Journal of Peptide Science, Vol. 2, 364–370 (1996)

Towards the Understanding of the Folding of Methylene Units in the Glutamine Residue

CARLOS ALEMÁN[1], M. CRISTINA VEGA[2], ELOÍSA NAVARRO[1] and JORDI PUIGGALÍ[1],

[1]Departament d'Enginyeria Química, ETS d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Barcelona, Spain
[2]Departament de Biologia Molecular i Cel.lular, Centre d'Investigació i Desenvolupament–CSIC, Barcelona, Spain

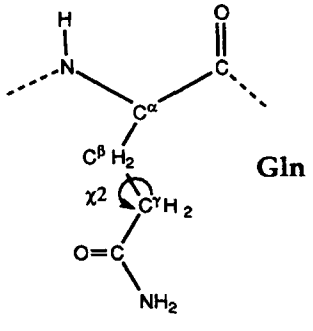
Received 8 January 1996
Accepted 18 March 1996

Abstract: The conformational preferences of the methylenic sequence in the side chain of the glutamine residue were investigated by *ab initio* and semi-empirical quantum mechanical calculations and examination of both the Brookhaven Protein Databank and Cambridge Structural Data Base. The results were analysed on the basis of our previous findings about the folding of methylene groups in aliphatic segments. Both energy calculations and the crystallographic structure of small peptides indicate that methylene units of the glutamine residue tend to fold in a *gauche* conformation. In contrast, such groups usually adopt an all-*trans* conformation in proteins due basically to the entropic and solvent contributions. These results have been demonstrated by computing the entropic correction to the free energy and evaluating the solvent effects through SCRF calculations

Keywords: conformation; glutamine; folding; simulation

INTRODUCTION

Recent studies using X-ray crystallography and quantum mechanical calculations revealed that central methylene units of the aliphatic segment in succinamide, glutaramide and adipamide analogues do not keep a *trans* (**T**) conformation and tend to fold in a *gauche* (**G**) conformation [1–3]. The overall results suggest that the carbonyl group induces the rotation towards the **G** conformation of the bond defined by the first and second carbon atoms next to the carbonyl carbon. This has been supported by a recent study about the conformational preferences of a set of ketones [4], which indicates that the carbonyl group plays a leading role in this pattern of folding rather than the amide group. These results are very important for biorganic chemistry, since in the solid state the methylene segments are usually considered in *all-trans* conformation. In this work, we try to apply our knowledge about the folding of methylene units to understand some unusual conformational features found in the structure of peptides and proteins. More specifically, we are particularly interested in the conformational preferences of the χ_2 dihedral angle of the side chain of glutamine (Gln) residue, which has two methylene units before the carbonyl group.



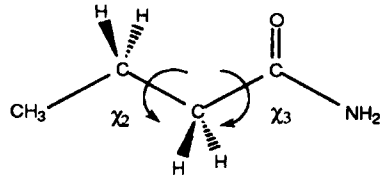
Abbreviations: AM1, Austin Model 1; SCRF, self-consistent reaction field.

Address for correspondence: Drs Jordi Puiggalí or Carlos Alemán, Departament d'Enginyeria Química, ETS d'Enginyers Industrials de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, E-08028, Spain. Tel. (34-3) 4016684; fax (34-3) 4016600.

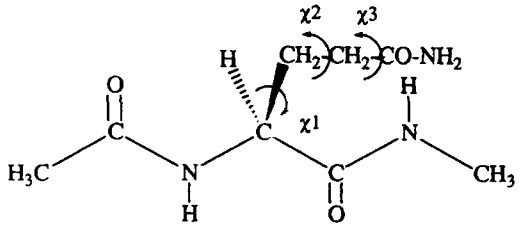
CCC 1075-2617/96/060364-07

Conformational energy studies of the Gln residue using methods based on empirical potentials indicate that the χ_2 dihedral angle, which is defined by the sequence $C^{\alpha}-C^{\beta}-C^{\gamma}-C(=O)$, either adopts the **T** as unique conformation in all the minimum energy structures [5] or has a **G** minimum considerably unfavoured with respect to the **T** [6,7]. This makes it difficult to rationalize the relatively frequent **G** conformations found in peptides and proteins by X-ray crystallography, this feature usually being attributed to the mobility of the Gln side chains [8,9] and to the formation of inter- or intramolecular hydrogen bonds [10]

In order to provide a better understanding of the conformational preferences of the χ_2 dihedral angle of the Gln residue we have performed high-level *ab initio* calculations of compound **I**, which can be envisaged as a model molecule of the side chain of Gln. The study of the full residue has been performed on compound **II** at the AM1 [11] semi-empirical level, since the large amount of computer resources required precludes its study at the *ab initio* level. However, the similarity between *ab initio* and AM1 results has been confirmed by comparing the relative energies for the different conformations of **I**. Furthermore, comparisons have been made with X-ray diffraction results on single crystals of small peptides and proteins containing the Gln residue. It should be emphasized that the purpose of this work has been the understanding of the conformational behaviour of the χ_2 dihedral angle of Gln, and not the conformational preferences of the full residue. The results allow us to make a comparison between the folding of methylene units in compounds with biological interest and that previously found in small organic compounds [1–4].



I



II

METHODS

Geometry optimizations of **I** were performed at the Hartree–Fock (HF) level with the 6–31G(d) basis set [12]. All the optimized structures were checked by analysis of harmonic vibrational frequencies obtained from diagonalization of force constant matrices. Excluding translational and rotational motions, only positive eigenvalues of the Hessian matrix were obtained, proving that the calculated conformer geometries are minima on the HF/6–31G(d) potential energy surface of **I**. The Møller–Plesset (MP) perturbation treatment [13] was used to compute electron correlation corrections to the energy. MP energies were computed at the full-fourth order level (MP4(SDTQ)) using the 6–31G(d) basis set.

Dipeptide **II** was considered in 54 conformations. The structures selected are not the result of a complete conformational search. Rather they were selected because their molecular geometries are near to characteristic regions of the conformational space that are of interest, as discussed below. Owing to the size of the molecule the calculations were performed at the AM1 semi-empirical level [11].

The effect of the aqueous solvent on the conformations of **I** and **II** was examined following the self-consistent reaction field (SCRF) strategy developed by Miertus and coworkers [14,15] and adapted to the AM1 semi-empirical method by Orozco and coworkers [16,17]. Atomic coordinates for the Gln residue in proteins and small peptides were obtained from the Brookhaven Protein Databank [18] and the Cambridge Structural Data Base [19] respectively.

Ab initio calculations were performed with Gaussian-92 [20]. Semi-empirical calculations were carried out with a modified version [21] of the MOPAC computer program [22]. All the calculations were performed on a CRAY-YMP at the CEntre de Supercomputació de CAtalunya (CESCA).

RESULTS AND DISCUSSION

Following the conformational multidimensional analysis one would expect to find nine minima on the potential energy surface $E=E(\chi_2,\chi_3)$ of **I**. All these structures were taken as starting geometries in HF/6–31G(d) geometry optimizations. Table 1 displays the conformational angles and energies for the six characterized minima. However, a detailed inspection to the results reveals that they are three

Table 1. Minimum Energy Conformations of the Model Compound **I**

χ_2[a]	χ_3[a]	Conformation[b]	HF/6-31G(d)[c]	MP4/6-31G(d)[d]
179.0	169.6	**TT**	0.0	0.4
−179.0	−169.6	**TT**	0.0	0.4
69.3	−157.1	$\mathbf{G^+T}$	< 0.1	0.0
−69.3	157.0	$\mathbf{G^-T}$	< 0.1	0.0
63.8	81.2	$\mathbf{G^+G^+}$	1.2	1.0
−63.8	−81.2	$\mathbf{G^-G^-}$	1.2	1.0

[a]Torsional angles in degrees.
[b]Frequency analysis were performed to verify the nature of the minimum state of the stationary points located during geometry optimizations.
[c]Relative energies (in kcal/mol) computed from wave functions using the 6–31G(d) basis set.
[d]Relative energies (in kcal/mol) computed at the full fourth-order level of MP (MP4(SDTQ)level) using the 6–31G(d) basis set.

minima twofold degenerate. The lowest energy minima correspond to the **TT** and **GT** conformations, which are almost isoenergetic at the HF/6–31G(d) computational level. Note that in the latter minimum the χ_2 dihedral angle adopts a **G** conformation, as expected from our previous results on a similar methylenic sequence [1–4]. On the other hand, the **GG** conformation is 1.2 kcal/mol unfavoured with respect to the all-*trans*. The addition of the electron correlation into energy calculations seems to be of great importance, since a stabilization of 0.4 kcal/mol is now predicted for the **GT** conformation with respect to the **TT** one.

The thermodynamics in the gas phase was computed by correcting the differences in electronic energy to enthalpies at 298 K (ΔH^{gp}). The entropic correction to the enthalpy was computed from frequency analysis following the standard formulas. The free energy of reaction in water was evaluated by the classical thermodynamic scheme: $\Delta G^{rec-t,aq} = \Delta G^{gp} + \Delta\Delta G_{hyd}$, where ΔG^{gp} and $\Delta\Delta G_{hyd}$ are relative free energies in the gas phase and in aqueous solution respectively. Results are summarized in Table 2. As can be noted from an enthalpic point of view the **GT** conformation is slightly favoured with respect to the **TT** one, while the entropic corrections make almost isoenergetic the two conformations. Water has an important effect on the **T/G** equilibrium, since the **TT** is better hydrated that the **GT** by 0.5 kcal/mol. The **GG** is the conformation most favoured by the solvent, indicating that such conformation is slightly more polarizable than the **TG** one. In an attempt to predict the frequency of each conformation in **I**, we compute the conformer populations considering that the molar ratio of a given conformation to the most stable conformation is $\exp(-\Delta G^{rect,aq}/RT)$. The **TT** is the dominant conformer (55.0%), but the **GT** is also calculated to be present to a significant amount. Considering the populations of the **GT** and **GG** conformations, the **T/G** equilibrium for the χ_2 angle of **I** has a ratio of 1.2/1.

Table 2 Thermodynamics and Conformer Populations of **I**

Conformation	ΔE^a	ΔH^{gpb}	ΔG^{gpc}	$\Delta\Delta G_{hyd}{}^d$	ΔG^{recte}	$P(\%)^f$
TT	0.4	0.3	0.1	0.1	0.0	55.0
GT	0.0	0.0	0.0	0.6	0.4	29.9
GG	1.0	1.1	1.0	0.0	0.8	15.1

[a]Relative electronic energies (kcal/mol) computed at the MP4(SDTQ) using the 6–31G(d) basis set.
[b]Gas phase enthaly differences (kcal/mol) at 298 K.
[c]Gas phase free energy differences (kcal/mol) at 298 K.
[d]Differential free energies of solvation (kcal/mol) at 298 K.
[e]Differential free energies of reaction (kcal/mol) at 298 K.
[f]Populations of the three conformations considering the six minima displayed in Table 1 and the free energies of reaction computed at 298 K.

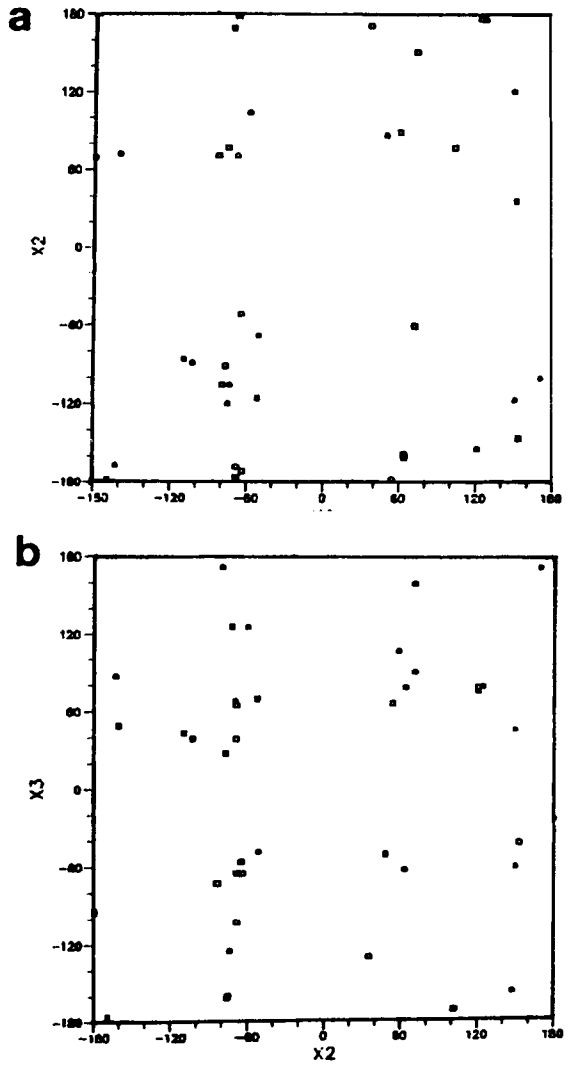


Figure 1 Distribution of the χ_1, χ_2 and χ_3 minimum energy values obtained from AM1 calculations on **II**: (a) χ_1 *vs* χ_2 and (b) χ_2 *vs* χ_3.

An energetic analysis of the Gln residue has been performed on monopeptide **II**, where the backbone dihedral angles φ and ψ were kept fixed at both the fully extended ($\varphi = \psi = 180°$) and helical ($\varphi = \psi = -60°$) conformations. Note that the fully extended conformation corresponds to the C_5 (five-membered hydrogen bonded ring) structure, which is usually characterized as the lowest-energy minimum on monopeptides [23–25]. On the other hand, the backbone dihedral angles used for the helical conformation are close to those of the standard α-helix ($\varphi = 55°$ and $\psi = -45°$) [26]. Side-chain dihedral angles χ_1, χ_2 and χ_3 were systematically varied in order to generate all the conformations predicted by

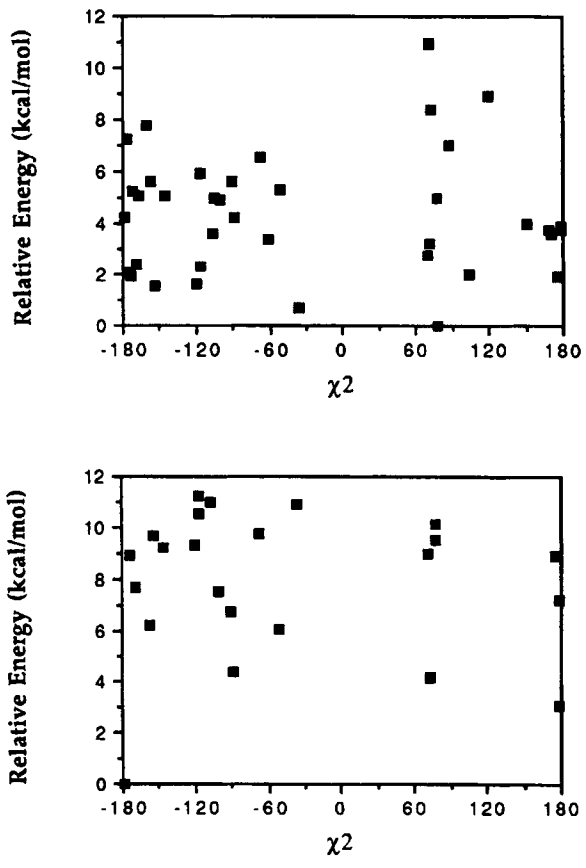


Figure 2 Schematic representation of the relative energies in both the (a) gas phase and (b) aqueous solution *vs.* the value of the dihedral angle χ_2 in the minimum energy conformations found for the model compound **II**.

the multidimensional analysis. Using the 54 generated conformations as starting points, 38 minima were found at the semiempirical AM1 level. Figure 1 represents the position of such minima on the surfaces χ_1 *vs.* χ_2 and χ_2 *vs.* χ_3. Note that the minimum energy conformations found for **II** are distributed along the usually allowed regions of the maps. Analysis of the data for each of the two backbone conformations indicates that backbone conformation does not have any effect on the side-chain conformational distribution of **II**.

The distribution of χ_2 *vs.* relative energies is represented in Figure 2 for both the gas phase and aqueous solution. It is worth noting that the number of minima in which χ_2 adopts a conformation near the **T** is greater than that with a **G** conformation. However, in the gas phase (Figure 2(a)) the **G** conformation is considerably favoured from an enthalpic point of view (about 1.5 kcal/mol), in good agreement with *ab initio* calculations on **I**. Figure 2(b)

shows the distribution of χ_2 *vs.* $\Delta G^{rect,aq}$, where, in order to clarify the scheme, only those conformations within the range 0–12 kcal/mol have been represented. As in the previous model compound the effect of the solvent was evaluated through SCRF calculations. The solvent turns out to change the relative energy order of the conformers. Thus, the most stable structure corresponds to a **T** conformation, whereas **G** conformers are not stabilized by the solvent. Furthermore, the number of low-energy structures with a **T** conformation is greater than the number of structures with a **G** conformation.

Analysis of the results allow us to obtain a picture of the conformational preferences of the methylenic sequence in the side chain of Gln. The results strongly indicate that the dihedral angle χ_2 of Gln can adopt a **G** conformation since it is favoured from an enthalpic point of view. However, it is not stabilized by either entropic or solvent effects. In order to give experimental support to our results, we have analysed the dihedral angle χ_2 of the Gln residues contained in some proteins whose X-ray structure was recently deposited in the Brookhaven Protein Databank [18]. Data were extracted from well-refined structures, with resolutions mostly better than 2.5 Å and *R*-factor of less than 20%. Mutants and complexes of an already analysed protein were not included in the statistics. A set of 38 crystal structures was recovered, which includes a total of 290 Gln residues. Since the focus of our study is the folding of methylene units in Gln only the χ_2 dihedral angle was analysed. Thus, other dihedral angles (φ, ψ, χ_1 and χ_3) equally important for a complete understanding of the conformational preferences of the Gln residue were not examined here.

Results are displayed in Figure 3. As can be noted, the **T** is the most frequent rotamer, although the **G** one is also usual in proteins. The ratio of the **T**/**G** equilibrium is 1.8/1, which is larger than the value predicted from our calculations. However, it is worth noting that computations and structural data are in qualitative agreement, indicating a large population of **T** when entropic and solvent effects are considered. The underestimation of the theoretical value must be basically attributed to the deficiencies of our simulation, *i.e.* use of a reduced model and the absence of specific interactions like hydrogen bonds with the solvent. On the other hand, inspection of the Cambridge Structural Data Base [19] showed six rotamers with a **G** conformation and only one with a **T** conformation. Thus, in small peptides the formation of intra- [8,27,28] or intermolecular [8,29,30] hydrogen bonds is a factor more important than the entropic and solvent contributions. In this cases, the enthalpic contribution plays a crucial role favouring the **G** conformation over the **T**.

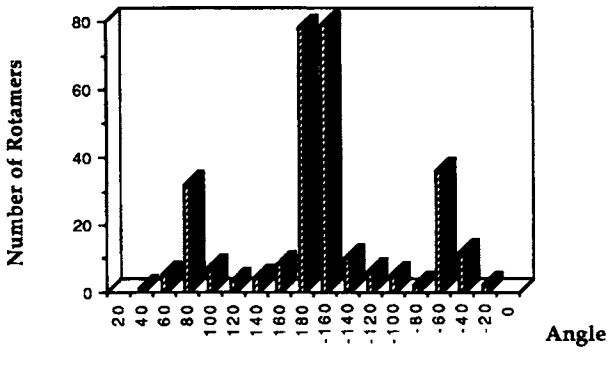


Figure 3 Distribution of the χ_2 angle of the Gln residue found in protein structures.

CONCLUSIONS

In summary, both *ab initio* and semi-empirical quantum mechanical calculations have provided an understanding of the conformational preferences of the methylenic sequence in the side chain of Gln. The agreement with previous theoretical [1,2,4] and experimental [2–4] results about the folding of methylene groups lend confidence to the present simulations. As can be noted, the present results are in poor agreement with force-field predictions. Some possible origin for the discrepancies between quantum mechanical data and force fields would be a bad parametrization of the Gln residue. In our opinion an erroneous parametrization of the force field could have a dramatic effect in the structural determination of proteins with a low resolution. From these calculations and our previous results several conclusions about the folding of methylene units in the Gln residue can be reached:

(1) The presence of a carbonyl group induces the rotation towards the **G** conformation of the bond defined by the first and second carbon atoms next to the carbonyl carbon.

(2) The **G** conformation is the global minimum of the χ_2 angle in the Gln residue. This explains the large number of Gln residues with this conformation found in both small peptides and proteins. An important effort must be performed with the aim of improving the parametrization of the empirical forcefields. This would give a better description of the conformational

preferences of the different residues and therefore help in the interpretation of the experimental data. (3) The dihedral angle χ_2 of Gln has an intrinsic tendency to adopt a **G** conformation. However, entropic contributions and solvent effects stabilize the **T** conformation.

Acknowledgements

We acknowledge the Centre de Supercomputació de Catalunya (CESCA) for providing computing facilities. This work has been supported by the DGICYT (under grant PB-931067). C. V. and E. N. acknowledge a predoctoral fellowship from the Departament d'Ensenyament de la Generalitat de Catalunya. Authors are indebted to Drs M. Orozco and F. J. Luque for making available their program with the AM1 Hamiltonian adapted to perform SCRF calculations. We also thank Dr I. Fita for reading the manuscript.

REFERENCES

1. C. Alemán, E. Navarro and J. Puiggalí (1995). Conformational analysis of succinamide analogues. *J. Org. Chem. 60*, 6135–6140.
2. E. Navarro, C. Alemán and J. Puiggalí (1995). Folding of methylene groups in linear glutaramide analogues. *J. Am. Chem. Soc. 117*, 7307–7310.
3. E. Navarro, V. Tereshko, J. A. Subirana and J. Puiggalí (1995). Incorporation of diacids into the polyglycine II structure: model studies. *Biopolymers, 36*, 711.
4. C. Alemán, E. Navarro and J. Puiggalií (1995). A comparison between diketones and diamides: Effects of carbonyl groups on the conformational preferences of small aliphatic segments. Submitted results.
5. R. Susnow, C. Schutt, H. Rabitz and S. Subramanian (1994). Conformational study of dipeptides: a sensitivity analysis approach. *J. Comp. Chem. 15*, 947–962.
6. K. H. Lee, D. Xie, E. Freire and L. M. Amzel (1994). Estimation of changes in side chain configuration entropy in binding and folding: general methods and application to helix formation. *Proteins 20*, 68–34.
7. P. N. Lewis, F. A. Momany and H. A. Scheraga (1973). Conformational energy analysis of the *N*-acetyl-*N'*-methylamides of the twenty naturally occurring amino acids. *Israel: J. Chem. 2*, 121–152.
8. I. Karle, J. L. Flippen-Anderson, S. Agarwalla and P. Balaram (1991). Crystal structure of [Leu^1]zervamicin, a membrane ion-channel peptide: implications for gating mechanism. *Proc. Natl. Acad. Sci. USA 88*, 5307–5311.
9. I. Karle, J. L. Flippen-Anderson, S. Agarwalla and P. Balaram (1994). Conformation of the flexible bent helix of Leu^1-zervamicin in crystal C and a possible gating action for ion passage. *Biopolymers 34*, 721–735.
10. M. Iqbal and P. Balaram (1981). Membrane channel forming polypeptides. 270 MHz proton magnetic resonance studies of the aggregation of the 11–21 fragment of suzukacillin in organic solvents. *Biochemistry 20*, 7278–7284.
11. M. J. S. Dewar, E. G. Zoebisch, E. F. Healey and J. J. P. Stewart (1985). Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc. 107*, 3902–3911.
12. P. C. Hariharam and J. A. Pople (1973). Influence of polarization functions on MO hydrogenations energies. *Theor. Chim. Acta 28*, 213–222.
13. C. Møller and M. S. Plesset (1934). Approximation treatment for many electron systems. *Phys Rev. 46*, 618–622.
14. S. Miertus, E. Scrocco and J. Tomasi (1981). Electrostatic interaction of a solute with a continuum. A direct utilization of *ab initio* molecular potentials for the prevision of solvents effects. *Chem. Phys. 55*, 117–127.
15. S. Miertus and J. Tomasi (1982). Approximate evaluations of the electrostatic free energy and internal changes in solution processes. *Chem. Phys. 65*, 239–245.
16. M. J. Negre, M. Orozco and F. J. Luque (1992). A new strategy for the representation of environment effects in semiempirical calculations based on Dewar's Hamiltonians. *Chem. Phys. Lett. 196*, 27–36.
17. F. J. Luque, M. J. Negre and M. Orozco (1993). AM1-SCRF approach to the study of changes in molecular properties induced by the solvent. *J. Phys. Chem. 97*, 4386–4391.
18. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, G. F. Mayer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi (1978). The protein data bank: a computer-based archival file for macromolecular structures. *Biophys. J. 185*, 584–591.
19. F. H. Allen, O. Kennard and R. Taylor (1983). Systematic analysis of structural data as a research technique in organic chemistry. *Acc. Chem. Res. 16*, 146–153.
20. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. González, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart and J. A. Pople (1992). Gaussian-92, Revision F.4., Gaussian Inc., Pittsburgh, P.A.
21. F. J. Luque and M. Orozco (1992). MOPAC/CMS computer program. University of Barcelona, Barcelona, Spain.
22. J. J. P. Stewart (1983). MOPAC computer program. *OCPE Bull. 3*, 101.
23. C. Alemán and J. Casanovas (1994). *Ab initio* SCF and force-field calculations on low energy conformers of 2-

acetylamino-2,*N*-dimethylpropanamide. *J. Chem. Soc. Perkin Trans 2*, 563–568.

24. H. -J. Böhm and S. Brode (1991). *Ab initio* SCF calculations on low-energy conformers of *N*-acetyl-*N'*-methylalaninamide and *N*-acetyl-*N'*-methylglycinamide. *J. Am. Chem. Soc. 113*, 7129–7135.
25. V. Barone and F. Fraternali (1990). Sensitivity of peptide conformation to methods and geometrical parameters. A comparative *ab initio* and molecular mechanics study of oligomers of α-aminoisobutyric acid. *Macromolecules 23*, 2038–2044.
26. C. Toniolo and E. Benedetti (1991). The polypeptide 3_{10}-helix. *Trends Biochem. Sci. 16*, 350–353.
27. D. A. Langs, G. D. Smith, J. J. Stezowski and R. E. Hughes (1986). Structure of pressinoic acid: the cyclic moiety of vasopressin. *Science 232*, 1240–1242.
28. A. Aubry, J. Protas and M. Marraud (1977). Conformation moléculaire à l'etat solide de la *N*-pivalyl-*N'*-méthyl-L-glutaminyl-méthylamide. *Acta Crystallogr. B33*, 2534–2539.
29. R. Bosch, G. Jung, H. Schmitt and W. Winter (1985). Crystal structure of Boc-Leu-Aib-Pro-Val-Aib-Aib-Glu(OBzl)-Gln-Phl×H_2O, the C-terminal nonapeptide of the voltage-dependent ionophore alamethicin. *Biopolymers 24*, 979–999.
30. S. P. Wood, I. J. Tickle, A. M. Treharne, J. E. Pitts, Y. Mascarenhas, J. Y. Li, J. Husain, S. Cooper, T. L. Blundell, V. J. Hruby, A. Buku, A. J. Fischman and H. R. Wyssbrod (1986). Crystal structure analysis of deamino-oxytocin: Conformational flexibility and receptor binding. *Science 232*, 633–636.