_D\gamma_L$, $\varepsilon_D\delta_L$
III	−60	−30	−60	−30	$\alpha_L\alpha_L$
III′	60	30	60	30	$\alpha_D\alpha_D$
VI *a*	−60	120	−90	0	$\varepsilon_L\alpha_L$, $\varepsilon_L\gamma_L$, $\varepsilon_L\delta_L$
VI *b*	−120	120	−60	0	$\beta_L\alpha_L$, $\beta_L\gamma_L$, $\beta_L\delta_L$
VIII	−60	−30	−120	120	$\alpha_L\beta_L$

subunits are the α-helices, β-pleated sheets and the poly-prolin II secondary structures [11]. These units consist of ϕ and ψ values in a monotonically repeated form. In an α-helix $\phi \approx -54°$ and $\psi \approx -45°$ for all amino acid residues within a relatively small torsional angle tolerance. These torsional angle variables are observed around $\phi \approx -150°$ and $\psi \approx +150°$ for β-pleated sheets [20–25]. In a polyprolin II secondary structure incorporating k successive amino acid residues the main chain conformation can be labelled as $[\phi \approx -60°$ and $\psi \approx +120°]_k$. The $(\alpha_L)_n$, $(\beta_L)_n$ and $(\varepsilon_L)_n$ symbols can be introduced for the description of α-helices, β-pleated sheets and poly-*L*-proline II secondary structural elements, respectively, emphasizing the 'homo-conformer' character of these polymers. Most of the hairpin conformations or β-turns consist of two different types of diamide units. Consequently, most of these secondary structures are 'ordered', but 'aperiodic' conformations. For example, to describe accurately the relative orientation of the three consecutive amide groups in the most familiar β-turns [26], the definition of the ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2} and ψ_{i+2} torsional angles is required. (For these backbone torsional angle values, associated with the different β-turn types see table 1.)

In both rigid and non-rigid molecules, composed from n atoms an explicit knowledge of the $3n$-6 internal coordinates is required to determine their structure. Although some of these internal coordinates are systematically monitored during



Scheme 2.

conformation analyses, the calculation of a minimal energy conformation involves the relaxation of all the required $3n$-6 internal coordinates. Investigating the conformational properties of a peptide molecule, most frequently the variation of specific torsional angles (ϕ_i, ψ_i, ω_i, χ_i^1, χ_i^2, etc. (scheme 2) is of a primary interest, where the ϕ and ψ variables describe the main-chain conformation of a polypeptide.

When modelling a larger peptide unit, its backbone conformational energy hypersurface ($E = E[(\mathbf{x})]$) containing k amino acid, (where $\mathbf{x} = (\phi_1, \psi_1, \ldots \phi_k, \psi_k)$), is often subdivided into smaller substances. Considering for example three consecutive amide groups as the 'building unit' of the macromolecule, this approach results in the following partitioning of the overall conformational potential energy function:

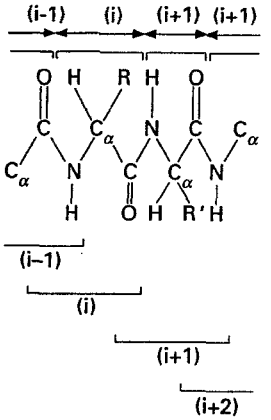
$$E = E\{\phi_1, \psi_1, \ldots, \phi_i, \psi_i, \ldots, \phi_k, \psi_k\} \Rightarrow \begin{cases} E\{\phi_1, \psi_1, \phi_2, \psi_2\} \\ E\{\phi_2, \psi_2, \phi_3, \psi_3\} \\ \quad \vdots \\ E\{\phi_{i-1}, \psi_{i-1}, \phi_i, \psi_i\} \\ E\{\phi_i, \psi_i, \phi_{i+1}, \psi_{i+1}\} \\ \quad \vdots \\ E\{\phi_{k-2}, \psi_{k-2}, \phi_{k-1}, \psi_{k-1}\} \\ E\{\phi_{k-1}, \psi_{k-1}, \phi_k, \psi_k\}. \end{cases} \tag{1}$$

Analysing the architecture of proteins it is reasonable to accept that the partitioning to the smallest structural unit is based on diamide systems which may mimicked by HCO–*Xxx*–NH_2, CH_3CO–*Xxx*–$NHCH_3$, etc. This approach results in the following partitioning of the overall conformational potential energy function:

$$E = E\{\phi_1, \psi_1, \ldots, \phi_i, \psi_i, \ldots, \phi_k, \psi_k\} \Rightarrow \begin{cases} E\{\phi_1, \psi_1\} \\ E\{\phi_2, \psi_2\} \\ \quad \vdots \\ E\{\phi_{i-1}, \psi_{i-1}\} \\ E\{\phi_i, \psi_i\} \\ \quad \vdots \\ E\{\phi_{k-1}, \psi_{k-1}\} \\ E\{\phi_k, \psi_k\}. \end{cases} \tag{2}$$

From the early sixties these 'Ramachandran type' [27] potential energy surfaces

Chemical partitioning of a protein backbone fragment



Conformational partitioning of a protein backbone fragment

Scheme 3.

(PES) ($E = E(\phi_i, \psi_i)$] have continuously formed the basis of peptide conformation analyses. Thus the complicated problem of the protein 3D structure involving $2k$ independent variables has been subdivided to simpler problems: namely, to k potential energy surfaces of two independent variables (2D). In this 2D-Ramachandran PES approach the investigation of the local backbone conformation is isolated from other 'structure influencing' effects [20–25]. In consequence, two types of interactions can be distinguished. (1) The structure modifying effects of the nearest-neighbour residues are separated from local interactions. (2) The interferences between the targeted unit and far laying molecular fragments (also called long-range interactions) are also detached from local factors in the 2D-Ramachandran concept. Although it is hard to define the optimal model size in general, it is clear that for a conformational investigation even in the simplest model outlined above it is rather the consecutive diamide system that should be studied instead of the individual amino acid residues in the sequence.

This strategy of subdividing a $2n$-dimensional space to n two-dimensional subspaces has been implicitly followed by most protein chemists during the past 40 years [11, 20–25]. Nevertheless, if the interest is focused on residue i and $i + 1$, it is recommended that the conformational properties of the adjacent triamide system incorporating both the i and $i + 1$ amino acid residues, should be studied together.

The structural analysis of protein loops or shorter peptides is often impossible even with high field nuclear magnetic resonance (NMR) spectrometers [28–33], since multiple conformers are simultaneously involved in the structure within a time-scale typically faster than resolved by NMR [31–34]. The key of the NMR based structure analyses is a set of interproton distances (nuclear overhauser effect (NOE) data) [42, 43] typically arising from a time averaged structure [38–41]. The assignment and interpretation of the normal mode vibrations recorded by an infrared (IR) spectrometer [44, 45], requires also the knowledge of the individual conformers. Beside the traditional force field approaches [46–52], these structural data of a molecule can now be also computed using *ab initio* methods [53, 54]. Generally, these lengthy but accurate calculations result in far more adequate results than those obtained previously by using parameterized molecular mechanics (MM), [46–48, 50–52] molecular dynamics (MD)

[49] and or semi-empirical molecular computations. Today the structure of any or all of the individual conformers of a peptide with a reasonable size can be determined with a higher degree of confidence from *ab initio* type calculations then ever before. These methods are free from external parameterizations introduced in the past in semi-empirical and empirical methods which became the source of misleading conclusions.

The multidimensional conformational analysis (MDCA) is a qualitative tool [53–55] to find the approximate location of the critical points on a potential energy surface. The full Ramachandran map is ($E = E(\omega, \phi, \psi, \omega')$) where the torsional angles are defined according to IUPAC-IUB convention. Since the torsional values of ω and ω' are typically 180° for *trans* amide bonds (and 0° for *cis* peptides), the conformation energy expression can be simplified ($E_{\omega = \omega' = \text{const.}} = E(\phi, \psi)$) as a function of two torsional variables. It is only the method of MDCA that predicts the existence of nine backbone conformational minima on the simplified, i.e. the 2D-Ramachandran map. Figure 1 shows the occurrence of the nine backbone minima (α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L, ε_D,) in an idealized fashion, one minimum in one segment. The IUPAC-IUB convention recommends to vary both the ϕ and the ψ in between $-180°$ and $+180°$. It is more convenient, however, to identify minima with their conformational region or catchment region by using the $0° \leqslant \phi \leqslant 360°$ and $0° \leqslant \psi \leqslant 360°$ cut, since in such a representation all minima with their complete catchment regions can fall in the same periodical unit [56]. (Figure 1 (*b*) is reported according to the IUPAC-IUB convention, while figure 1 (*a*) is plotted according to a topology oriented mode.)

Each of the nine minima specified on figure 1 are legitimate both in terms of multidimensional conformational analysis as well as on the basis of X-ray analysed structures in larger systems. Typical ϕ and ψ values of the nine different backbone conformation prototypes are reported in table 2. In order to emphasize this, the 3D structure of a total of 78 selected globular proteins were analysed by us using X-ray crystallographical data and the nine different backbone conformational types [56] (α_L, α_D, β_L, δ_L, δ_D, ε_L, ε_D, γ_L, γ_D) were also tabulated against the 20 natural amino acid residues [57]. Their occurrence (*a*) as well as the relative deviations from the expected conformational angle values (*b*) are summarized in table 3. Analysing a total of 11 793 amino acid residues only eight combinations from the 20*9 = 180 amino acid-conformation possibilities were not found. These exceptions were the α_D type conformation by the Ile and Pro, the γ_D by the Pro, Phe, Tyr, Trp and His, and the ε_D by the Trp. On the other hand, all the remaining 172 amino acid-conformation combinations were found, demonstrating that the adoption of the above nine backbone conformation types is not really amino acid specific. In contrast within the conformation analyses of some diamide systems (e.g. For–*L*–Ala–NH_2, For–*L*–Val–NH_2), [56–58] using *ab initio* type calculations as well as different spectroscopical methods. The lack of certain diamide backbone conformation types was also observed. For example, the α_L and ε_L peptide conformations were not observed in the case of alanine diamides, although this amino acid is known as an 'α-helix former' in proteins. The α_L conformation of the alanine diamide is not even a minimum on the Ramachandran type potential energy surface according to *ab initio* calculations; in fact, it is only a point on the side of the energy hill.

Our intention was to resolve these apparent 'contradictions' and also to search those simplest structural units that could form a complete set of conformations applicable to describe the three-dimensional structure of proteins. It will be shown below that regardless of whether the secondary structure of a protein is *periodic* or *aperiodic* [20–23] or if it was previously classified as *typical* or *atypical*, the nine different

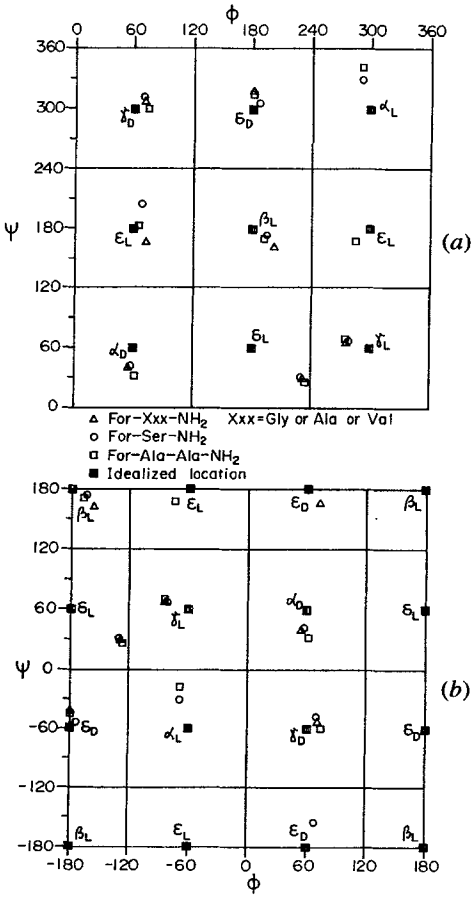


Figure 1. (*a*) Topological representation ($0° \leqslant \phi \leqslant 360°$, $0° \leqslant \psi \leqslant +360°$) of the potential energy surface of an amino acid diamide. Minima specified by their names usually referred to by subscripted Greek letter (α_L, α_D, β_L, δ_L, δ_D, ε_L, ε_D, γ_L and γ_D). The positions of the idealized form (open circles) are supplanted by the position of *ab initio* calculations performed on single amino acid diamides and diamino acid trimides. (*b*) The standard representation ($-180° \leqslant \phi \leqslant 180°$, $-180° \leqslant \psi \leqslant +180°$) of the PES presented in figure 1 (*a*).

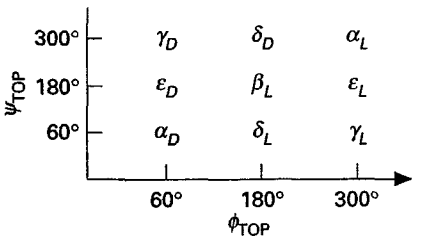
Table 2. Characteristic ϕ and ψ values of the nine conformational centres.

B. B. conf.	ϕ	ψ
α_D	61·8	31·9
α_L	−68·6	−17·5
β_L	−167·6	169·9
γ_L	−84·5	68·7
γ_D	74·3	−59·5
δ_L	−126·2	26·5
δ_D	−179·6	−43·7
ε_L	−74·7	167·8
ε_D	64·7	−178·3

Table 3. The number of the observed nine conformational centres (α_L, α_D, β_L, δ_L, δ_D, ε_L, ε_D, γ_L and γ_D) (*a*) and their deviations (*b*) in selected 78 globular proteins.

(*a*)	α_L	α_D	β_L	δ_L	δ_D	ε_L	ε_D	γ_L	γ_D	Total
GLY	177	277	124	30	30	134	190	16	78	1056
ALA	547	13	110	40	15	203	5	44	7	984
VAL	289	10	159	30	31	229	2	108	2	860
LEU	410	11	84	39	17	221	12	91	5	890
ILE	215	0	87	25	5	132	4	81	1	550
PRO	214	0	2	4	15	288	1	22	0	546
SER	355	19	186	72	30	236	10	53	8	969
THR	256	3	131	62	22	223	8	58	2	765
CYS	112	5	60	21	9	67	2	29	1	306
MET	71	3	29	10	5	37	1	14	1	171
PHE	172	6	78	22	9	97	1	39	0	424
TYR	136	14	94	36	10	130	4	43	0	467
TRP	75	1	33	16	5	47	0	13	0	190
ASP	285	22	42	68	17	118	3	85	11	651
GLU	304	9	47	33	21	111	5	40	3	573
ASN	171	70	53	70	17	96	8	77	5	567
GLN	195	10	65	20	14	105	5	20	1	435
LYS	367	22	84	39	20	144	8	52	6	742
ARG	152	11	61	23	14	80	3	26	4	374
HIS	96	9	52	31	8	45	2	30	0	273
Total	4599	515	1581	691	314	2743	274	941	135	11793

(*b*)	α_L	α_D	β_L	δ_L	δ_D	ε_L	ε_D	γ_L	γ_D	dev.
GLY	24	31	29	37	65	29	50	32	45	38
ALA	18	30	32	34	46	34	71	40	48	39
VAL	20	38	45	37	55	43	39	50	44	41
LEU	20	35	44	33	50	36	51	44	31	38
ILE	23	0	44	33	51	43	57	47	54	44
PRO	17	0	45	29	21	27	70	38	0	36
SER	21	40	35	32	52	33	53	43	44	39
THR	23	45	41	34	58	38	66	48	14	41
CYS	22	35	37	34	52	36	49	48	44	40
MET	21	11	39	42	61	36	51	33	54	39
PHE	22	14	34	31	62	37	49	38	0	36
TYR	23	18	39	32	59	39	64	44	0	40
TRP	20	60	41	22	40	40	0	40	0	38
ASP	22	21	33	32	47	34	66	37	40	37
GLU	20	34	39	33	50	38	56	38	50	40
ASN	22	18	40	30	56	37	56	43	22	36
GLN	21	37	41	32	50	37	68	43	55	43
LYS	21	32	44	33	54	36	52	41	52	40
ARG	20	26	40	32	56	37	48	46	38	38
HIS	21	14	36	35	46	39	37	44	0	34
dev.	21	30	39	33	52	36	55	42	42	



Scheme 4.

backbone conformation types can always describe the backbone conformation of protein building units since these minima are arising from the relative orientation of successive amino acid residues, forming the skeleton of a protein molecule. The most important question for us was the determination of the precise location of these nine minima, schematically shown in Scheme 4.

2. The force field and the *ab initio* concept in peptide conformation analyses

Sophisticated *ab initio* computational methods using a flexible atomic orbital basis sets can reproduce many features of the conformational potential energy hypersurface (PEHS) to practically any desired degree of accuracy. Such computations, with the aid of gradient-optimization techniques, may be used to determine the location (molecular geometry) of the critical points such as conformational energy minima on the conformational PEHS. Both the total and relative energies are generated by these computations, but one can also determine the shape (i.e. the steepness or shallowness) of the PEHS in the vicinity of the critical points (minima) from the computed force constants (or fundamental vibrational frequencies). Such computations are feasible for relatively small molecules. The limitations of the method, have changed with calendar time, but by now, at the middle of 1990's fairly accurate computations are possible to be made on alanine diamide, HCO–NH–$CHCH_3$–$CONH_2$ ($C_4H_8N_2O_2$). On the other hand, the accuracy has also been gradually reduced with the increased number of amino acid residues in an oligopeptde (HCO–[NH–$CHCH_3$–CO]$_n$–NH_2 [$C_{3n+1}H_{5n+3}N_{n+1}O_{n+1}$]).

The same kind of information may also be obtained for relatively small molecules from accurate experimental observations. Molecular geometry can be obtained from rotational spectroscopy [59–67], electron or neutron diffraction as well as from X-ray crystallography. Thermodynamic stabilities may come from enthalpies of formation. Finally, the steepness or shallowness of the PEHS can be obtained from vibrational spectroscopy. These accumulated experimental energetic and conformational data are built in a suitable force field (FF) [46–52] program as external parameters. In such a way knowledge accumulated for small molecules is 'extrapolated' into larger systems. In this context data obtained from small organic compounds considered as building blocks of a macromolecule, are in fact regarded to be transferable to larger systems. Our philosophy is close to such an approach! Thus the determination of the conformational properties of diamide and triamide systems may result in an applicable knowledge to describe the conformation of a longer peptide. The increased full speed and storage capacity of personal computers and work stations made the routine application of a quantum chemical program possible even for shorter oligopeptides. Although severe limitations concerning the size of the molecule as well as the applied basis set for self-consistent field (SCF) type computations are still existing, the *ab initio* concept has already been applicable for di- and triamide systems [68–78].

An intrinsic problem of *ab initio* computations is to find a small but relatively accurate basis set yielding a reasonably good result even without the inclusion of electron correlation. An attempt has been made by us to monitor the fluctuation of the ϕ, ψ values in the γ_L and the β_L backbone orientations of the For–*L*–Ala–NH_2 molecule as a function of the applied *ab initio* method [79]. For SCF type geometry optimizations a total of 11 basis sets were used: 3–21G, 4–21G, 4-31G, 6–31G, 6–311G, 6–31 + + G, 6–31G**, 6–31 + + G**, 6–311G**, 6–311 + + G and 6–311 + + G**. The accuracy of the computed wavefunction was increased by geometry optimization carried out with the inclusion of electron correlation using a second ordered perturbation method (MP2) with the same 11 basis sets. Note that the symbol for the basis set 6–311 + + G**, i.e. 6–311 + + G(d, p) implies a minimal core representation with the aid of 6 s-type Gaussians, and 311 implies that 5 s and p type Gaussians are contracted in the ratio of 3 : 1 : 1 to a triple-zeta quality representation of the valence electron shell. The + + sign indicates that diffuse Gaussians were added to all heavy atoms (C, N, O) as well as to all hydrogens. The ** superscript specifies that d-type polarization functions for the heavy atoms (C, N) and p-type polarization functions for the hydrogen atoms were included in the basis set. These polarizations functions are necessary for the computation of accurate torsional angles. Figure 2 reveals that the backbone torsional angles vary depending on the level of 'accuracy', but the alteration is smaller than $\pm 11°$. The variation of the relative energy with the accuracy of the calculation is shown in figure 3. Not considering the relative energy differences obtained with an MP2 calculation on an unreasonably small basis set (3–21G and 4–21G), the relative energy differences of the two conformers fluctuated around $0{\cdot}7 \pm 0{\cdot}6$ kcal mol^{-1}. Surprisingly, the HF/3–21G energy difference was close to data obtained from MP2/6–31 + + G** and MP2/6–311 + + G** calculations. This, nevertheless, is nothing more than fortuitous cancellation of errors of course, justifying our *a priori* choosing the 3–21G basis set as suitable for studying oligopeptides: (HCO–(NH–$CHCH_3$–CO)$_n$–NH_2).

3. The stability of the α_L backbone conformation in function of its molecular environment

According to selected crystallographical data amino acid diamides adopt minimal energy conformations close to β_L, γ_L and δ_L. These three energetically low lying minima are close to each other in terms of ϕ, ψ torsional angle values. Based on *ab initio* calculations on diamide systems the β_L and δ_L minima are shifted on the PES away from their idealized locations towards the γ_L conformation [56, 75, 76, 80]. Due to this shift all the three minima are located in a 'common region', sometimes referred to as the 'grand canyon' of the Ramachandran PES (figure 4). In solutions the presence of a single conformer for amino acid diamides has not been confirmed by NMR, by circular dichroism (CD) or infrared (IR) spectroscopy either. In contrast, a conformational mixture of an 'undefined' number of structures is typically present, which is usually classified as 'random' [30, 31]. The lack of any solution state evidence supporting the stability of the α_L and the ε_L backbone conformations in the case of diamide systems initiated the investigations [56, 72, 75, 76, 80] to determine, whether such a main-chain orientation is intrinsically stable or unstable. The lack of the above two minima (α_L and ε_L) in the simplest peptide model (*P*–CONH–CH*R*–CONH–*Q*) has had far reaching consequences. This may have led scientists to the conclusion that polypeptide backbone conformation simply cannot be mimicked by diamide models at all. On top of that, this would have meant that from a conformational point of view, proteins cannot be regarded

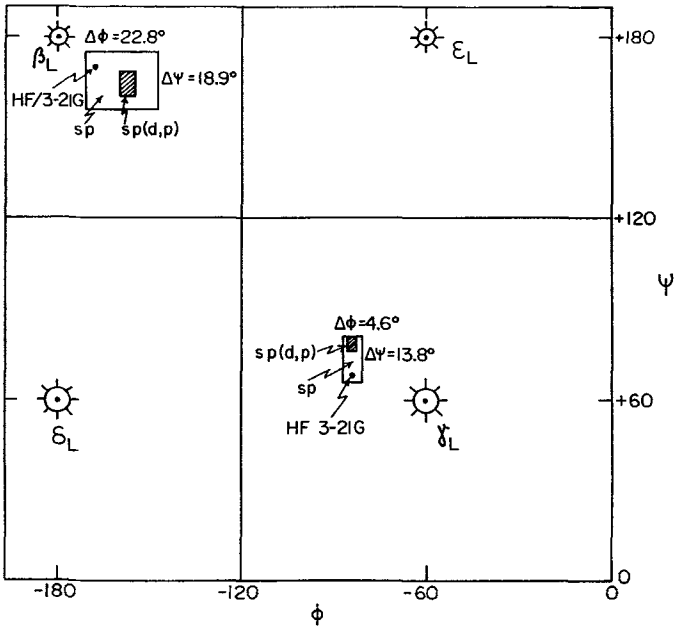


Figure 2. Domains of the ϕ, ψ torsional angles computed for the γ_L and β_L conformers of the For–*L*–Ala–NH_2 molecule, using sp and sp(d, p) basis sets. The ϕ, ψ torsional angle pairs obtained at the level of HF/3–21G are marked by solid dots to show the deviation from the more accurate sp(d, p) results.

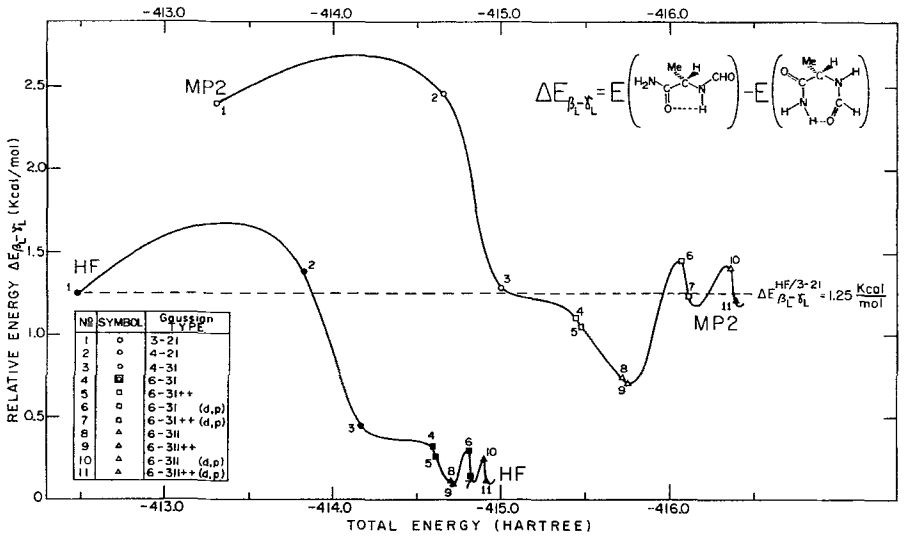


Figure 3. The relative energy differences of the γ_L and β_L conformers of the For–*L*–Ala–NH_2 molecule in function of the accuracy of the applied *ab initio* method measured by the total energy of the γ_L conformer.

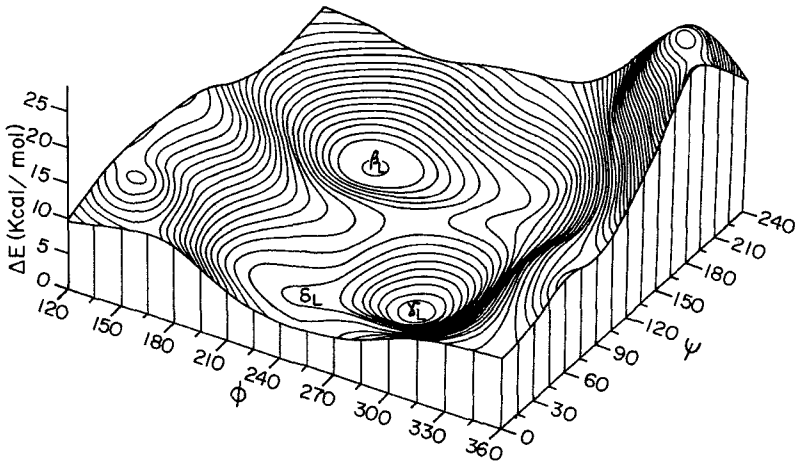


Figure 4. The β_L, γ_L and δ_L regions of the potential energy surface associated with the For–*L*–Ala–NH_2 molecule. This section was called the 'Grand Canyon' region of the 2D-Ramachandran map (HF/3–21G).

as simply the polymers of the –CONH–CH*R*–CONH– systems, but must be built up from longer substructures.

3.1. *The α_L backbone conformation in diamide systems*

In agreement with qualitative multidimensional conformational analyses (MDCA) and with different force field computational methods a total of nine different backbone conformations are expected on a 2D-Ramachandran type potential energy surface (PES). The MDCA results are illustrated schematically in figure 5, showing two full cycles of rotation (from $-360°$ to $+360°$). However, *ab initio* computations performed on small peptides, such as the For–*L*–Ala–NH_2 or the Ac–*L*–Ala–NHMe show that, the right-handed helical conformation, denoted here as α_L, is not a minimum energy conformation at the HF/3–21G level of theory (scheme 5). The Ramachandran map (figure 6) is presented for one cycle of rotation (from 0° to 360°) in a pseudo three-dimensional representation [56, 57]. The energy contour diagram of the 2D-Ramachandran map computed at the HF/3–21G level of theory is given in figure 7. It is clear from this figure that there is a principal mountain ridge along the conrotary mode of motion. The shape of this diagonal mountain ridge, also depicted in figure 8, clearly influences the existence of its four neighbouring minima: α_L, α_D, ε_L and ε_D. In the case of the achiral glycine diamide the mountain ridge is symmetrical and thus all four minima are absent. The chiral structure makes the PES chiral too, which implies that only two of these four minima are annihilated in each of the enantiomers. In the case of For–*L*–Ala–NH_2 the α_L and ε_L conformations are missing, and in the case of For–*D*–Ala–NH_2 the α_D and the ε_D conformations are annihilated as shown by scheme 6 and discussed previously [56].

A detailed search on the 2D-Ramachandran type potential energy surface associated with single amino acid diamides revealed a mountain side region where the α_L conformation is expected to be located without finding even a shallow minimum [80] (figure 9). In contrast, the classical force field (FF) programs do result in the α_L type backbone orientation ($\phi \approx -60 \pm 30°$ and $\psi \approx -30 \pm 30°$) even for a diamide system such as the For–*L*–Ala–NH_2 molecule. This clearly illustrates that the FF approaches

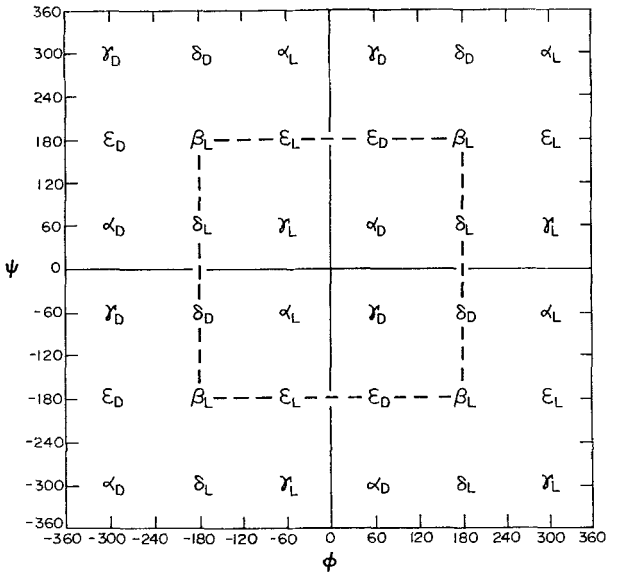
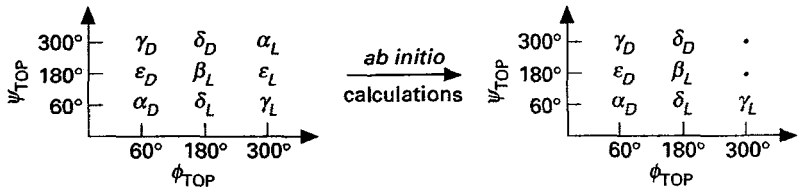


Figure 5. Schematic illustration of the results of multidimensional conformation analyses (MDCA) for a P–CONH–CH*R*–CONH–*Q* type molecule, showing two full cycles of rotation (from $-360°$ to $+360°$).



Scheme 5.

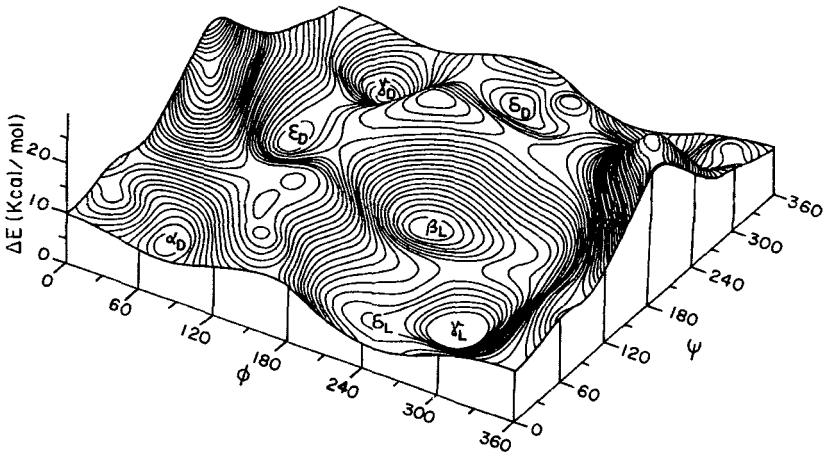


Figure 6. The 2D-Ramachandran type potential energy surface (PES) (one full cyle of rotation is reported (from $0°$ to $+360°$)), associated with For–*L*–Ala–NH_2. (Computed by HF/3–21G.) Note that the α_L and the ε_L conformations are missing from the nine expected minima.

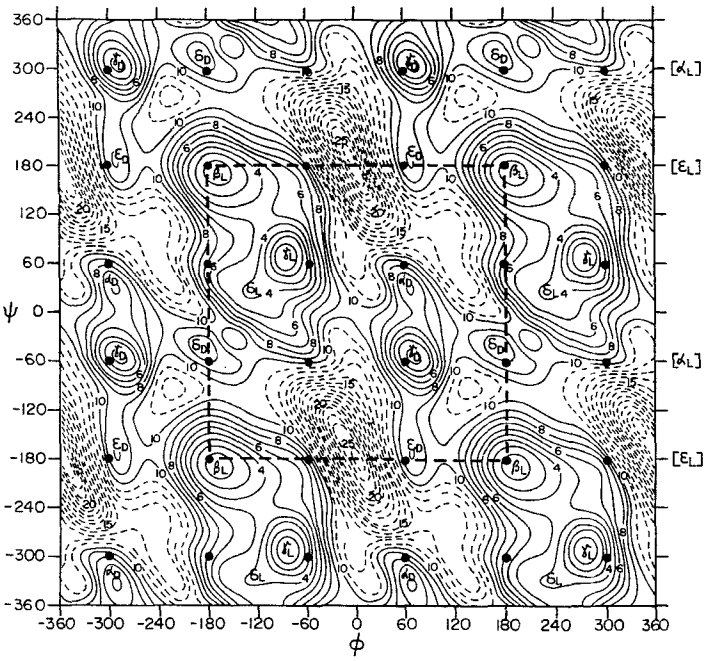


Figure 7. Energy contour diagrams (HF/3–21G) of the 2D-Ramachandran map, showing two cycles of rotation (from $-360°$ to $+360°$) of For–*L*–Ala–NH_2. Contour lines up to 10 kcal mole^{-1} are solid lines, above 10 kcal mole^{-1} there are broken lines. Contour lines are drawn with 1 kcal mole^{-1} increments from 0 to 25 kcal mole^{-1}.

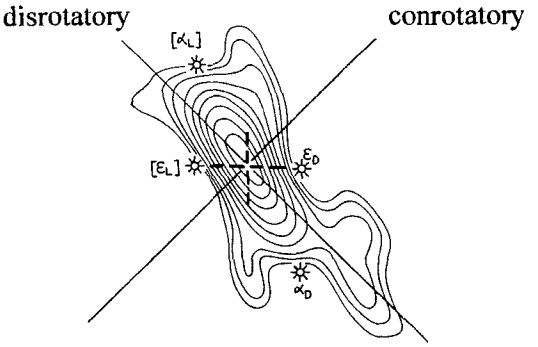
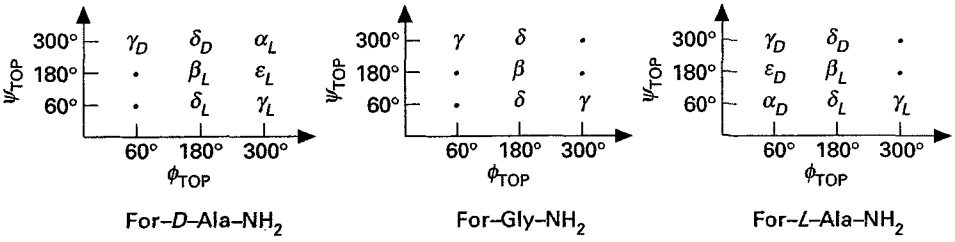


Figure 8. Principal mountain ridge of the 2D-Ramachandran map of For–*L*–Ala–NH_2 oriented along the disrotatory mode of motion: (ϕ–ψ). Note that the existence of the four minima (α_L, α_D, ε_L and ε_D) depends on the actual shape of this diagonal mountain ridge.

are biased toward the description of protein conformations and therefore they cannot describe faithfully the conformations of small peptides.

First Schäfer and coworkers [72], when mapping low-energy pathways on the $E(\phi,\psi)$ of Ac–*L*–Ala–NHMe, found that no minimum exists in the neighbourhood of $\phi = -54°$, $\psi = -45°$, contrary to the 'normal expectation'. Recently Head Gordon *et al.* [75] performed a mapping of the above potential energy surface and counted seven minima on the total conformational area $-180° \leqslant \phi \leqslant +180°$,



Scheme 6.

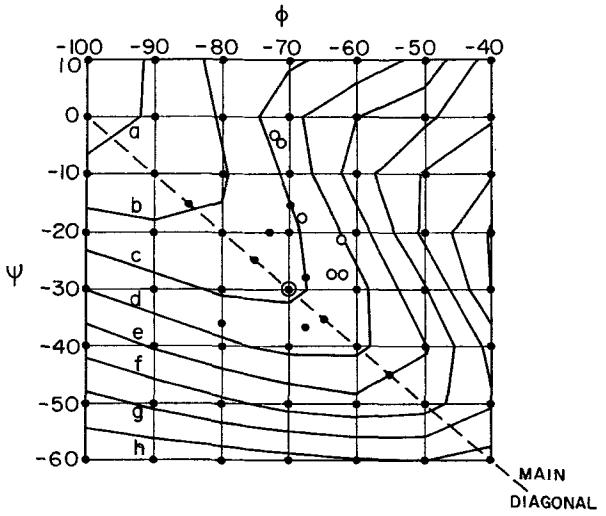
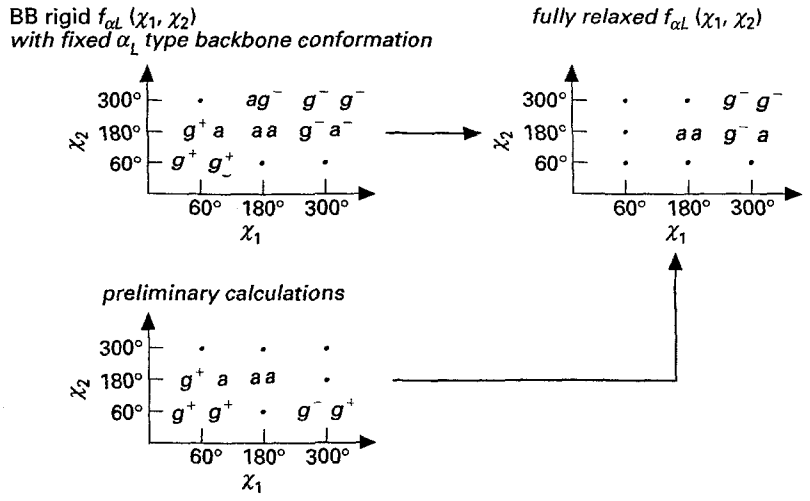


Figure 9. The α_L region of the 2D-Ramachandran type potential energy surface (PES) calculated by *ab initio* methods for the For–*L*–Ala–NH_2 molecule.

$-180° \leqslant \psi \leqslant +180°$. Here again the α_L (and the ε_L HF/3–21G) minima were not among them [56]. The loss of these minima may be related with the destabilizing interaction between these two adjacent amide groups at such relative orientations. The annihilation of the α_L (and ε_L) minima were also observed for For–*L*–Val–NH_2 [58] at any investigated iso-propyl side-chain conformations. None of the analysed χ_1 (N–C^{α}–C^{β}–C^{γ}) torsional angle values resulted in the stabilization of the two lacking backbone minima discussed above. A preliminary optimization of the 'helix-like' backbone conformation (α_L) of the For–Gly–NH_2 molecule (the α_L is identical to the α_D in the case of the achiral glycine residue) was first identified as a minimal energy conformation [56], but careful optimization [80] revealed that it is not really a minimum. Therefore the smaller –H and $-CH_3$ as well as the larger $-CH(CH_3)_2$ apolar side chain types resulted in the annihilation of the legitimate α_L backbone conformation. Thus, the most likely hydrocarbon type side chains cannot stabilize the intrinsically unstable α_L backbone conformation of simple diamides of amino acids.

On the other hand, a specific backbone/side-chain effect in the case of the polar side chain containing amino acid (such as For–*L*–Ser–NH_2) may be expected to provide a favourable interaction and could stabilize the α_L orientation [81, 82]. Based on Scheraga's molecular mechanic structures our preliminary SCF investigation [81 *a*] suggested that the α_L backbone minima could be present in the case of the



Scheme 7.

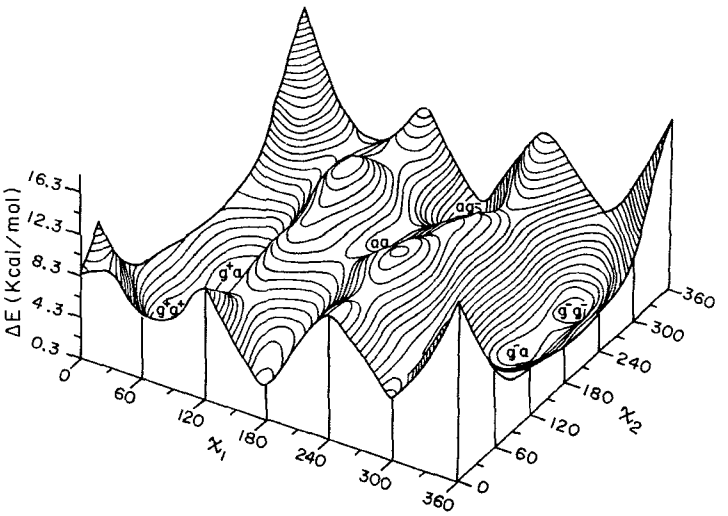


Figure 10. The rigid side-chain potential energy surface associated with an α_L type backbone orientation ($f_{\alpha L}(\chi_1, \chi_2)$) of the For–*L*–Ser–$NH_2$ molecule determined by the *ab initio* method. (The 3–21G basis set was applied.)

For–*L*–Ser–Nh_2 molecule. The SCF total energy of the previously found α_L related backbone structures was high compared to the appropriate γ_L and/or β_L structures. We decided to reinvestigate the stability of the α_L-type geometries of For–*L*–Ser–NH_2 with all possible nine side chain conformations (scheme 7) applying more severe convergence criteria [81 *b*].

The stability investigation of the For–*L*–Ser–NH_2 conformers using *ab initio* methods, associated with an α_L type backbone orientation was performed using the $\phi = -60°$, $\psi = -40°$ cross-section of the $E = E(\phi, \psi, \chi_1, \chi_2)$ 'hypercube'. The rigid SCF $f_{\alpha_L}(\chi_1, \chi_2)$ (see figure 10.) revealed five (5) maxima, eleven (11) saddle point as well

as six (6) minima ($[g^+g^+]$, $[g^+a]$, $[ag^-]$, $[aa]$, $[g^-a]$ and $[g^-g^-]$). When the backbone conformation of these six minima were also totally relaxed, three (3) from the six (6) of them migrated to another backbone conformation type.

It is therefore not so surprising that using less rigorous convergence criteria the $[g^+a]$, $[g^+g^+]$ and $[g^-g^+]$ points were previously mistaken for minima, since these conformations appear even on a rigid surface on the other hand, the remaining three ($[aa]$, $[g^-a]$ and $\{g^-g^-\}$) α_L type backbone conformations are true minima (figure 11 (*a–c*)) not only because the gradient length is lower than 10^{-6} a.u., but also because no negative force constants could be found. These structures are unique examples that even a simple amino acid diamide may adopt the 'conformational monomer' of the right handed helical structure i.e. the α_L. The backbone conformation angles as well as selected interatomic distances are reported in table 4. Two forms, the $\alpha_L[g^-g^-]$ and $\alpha_L[g^-a]$ out of the three cases ($\phi \approx -60°$, $\chi_1 \approx -50°$ orientation) make possible the formation of a weak hydrogen bond, where the side chain oxygen adopts the amid proton of the serine residue. (Such a backbone/side-chain interaction served as an example observed in the solid state structure analyses of the $(CH_3)_3CO$–Pro–Ser–$NHCH_3$ [15, 16] and other molecules [18] and was presumed to exist in aprotic solvents on the basis of the IR spectroscopical data.) On the other hand, in the case of the third α_L minimum associated with an $[a, a]$ type side chain conformation, no sign of any hydrogen bonding (Bronsted complex) interaction could be observed. The side chain is 'placed' in between the two amide groups.

Mayer type bond-order calculations were carried out on structure $[g^-g^-]$, $[g^-a]$ and $[aa]$. Structures $[g^-g^-]$ and $[g^-a]$ showed that the intramolecular hydrogen bond was 4·1% of a single bond in both cases; they differed from each other only in the second decimal places. This is a typical value for intramolecular hydrogen bonds in peptides [82]. In contrast to that we found in structure $[aa]$ that the extent of the interaction between the side chain oxygen and the carbon of the $-CONH_2$ group was 1·3% of a single bond. This value is about one third of the extent of a hydrogen bond (typically 4·0%), nevertheless this is significant, since the Mayer bond order for non-interacting atoms is usually less than 0·001%. On the basis of this we concluded that the stabilization of structure $[aa]$ occurs via intramolecular Lewis type (charge transfer) complex formation.

Since three (3) from the six (6) minima assigned on the rigid $f_{\phi=-60,\psi=-40}(\chi_1, \chi_2)$ map migrated when fully relaxed, one could think that these minima ($[aa]$, $[g^-a]$ and $[g^-g^-]$) are the results of the selected basis set type [3–21G]. Therefore, and because a function of the number of Gaussians is increasing with the basis set, these may vanish. Fortunately, a systematic increase of the basis set size from (3–21G via 4–21G to 6–31G*) has shown only a minor influence on the ϕ, ψ, χ_1 and $\chi_2(\approx \pm 10°)$ on the precise location of these minima and no case was observed where a minimum was annihilated.

Investigating the conformational properties of the two closest side-chain types (–CH[CH_3]–OH and $-CH_2$–SH) in the hydroxy methyl group of the serine residue, surprisingly two different types of results were obtained. The 'extra' methyl (Thr) group located at the β carbon atom of the side chain does not influence the relative orientation of the main chain at an α_L conformation. Minimizations performed on the For–*L*–Thr–NH_2 molecule at $\alpha_L[g^-g^-]$, $\alpha_L[g^-a]$ and $\alpha_L[aa]$ conformations resulted in almost identical three structures as obtained for the For–*L*–Ser–NH_2 (table 4). On the other hand, the same type of minimization of the For–*L*–Cys–NH_2 structures

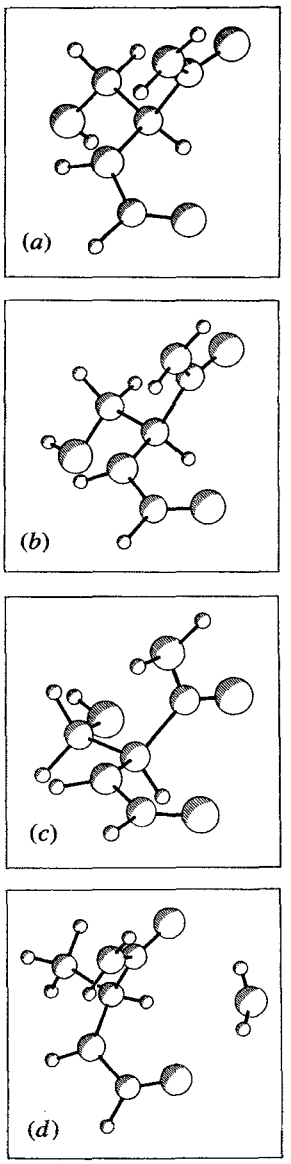


Figure 11. Relaxed backbone potential energy surface $f(\phi, \psi)$ sections associated with an α_L type backbone orientation of the For–*L*–Ser–NH_2 molecule determined by the *ab initio* method. (The 3–21G basis set was applied.) (*a*) Minimum around the ($\phi = -76{\cdot}8°$, $\psi = -18{\cdot}7°$, $\chi_1 = -52{\cdot}0°$ and $\chi_2 = -74{\cdot}6°$) point labelled as $\alpha_L[g-g-]$ conformer. (*b*) Minimum around the ($\phi = -76{\cdot}4°$, $\psi = -16{\cdot}8°$, $\chi_1 = -52{\cdot}3°$ and $\chi_2 = -176{\cdot}8°$) point labelled as $\alpha_L[g-a]$ conformer. (*c*) Minimum around the ($\phi = -69{\cdot}1°$, $\psi = -39{\cdot}9°$, $\chi_1 - 173{\cdot}3°$ and $\chi_2 = -165{\cdot}5°$) point labelled as $\alpha_L[a\,a]$ conformer. (*d*) the *ab initio* optimized structure of the For–*L*–Ala–$NH_2.H_2O$ molecular complex. Two intermolecular hydrogen bonds ($d_{CO\ldots HOH} = 2{\cdot}06$ Å and $d_{CO\ldots HOH} = 2{\cdot}09$ Å) are present, resulting in $\phi = -69{\cdot}8°$, $\psi = -36{\cdot}8°$ values.

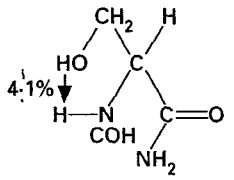
Table 4. Optimized *ab initio* structures of For–*L*–Ser–NH_2 For–*L*–Thr–NH_2 at α_L-type backbone conformations, using a standard 3–21G basis set.

	For–*L*–Ser–NH_2			For–*L*–Thr–NH_2		
INI.[a]	$\alpha_L[g^-g^-]$	$\alpha_L[g^-a]$	$\alpha_L[aa]$	$\alpha_L[g^-g^-]$	$\alpha_L[g^-a]$	$\alpha_L[aa]$
CONV.[b]	$\alpha_L[g^-g^-]$	$\alpha_L[g^-a]$	$\alpha_L[aa]$	$\alpha_L[g^-g^-]$	$\alpha_L[g^-a]$	$\alpha_L[aa]$
ω^1	−171·1	−169·4	−174·6	−171·5	−169·9	−174·2
ϕ^1	−72·0	−70·5	−62·4	−70·8	−69·3	−69·3
ψ^1	−23·7	−24·9	−42·8	−26·5	−27·8	−32·0
ω^2	+157·0	−179·8	−179·7	−179·3	−179·8	−177·5
χ^1	−45·6	−47·5	−179·9	−37·5	−40·3	−171·5
χ^2	−77·5	−174·5	−168·7	−83·7	−172·8	−172·1
MAX. FORCE	<1E−6	<1E−6	<1E−6	<1E−6	<1E−6	<1E−6

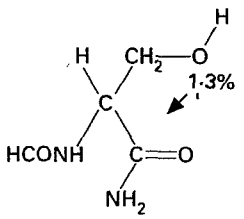
Torsion angles (ω, ϕ, ψ, χ^1 and χ^2) in degrees according to IUPAC–IUB, distances in Angstrom, forces in au.

[a] INI: Initial backbone conformation (Calculated by ECEPP/2).

[b] CONV: Converged backbone conformation.



Stabilization *by* an intramolecular H-bonded complexation (hydrogen bond is 4·1% of a single bond in $[g^-g^-]$ and $[g^-a]$.



Stabilization *without* an intramolecular H-bonded complexation (charge transfer bond is 1·3% of a single bond in [*aa*]).

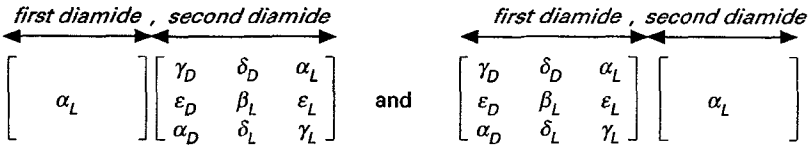
Scheme 8.

ended in totally different geometries. The following shifts were observed: $\alpha_L[g^-g^-] \rightarrow \gamma_L[ag^+]$, and $\alpha_L[g^-a] \rightarrow \delta_L[g^-a]$ as well as the $\alpha_L[aa] \rightarrow \delta_L[g^-g^-]$.

These data obtained for For–*L*–Ser–NH_2 and For–*L*–Thr–NH_2 support the idea that for a polar side chain containing amino acid the intrinsically unstable conformational monomer of a 'helix-like' structure can be stabilized even before this structure gets stabilized through the repetitive $(i, i+3)$ or $(i, i+4)$ type backbone/backbone H-bond network system characteristic for helices. On the other hand, *ab initio* data revealed the crucial role of the $(i, i+3)$ type H-bonds if the main chain conformation is α_L (i.e. $-[\alpha_L]_n-$), for amino acids with no side chain (Gly) or with an apolar one (e.g. Ala or Val) as observed for the For–*L*–Ala–*L*–Ala–NH_2 triamide or for longer For–$(Ala)_n$–NH_2 oligopeptides.

3.2. *The α_L backbone conformation in triamide systems*

Analysing the main chain folding possibilities of the For–*L*–Ala–*L*–Ala–NH_2 [83, 84] molecule, where only one subunit has an α_L-type backbone structure, eight different $\alpha_L x_L$ or $\alpha_L x_D$-type formations with an additional eight $x_L\alpha_L$ or $x_D\alpha_L$ backbone conformations may exist (scheme 9). Beside these 16 theoretically predicted structures,



Scheme 9.

the $\alpha_L\alpha_L$-type relative orientation of the two successive amino acid residues may also be possible.

Analysing these 17 different relaxed conformations of the Ac–Ala–Ala–$NHCH_3$ molecule, by using a force field approach, only six (6) from the 17 backbone conformations discussed above incorporate some sort of a favourable backbone/backbone interaction. Four (4) of these six (6) possible geometries incorporate a seven member ring (C_7 type) H-bond ($\alpha_L\gamma_L$, $\alpha_L\gamma_D$, $\gamma_L\alpha_L$, and $\gamma_D\alpha_L$). The remaining two promising candidate structures, the $\alpha_L\alpha_L$, and $\alpha_L\delta_L$ contain a ten member ($1 \leftarrow 4$ type) H-bond†. According to the *ab initio* computations only the $\alpha_L\delta_L$ structure was found to be minimal energy conformation incorporating a stabilizing intramolecular Bronsted complex; a ten member ($1 \leftarrow 4$ type) H-bond. This led to the conclusion that even an amino acid with a polar side-chain (e.g. $-CH_3$) may adopt an α_L sub-conformation in a triamide system if stabilized by a favourable 'intra-backbone' hydrogen bond. The $\alpha_L\delta_L$ conformation ($\phi_1 = -68{\cdot}6°$, $\psi_1 = -17{\cdot}5°$, $\phi_2 = -113{\cdot}1°$ and $\psi_2 = +21{\cdot}3°$) is the so-called type I β-turn. This $\alpha_L\delta_L$ structure was the first reported [84 *a*] *ab initio* conformation incorporating the α_L 'sub-unit' for an amino acid with an apolar type side chain. Although the α_L backbone conformation is not a minimum on the 2D-Ramachandran PES associated with a diamide system of an apolar amino acid, it can be assigned as a sub-conformation in a triamide system, since the $\alpha_L\delta_L$ conformation is an energy minimum on the 4D-Ramachandran PEHS.

3.3. *The α_L backbone conformation in a bimolecular complex*

Beside the favourable backbone/side-chain (e.g. For–*L*–Ser–NH_2) and backbone/backbone effect (e.g. the stabilizing intramolecular hydrogen bond observed in the $\alpha_L\delta_L$ conformation of the For–*L*–Ala–*L*–Ala–NH_2 molecule), the third type of stabilizing phenomena may originate from an intermolecular complex formation. The simplest model is an alanine diamide with an optimally positioned water molecule. In fact the *ab initio* optimization of the For–*L*–Ala–$NH_2.H_2O$ molecular complex resulted in an α_L minimum [85], where two intermolecular hydrogen bonds ($d_{CO\ldots HOH} = 2{\cdot}06$ Å and $d_{CO\ldots HOH} = 2{\cdot}09$ Å) tighten the backbone of the diamide at $\phi = -69{\cdot}8°$, $\psi = -36{\cdot}8°$ values (figure 11 (*d*)).

4. The stability of the ε_L backbone conformation as a function of its molecular environment

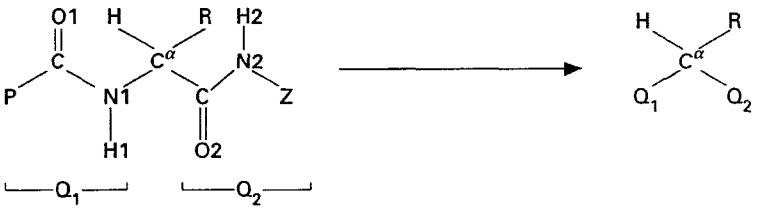
The ε_L subconformation which is the other 'missing' backbone conformational type

† The $\alpha_L\alpha_L$ conformation can be a fragment of a helical segment in proteins or when isolated from similar secondary structures, it may be called type III β-turn. the $\alpha_L\delta_L$ conformation is the so called type I β-turn. According to the original definitions of the type I and the type III β-turn backbone conformations, any backbone torsional angle of these two structures differs less than 30° (see table 1), therefore the type III β-turn is often regarded as a specific case of the type I β-turn.

[56, 75, 76, 80] has not been located up until now in amino acid diamides. *Ab initio* calculations showed that the unfavourable eclipsed H–N–C^{α}–C^{β} orientation ($\approx -20°$) in the ε_L backbone conformation results in the annihilation of such a backbone geometry in For–Gly–NH_2, For–Ala–NH_2, For–Val–NH_2 and probably in the other amino acid diamides as well [56, 58, 75, 80]. In contrast, increasing the peptide chain length with an additional amino acid residue resulted in the stabilization of the ε_L substructure in three conformations ($\gamma_L\varepsilon_L$, $\gamma_D\varepsilon_L$ and $\delta_D\varepsilon_L$). While $\delta_D\varepsilon_L$ and $\gamma_D\varepsilon_L$ geometries are perfect β-turns [84], in accordance with both the distance and the angularity criteria [83], the $\gamma_L\varepsilon_L$ has an extended backbone conformation. Although these are only three conformations from the 17 possibilities incorporating the ε_L subunit, they represent solid evidence that a 'favourable' backbone/backbone interaction may 're-stabilize' the 'annihilated' ε_L minimum. Note that there is no direct influence (like an intramolecular hydrogen bond) between the third and the second amide groups of the molecule, oriented in an ε_L conformation. We are not certain about the nature of the interaction that could stabilize the ε_L conformation substructure in triamide systems however, it is conceivable that the $(\delta+)C\equiv O(\delta-)$ polarity of the carbonyl group in the central amide moiety may have some direct or indirect stabilizing effect. The increased polarity of the central carbonyl group may have a stabilizing influence on the 'second half' of the molecule with an ε_L conformation, yielding the stable $\gamma_L\varepsilon_L$, $\gamma_D\varepsilon_L$ and $\delta_D\varepsilon_L$ structures [84] (table 5). This hypothesis is supported by the fact that all three conformers incorporating the ε_L subconformation were of the $x\varepsilon_L$ (and not $\varepsilon_L x$) type. In the remaining $17-3=14$ initial geometries incorporating the ε_L substructure ($\varepsilon_L x_L$, $x_D\varepsilon_L$, $x_L\varepsilon_L$ and $\varepsilon_L x_D$), generally the $\varepsilon_L \rightarrow \beta_L$ conformational shift was observed [84], resulting in the $\beta_L x_L$, $x_D\beta_L$ and $x_L\beta_L$ conformations. However, when the initial geometry falls in the $\varepsilon_L x_D$ category, the $\delta_D x_D$ backbone conformation was obtained ($\varepsilon_L x_D \rightarrow \delta_D x_D$).

5. The conformational effects of the nearest neighbouring 'groups'

Concentrating strictly on the immediate environment of the chiral α-carbon atom in a peptide, two amide groups (Q_1 and Q_2) and a side chain (R) can be identified beside the hydrogen atom.



Scheme 10.

Since all these groups are in a 'geminal position' relative to each other, it is not surprising that a strong stereo-electron influence is operative between them [20–25]. When the investigation concentrates on the interaction between the preceding (Q_1) and the following (Q_2) amid groups it is called 'backbone/backbone' modifying effect. On the other hand, if the structural influence of the side-chain (R) is analysed on one or both of the amide groups (Q_1 or Q_2), the phenomenon is often related to 'backbone/side-chain' effect.

5.1. *An example for the backbone/side-chain effect*

In systematic conformation analyses of For–Gly–NH_2, For–*L*–Ala–NH_2, For–*D*–Ala–NH_2 and For–*L*–Val–NH_2 derivatives, the serine (an additional natural amino acid

Table 5. The three (3) $x\varepsilon_L$ type backbone conformations of the For–*L*–Ala–*L*–Ala–NH_2 molecule using an *ab initio* (HF/3–21G) computation.

INI.[a] / CONV.[b]	$\delta_D\varepsilon_L$† / $\delta_D\varepsilon_L$	$\gamma_L\varepsilon_L$ / $\gamma_L\varepsilon_L$	$\gamma_D\varepsilon_L$ / $\gamma_D\varepsilon_L$
ω^1	+ 175·2	− 174·8	− 177·2
ϕ^1	− 174·2	− 83·9	+ 72·6
ψ^1	− 55·0	+ 65·3	− 66·7
ω^2	+ 154·1	+ 172·9	+ 160·6
ϕ^2	− 79·0	− 71·4	− 73·7
ψ^2	+ 171·7	+ 162·8	+ 168·8
ω^3	+ 179·7	+ 178·7	+ 179·0
model dis[c]	6·36	8·53	5·87
τ	+ 45·2	+ 166·0	+ 57·0
crit dis[d]	6·42	9·32	6·09
τ	+ 43·3	+ 166·6	+ 53·2
O1 ... HN4	8·11	6·67	5·99
O1 ... N4	7·37	6·32	5·34
O1.HN4.N4	− 39·0	− 65·4	− 45·9
O1 ... HN3	4·46	2·03	1·90
O1 ... N3	4·85	2·88	2·80
O1.HN3.N3	+ 107·7	+ 141·0	+ 147·8
O2 ... HN4	4·53	4·14	4·35
O2 ... N4	4·31	3·94	4·12
O2.HN4.N4	− 71·0	− 71·5	− 70·1
MAX. FORCE	1·3 E–4	1·5 E–5	1·6 E–4
RMS. FORCE	3·8 E–5	4·1 E–5	4·7 E–5
E	0·954315	0·954742	0·953218
ΔE	5·88	5·61	6·57
STRCT.[e]	31	331	55

† The $\delta_D\varepsilon_L$ conformation contains an eight member intramolecular H-bond, where O3 ... H2 = 2·07 Å, O3 ... N2 = 3·05 Å with an H-bond angle (O3–H2–N2) = + 163·1°. Torsion angles (ω, ϕ, ψ) in degrees, distances in Angstrom, forces in a.u. energy (E) in hartrees and the energy difference (ΔE) in kcal mole^{-1} compared to $E(\gamma_L\gamma_L) = -656.963681$.

[a] INI: Initial backbone conformation (calculated by ECEPP/2).

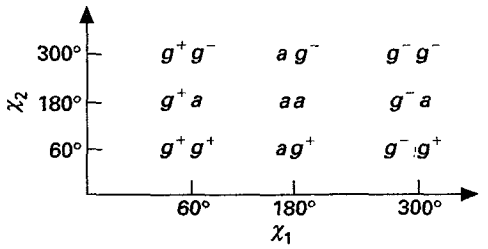
[b] CONV: Converged backbone conformation.

[c] MODEL DIS. C_i^α and C_{i+3}^α in accordance to classical β-turn definition must be shorter than 7 Å. In For–Ala–Ala–NH_2 the two C^α atoms are replaced by hydrogens (H1 and H4*) (see scheme 14), therefore the model distance (d_{model}) is shorter than C_i^α–C_{i+3}^α (figure 1), no more than 1·1 Å. In such a case τ is H1–C2$^\alpha$–C3$^\alpha$–H4*.

[d] CRIT DIS. critical distances for β-turn assignment (C_i^α–C_{i+3}^α) were extrapolated using *ab initio* resulted bond lengths and bond angles, on the basis of the determined N–H and C′–H distances. In such a case τ is C1$^\alpha$–C2$^\alpha$–C3$^\alpha$–C4$^\alpha$.

[e] Number of structures assigned in the 78 selected proteins.

frequently found in proteins) diamide is an optimal candidate for analysis. The hydroxy methyl group (the side-chain of the serine) can be involved *directly* in enzymatic reactions as demonstrated for the serine proteases family (trypsine, chymotrypsine, etc.), or can have an *indirect* influence on the protein by stabilizing different secondary structural elements [12, 13, 20–22]. Serine is often found in the α-helical region of globular proteins as well as in the $(i+2)$ positions of a β-turn, where a



Scheme 11.

backbone/side-chain stabilizing effect is expected. (For the definition of the ϕ, ψ, χ^1 and χ^2 torsional angle see scheme 2.)

As demonstrated previously (scheme 4) for an amino acid diamide, a total of nine legitimate backbone minima (α_L, α_D, β_L, γ_L, γ_D, ε_L, ε_D, δ_L, δ_D) occur [56]. The side chain may 'split' these backbone conformations into additional minimal energy conformations. In the case of amino acid residues with χ_1 and χ_2 side chain torsional angles not less than $3 \times 3 = 9$ legitimate minima are expected on the basis on MDCA (scheme 11). These nine side-chain conformations may be associated with each of the previously mentioned typical backbone conformations leading to a grand total of $9 \times 9 = 81$ legitimate conformations. A systematic side-chain conformational mapping [81] with α_L, β_L, γ_L, δ_L, ε_L, α_D, γ_D, δ_D, and ε_D backbone conformations was expected to reveal the most important structures of For–*L*–Ser–NH_2. To reduce the possibility to skip over minima of high energy or to overpass structures located in a 'hidden' conformation valley, nine complete grids (a total of 9×169 *ab initio* grid points) were computed. Using *ab initio* type (3–21G) calculations (figure 12 and table 6) a total of $3\alpha_L$, $5\beta_L$, $6\gamma_L$, $5\delta_L$, $4\alpha_D$, $9\gamma_D$, $6\delta_D$, and $6\varepsilon_D$, but no ε_L fully relaxed structures (a total of 44 conformers) were assigned for the For–*L*–Ser–NH_2 molecule [81]. Total relaxation was started from selected grid points (typically minima) as well as from locations predicted by MDCA [53–55].

These *ab initio* calculations are perfect examples of the 'backbone/side-chain' effect, since certain backbone orientations can be stabilized, while others can be destabilized by specific side-chain conformations. On the other hand, plotting all 44 side chain conformations (figure 12), and all the nine theoretically predicted legitimate χ_1 and χ_2 combinations can occur.

5.2. *An example for backbone/backbone effect*

The conformational analysis of small peptides has not only challenged spectroscopic aspects [88–105], but also has a significant role in the understanding of protein folding [11, 12, 20–25]. In vacuo each peptide residue may adopt only a limited number of discrete backbone torsional angle combinations according to theoretical calculations [57, 106]. These molecular calculations may yield all possible minimum energy structures, regardless of their relative energy.

The identification of *all the possible conformers* is not, but the actual number of intrinsically stable conformations is still problematic, since it is not known (1) how many minima are located on an *nD-potential energy surface* (*nD-PES*), and (2) where these minima are located on the nD-PES. In contrast to the few recognized structures of triamide systems known as β-turns or hairpin conformations (table 1), a total of 81 (9×9) distinctly different conformations are predicted by MDCA as shown in

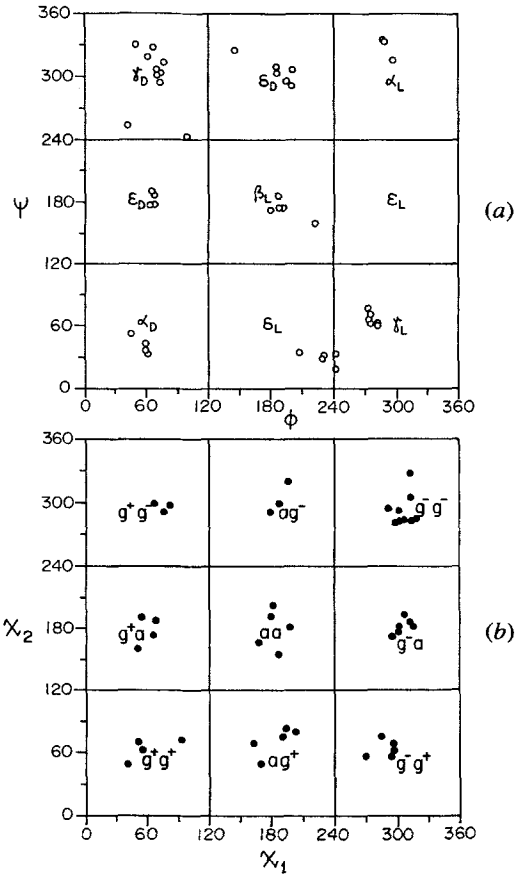


Figure 12. A total of 44 different minima of the For–*L*–Ser–NH_2 molecule (generated from 9×169 *ab initio* points). (*a*) The location of these $3\alpha_L$, $5\beta_L$, $6\gamma_L$, $5\delta_L$, $4\alpha_D$, $9\gamma_D$, $6\delta_D$, $6\varepsilon_D$ type conformations on a $E = E(\phi, \psi)$ surface. (*b*) The distribution of the side chain conformations.

scheme 12. MDCA, without the consideration of destabilizing or stabilizing interactions, predicts a total of 81 legitimate minima (scheme 12). The analysis of triamide substructures in a large protein data base where all stabilizing interactions are present, suggests the existence of all the 81 conformations. On the other hand, molecular mechanics (MM) calculations confirmed 75 out of the 81 MDCA legitimized conformations [53–55]. The six non-existing conformations were found according to the MM calculations [57, 83].

The upper limit (81) of the number of legitimate backbone conformations on the four-dimensional (4D) Ramachandran map associated with For–*L*–Ala–*L*–Ala–NH_2 was previously determined by MDCA [57, 83]. According to *ab initio* (SCF) geometry optimizations (using a 3–21G basis set) only 49 intrinsically stable conformations were found [84], while 32 (marked by * on scheme 12) migrated to one of the above 49 stable structures. All triamides incorporating the α_L and/or the ε_L substructures, such as $[(\alpha_L x_L, \alpha_L x_D)$, $(x_L \alpha_L, x_D \alpha_L)$, $(\varepsilon_L x_L, \varepsilon_L x_D)$ and $(x_D \varepsilon_L, x_L \varepsilon_L)]$, were investigated with precaution. From these 34 possible structures only four conformations ($\alpha_L \delta_L$, $\gamma_L \varepsilon_L$, $\gamma_D \varepsilon_L$ and $\delta_D \varepsilon_L$) turned out to be intrinsically stable [84]. These four unexpectedly stable

Table 6. The 44 optimized *ab initio* structures of For–*L*–Ser–NH_2 molecule using 3–21G basis set.

BB	SC	ϕ_{Top}	ψ_{Top}	χ_1	χ_2	ΔE
α_D	g^-a	60·3	37·6	302·0	181·0	12·9
α_D	g^+g^+	46·4	53·6	56·2	62·1	12·1
α_D	ag^-	62·3	34·1	291·9	294·6	9·1
α_D	ag^+	60·1	43·8	203·1	79·1	20·5
α_L	g^-g^-	288·0	336·3	314·4	282·5	12·5
α_L	g^-a	289·5	335·1	312·5	185·5	16·9
α_L	aa	297·6	317·2	179·9	191·3	20·6
β_L	g^-g^+	181·0	172·9	270·8	55·5	10·5
β_L	g^+a	189·4	174·9	68·0	187·1	11·2
β_L	aa	188·2	186·6	187·0	155·1	3·8
β_L	g^-a	222·7	160·0	295·5	171·5	15·4
β_L	g^+g^-	193·3	174·8	67·4	299·5	9·1
δ_D	g^-g^-	202·1	308·2	312·5	327·8	15·7
δ_D	g^-g^+	146·3	326·1	284·8	74·4	12·3
δ_D	aa	187·6	304·9	168·6	165·9	17·2
δ_D	ag^+	186·7	310·6	163·9	68·2	15·7
δ_D	g^+g^-	200·7	292·4	307·7	283·9	10·5
δ_D	g^+a	196·3	296·7	55·1	190·2	11·2
δ_L	g^-g^+	242·0	34·4	296·9	61·0	14·0
δ_L	g^+a	241·9	20·2	51·4	159·8	14·0
δ_L	ag^-	231·5	32·9	187·7	299·5	8·3
δ_L	g^-g^-	208·1	35·6	313·2	305·3	11·0
δ_L	g^-a	230·1	29·8	306·9	192·5	13·3
ε_D	g^-g^-	68·9	178·2	301·9	292·7	20·5
ε_D	g^-g^+	64·5	177·9	296·5	67·4	16·3
ε_D	aa	68·4	187·1	197·5	180·7	9·4
ε_D	ag^+	66·9	191·0	193·6	82·4	10·1
ε_D	g^+g^-	99·8	243·1	76·3	291·3	4·9
ε_D	g^+a	43·0	254·5	92·8	71·6	18·5
γ_D	g^-g^-	74·7	304·8	302·4	281·8	12·9
γ_D	g^-a	75·4	303·9	301·4	176·0	12·0
γ_D	g^-g^+	72·2	302·5	298·9	280·9	12·5
γ_D	ag^-	67·5	328·8	196·2	320·1	10·6
γ_D	aa	74·0	295·0	182·5	202·0	12·7
γ_D	ag^+	71·3	307·8	170·8	49·0	12·0
γ_D	g^+g^-	78	314·8	81·9	297·8	9·4
γ_D	g^+a	51·9	331·3	65·8	173·3	17·1
γ_D	g^+g^+	62·9	319·7	41·7	48·7	14·0
γ_L	g^-g^+	274·6	67·4	294·7	55·2	10·5
γ_L	g^+g^+	276·4	71·5	51·9	69·8	0
γ_L	ag^-	276·6	62·7	179·8	291·3	4·8
γ_L	ag^+	273·5	77·8	190·8	74·8	12·5
γ_L	g^-g^-	282·6	63·4	318·8	284·5	7·8
γ_L	g^-a	282·9	62·1	315·9	181·5	0·1

$\gamma_D\gamma_D$	$\gamma_D\delta_D$	$\gamma_D\alpha_L$	$\delta_D\gamma_D$	$\delta_D\delta_D$	$\delta_D\alpha_L$	$\alpha_L\gamma_D$	$\alpha_L\delta_D$	$\alpha_L\alpha_L$
$\gamma_D\varepsilon_D$	$\gamma_D\beta_L$	$\gamma_D\varepsilon_L$	$\delta_D\varepsilon_D$	$\delta_D\beta_L$	$\delta_D\varepsilon_L$	$\alpha_L\varepsilon_D$	$\alpha_L\beta_L$	$\alpha_L\varepsilon_L$
$\gamma_D\alpha_D$	$\gamma_D\delta_L$	$\gamma_D\gamma_L$	$\delta_D\alpha_D$	$\delta_D\delta_L$	$\delta_D\gamma_L$	$\alpha_L\alpha_D$	$\alpha_L\delta_L$	$\alpha_L\gamma_L$
$\varepsilon_D\gamma_D$	$\varepsilon_D\delta_D$	$\varepsilon_D\alpha_L$	$\beta_L\gamma_D$	$\beta_L\delta_D$	$\beta_L\alpha_L$	$\varepsilon_L\gamma_D$	$\varepsilon_L\delta_D$	$\varepsilon_L\alpha_L$
$\varepsilon_D\varepsilon_D$	$\varepsilon_D\beta_L$	$\varepsilon_D\varepsilon_L$	$\beta_L\varepsilon_D$	$\beta_L\beta_L$	$\beta_L\varepsilon_L$	$\varepsilon_L\varepsilon_D$	$\varepsilon_L\beta_L$	$\varepsilon_L\varepsilon_L$
$\varepsilon_D\alpha_D$	$\varepsilon_D\delta_L$	$\varepsilon_D\gamma_L$	$\beta_L\alpha_D$	$\beta_L\delta_L$	$\beta_L\gamma_L$	$\varepsilon_L\alpha_D$	$\varepsilon_L\delta_L$	$\varepsilon_L\gamma_L$
$\alpha_D\gamma_D$	$\alpha_D\delta_D$	$\alpha_D\alpha_L$	$\delta_L\gamma_D$	$\delta_L\delta_D$	$\delta_L\alpha_L$	$\gamma_L\gamma_D$	$\gamma_L\delta_D$	$\gamma_L\alpha_L$
$\alpha_D\varepsilon_D$	$\alpha_D\beta_L$	$\alpha_D\varepsilon_L$	$\delta_L\varepsilon_D$	$\delta_L\beta_L$	$\delta_L\varepsilon_L$	$\gamma_L\varepsilon_D$	$\gamma_L\beta_L$	$\gamma_L\varepsilon_L$
$\alpha_D\alpha_D$	$\alpha_D\delta_L$	$\alpha_D\gamma_L$	$\delta_L\alpha_D$	$\delta_L\delta_L$	$\delta_L\gamma_L$	$\gamma_L\alpha_D$	$\gamma_L\delta_L$	$\gamma_L\gamma_L$

structure minimization at ↓ *HF/3-21 G level of theory*

$\gamma_D\gamma_D$	$\gamma_D\delta_D$	$\gamma_D\delta_L^*$	$\delta_D\gamma_D$	$\delta_D\delta_D$	$\delta_D\gamma_L^*$	$\delta_L\gamma_D^*$	$\delta_L\delta_D^*$	$\alpha_L\delta_L^*$
$\gamma_D\varepsilon_D$	$\gamma_D\beta_L$	$\gamma_D\varepsilon_L$	$\delta_D\varepsilon_D$	$\delta_D\beta_L$	$\delta_D\varepsilon_L$	$\delta_L\varepsilon_D^*$	$\delta_L\beta_L^*$	$\delta_L\beta_L^*$
$\varepsilon_D\alpha_D^*$	$\gamma_D\delta_L$	$\gamma_D\gamma_L$	$\delta_D\alpha_D$	$\delta_D\gamma_L^*$	$\delta_D\gamma_L$	$\delta_L\alpha_D^*$	$\alpha_L\delta_L$	$\delta_L\gamma_L$
$\varepsilon_D\gamma_D$	$\varepsilon_D\delta_D$	$\varepsilon_D\delta_L^*$	$\beta_L\gamma_D$	$\beta_L\delta_D$	$\beta_L\delta_L^*$	$\delta_D\gamma_D^*$	$\delta_D\delta_D^*$	$\beta_L\delta_L^*$
$\varepsilon_D\varepsilon_D$	$\varepsilon_D\beta_L$	$\varepsilon_D\gamma_L^*$	$\beta_L\varepsilon_D$	$\beta_L\beta_L$	$\beta_L\beta_L^*$	$\delta_D\varepsilon_D^*$	$\beta_L\beta_L^*$	$\beta_L\beta_L^*$
$\varepsilon_D\alpha_D$	$\varepsilon_D\delta_L$	$\varepsilon_D\gamma_L$	$\beta_L\alpha_D$	$\beta_L\delta_L$	$\beta_L\gamma_L$	$\delta_D\alpha_D^*$	$\beta_L\delta_L^*$	$\beta_L\gamma_L^*$
$\alpha_D\gamma_D$	$\alpha_D\delta_D$	$\alpha_D\alpha_L^*$	$\delta_L\gamma_D$	$\delta_L\delta_D$	$\alpha_L\delta_L^*$	$\gamma_L\gamma_D$	$\gamma_L\delta_D$	$\gamma_L\delta_L^*$
$\gamma_D\varepsilon_D^*$	$\alpha_D\beta_L$	$\alpha_D\beta_L^*$	$\delta_L\varepsilon_D$	$\delta_L\beta_L$	$\delta_L\beta_L^*$	$\gamma_L\varepsilon_D$	$\gamma_L\beta_L$	$\gamma_L\varepsilon_L$
$\alpha_D\alpha_D$	$\alpha_D\delta_L$	$\alpha_D\gamma_L$	$\delta_L\alpha_D$	$\alpha_L\delta_L^*$	$\delta_L\gamma_L$	$\gamma_L\alpha_D$	$\gamma_L\delta_L$	$\gamma_L\gamma_L$

Scheme 12.

triamide conformations incorporated, for the first time, as stable substructures, the α_L and the ε_L subunits, previously annihilated on the 2D-Ramachandran map. The 49 stable triamide conformations yield a total of $49 \times 2 = 98$ single amino acid residue subconformations classified as α_L, β_L, δ_L etc. Scheme 13 demonstrates the grouping of the triamide backbone geometries in order to extract typical ϕ and ψ torsional angle pairs of the nine legitimate backbone conformations.

Figure 13 and table 7 show the distribution of (ϕ, ψ) angle pairs as found in triamides. (Data in figure 13 are reported in the conventional $-180° \leqslant \phi \leqslant 180°$ and $-180° \leqslant \psi \leqslant 180°$ representations, while a topologically more useful plot is shown in figure 13.) The arrangement of the 98 calculated backbone subconformations shows the existence of nine clusters as expected (figure 13), where only two of the δ_L type conformations are shifted to one of the square representing the borders of the appropriate idealized catchment region. However, even these two cases are due to the fact that the average δ_L position ($\phi = -126{\cdot}2°, \psi = +26{\cdot}5°$) is shifted substantially away from the idealized position ($\phi = -180°, \psi = +60°$) of δ_L towards the lower right hand corner of the square. With the exception of two points, one δ_L and one ε_D, all the 98 points are within a 30° radius. These two points are called 'ghost' conformations and are marked by a star such as δ_L^* and ε_D^*.

Peptides, incorporating three amide groups, such as the For–*L*–Ala–*L*–Ala–NH_2 molecule (see scheme 14), form a representative element of the *diamino acid triamide* systems. Among their main chain conformers some have a 'hair-pin' like conformation

$\alpha_D x$ and $x\alpha_D \rightarrow \alpha_D$,
$\alpha_L \delta_L \rightarrow \alpha_L$,
$\beta_L x$ and $x\beta_L \rightarrow \beta_L$,
$\gamma_D x$ and $x\gamma_D \rightarrow \gamma_D$,
$\gamma_L x$ and $x\gamma_L \rightarrow \gamma_L$,
$\delta_D x$ and $x\delta_D \rightarrow \delta_D$,
$\delta_L x$ and $x\delta_L \rightarrow \delta_L$,
$\varepsilon_D x$ and $x\varepsilon_D \rightarrow \varepsilon_D$,
$\gamma_L \varepsilon_L$, $\gamma_D \varepsilon_L$ and $\delta_D \varepsilon_L \rightarrow \varepsilon_L$.

Scheme 13.

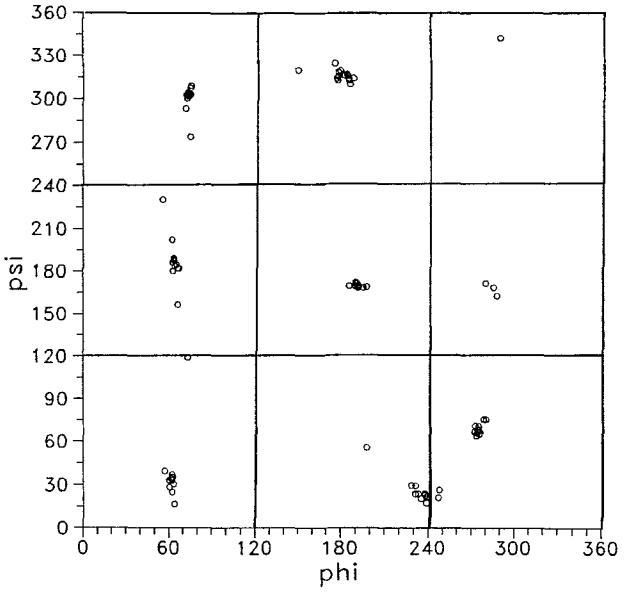


Figure 13. The 49 stable triamide conformations of the For–*L*–Ala–*L*–Ala–NH_2 molecule yielded a total of $49 \times 2 = 98$ diamide type subconformations. The location of these conformations are reported according to a topological ($0° \leq \phi \leq +360°$, $0° \leq \psi \leq +360°$).

(such as type I and/or type II β-turns, etc.), while others look more or less extended [83, 84]. Based on the X-ray structure analyses of proteins, only type I, II and III β-turns are frequently assigned.

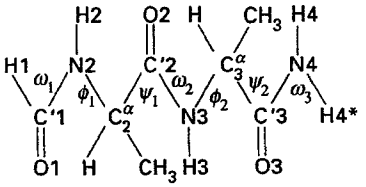
Their conformational mirror images (the conformational enantiomers such as the I′, II′ and III′ type β-turns) as well as the type VI *a*, *b* and VIII β-turns have only sporadic occurrences. There are at least two different ways to distinguish β-turns from untwisted secondary structural elements. The traditional method applied in proteins is to define four consecutive amino acids in the sequence ($i, i+1, i+2$ and $i+3$ positions) labelled as 1, 2, 3 and 4 and determine the C_1^{α}–C_4^{α} distance. For a hairpin or β-turn structure this value ($d_{\text{crit.}}$) must be shorter than 7 Å. (Otherwise the structural unit is quoted as extended like conformation.) Although hydrogen bonding is not a pre-requisite of a β-turn, the so-called $1 < -4$ type hydrogen bond is frequently assigned when X-ray

Table 7. The 49 optimized *ab initio* structures of For–Ala–Ala–NH_2 molecule using 3–21G basis set.

BB sub. conf.	$\phi_{Top.}$	$\psi_{Top.}$
$\alpha_L\delta_L \rightarrow \alpha_L$	291·4	342·5
$\alpha_D\alpha_D \rightarrow \alpha_D$	60·4	28·3
$\alpha_D\alpha_D \rightarrow \alpha_D$	62·3	24·9
$\alpha_D\beta_L \rightarrow \alpha_D$	60·2	33·0
$\alpha_D\delta_D \rightarrow \alpha_D$	64·1	16·8
$\alpha_D\delta_L \rightarrow \alpha_D$	57·0	39·4
$\alpha_D\gamma_D \rightarrow \alpha_D$	62·3	37·0
$\alpha_D\gamma_L \rightarrow \alpha_D$	63·4	30·4
$\beta_L\alpha_D \rightarrow \alpha_D$	62·1	35·3
$\delta_D\alpha_D \rightarrow \alpha_D$	61·5	34·3
$\delta_L\alpha_D \rightarrow \alpha_D$	63·1	35·1
$\varepsilon_D\alpha_D \rightarrow \alpha_D$	62·7	35·3
$\gamma_L\alpha_D \rightarrow \alpha_D$	62·4	33·0
$\alpha_D\beta_L \rightarrow \beta_L$	186·4	169·8
$\beta_L\alpha_D \rightarrow \beta_L$	192·4	168·4
$\beta_L\beta_L \rightarrow \beta_L$	191·4	171·1
$\beta_L\beta_L \rightarrow \beta_L$	192·4	170·3
$\beta_L\delta_D \rightarrow \beta_L$	192·6	169·2
$\beta_L\delta_L \rightarrow \beta_L$	191·8	171·6
$\beta_L\varepsilon_D \rightarrow \beta_L$	191·0	172·4
$\beta_L\gamma_D \rightarrow \beta_L$	192·8	168·4
$\beta_L\gamma_L \rightarrow \beta_L$	192·3	169·3
$\delta_D\beta_L \rightarrow \beta_L$	192·6	170·2
$\delta_L\beta_L \rightarrow \beta_L$	190·6	169·6
$\varepsilon_D\beta_L \rightarrow \beta_L$	193·1	169·8
$\gamma_D\beta_L \rightarrow \beta_L$	198·4	169·2
$\gamma_L\beta_L \rightarrow \beta_L$	196·0	168·6
$\delta_D\alpha_D \rightarrow \delta_D$	178·6	313·0
$\delta_D\beta_L \rightarrow \delta_D$	183·4	316·4
$\delta_D\delta_D \rightarrow \delta_D$	178·0	314·6
$\delta_D\delta_D \rightarrow \delta_D$	187·1	310·3
$\delta_D\varepsilon_D \rightarrow \delta_D$	180·6	319·7
$\delta_D\varepsilon_L \rightarrow \delta_D$	185·8	316·0
$\delta_D\gamma_D \rightarrow \delta_D$	179·2	318·6
$\delta_D\gamma_L \rightarrow \delta_D$	178·9	315·7
$\alpha_D\delta_D \rightarrow \delta_D$	150·8	320·0
$\beta_L\delta_D \rightarrow \delta_D$	186·3	312·9
$\delta_L\delta_D \rightarrow \delta_D$	186·3	314·4
$\varepsilon_D\delta_D \rightarrow \delta_D$	185·0	317·1
$\gamma_D\delta_D \rightarrow \delta_D$	189·5	314·6
$\gamma_L\delta_D \rightarrow \delta_D$	176·7	324·8
$\alpha_D\delta_L \rightarrow \delta_L$	228·6	29·7
$\alpha_L\delta_L \rightarrow \delta_L$	246·9	21·3
$\beta_L\delta_L \rightarrow \delta_L$	231·4	29·6
$\varepsilon_D\delta_L \rightarrow \delta_L$	247·6	26·6
$\gamma_D\delta_L \rightarrow \delta_L$	237·9	22·9
$\gamma_L\delta_L \rightarrow \delta_L$	231·2	23·8
$\delta_L\alpha_D \rightarrow \delta_L$	237·6	24·0
$\delta_L\beta_L \rightarrow \delta_L$	238·7	17·6

Table 7 (*continued*).

BB sub. conf.	$\phi_{Top.}$	$\psi_{Top.}$
$\delta_L\delta_D \rightarrow \delta_L$	233·2	23·8
$\delta_L\varepsilon_D \rightarrow \delta_L$	198·2	55·7
$\delta_L\gamma_D \rightarrow \delta_L$	238·8	21·8
$\delta_L\gamma_L \rightarrow \delta_L$	235·1	20·6
$\beta_L\varepsilon_D \rightarrow \varepsilon_D$	65·7	184·5
$\delta_D\varepsilon_D \rightarrow \varepsilon_D$	63·8	188·4
$\delta_L\varepsilon_D \rightarrow \varepsilon_D$	62·6	186·2
$\varepsilon_D\varepsilon_D \rightarrow \varepsilon_D$	63·7	189·4
$\gamma_D\varepsilon_D \rightarrow \varepsilon_D$	63·3	180·1
$\gamma_L\varepsilon_D \rightarrow \varepsilon_D$	64·3	183·5
$\varepsilon_D\alpha_D \rightarrow \varepsilon_D$	73·8	119·1
$\varepsilon_D\beta_L \rightarrow \varepsilon_D$	66·7	156·5
$\varepsilon_D\delta_D \rightarrow \varepsilon_D$	62·6	202·3
$\varepsilon_D\delta_L \rightarrow \varepsilon_D$	56·1	230·3
$\varepsilon_D\varepsilon_D \rightarrow \varepsilon_D$	67·7	181·8
$\varepsilon_D\gamma_D \rightarrow \varepsilon_D$	66·9	181·9
$\varepsilon_D\gamma_L \rightarrow \varepsilon_D$	64·0	187·2
$\delta_D\varepsilon_L \rightarrow \varepsilon_L$	281·0	171·7
$\gamma_D\varepsilon_L \rightarrow \varepsilon_L$	286·3	168·8
$\gamma_L\varepsilon_L \rightarrow \varepsilon_L$	288·6	162·8
$\alpha_D\gamma_D \rightarrow \gamma_D$	74·1	302·0
$\beta_L\gamma_D \rightarrow \gamma_D$	75·6	302·8
$\delta_D\gamma_D \rightarrow \gamma_D$	74·3	302·3
$\delta_L\gamma_D \rightarrow \gamma_D$	74·9	302·0
$\varepsilon_D\gamma_D \rightarrow \gamma_D$	75·8	274·0
$\gamma_D\gamma_D \rightarrow \gamma_D$	74·4	303·3
$\gamma_L\gamma_D \rightarrow \gamma_D$	72·7	302·7
$\gamma_D\beta_L \rightarrow \gamma_D$	76·1	308·8
$\gamma_D\delta_D \rightarrow \gamma_D$	73·8	302·4
$\gamma_D\delta_L \rightarrow \gamma_D$	75·5	307·3
$\gamma_D\varepsilon_D \rightarrow \gamma_D$	73·2	300·3
$\gamma_D\varepsilon_L \rightarrow \gamma_D$	72·6	293·3
$\gamma_D\gamma_D \rightarrow \gamma_D$	73·9	304·3
$\gamma_D\gamma_L \rightarrow \gamma_D$	73·8	301·7
$\alpha_D\gamma_L \rightarrow \gamma_L$	272·7	67·4
$\beta_L\gamma_L \rightarrow \gamma_L$	274·9	68·2
$\delta_D\gamma_L \rightarrow \gamma_L$	274·5	68·8
$\delta_L\gamma_L \rightarrow \gamma_L$	275·3	67·0
$\varepsilon_D\gamma_L \rightarrow \gamma_L$	274·0	66·2
$\gamma_D\gamma_L \rightarrow \gamma_L$	276·4	67·1
$\gamma_L\alpha_D \rightarrow \gamma_L$	274·2	64·0
$\gamma_L\beta_L \rightarrow \gamma_L$	273·2	71·4
$\gamma_L\delta_D \rightarrow \gamma_L$	280·7	75·8
$\gamma_L\delta_L \rightarrow \gamma_L$	275·4	71·0
$\gamma_L\varepsilon_D \rightarrow \gamma_L$	279·1	75·8
$\gamma_L\varepsilon_L \rightarrow \gamma_L$	276·1	65·3
$\gamma_L\gamma_D \rightarrow \gamma_L$	275·5	68·6
$\gamma_L\gamma_L \rightarrow \gamma_L$	275·8	67·0
$\gamma_L\gamma_L \rightarrow \gamma_L$	275·1	66·4



Scheme 14.

$\gamma_D\gamma_D$	$\gamma_D\delta_D$	$\gamma_D\alpha_L$	$\delta_D\gamma_D$	$\delta_D\delta_D^T$	$\delta_D\alpha_L$	$\alpha_L\gamma_D$	$\alpha_L\delta_D$	$\alpha_L\alpha_L$
$\gamma_D\varepsilon_D$	$\gamma_D\beta_L^t$	$\gamma_D\varepsilon_L^T$	$\delta_D\varepsilon_D$	$\delta_D\beta_L^t$	$\delta_D\varepsilon_L^T$	$\alpha_L\varepsilon_D$	$\alpha_L\beta_L$	$\alpha_L\varepsilon_L$
$\gamma_D\alpha_D$	$\gamma_D\delta_L^T$	$\gamma_D\gamma_L^T$	$\delta_D\alpha_D$	$\delta_D\delta_L$	$\delta_D\gamma_L^t$	$\alpha_L\alpha_D$	$\alpha_L\delta_L^T$	$\alpha_L\gamma_L$
$\varepsilon_D\gamma_D^t$	$\varepsilon_D\delta_D$	$\varepsilon_D\alpha_L$	$\beta_L\gamma_D^t$	$\beta_L\delta_D$	$\beta_L\alpha_L$	$\varepsilon_L\gamma_D$	$\varepsilon_L\delta_D$	$\varepsilon_L\alpha_L$
$\varepsilon_D\varepsilon_D^t$	$\varepsilon_D\beta_L$	$\varepsilon_D\varepsilon_L$	$\beta_L\varepsilon_D^t$	$\beta_L\beta_L$	$\beta_L\varepsilon_L$	$\varepsilon_L\varepsilon_D$	$\varepsilon_L\beta_L$	$\varepsilon_L\varepsilon_L$
$\varepsilon_D\alpha_D$	$\varepsilon_D\delta_L^T$	$\varepsilon_D\gamma_L^T$	$\beta_L\alpha_D^t$	$\beta_L\delta_L$	$\beta_L\gamma_L^t$	$\varepsilon_L\alpha_D$	$\varepsilon_L\delta_L$	$\varepsilon_L\gamma_L$
$\alpha_D\gamma_D^T$	$\alpha_D\delta_D^T$	$\alpha_D\alpha_L$	$\delta_L\gamma_D$	$\delta_L\delta_D^T$	$\delta_L\alpha_L$	$\gamma_L\gamma_D^T$	$\gamma_L\delta_D^T$	$\gamma_L\alpha_L$
$\alpha_D\varepsilon_D$	$\alpha_D\beta_L^t$	$\alpha_D\varepsilon_L$	$\delta_L\varepsilon_D^T$	$\delta_L\beta_L^t$	$\delta_L\varepsilon_L$	$\gamma_L\varepsilon_D^T$	$\gamma_L\beta_L^t$	$\gamma_L\varepsilon_L$
$\alpha_D\alpha_D^T$	$\alpha_D\delta_L$	$\alpha_D\gamma_L$	$\delta_L\alpha_D$	$\delta_L\delta_L$	$\delta_L\gamma_L$	$\gamma_L\alpha_D^T$	$\gamma_L\delta_L$	$\gamma_L\gamma_L^t$

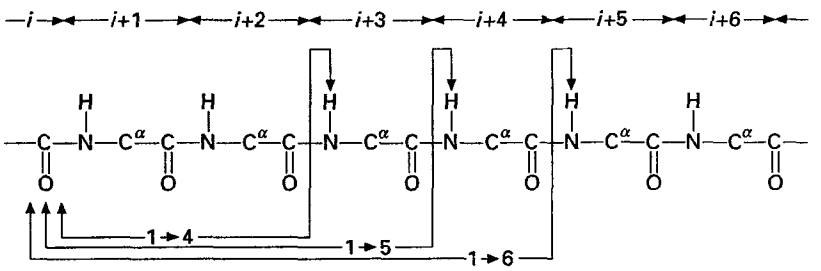
Scheme 15.

determined protein structures are analysed. To identify such an H-bond pattern where the NH of the $i+3$ residue points toward the carbonyl oxygen of the ith residue, the conformation of four residues is required. On the other hand, when a β-turn type is characterized throughout its main chain conformation, the backbone torsional angle values of the two middle amino acid residues (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2} and ψ_{i+2}) are sufficient.

To classify a structural unit as a β-turn, the second option involves the determination of the relative orientation of the three consecutive amide bonds. The C_i^α–C_{i+1}^α–C_{i+2}^α–C_{i+3}^α torsional angle (labelled τ) may have a value between $-180° \leqslant \tau \leqslant +180°$, where conformations with $-90° \leqslant \tau \leqslant +90°$ categorize as β-turns. For example the $\gamma_L\gamma_L$ conformation, which has been suspected on the basis of an educated guess to be the global minimum on a 4D-Ramachandran map, has a $\tau = 168{\cdot}4°$ value which can quantitatively express the degree of unfolding. This conformation is stabilized by two backbone type H-bonds, each involving a seven member pseudo-ring (C_7). In accordance with the criterion that the C_1^α–C_4^α distance must be shorter than 7 Å, only a range of τ around 0° will result in the so called hairpin conformation. A total of 30 from these 81 MDCA predicted conformations qualify as β-turns or reverse turns according to *ab initio* (SCF/3–21G) calculations. Structures marked by the T superscript qualify as β-turns according to both the τ and the d criteria. On the other hand, in conformers with the t superscript either the τ or the d criterion is not fulfilled.

6. The effects of far-lying peptide units

In the secondary structural element sets (helices, β-turns, β-sheets etc.) [11, 12, 20–25] composing the three-dimensional structure of globular proteins, three forms ($3_{10}-$, $\alpha-$ and $\Pi-$) of the right-handed helical structure are noted, although the Π-helix has not been observed during X-ray data analyses. The $3_{10}-$, the $\alpha-$ and the Π-helices are the three different forms of the right handed helical structure, usually referred to



Scheme 16.

in the literature. This 3_{10}-helix is a 'slimmer' structure ($\phi = -60°$, $\psi = -30°$) and, compared to the 'normal' α-helix ($\phi = -54°, \psi = -45°$), the II-helix ($\phi = -45°, \psi = -55°$) is a 'huskier' one.

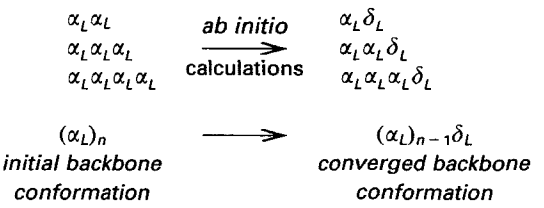
An important and common structure pattern of helical units is their periodic hydrogen bond network. The first type (the 3_{10}-helix) contains a series of H-bond systems incorporating 10-atoms ($i + 4$), while the second type (the α-helix) has 13-($i + 5$), and the third form (the Π-helix) include 16-atoms ($i + 6$) in the repetitive H-bond network system. According to X-ray data analyses of globular proteins the 3_{10}-helix can be found as an individual secondary structural element, but also located at the C or N terminus of the α-helices as the first or the last turn in an α-helix. In this steeper or slimmer helix with a ϕ torsional angle value around $-60°$ and ψ around $-30°$, the H-bond network connects the i and the $i + 3$ amino acid residue. In a 'standard' α-helix ($\phi \approx -54°, \psi \approx -45°$) the i and $i + 4$ amino acid residues are H-bonded in the polypeptide chain. How and why a helix of any type is adopted by the folded protein, and the question of 'helix-signal(s)' in proteins are continuously investigated [107–109]. Selected oligopeptides are also heavily investigated by spectroscopical methods and by molecular dynamics to find the conformational properties of isolated helix-like segments, and very often the conformational distortion of the helix ending units was observed.

A systematic study on periodic (or homo-) conformations [110–112] (e.g. $\alpha_D\alpha_D\alpha_D$, $\alpha_L\alpha_L\alpha_L$, $\gamma_D\gamma_D\gamma_D$ etc.) of the For–Ala–Ala–Ala–NH_2 tripeptide revealed that the geometry optimization started from the $\alpha_L\alpha_L\alpha_L$ conformation converged to the $\alpha_L\alpha_L\delta_L$ backbone conformation. The extension of the amino acid chain length into longer helical conformations of For–$(Ala)_n$–NH_2 molecule resulted in similar conformational consequences. All these backbone conformations incorporating the δ_L sub-conformation at the carboxyl-end of the oligopeptide chain, have intrinsic stability. Thus, the $(\alpha_L)_n$ initial conformations for various n-values had always flip to $(\alpha_L)_{n-1}\delta_L$ (see scheme 17) during geometry optimization. Analysing the H-bond network system of the optimized $(\alpha_L)_4\delta_L$ conformation of the For–$(L$–Ala$)_5$–NH_2 molecule [110, 113], a 3_{10}-helix was assigned. *Ab initio* calculations confirmed the expectation that the formation of a helical segment is strongly coupled with the build up of a systematic H-bond network system. The geometries of the six (6) α_L subconformations (table 8) specified at the right hand side of scheme 17 are marked in as open circles in figure 9.

7. Application of the conformational results

7.1. *Main-chain folding of peptides*

The importance of some triamide conformations, especially type I (or III) and type II β-turns, is continuously emphasized in the literature [61–63, 88–100]. Indeed,



Scheme 17.

Table 8. Relaxed $(\alpha_L)_{n-1}\delta_L$ conformations* of For–$(L$–Ala$)_n$–NH_2 computed at the HF/3–21G level of theory.

n Conf.	1	2	3	4	BB†
ϕ_{n-3}				−61·9	
ψ_{n-3}				−27·6	α_L
ϕ_{n-2}			−63·5	−61·9	
ψ_{n-2}			−26·8	−21·1	α_L
ϕ_{n-1}		−68·0	−71·3	−72·6	
ψ_{n-1}		−17·5	−4·6	−3·9	α_L
ϕ_n	−128·1	−113·1	−105·3	−106·0	
ψ_n	+29·8	+21·3	+12·1	+13·6	δ_L
Max. force (a.u.)	$1{\cdot}9\times10^{-6}$	$1{\cdot}5\times10^{-5}$	$6{\cdot}0\times10^{-6}$	$6{\cdot}7\times10^{-5}$	
R.M.S. force (a.u.)	$6{\cdot}5\times10^{-7}$	$5{\cdot}8\times10^{-6}$	$2{\cdot}0\times10^{-6}$	$1{\cdot}7\times10^{-5}$	

*Torsional angles are given in degrees.
†Backbone conformation.

these conformations are frequently assigned to secondary structural elements of globular proteins. Although the classification of β-turns is traditionally given on the basis of the backbone torsional angle values (ϕ_i, ψ_i) [114], the degree of folding or unfolding of a β-turn can be defined in a simpler way, based on the twisting of the hairpin conformation [83, 84]. As discussed in section 5.2., we have recently re-introduced the C^α_i–C^α_{i+1}–C^α_{i+2}–C^α_{i+3} torsional angle labelled as τ which describes the overall angularity of the backbone conformation with values $-180° \leqslant \tau + 180°$. The global minimum (the $\gamma_L\gamma_L$ conformation) has a $\tau_{\gamma_L\gamma_L} = +168{\cdot}4°$ which can quantitatively describe the degree of backbone unfolding since for a perfectly unfolded conformation τ must by $\pm 180°$. Considering the traditional criterion that the C^α_1–C^α_4 distance (d) must be shorter than 7 Å, only a fraction (18) of the total number of β-turn conformations (30 structures altogether) could be assigned as β-turns. This is because all the 49 For–L–Ala–L–Ala–NH_2 conformations [84] of the model compound resulted in 30 structures within the $-90° \leqslant \tau \leqslant +90°$ angularity range and 19 backbone geometries with $-180° \leqslant \tau \leqslant -90°$ or $+90° \leqslant \tau \leqslant +180°$ values. Depending on the d or τ type of 'conformation selection rule', both groups of hairpin geometries of For–Ala–Ala–NH_2 consist of a large number of elements. A total of 18 structures are assigned as β-turns if both the angularity and the distance criteria are applied simultaneously, but as many as 30 conformations are considered folded structures when only the angularity (τ) selection rule ($-90° \leqslant \tau \leqslant +90°$) is considered. All the calculated conformations had been observed previously by X-ray data-base analysis of globular proteins [115].

However, the extraction of all the β-turn conformations fulfilling the angularity ($-90° \leqslant \tau \leqslant +90°$) or distance ($d \leqslant 7$ Å) criteria is not possible by an X-ray data-base analysis. All existing β-turn conformations can be assigned on the corresponding 4D-Ramachandran type PES, however, many of these conformations that do occur in globular proteins cannot be extracted from X-ray data, because some of these structures have a rare occurrence.

The functional role of intramolecular hydrogen bonds in the formation of folded conformations is a subject of some controversy. The hydrogen bond is often claimed as the 'driving force' of the main chain-folding. If an α_L substructure is incorporated in a β-turn conformation (e.g. $\alpha_L\delta_L$ in type I β-turn), a favourable H-bond interaction is required to stabilize such a β-turn [11, 12, 114]. By contrast, for β-turns not containing an α_L conformational subunit, the existence of such an $1 < -4$ H-bond is not required [84]. For example, the $\alpha_D\alpha_D$ hairpin conformation contains a $1 < -4$ type H-bond, but in the $\delta_L\delta_D$ conformation (type II' β-turn) such an intramolecular H-bond is not present. Therefore it may be concluded on the basis of these *ab initio* calculations, that while the $1 < -4$ H-bond may be present in β-turns, it is not a necessary condition for the intrinsic stability of these structures. Only five structures ($\alpha_L\delta_L$, $\alpha_D\alpha_D$, $\alpha_D\delta_D$, $\varepsilon_D\delta_L$ and $\gamma_L\delta_D$) of the 30 β-turns located in this study incorporate the intramolecular $1 < -4$ H-bond ($d_{O\ldots HN} \leqslant 2{\cdot}2$ Å).

7.2. *Spectroscopic aspects of the* ab initio *calculations*

The X-ray analysis of crystalline compounds where typically a single conformer is observed can yield extremely valuable information [116–125]. However, the solid state structure of a peptide is often modified by 'crystal forces'. Therefore, these geometries are not necessarily identical to the dominant conformation adopted in the solutions, or in the gas phase. Although calculations can yield the geometry and the energy of any minimum energy structure in the environment-free state, the spectroscopic identification of these conformers or at least conformational regions, is still problematic in solutions. The application of the results obtained by *ab initio* calculations may have some relevance to conformational analyses performed by vibrational, circular dichroism, microwave, electron diffraction and NMR spectroscopy [28–33]. The time-scale of the experiment performed in solution has a cornerstone role. The applicability of *ab initio* structures can be different when the time-scale is long (e.g. NMR) or when spectroscopy reflects a time-resolved structure set (e.g. circular dichroism (CD), Fourier transform-infrared (FT-IR)).

Nuclear Overhauser effect (NOE) [42, 43] such as the ^{1}H–{^{1}H}–NOE experiment is the most frequently applied NMR measurement to determine the folding pattern of peptide or protein backbones. Previously backbone torsional angle constraints arising from the $^3J_{HH}$ spin–spin coupling constants [126, 127] as well as data obtained from the mobility analysis of the amide protons [128–133] were also involved in structure determinations. It has been shown that interproton distances (d_i) can be determined on the basis of quantitative NOE even for multiple spin systems, to an accuracy of $\pm 0{\cdot}1$ Å. For the interpretation of these experimentally determined distances, selected interproton distances are also calculated (usually by molecular mechanics (MM) and/or molecular dynamics (MD) [134–142]), but the empirical nature of the applied parameters used in MM and/or MD computations may introduce significant uncertainties in the distance data set [134–142]. Figure 14 summarizes 10 possible marker distances of the For–Ala–Ala–NH_2 molecule, useful for a quantitative structure assignment by NMR. The $d_{N\alpha}$ ($;d_{NH(i)/H\alpha(i)}$), d_{NN} ($;d_{NH(i)/NH(i+1)}$), and $d_{\alpha N}$($;d_{H\alpha(i)/NH(i+1)}$)

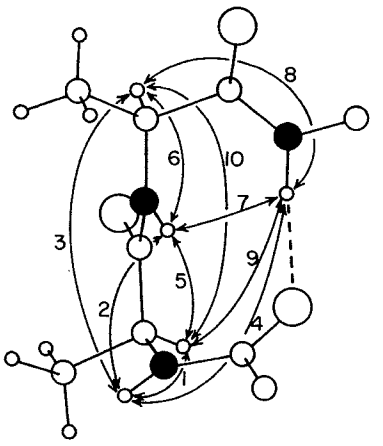


Figure 14. The ten marker interproton distances (d_i's) of For–Ala–Ala–NH_2; (1); NH_{i+1}–H^{α}_{i+1}, (2); NH_{i+1}–NH_{i+2}, (3); NH_{i+1}–H^{α}_{i+2}, (4); NH_{i+1}–NH_{i+3}, (5); NH_{i+2}–H^{α}_{i+1}, (6); NH_{i+2}–H^{α}_{i+2}, (7); NH_{i+2}–NH_{i+3}, (8); NH_{i+3}–H^{α}_{i+2}, (9); NH_{i+3}–H^{α}_{i+1}, (10); H^{α}_{i+1}–H^{α}_{i+2}. The distance values are different for each conformer.

interproton distances have a specific role in the conformation assignment performed by NOE's [84]. While the $d_{N\alpha}$ is a correlated spectroscopy (COSY) type connectivity, the $d_{\alpha N}$ and the d_{NN} are sequential NOE information. The conformation of any diamide-type subunit of a protein backbone can be characterized by one of the nine conformational types (see below). The assignment of these substructures using a combination of the three well known $d_{N\alpha}$, $d_{\alpha N}$ and d_{NN} distances has great importance. Although the determination of interproton distances based on ^1H–{^1H}–NOE or 2D-nuclear Overhauser and exchange spectroscopy (NOESY) is problematic, the secondary structure of a polypeptide can now be assigned not only in terms, of helices, sheets or turns but also as accurately as α_L, β_L, γ_L, δ_L, ε_L, etc. [84].

In the case of a single conformation these d_j's (equation (3)) are typical for the investigated conformation and are therefore regarded as marker distances. However, in the case of a conformational mixture these marker distances d_{ji} (that is the interproton distance in the ith conformer) have to be determined by computation for each of the conformers [84,91,92]. The experimentally determined interproton distances in the case of conformational mixtures reflect the average interproton distance due to the time-scale of the NMR experiment, which makes structure determination of small or medium size peptides extremely difficult. Consequently, the application of NOE based structural constraints for small or middle size peptides has serious limitations and requires sophisticated calculations [91,92]. The reciprocal of the sixth power of the average interproton distance (d_j) which is an experimental distance, is the weighted average of the reciprocal of the sixth power of the individual marker distances (d_{ji} which are calculated distances as shown by the following equation:

$$\frac{1}{d_j^6} = \sum_{i=1}^{\text{all. conf}} p_i \frac{1}{d_{ji}^6}. \tag{3}$$

The weighting coefficient (p_i) is the probability occurrence of the ith conformation in the mixture. Consequently, the spectroscopic determination of peptide conformations requires the quantitative knowledge of these marker distances (d_{ji}) in all the participating conformers (table 11).

The IR and UV type spectra containing time-resolved structural information are rather complex due to the presence of multiple conformations and/or the dynamic aspects of backbone conformation in solution. Moreover, the IR and/or CD spectra provide no insight to structure-sequence correlation. Therefore, the deconvolution [92, 143, 144] of the complex spectra must be followed by the assignment of the component curves to individual conformations, using external data (e.g. calculations resulting in probability occurrences (p_j), NOE structural constraints, etc.).

7.3. *Conformers to describe the 3D backbone structure of proteins*

Up to now, no relationship between the amino acid sequence of proteins and their three-dimensional conformation has been found. Although many different approaches have been attempted resulting in a variety of prediction algorithms, no reliable *a priori* prediction of protein 3D-structure from its amino acid sequence is in sight. A less glamorous but more realistic task is to classify the already determined backbone conformations. The determination of several 'conformational centres' on the Ramachandran map led to the knowledge of all the possible discrete backbone conformations in terms of which protein conformations may be represented. The fact that only a few discrete backbone conformations exist is equivalent to the recognition that certain conformations are playing some 'key' role during protein folding. In the past the division of the $-180° \leqslant \phi \leqslant +180°$ and $-180° \leqslant \psi \leqslant +180°$ spaces was achieved by the fragmentation of the X-ray determined protein structures (method *A*) [135, 136], as well as on the basis of the conformations determined by MM calculations (method *B*) [22]. The frequency ratios of subconformations in the investigations protein data set influenced the statistical analyses, introducing inaccuracies into method *A*. On the other hand, the parameterization of the force fields, partially based on X-ray data, can induce errors in the geometrical and energetic description of the resulting structures. Although the two methods differ from each other, their arbitrary character may well be the cause of their limitations. We have proposed a third method (method *C*) for the division of the ϕ, ψ surface on the basis of multidimensional conformational analysis combined with *ab initio* calculations. Since all published *ab initio* calculations on amino acid diamides, regardless to the applied basis set or calculation type, resulted in nine (or fewer) backbone conformations, this value, is treated as an upper limit. Although the Hartree–Fock calculations have also limitations, they can describe conformational properties quite accurately, which is not the case for the semi-empirical or force field methods. Consequently, the 2D-Ramachandran map ($0° \leqslant \phi \leqslant +360°$, $0° \leqslant \psi \leqslant +360°$) has been divided into nine conformation-regions and their minima have been labelled ($\alpha_L, \alpha_D, \beta_L$, etc.). The conformation centres are already specified in table 2.

The backbone distorting effect of the side chains can also be ignored to some extent since the analysis of For–Gly–NH_2, For–Ala–NH_2 and For–Val–NH_2 amino acid diamides yielded similar locations of backbone conformations [146] (table 9). The 44 different structures of the For–*L*–Ser–NH_2 yielded also nine clusters of backbone conformation types [82]. All the optimized 49 different conformations of For–Ala–Ala–NH_2 are also the combination of these nine ($\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$ and ε_D) legitimate conformations [84]. However, the application of these minima as 'conformational centres' on the PES was previously questionable, because two of the nine minima were annihilated in the case of simple amino acid diamides. This apparent paradox is now resolved. Also, the degree of conformational distorting effects of the nearest neighbours was unknown. The analysis of the torsional angle distributions of the $12\alpha_D$, $14\beta_L$, $15\gamma_L$, $14\gamma_D$, $12\delta_L$, $14\delta_D$, $3\varepsilon_L$, $13\varepsilon_D$ and $1\alpha_L$ subconformations yielded a set of average or central

Table 9. The 49 optimized *ab initio* structures of For–Gly–NH_2 For–Ala–NH_2 For–Val–NH_2 molecules using 3–21G basis set.

BB	Aaa	SC	$\phi_{Top.}$	$\psi_{Top.}$
α_D	Ala	g	63·8	32·7
α_D	Val	g^+	50·0	43·1
α_D	Val	a	60·2	40·9
α_D	Val	g^-	47·2	44·6
β_L	Gly		180·0	180·0
β_L	Ala	g	191·6	170·9
β_L	Val	g^+	197·7	156·8
β_L	Val	a	224·0	142·8
β_L	Val	g^-	218·4	163·7
δ_D	Gly		126·0	334·5
δ_D	Ala	g	181·4	316·0
δ_D	Val	g^+	183·8	326·2
δ_D	Val	a	222·4	299·9
δ_D	Val	g^-	190·1	313·1
δ_L	Gly		234·0	25·5
δ_L	Ala	g	232·2	30·0
δ_L	Val	g^+	225·9	35·3
δ_L	Val	g^-	236·3	28·4
ε_D	Ala	g	67·6	181·9
ε_D	Val	g^+	76·2	162·3
ε_D	Val	a	75·1	152·8
ε_D	Val	g^-	70·9	170·6
γ_D	Gly		83·9	292·2
γ_D	Ala	g	73·9	303·3
γ_D	Val	g^+	62·9	320·8
γ_D	Val	a	74·0	299·0
γ_D	Val	g^-	59·6	321·4
γ_L	Gly		276·1	67·8
γ_L	Ala	g	275·6	67·7
γ_L	Val	g^+	275·0	66·1
γ_L	Val	a	276·7	71·6
γ_L	Val	g^-	275·0	63·2

conformations' with high confidence (see table 2). Although the conformation dependent alteration of some substructures was observed (e.g. four of the $14\varepsilon_D$ substructures deviate significantly from the average value) in general they are located close to each other. The result obtained from the study involving several oligopeptides also underline the above conclusion [82, 110–112]. All these data are summarized in table 10.

As an example, the assignment of the 3D structures of a Cytochrome C (1CCR) fragment (scheme 18) is reported. When using only the classical terms to describe the folding of the backbone (α-helices, β-turns and β-pleated sheets), the 22–23, 26–29 and 32–40 regions of this protein backbone cannot be assigned by this traditional method. The 'structure-describing' method presented above called ACAP [145] (amino acid conformation assignment in proteins) makes the description of the overall backbone structure possible [146].

Three points should be emphasized in closing:

(1) the observation that certain substructures have ϕ, ψ torsional angle pairs far away ($>40°$) from the 'conformational centres' does not indicate the failure

Table 10. The location of the nine conformational centres on the basis of three different type of molecules.

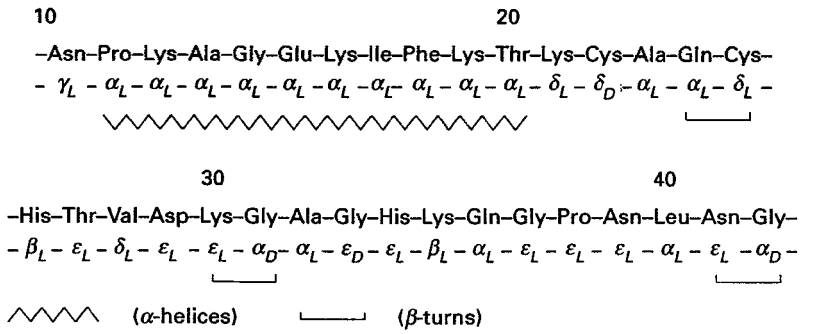
		ϕ	ψ	$\phi_{Top.}$	$\psi_{Top.}$	σ_ϕ	σ_ψ	n
α_D	For–*Xxx*–NH_2[a]	55·3	40·3	55·3	40·3	6·9	4·6	4
	For–Ser–NH_2	57·3	42·3	57·3	42·3	6·3	7·4	4
	For–Ala_2–NH_2	61·8	31·9	61·8	31·9	1·8	5·9	12
α_L	For–*Xxx*–NH_2	—	—	—	—	—	—	0
	For–Ser–NH_2	− 68·3	− 30·5	291·7	329·5	4·2	8·7	3
	For–Ala_2–NH_2	− 68·6	− 17·5	291·4	342·5	—[b]	—	1
β_L	For–*Xxx*–NH_2	157·7	162·8	202·3	162·8	16·5	12·6	5
	For–Ser–NH_2	165·1	173·8	194·9	173·8	14·5	8·5	5
	For–Ala_2–NH_2	167·6	169·9	192·4	169·9	2·6	1·1	14
δ_D	For–*Xxx*–NH_2	− 179·3	− 42·1	180·7	317·9	31·1	11·8	5
	For–Ser–NH_2	− 173·4	− 53·3	186·6	306·5	19·0	10·8	6
	For–Ala_2–NH_2	− 179·6	− 43·7	180·4	316·3	9·1	3·5	14
δ_L	For–*Xxx*–NH_2	− 127·9	29·8	232·1	29·8	3·9	3·6	4
	For–Ser–NH_2	− 129·3	30·6	230·7	30·6	12·4	5·5	5
	For–Ala_2–NH_2	− 126·2	26·5	233·8	26·5	12·1	9·4	12
ε_D	For–*Xxx*–NH_2	72·5	166·9	72·5	166·9	3·4	10·7	4
	For–Ser–NH_2	68·6	− 154·7	68·6	205·3	16·6	31·3	6
	For–Ala_2–NH_2	64·7	− 178·6	64·7	182·4	3·8	24·0	13
ε_L	For–*Xxx*–NH_2	—	—	—	—	—	—	0
	For–Ser–NH_2	—	—	—	—	—	—	0
	For–Ala_2–NH_2	− 74·7	167·8	285·3	167·8	3·2	3·7	3
γ_D	For–*Xxx*–NH_2	70·9	− 52·7	70·9	307·3	8·7	11·8	5
	For–Ser–NH_2	69·8	− 47·9	69·8	312·1	7·6	11·7	9
	For–Ala_2–NH_2	74·3	− 59·5	74·3	300·5	1·1	8·4	14
γ_L	For–*Xxx*–NH_2	− 84·3	67·3	275·7	67·3	0·6	2·7	5
	For–Ser–NH_2	− 82·2	67·5	277·8	67·5	3·7	5·6	6
	For–Ala_2–NH_2	− 84·5	68·7	275·5	68·7	2·0	3·4	15

$$\sigma = \left[\sum_{i=1}^{n} (x_i - \underline{x}_i)^2\right]^{1/2},$$

$$\underline{x}_i = \left[n^{-1} \sum_{i=1}^{n} (x_i)\right].$$

[a] The conformational parameters of the For–*Xxx*–NH_2 are obtained from the averaging of the For–Gly–NH_2, For–Ala–NH_2 and For–Val–NH_2 structual data.

[b] No σ was calculated since only one conformer was evaluated.



Scheme 18.

Table 11. The 10 marker interproton distances in the 49 different conformers of For–Ala–Ala–NH_2.

	Backbone proton distances[a]									
	NH_{i+1} H^{α}_{i+1}	NH_{i+1} NH_{i+2}	NH_{i+1} H^{α}_{i+2}	NH_{i+1} NH_{i+3}	NH_{i+2} H^{α}_{i+1}	NH_{i+2} H^{α}_{i+2}	NH_{i+2} NH_{i+3}	NH_{i+3} H^{α}_{i+2}	NH_{i+3} H^{α}_{i+1}	H^{α}_{i+1} H^{α}_{i+2}
Abbreviation[b]	$d_{N\alpha}$	d_{NN}			$d_{\alpha N}$	$d_{N\alpha}$	d_{NN}	$d_{\alpha N}$		
$\alpha_D\alpha_D$	2·27	2·88	5·12	4·48	2·91	2·25	2·82	3·09	4·48	4·84
$\alpha_D\beta_L$	2·27	2·81	5·40	7·10	2·90	2·85	4·29	2·51	6·61	4·38
$\alpha_D\delta_D$	2·25	2·84	5·33	4·28	2·99	2·64	2·35	3·59	4·78	4·64
$\alpha_D\delta_L$	2·29	2·87	5·40	4·99	2·81	2·93	2·72	2·92	3·70	4·45
$\alpha_D\gamma_D$	2·28	2·80	5·02	5·53	2·80	2·24	3·64	3·59	4·61	4·65
$\alpha_D\gamma_L$	2·27	2·83	5·35	5·66	2·90	2·89	3·70	2·43	4·00	4·60
$\alpha_L\delta_L$	2·82	2·83	5·39	4·39	3·44	2·94	2·57	3·10	3·81	4·44
$\beta_L\alpha_D$	2·85	4·28	5·67	4·52	2·49	2·26	2·84	2·88	3·56	4·57
$\beta_L\beta_L$	2·84	4·31	4·69	7·15	2·55	2·84	4·29	2·52	6·60	4·62
$\beta_L\delta_D$	2·84	4·28	4·83	5·96	2·50	2·79	2·28	3·60	4·09	4·65
$\beta_L\delta_L$	2·84	4·27	4·42	5·03	2·46	2·94	2·64	2·94	4·40	4·43
$\beta_L\varepsilon_D$	2·83	4·24	5·65	6·56	2·43	2·25	4·60	2·76	6·32	4·54
$\beta_L\gamma_D$	2·85	4·24	5·71	3·74	2·40	2·22	3·66	3·59	4·31	4·54
$\beta_L\gamma_L$	2·84	4·28	4·50	3·93	2·50	2·90	3·71	2·43	4·75	4·33
$\delta_D\alpha_D$	2·71	2·30	4·15	4·76	3·55	2·26	2·80	2·90	4·46	5·33
$\delta_D\beta_L$	2·79	2·19	3·95	5·81	3·60	2·86	4·31	2·53	7·05	4·66
$\delta_D\delta_D$	2·71	2·26	3·89	4·51	3·54	2·82	2·28	3·61	5·51	4·72
$\delta_D\varepsilon_D$	2·74	2·33	4·09	6·42	3·52	2·26	4·60	2·83	6·69	5·33
$\delta_D\varepsilon_L$	2·78	2·27	4·00	5·01	3·53	2·93	4·52	2·54	6·49	4·61
$\delta_D\gamma_D$	2·72	2·25	94	4·96	3·51	2·24	3·64	3·60	4·23	5·31
$\delta_D\gamma_L$	2·72	2·27	4·52	4·10	3·55	2·89	3·74	2·42	4·16	4·64
$\delta_L\alpha_D$	2·94	2·67	4·72	3·53	2·94	2·27	2·86	2·88	4·75	4·85
$\delta_L\beta_L$	2·94	2·50	5·00	6·65	3·09	2·83	4·29	2·51	6·65	4·36
$\delta_L\delta_D$	2·94	2·62	5·08	3·69	2·92	2·78	2·28	3·59	5·11	4·30
$\delta_L\varepsilon_D$	2·90	2·49	3·96	5·10	2·52	2·26	4·57	2·80	6·60	4·66
$\delta_L\gamma_D$	2·94	2·54	4·67	4·77	2·98	2·23	3·67	3·59	4·58	4·81
$\delta_L\gamma_L$	2·94	2·57	4·60	5·42	3·00	2·91	3·67	2·44	3·96	4·64
$\varepsilon_D\alpha_D$	2·23	3·66	5·64	5·69	3·57	2·27	2·82	2·86	4·41	5·32
$\varepsilon_D\beta_L$	2·26	4·60	5·55	7·91	2·73	2·83	4·28	2·51	6·66	4·64
$\varepsilon_D\delta_D$	2·26	4·65	5·61	6·34	3·06	2·73	2·27	3·59	4·35	4·79
$\varepsilon_D\delta_L$	2·26	4·62	5·38	4·71	3·31	2·93	2·60	3·02	4·60	4·45
$\varepsilon_D\varepsilon_D$	2·26	4·48	6·14	7·06	2·58	2·24	4·63	2·85	6·16	4·56
$\varepsilon_D\gamma_D$	2·26	4·50	6·26	4·52	2·60	2·21	3·67	3·59	4·14	4·66
$\varepsilon_D\gamma_L$	2·26	4·62	5·26	4·93	2·79	2·91	3·63	2·46	4·69	4·27
$\gamma_D\beta_L$	2·24	3·62	5·12	7·25	3·63	2·89	4·38	2·49	6·94	4·54
$\gamma_D\delta_D$	2·23	3·62	5·14	5·69	3·57	2·84	2·14	3·60	5·43	4·67
$\gamma_D\delta_L$	2·23	3·64	5·41	3·96	3·57	2·95	2·57	3·03	4·43	4·47
$\gamma_D\varepsilon_D$	2·23	3·67	5·73	7·76	3·55	2·26	4·60	2·68	6·70	5·31
$\gamma_D\varepsilon_L$	2·25	3·61	5·24	6·40	3·56	2·94	4·56	2·47	6·47	4·63
$\gamma_D\gamma_D$	2·23	3·63	5·50	5·64	3·54	2·23	3·68	3·59	4·10	5·31
$\gamma_D\gamma_L$	2·23	3·66	5·61	4·80	3·58	2·91	3·70	2·44	4·09	4·58
$\gamma_L\alpha_D$	2·90	3·67	5·52	3·62	2·42	2·27	2·77	2·90	4·33	4·52
$\gamma_L\beta_L$	2·91	3·78	5·46	7·50	2·44	2·88	4·36	2·48	6·39	4·31
$\gamma_L\delta_D$	2·90	3·82	5·69	4·72	2·26	2·82	1·94	3·62	3·95	4·41
$\gamma_L\delta_L$	2·90	3·68	5·12	5·73	2·40	2·95	2·55	3·02	3·46	4·40
$\gamma_L\varepsilon_D$	2·93	3·61	5·26	6·23	2·30	2·26	4·56	2·75	6·50	4·47
$\gamma_L\varepsilon_L$	2·90	3·70	5·06	7·29	2·43	2·92	4·66	2·36	6·19	4·54
$\gamma_L\gamma_D$	2·90	3·64	5·59	4·50	2·38	2·23	3·68	3·59	4·65	4·42
$\gamma_L\gamma_L$	2·90	3·69	4·91	5·55	2·41	2·91	3·71	2·45	4·26	4·53

[a] Backbone interproton distances calculated by *ab initio* (HF/3–21G) geometry optimization.
[b] Standard NMR type abbreviations of the marker distances.

of the method. In fact, it is diagnostic that some unusual stabilizing or destabililzing interaction is operative at that particular amino acid residue. Therefore we think that 'ghost' structures, assigned in proteins (figure 13) are diagnostic of special intra- or intermolecular interactions.

(2) the increase of the polypeptide chain length as to considering the For–$(Ala)_i$–NH_2 $(1 \leqslant i \leqslant 4)$ molecule, the average $\{\phi, \psi\}$ torsional angle pairs do not change dramatically with i. The increase in basis set size and/or the inclusion of electron correlation did not result in significant conformational changes as demonstrated previously. The observed alteration on the $\{\phi, \psi\}$ PES was small for selected backbone structures. Certainly the precise determination of the nine conformational centres depends upon the applied basis set and could well be modified in the future, when high computational power is available, but the basis concept of the conformational assignment will remain intact;

(3) all the 'conformational centres' were determined on the basis of *ab initio* calculations performed on For–$(Ala)_i$–NH_2 $(1 \leqslant i \leqslant 4)$ molecules. It is obvious that the very simple side chain of such a model compound cannot be representative of the large variety of polar and apolar side chains. However, polyalanine models are the best systems for the investigation of purely backbone/backbone type interactions. From this perspective all backbone/side-chain interactions can be treated as indicative backbone modifying effects. Consequently, when the conformation of an amino acid residue is classified by the ACAP algorithm, and a noticeable deviation is observed, we may be certain that the resulted deviation reflects, as a global indicator, all types of interactions that are exerted on the backbone of the protein.

Acknowledgments

First of all the authors would like to express their gratitude to their co-authors in previous publications related to the present review: János Ángyán, Pál Császár, Gábor Endrédi, Ödön Farkas, János Ladik, Márton Kajtár, John-Frank Marcoccia, Michael A. McAllister, Jean-Luis Rivail and Wladia Viviani in alphabetical order.

The continued financial support of the NSERC of Canada is gratefully acknowledged.

This research was also supported in part by a grant from the Hungarian Scientific Research Foundation (OTKA no. III-2245 and F 013799), as well as by CFA. For a free copy of the ACAP program ask the authors.

References

[1] Howard, J. C., Ali, A., Scheraga, H. A., and Momany, F. A., 1975, *Macromol.*, **8,** 607.
[2] Wolfe, S., Bruder, S., Weaver, D. F., and Yang, K., 1988, *Can. J. Chem.*, **66,** 2703.
[3] Paine, G. H., and Scheraga, H. A., 1986, *Biopolymers*, **25,** 1547.
[4] Ripoll, D. R., and Scheraga, H. A., 1988, *Biopolymers*, **27,** 1283.
[5] Némethy, G., Miller, M. H., and Scheraga, H. A., 1980, *Macromol.*, **13,** 914.
[6] Schulz, G. E., 1988, *Ann. Rev. biophys. Chem.*, **17,** 1.
[7] Bruccoleri, R. E., and Karplus, M., 1987, *Biopolymers*, **26,** 137.
[8] Lambert, M. H., and Scheraga, H. A., 1989, *J. Comput. Chem.*, **10,** 798.
[9] Lambert, M. H., and Scheraga, H. A., 1989, *J. Comput. Chem.*, **10,** 770.
[10] Lambert, M. H., and Scheraga, H. A., 1989, *J. Comput. Chem.*, **10,** 817.
[11] Liquori, A. M., 1969, *Quant. Rev. Biophys.*, **2,** 65.
[12] Gething, M.-J., and Sambrook, J., 1992, *Nature*, **335,** 33.
[13] Lattmann, E. E., and Rose, G. D., 1993, *Proc. natl. Acad. Sci.*, **90,** 439.
[14] Mirsky, A. E., and Pauling, L., 1936, *Proc. natl. Acad. Sci.*, **22,** 439.

[15] AUBRY, A., MARRAUD, M., PROTAS, J., and NEEL, J., 1974, *C. R. Acad. Sci. Paris* c, **287,** 163.
[16] AUBRY, A., GHERMANI, M., and MARRAUD, M., 1984, *Int. J. Peptide, Protein Res.*, **23,** 113.
[17] STROUP, A. N., COLE, L. B., DHINGRA, M. M., and GIERACH, L. M., 1990, *Int. J. Peptide, Protein Res.*, **36,** 531.
[18] PERCZEL, A., FOXMAN, B. M., and FASMAN, G. D., 1992, *Proc. nat. Acad. Sci.*, **89,** 8210.
[19] RAJ. P. A., SONI, S. D., RAMASUBBU, N., BHANDARY, K. K., and LEVINE, M. J., 1990, *Biopolymers*, **30,** 73.
[20] KARLE, I. L., 1977, *Peptide Structure and Biological Function*, edited by E. Gross and J. Meienhofer (New York: Wiley), p. 681.
[21] KARLE, I. L., 1981, *Peptides*, edited by E. Gross and J. Meienhofer (New York: Academic), p. 1.
[22] KARLE, I. L., KARLE, J., MASTROPAOLO, D., CAMERMAN, A., and CAMERMAN, N., 1983, *Peptides. Structure and Function*, edited by V. J. Hruby and D. H. Rich (Rockford, Illinois: Pierce Chem. Co.), p. 291.
[23] KAISER, E. T., and KEZDY, F. J., 1984, *Science*, **223,** 249.
[24] WEBER, C., WIDER, G., VAN FREYBERG, B., TRABER, R., BRAUN, W., WIDMER, H., and WÜTRICH, K., 1991, *Biochem.*, **30,** 3563.
[25] FESIK, S. W., GAMPE, R. T., EATON, H. L., GEMMECKER, G., OLEJNICZAK, E. T., NERI, P., HOLYMAN, T. F., EGAN, D. A., EDALJI, R., SIMMER, R., HELFRICH, R., HOCHLOWSKI, J., and JACKSON, M., 1991, *Biochem.*, **30,** 6574.
[26] VENKATACHALAM, C. M., 1968, *Biopolymers*, **6,** 1425.
[27] RAMACHANDRAN, G. N., RAMAKRISHNAN, C., SASISEKHARAN, V., 1963, *J. molec. Biol.*, **7,** 95.
[28] WÜTHRICH, K., 1986, *NMR of Protein and Nucleic Acids* (New York: Wiley).
[29] WÜTHRICH, K., 1990, *J. biol. Chem.*, **265,** 22059.
[30] DYSON, H. J., RANCE, M., HOUGHTEN, R. A., LERNER, A., and WRIGHT, P. E., 1988, *J. molec. Biol.*, **201,** 161.
[31] DAYSON, H. J., and WRIGHT, P. E., 1991, *Ann. Rev. biophys. Chem.*, **20,** 519.
[32] WILLIAMSON, M. P., 1992, *Methods in Molecular Biology*, edited by C. Jones, B. Mulloy and A. H. Thomas (New Jersey: Humana Press), vol. 7.
[33] WILLIAMSON, M. P., and WALTHO, J. P., 1992, *Chem. Soc. Rev.*, **227.**
[34] WISHART, D. S., SYKES, B. D., and RICHARDS, F. M., 1991, *FEBS*, **293,** 72.
[35] WISHART, D. S., SYKES, B. D., and RICHARDS, F. M., 1991, *J. molec. Biol.*, **222,** 311.
[36] WISHART, D. S., SYKES, B. D., and RICHARDS, F. M., 1992, *Biochem.*, **31,** 1647.
[37] BRAUN, W., 1987, *Quart Rev. Biophys.*, **19,** 115.
[38] WALTHO, J. P., FEHER, V. A., MERUTKA, G., DYSON, H. J., and WRIGHT, P. E., 1993, *Biochem*, **32,** 6337.
[39] WALTHO, J. P., FEHER, V. A., and WRIGHT, P. E., 1990, *Current Research in Protein Chemistry*, edited by J. J. Villafranca (New York: Academic), p. 283.
[40] DYSON, H. J., MERUTKA, G., WALTHO, J. P., LERNER, R. A., and WRIGHT, P. E., 1992, *J. molec. Biol.*, **226,** 795.
[41] DYSON, H. J., SAYRE, J. R., MERUTKA, G., SHIN, H. C., LERNER, R. A., and WRIGHT, P. E., 1992, *J. molec. Biol.*, **226,** 819.
[42] NOGGLE, J. H., and SCHIRMER, R. E., 1971, *The Nuclear Overhauser Effect, Chemical Applications* (New York: Academic).
[43] NEUHAUS, D., and WILLIAMSON, M., 1989, *The Nuclear Overhauser Effect in Structural and Conformational Analysis* (Weinheim: VCH).
[44] ATAKA, S., TAKEUCHI, H., and TASUMI, M. J., 1984, *J. molec. Struct.*, **113,** 147.
[45] DRAKENBERG, T., and FORSEN, S., 1971, *J. chem. Soc., Chem. Commun.*, 1404.
[46] HAGLER, A. T., LIFSON, S., and DAUBER, P., 1979, *J. Am. chem. Soc.*, **101,** 5121.
[47] WEINER, S. J., KOLLMAN, P. A., CASE, D. A., SINGH, U. C., GHIO, C., ALAGONA, G., PROFETA, S. JR.., and WEINER, P., 1984, *J. Am. chem. Soc.*, **106,** 765.
[48] WEINER, S. J., SINGH, U. C., O'DONNELL, T. J., and KOLLMAN, P. A., 1984, *J. Am. chem. Soc.*, **106,** 6243.
[49] BROOKS, B. B., BRUCCELORI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., and KARPLUS, M., 1983, *J. comput. Chem.*, **4,** 187.
[50] VÁSQUEZ, M., and SCHERAGA, H. A., 1988, *J. biomol, Struct. Dyn.*, **5,** 705.

[51] NÉMETHY, G., and SCHERAGA, H. A., 1965, *Biopolymers*, **3,** 155.
[52] MOMANY, F. A., MCGUIRE, R. F., BURGESS, A. W., and SCHERAGA, H. A., 1975, *J. phys. Chem.*, **79,** 2361.
[53] SCHAFER, L., NEWTON, S. Q., CAO, M., PEETERS, A., VANALSENOY, C., WOLINSKI, K., and MOMANY, F. A., 1993, *J. Am. chem. Soc.*, **115,** 272.
[54] (*a*) PETERSON, M. R., and CSIZMADIA, I. G., 1982, *Progress of Theoretical Organic Chemistry* (Amsterdam: Elsevier), 3, 190; (*b*) MEZEY, P. G., (editor), 1987, *Potential Energy Hypersurfaces* (Amsterdam: Elsevier).
[55] CSIZMADIA, I. G., and BERTRAN, J. D., (editors), 1989, *New Theoretical Concept for Understanding Organic Reactions* (Dordrecht: Reidel), p. 1.
[56] PERCZEL, A., ANGYAN, J. G., KAJTAR, M., VIVIANI, W., RIVAIL, J.-L., MARCOCCIA, J.-F., and CSIZMADIA, I. G., 1991, *J. Am. chem. Soc.*, **113,** 6256.
[57] PERCZEL, A., KAJTÁR, M., MARCOCCIA, J.-F., and CSIZMADIA, I. G., 1991, *J. molec. Struct. (Theochem)*, **232,** 291.
[58] VIVIANI, W., RIVAIL, J.-L., PERCZEL, A., and CSIZMADIA, I. G., 1993, *J. Am. chem. Soc.*, **115,** 8321.
[59] KARLE, I. L., 1989, *Biopolymers*, **28,** 1.
[60] ROSE, G. D., GIERASCH, L. M., and SMITH, J. A., 1985, *Advances in protein chemistry*, **37,** 1.
[61] AUBRY, A., MARRAUD, M., CUNG, M. T., and PROTAS, J., 1985, *Comp. Acad. Sci. Paris*, c, **280,** 861.
[62] BOUSSARD, G., MARRAUD, M., and AUBRY, A., 1986, *Int. J. Peptide, Protein Res*, **28,** 508.
[63] BOUSSARD, G., MARRAUD, M., and AUBRY, A., 1979, *Biopolymers*, **18,** 1297.
[64] BOUSSARD, G., and MARRAUD, M., 1985, *J. Am. chem. Soc.*, **107,** 1825.
[65] PANNEERSELVAM, K., and CHACKO, K. K., 1990, *Int. J. Peptide, Protein Res.*, **35,** 460.
[66] BRAKMACHARI, S. K., BHAT, T. N., SUDHAKAR, V., VIJAYAN, M., RAPAKA, R. S., BHATNAGAR, R. S., and ANANTHANARAYANAN, V. S., 1981, *J. Am. chem. Soc.*, **103,** 1703.
[67] ANANTHANARAYANAN, V. S., and CAMERON, T. S., 1988, *Int. J. Peptide, Protein Res.*, **31,** 399.
[68] SHAFTER, L., SELLERS, H. L., LOVAS, F. J., and SUENRAM, R. D., 1980, *J. Am. chem. Soc.*, **102,** 6566.
[69] BONACCORSI, R., PALLA, P., and TOMASI, J., 1984, *J. Am. chem. Soc.*, **106,** 1945.
[70] SUKUMAR, N., and SEGAL, G. A., 1986, *J. Am. chem. Soc.*, **108,** 6880.
[71] SCHAFER, L., VAN ALSENOY, C., and SCARSDALE, J. N., 1982, *J. chem. Phys.*, **76,** 1439.
[72] SCHAFER, L., KLIMKOWSKI, V. J., MOMANY, F. A., CHUMAN, H., and VAN ALSENOY, C., 1984, *Biopolymers*, **23,** 2335.
[73] KLIMKOWSKI, V. J., SCHAFER, L., MOMANY, F. A., and VAN ALSENOY, C., 1985, *J. molec. Struct.*, **124,** 143.
[74] SCARSDALE, J. N., VAN ALSENOY, C., KLIMKOWSKI, V. J., SCHAFER, L., and MOMANY, F. A., 1983, *J. Am. chem. Soc.*, **105,** 3438.
[75] (*a*) HEAD GORDON, T., HEAD GORDON, M., FRISCH, M. J., BROOKS II, C., and POPLE, J. A., 1989, *Int. J. quant. Chem. Quantum Biology Symposium*, **16,** 311; (*b*) HEAD GORDON, T., HEAD GORDON, M., FRISCH, M. J., BROOKS II, C., and POPLE, J. A., 1991, *J. Am. chem. Soc.*, **113,** 5989.
[76] BOHM, H.-J., and BRODE, S., 1991, *J. Am. chem. Soc.*, **113,** 7129.
[77] SIAM, K., KLIMKOWSKI, V. J., VAN ALSENOY, C., EWBANK, J. D., and SCHAFER, L., 1987, *J. molec. Struct.*, **152,** 261.
[78] SIAM, K., KULP, S. Q., EWBANK, J. D., SCHAFER, L., and VAN ALSENOY, C., 1989, *J. Molec. Struct.*, **184,** 143.
[79] ENDREDI, G., PERCZEL, A., MCALLISTER, M. A., CSONKA, G., LADIK, J., and CSIZMADIA, I. G., 1994, (in preparation).
[80] MCALLISTER, M. A., PERCZEL, A., CSÁSZÁR, P., VIVIANI, W., RIVAIL, J.-L., and CSIZMADIA, I. G., 1994, *J. molec. Struct.*
[81] (*a*) PERCZEL, A., DAUDEL, R., ÁNGYÁN, J., and CSIZMADIA, I. G., 1990, *Can. J. Chem.*, **68,** 1182; (*b*) PERCZEL, A., FARKAS, Ö., MARCOCCIA, J. F., and CSIZMADIA, I. G., 1994, *J. Am. chem. Soc.* (submitted for publication); (*c*) PERCZEL, Á., FARKAS, Ö., and CSIZMADIA, I. G., 1994, *J. Am. chem. Soc.* (in press).
[82] ENDREDI, G., LIEGNER, C.-M., MCALLISTER, M. A., PERCZEL, A., LADIK, J., and CSIZMADIA, I. G., 1994, *J. molec. Struct. (Theochem)*, **306,** 1.

[83] McAllister, M. A., Perczel, A., Császár, P., and Csizmadia, I. G., 1993, *J. molec. Struct. (Theochem)*, **288,** 181.
[84] (*a*) Perczel, A., McAllister, M. A., Csaszar, P., and Csizmadia, I. G., 1993, *J. Am. chem. Soc.*, **115,** 4849; (*b*) Perczel, A., McAllister, M. A., Csaszar, P., and Csizmadia, I. G., 1994, *Can. J. Chem.*
[85] Perczel, A., Farkas, Ö., and Csizmadia, I. G., 1994, *J. Am. chem. Soc.* (submitted for publication).
[86] Mezey, P. G., 1981, *Chem. Phys. Lett.*, **82,** 100.
[87] Mezey, P. G., 1982, *Chem. Phys. Lett.*, **86,** 562.
[88] Gierasch, L. M., Deber, C. M., Madison, V., Niu, C. H., and Blout, E. R., 1981, *Biochem.*, **20,** 4730.
[89] Bruch, M. D., Noggle, J. H., and Gierasch, L. M., 1985, *J. Am. Chem. Soc.*, **107,** 1400.
[90] Wright, P. E., Dyson, H. J., and Lerner, R. A., 1988, *Biochem.*, **27,** 7167.
[91] Stradley, S. J., Rizo, J., Bruch, M. D., Stroup, A. N., and Gierasch, L. M., 1990, *Biopolymers*, **29,** 263.
[92] Perczel, A., Hollósi, M., Foxman, B. M., and Fasman, G. D., 1991, *J. Am. chem. Soc.*, **113,** 9772.
[93] Deber, C. M., Torchia, D. A., and Blout, E. R., 1971, *J. Am. chem. Soc.*, **93,** 4893.
[94] Cerrini, S., Gavuzzo, E., Lucente, G., Luisi, G., Pinnen, F., and Radics, L., 1991, *Int. J. Pept. Prot. Res.*, **38,** 289.
[95] Falcomer, C. M., Meinwald, Y. C., Choudhary, I., Talluri, S., Miburn, P. J., Clardy, J., and Scheraga, H. A., 1992, *J. Am. chem. Soc.*, **114,** 4036.
[96] Bandekar, J., Evans, D. J., Krimm, S., Leach, S. J., Lee, S., McQuie, J. R., Minasian, E., Nemethy, G., Pottle, M. S., Scheraga, H. A., Stimson, E. R., and Woody, R. W., 1982, *Int. J. Peptide, Protein Res.*, **19,** 187.
[97] Stradley, S. J., Rizo, J., Bruch, M. D., Stroup, A. N., and Gierasch, L. M., 1990, *Biopolymers*, **29,** 263.
[98] Torchia, D. A., Wong, S. C. K., Deber, C. M., and Blout, E. R., 1972, *J. Am. chem. Soc.*, **94,** 616.
[99] Kessler, H., Klein, M., and Wagner, K., 1988, *Int. J. Peptide, Protein Res.*, **31,** 481.
[100] Otter, A., Scott, P. G., Xiaohoong, L., and Kotovych, G., 1989, *J. biomol. Struct. Dynam*, **7,** 455.
[101] Xiaohoong, L., Scott, P. G., Otter, A., and Kotovych, G., 1992, *Biopolymers*, **32,** 119.
[102] Perczel, A., and Fasman, G. D., 1992, *Protein. Sci.*, **1,** 210.
[103] Ciesla, D. J., Gilbert, D. E., and Feigon, J., 1991, *J. Am. chem. Soc.*, **113,** 3957.
[104] Kessler, H., Anders, U., and Schudok, M., 1991, *J. Am. chem. Soc.*, **112,** 5908.
[105] Pons, M., and Giralt, E., 1991, *J. Am. chem. Soc.*, **114,** 5049.
[106] Perczel, A., and Csizmadia, I. G., 1993, *J. molec. Struct.*, **286,** 75.
[107] Richardson, J. S., 1981, *Adv. Protein Chem.*, **34,** 167.
[108] Richardson, J. S., and Richardson, D. C., 1988, *Science*, **240,** 1648.
[109] Presta, L. G., and Rose, G. D., 1988, *Science*, **240,** 1632.
[110] Perczel, A., Endrédi, G., McAlister, M. A., Farkas, Ö., Csaszár, P., Ladik, J., and Csizmadia, I. G., 1994, *J. Am. chem. Soc.* (submitted for publication).
[111] Cheng, M., McGovern, M. E., Jin, T., Zhao, Da-C., McAllister, M. A., Farkas, Ö., Perczel, A., Csaszar, P., and Csizmadia, I. G., 1993, *J. molec. Struct.*, **309,** 151.
[112] Liegner, C.-M., Endredi, G., McAllister, M. A., Perczel, A., Ladik, J., and Csizmadia, I. G., 1993, *J. Am. chem. Soc.*, **115,** 8275.
[113] Schäfer, L., Newton, S. Q., Cao, M., Peeters, A., van Alsenoy, C., Wolinski, K., Momany, F. A., 1993, *J. Am. Chem. Soc.*, **115,** 272..
[114] Chou, P. Y., and Fasman, G. D., 1977, *J. molec. Biol.*, **115,** 135.
[115] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. L., Kennard, O., Shimanouchi, T., and Tasumi, M., 1977, *Eur. J. Biochem.*, **80,** 319, and 1977, *J. molec. Biol.*, **112,** 535.
[116] Karle, I. L., 1992, *Acta crystallogr.* B, **48,** 341.
[117] Blundell, T. L., and Johnson, L. N., 1976, *Protein Crystallography* (New York: Academic).
[118] Nobel, K., Altmann, E., Mutter, M., Bardi, R., Piazzesi, A. M., Crisma, M., Bonora, G. M., and Toniolo, C., 1991, *Biopolymers*, **31,** 1135.

[119] KARLE, I. L., FLIPPEN-ANDERSON, J. L., SUKUMAR, M., UMA, K., and BALARAM, P., 1991, *J. Am. chem. Soc.*, **113,** 3952.
[120] FISHER, E., and SUZUKI, U., 1905, *Ber.*, **38,** 4173.
[121] FISHER, E., and JACOBS, W. A., 1906, *Ber.*, **39,** 2942.
[122] PAVONE, V., LOMBARDI, A., D'AURIA, G., SAVINO, M., NASTRI, F., PAOLILLO, L., DIBLASIO, B., and PEDONE, C., 1992, *Biopolymers*, **32,** 173.
[123] PAVONE, V., LOMBARDI, A., YANG, X., PEDONE, C., and DIBLASIO, B., 1990, *Biopolymers*, **30,** 189.
[124] DIBLASIO, B., LOMBARDI, A., YANG, X., PEDONE, C., and PAVONE, V., 1991, *Biopolymers*, **31,** 1181.
[125] SCHWYZER, R., SIEBER, P., and GORUP, B., 1958, *Chimia*, **12,** 90.
[126] KARPLUS, M., 1959, *J. phys. Chem.*, **30,** 11.
[127] BYSTROV, V. F., 1976, *Prog. Nucl. Magn. Reson. Spectrosc.*, **10,** 41.
[128] PARDI, A., BILLETER, M., and WUTHRICH, K., 1984, *J. molec. Biol.*, **180,** 741.
[129] LINAS, M., and KLEIN, M. P., 1975, *J. Am. chem. Soc.*, **97,** 4731.
[130] MARRAUD, M., and AUBRY, A., 1984, *Int. J. Peptide, Protein Res.*, **23,** 123.
[131] OLSEN, H. B., GESMAR, H., and LED, J. L., 1992, *J. Am. chem. Soc.*, **115,** 1456.
[132] LIANG, G. B., RITO, C. J., and GELLMAN, S. H., 1992, *J. Am. chem. Soc.*, **114,** 4440.
[133] LIANG, G. B., RITO, C. J., and GELLMAN, S. H., 1992, *Biopolymers*, **32,** 293.
[134] LEWIS, N. P., MOMANY, F. A., and SCHERAGA, H. A., 1973, *Israel J. Chem.*, **11,** 121.
[135] ZIMMERMAN, S. S., POTTLE, M. S., NEMETHY, G., and SCHERAGA, H. A., 1977, *Macromolecules*, **10,** 1.
[136] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1977, *Proc. natl. Acad. Sci. USA*, **74,** 4126.
[137] ZIMMERMAN, S. S., POTTLE, M. S., NÉMETHY, G., and SCHERAGA, H. A., 1977, *Macromol.*, **10,** 1.
[138] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1977, *Biopolymers*, **16,** 811.
[139] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1978, *Biopolymers*, **17,** 1849.
[140] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1978, *Biopolymers*, **17,** 1886.
[141] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1978, *Biopolymers*, **17,** 1871.
[142] ZIMMERMAN, S. S., and SCHERAGA, H. A., 1976, *Macromol.*, **9,** 408.
[143] PERCZEL, A., HOLLÓSI, M., TUSNÁDY, G., and FASMAN, G. D., 1991, *Protein Eng.*, **4,** 669.
[144] PERCZEL, A., 1993, *Recent Experimental and Computational Advances in Molecular Spectroscopy* (Dordrecht: Kluwer), p. 63.
[145] PERCZEL, H., HUDÁKY, P. (in preparation).
[146] SZÉKELY, Z., PERCZEL, A., PENKE, B., and MOLNÁR, J., 1993, *J. molec. Struct.*, **286,** 165.