A MICROSCOPIC APPROACH TO THE STRUCTURE AND THERMODYNAMIC PROPERTIES OF PEPTIDES AND PROTEINS *

PIETRO AMODEO, VINCENZO BARONE and FRANCA FRATERNAL1

Dipartimento di Chimica, Università di Napoli, via Mezzocannone 4, 80134 Napoli (Italy) (Received in final form 3 October 1989)

ABSTRACT

Models and methods for theoretical evaluation of the structure and thermodynamic characteristics of large molecules are analysed with particular reference to biomolecules. Original results for amides and model peptides are reported to show the reliability of ab initio, semi-empirical and empirical methods, and the fields of application of approaches at different levels of sophistication.

INTRODUCTION

Reliable descriptions of the physico-chemical characteristics of molecules require characterisation of the structure and the thermodynamic properties for the isolated systems, and then proper inclusion of environmental (e.g. solvent) effects. The need for good in vacuo calculations is obviously the most fundamental, in that the testing of attempted solutions to problems of environmental effects presupposes satisfactory intramolecular potentialenergy surfaces. This is the main topic of the present study with special reference to the determination of reliable thermodynamic properties. The theoretical approach to this field complements experimental information by the following.

(1) The determination of thermodynamic quantities that cannot be obtained by experimental measurements: the gas-phase properties of nonvolatile molecules and the solution data of almost insoluble substances.

(2) The calculation of single-conformer contributions in systems characterised by one or more equilibria between different conformers. This feature is particularly important in the development of models for chemical and physico-chemical phenomena, because such equilibria appear to be rather common for even small and medium-sized biomolecules in solution.

(3) The resolution of individual group or residue contributions to the mean thermodynamic and spectroscopic observables. These are difficult to

^{*} Presented at the 10th AICAT, Pisa, Italy, 11-14 December 1988.

measure and interpret and give no more than 'unexpressive' mean values or, at the extreme, very complex signal patterns arising from the overlap of similar signals from several groups.

Theoretical calculation of the thermodynamic observables described in these three points requires increasing levels in accuracy of description of the potential energy surface (PES) of the tested system. Point (1), in fact, just requires a correct description of the PES around the absolute minimum of the system, whereas point (2) needs characterisation of the principal relative minima of the PES, together with the correct determination of the relative contribution of each minimum to the properties considered. Point (3) also demands accurate description of single-group contributions to the general molecular property.

The theoretical calculation of the thermodynamic observables can be divided into the following three steps.

(a) The determination of the principal molecular conformations for which the properties are to be calculated.

Selection of the theoretical method for this step depends on the dimension of the molecule and on the desired accuracy. When information on reactions involving the breaking of covalent bonds is required for small molecules (up to 50 atoms at the minimal basis-set Hartree-Fock level), the methods to be applied are ab initio quantum mechanical calculations [l]. Once an appropriate basis set is used in the computation, these methods provide rather accurate values for thermodynamic quantities and vibrational frequencies [2,3]. They constitute the only way to treat bond scission and electronic rearrangements. When the former is involved and the dimension of the system prevents the use of ab initio calculations, semi-empirical methods become the natural choice, as the latest generations of parametrisations (e.g. AM1 [4]) provide sufficiently accurate results for many different observables [5-71. Lastly, conformational processes of very large molecules can be adequately studied by the molecular mechanics (MM) approach [8-20], whose parameters can be guessed from refined computations of small models and improved with reference to experimental data [21,22].

(b) The calculation of rotational and vibrational frequencies for each considered point. The previous remarks also apply here, although the parametrisation of MM methods is usually less reliable [23,24].

(c) The determination of thermodynamic quantities from vibrational frequencies and momenta of inertia according to well-known relationships involving the partition function [25,26].

Some problems may arise when the mean "experimental" value of an observable at a given temperature derives from the contribution of several conformations. This could happen when the PES is rather flat around the absolute minimum and/or when several minima (conformers) contribute to the studied property, because of a quasi-degenerate distribution of minima in the PES. Both situations are quite frequent in the field of biomolecules.

Solution of this kind of problem mainly depends on the dimensions of the system, the number of degrees of freedom influencing the desired properties, and the characteristics of the system PES (number of minima and relative extension of low-energy regions around each minimum point).

For small- to medium-sized systems characterised by a small number of significant degrees of freedom, thermodynamic properties can be determined systematically by calculating them at each point of an n -dimensional grid, generated by a regular scan of the significant degrees of freedom. The desired properties can then be obtained from a Boltzmann average.

A simpler, though approximate, method can be applied when all the minimum points appreciably influencing the desired properties belong to narrow regions of the PES. Here the value of a thermodynamic property can be regarded as almost constant in all the points belonging to a given region, and its mean value can be obtained by simply averaging the values calculated for each minimum point, weighted by the area value of the PES region and by the relative Boltzmann factor.

When the dimensions of the system or the number of degrees of freedom prevent a systematic approach, statistical methods must be used to determine thermodynamic properties. Both Monte Carlo (MC) [27-321 and molecular dynamics (MD) [33-371 methods may be used to explore the configuration space of the system and to calculated Boltzmann averages of various thermodynamic parameters.

The potentialities of these approaches will now be illustrated by comparison between the computational methods in the determination of thermodynamic properties of simple amides, and in the characterisation of the structure and thermodynamic properties of model peptides, both by MM computations.

COMPUTATIONAL APPROACH

In this work we have used the MM method, based on the minimisation of an empirical function describing the molecular PES with respect to the whole set of internal coordinates of the system (EM calculations).

The AMBER force field derived by Kollman et al. [19,20] has been used. It contains stretching, bending, torsional, H-bond, non-bonded steric and electrostatic terms with no cross terms, such as bending-stretching coupling, often adopted in more complex force fields, used to calculate vibrational frequencies. The less accurate frequencies yielded by this simplification still give satisfactory values for the thermodynamic functions calculated from them.

Some updating developed in our laboratory was included in the original force field. This involves the charge distribution parameters and is exten-

Fig. 1. Temperature dependence of the entropy of formamide obtained by experimental and calculated harmonic frequencies.

sively described in ref. 22, where successful applications to conformational analysis of glycyl- and α , α -dialkylated peptides are reported.

The determination of the thermodynamic properties was divided into the following steps.

(1) A conformational map was elaborated for each system. The map was calculated, using a program developed by our group, by scanning on dihedral angles ϕ and ψ (see Fig. 1 and ref. 38 for definitions and conventions on these angles). All reported maps (Figs. 2 and 3) were obtained by 10 $^{\circ}$ increments on both angles in the range $[-180^{\circ}$ to 180 $^{\circ}$]. In these calculations, a "rigid" geometry is used, i.e. all the degrees of freedom different from ϕ and ψ are frozen at a standard value [39].

(2) After characterisation of PES regions, structures and energies of minimum points were determined by EM calculations, by successive application of a steepest descent method (until the gradient norm becomes lower than 10^{-3}) followed by a more sophisticated second-derivative-based search around the minimum region. This search employs a modified Newton-Raphson (MNR) algorithm [40].

(3) The thermodynamic properties of each system were then obtained by a Boltzmann average of the values corresponding to each energy minimum. Two classes of systems have been studied in detail.

(a) Simple amides: formamide (FAM), acetamide (AAM), N-methylformamide (NMF) and N-methylacetamide (NMA).

(b) Model peptides: Ac-Gly-NH₂ (I), Ac-Ala-NH₂ (II), Ac-Gly-NHCH₃ (III), Ac-Ala-NHCH₃ (IV), Ac-Aib-NHCH₃ (V) and Ac-Cprg-NHCH₃ (VI), where Ac = acetyl group, Gly = glycyl, Ala = alanyl, Aib = aminoisobutyryl and Cprg = cyclopropylglycyl residues.

The first class was chosen because simple amides are small enough to be studied by refined quantum mechanical methods and are well characterised spectroscopically and thermodynamically [41,42], In addition, their conformational freedom is limited to possible methyl rotations, so that the problems associated with large amplitude motions and the presence of different energy minima are avoided.

The second class is a valid example of systems which, though widely studied in solution and in solid state, are practically impossible to characterise experimentally in the gas phase. They are also well suited for testing methods to be employed in the study of flexible and multi-conformer systems.

The geometrical and thermodynamical parameters of formamide given by three theoretical methods are compared with experimental data in Tables 1 and 2. All three methods are sufficiently reliable with respect to the geometrical aspects, whereas the situation is less satisfactory for the vibra-

Fig. 2. (ϕ, ψ) map for Ac-Gly-NH₂ model. The contour lines are spaced 1, 2, 3, 4, 5, 7 and 9 kcal mol^{-1} over the 10 \degree point of lowest energy.

Fig. 3. (ϕ, ψ) maps for Ac-X-NHCH₃ model systems. The contour lines are spaced 1, 2, 3,4, 5, 7 and 9 kcal mol⁻¹ over the 10 \degree point of lowest energy.

tional frequencies. The semi-empirical AM1 method [4] provides reasonable results, whereas the ab initio results [43] must be scaled. The quality of all the computed frequencies is, however, sufficient for the prediction of thermodynamic properties. This is confirmed in the case of molecular mechanics by the plots of heat capacity and entropy versus absolute temperature, shown in Figs. 1 and 4. In addition, the thermodynamic parameters of various methyl-substituted amides, computed by the modified AMBER force field, are reported in Table 3.

Comparison between the empirical and experimental results shows a relatively satisfactory agreement, particularly for entropy values, over the whole temperature range considered. In both plots, a systematic and almost constant discrepancy between the empirical and experimental lines appears, the calculated values being in excess by less than 1% for entropy and less than 2% for heat capacity. These differences can certainly be reduced using a

TABLE 1

Geometric parameters computed for formamide by different theoretical methods: bond lengths are in A and valence angles in degrees

TABLE 2

Harmonic frequencies $(cm⁻¹)$ and thermodynamic properties (zero-point energy in kcal mol⁻¹, entropy in cal K^{-1} mol⁻¹, C_p in cal K^{-1} mol⁻¹) computed for formamide by different theoretical methods

 a Scaled by 0.9.

TABLE 3

Thermodynamic properties of various amides computed by the modified AMBER force field

more refined force field [44]. Their systematic nature, however, allows quite accurate results to be reached at the present level in the calculation of entropy and heat capacity differences.

Fig. 4. Temperature dependence of the specific heat of formamide obtained by experimental and calculated harmonic frequencies.

From another point of view, the very good quality of the AM1 results is particularly gratifying, since this method allows the study of reactive processes at a fraction of the cost of ab initio computations.

Peptides are "intrinsically" interesting (owing to their biological or pharmacological activity, structural properties, binding capabilities), but are also studied as "model systems" to mimic some properties of more complex molecules, whose main chemical and structural features are (or are thought to be) retained by those simpler systems.

Determination of the thermodynamic properties of "intrinsically" interesting systems requires only the choice of the computational method and, if necessary, a reasonable guess as to the initial molecular structure (when the data available are not sufficient to define such a structure univocally, or when several relative minima are to be detected).

In addition to the features just described, the study of "model systems" involves the rather complex (and often neglected) problem of the choice of the model molecule itself. Minor and apparently unimportant differences in side chains or even in the terminal groups of a peptide chain, in fact, may sometimes strongly affect structural and functional molecular properties.

In this section, the replacement of hydrogen atoms by methyl groups as substituents at the C^{α} atom or in the terminal groups is discussed. Even such simple substitutions affect different features of the system PES (see below), such as: (i) the number of conformational minima (conformers); (ii) the relative energy of the conformers; (iii) the extension of low-energy regions; and (iv) the height of conformational transition barriers. These results, therefore, suggest a cautious choice of model systems, which should be as similar as possible to the bigger molecules whose structural and functional behaviour is to be simulated. In particular, the chemical nature of the functional groups directly involved in interactions responsible for the molecular properties studied should be left unaltered to the greatest possible extent.

This principle should be kept well in mind, because the choice of model peptides, both in experimental and theoretical works, is usually based on other criteria. In experimental studies these are simplicity in synthesis or purification, higher solubility and conservability, and lower product or reagent cost, while in theoretical calculations they are computational time and simplification of the methods to be used by reduction of the number of degrees of freedom.

The (ϕ , ψ) maps obtained for several model peptides using the force field described above, and fixed bond lengths and valence angles, are shown in Figs. 2 and 3. The role played by the substituents at C^{α} is quite apparent because the conformational freedom of the backbone is progressively reduced by replacement of hydrogen atoms by methyl groups. The modifications induced by different terminal groups is by far less important, but by no means negligible, especially with regard to the thermodynamic properties (see below).

Low energy regions include the fully extended C₅ structure ($\phi \approx 180^{\circ}$, $\psi \approx 180^\circ$), the hydrogen-bonded C₇ structure ($\phi \approx \pm 90^\circ$, $\psi \approx \mp 90^\circ$) and the helical region ($\phi \approx \pm 60^{\circ}$, $\psi \approx \pm 30^{\circ}$). These structures are sketched in Figs. 5-7. The nomenclature adopted is discussed in ref. 17. The absolute minimum always corresponds to the C_7 structure, but the relative stabilities of different conformers are rather variable. Although the details of these results can be altered by different computational methods, the general considerations developed in the following discussion should remain valid.

The effect of substituents at C^{α} will be discussed in detail in a forthcoming paper. Here we concentrate on the more subtle effects induced by different terminal groups. Of course, models III and IV better "simulate" an inner unit in a peptide chain by preserving both peptide -CONH- **groups,** whose electronic distribution (and consequently related dipole moment and hydrogen-bond formation properties) differ significantly from that of the -CONH, group present in models I and II.

Conformational maps confirm such differences. Although their general features appear to be similar, both the extension of low-energy regions and the relative stability of the two conformers found for these systems are different. Furthermore, the flat region corresponding to helical conformations with a plateau in the model III map is absent in the model **I** map.

Fig. 5. Schematic drawing of the C_5 conformation of Ac-Aib-NHCH₃.

Fig. 6. Schematic drawing of the C_7 conformation of Ac-Aib-NHCH₃.

Fig. 7. Schematic drawing of the helix formed by $Ac-(Aib)_3-NHCH_3$.

Conformer	φ	ψ	τ	$E_{\rm tot}$	ΔE $(kJ \text{ mol}^{-1})$	
				$(kJ \text{ mol}^{-1})$		
CH ₃ CO-Gly-NH ₂						
C_5	180.0	180.0	108.1	-47.4	4.9	
C_7	-75.2	65.7	110.3	-52.3	$0.0\,$	
$CH_3CO-Gly-NHCH_3$						
C_5	180.0	180.0	108.5	-32.8	2.5	
C_7	-75.8	70.6	110.4	-35.3	0.0	
$CH3CO-Ala-NH2$						
C_5	158.5	-162.9	107.3	-48.0	4.8	
C_7^{ax}	68.5	-62.8	111.5	-51.9	0.9	
C_7^{eq}	-75.4	66.2	109.6	-52.8	0.0	
$CH3CO-Ala-NHCH3$						
$C_{\rm S}$	158.2	-163.0	107.6	-31.9	2.6	
R helix	-65.1	-19.2	113.0	-29.3	5.2	
L helix	54.3	31.1	113.1	-28.8	5.7	
C_7^{ax}	68.2	-67.3	111.7	-33.3	1,2	
C_7^{eq}	-75.9	71.2	109.7	-34.5	0.0	

Structural and energetical characteristics of model peptides

TABLE 4

These discrepancies are confirmed by full geometry optimisations (EM) of the different conformers (see Tables 4 and 5).

Even larger differences are found when EM results for models II and IV are compared. In the case of alanyl-peptides, the simple substitution of a methyl end-group determines a variation in the number of PES minima. Helical C_{10} and C_{13} structures only represent true minima in system IV, whereas in models I and III no minimum is present in that PES region, and in system II a flat plateau is found, but no local minimum.

Introduction of a methyl group seems to increase the relative stability of C_5 , C_{10} and C_{13} conformations in comparison to C_7 conformers. These, however, represent the absolute minima for all the systems studied.

A more detailed analysis of these results shows that the energy difference between the fully extended C_5 conformer and the absolute minimum (C_7) conformer) is almost halved on methyl addition, both in glycyl- and **in** alanyl-peptides. Also, the ΔE values in both cases are similar: about 20 kJ mol⁻¹ for systems I and III, and about 10 kJ mol⁻¹ for models II and IV.

A general feature that emerges from all the calculated thermodynamic observables is their relative sensitivity to conformational properties. From this point of view, entropy revealed itself to be a rather sensitive sampling property, showing quite large variations for different conformers of each model system. By contrast, heat capacity appears to be rather insensitive to any variation in the conformational parameters.

TABLE 5

Thermodynamic characteristics of different conformers of model peptides: nomenclature and units are the same as in previous tables; values obtained from Boltzmann averages among different minima are also given

Conformer	$CH_3CO-Gly-NH_2$				$CH3CO-Gly-NHCH3$			
	ΔE	ZPE	S^{Θ}	C_p	ΔE	ZPE	S^{Θ}	C_p
C_5	4.9	77.0	95.5	34.1	2.5	93.0	104.7	39.9
C_{7}	0.0	77.7	90.4	33.4	0.0	93.4	100.6	39.5
Average		77.7	90.4	33.4		93.4	100.6	39.5
	$CH_3CO-Ala-NH_2$				$CH_3CO-Ala-NHCH_3$			
	ΔE	ZPE	$S^{\, \Theta}$	C_p	ΔE	ZPE	$S^{\, \Theta}$	C_p
C_{5}	4.8	94.0	100.9	39.6	2.6	109.9	110.1	45.4
R helix					5.2	109.8	113.0	45.6
L helix					5.7	109.9	108.7	45.4
$C_7(ax)$	0.9	94.8	96.4	38.9	1.2	110.5	106.3	45.0
C_7 (eq)	0.0	94.6	96.7	39.1	0.0	110.2	106.9	45.2
Average		94.6	96.6	39.0		110.3	106.8	45.1

CONCLUDING REMARKS

The results obtained in this study and the comparison between experimental and calculated thermodynamic properties show a reasonable agreement allowing the use of calculated parameters instead of experimental data whenever the latter are not available. This is particularly true when one is interested in differences in thermodynamic properties, because the main error in the calculated parameters seems to have a systematic nature.

When greater accuracy is needed, calculated vibrational frequencies must be improved. As the force field used in the present study is very simple and has a "general purpose" nature (i.e. it is not particularly suitable for the calculation of thermodynamic properties), substantial improvements in frequency accuracy may be achieved both by refining force field parameters and (if this is not enough) by switching to a more sophisticated force field. Force fields usually employed in this context include coupling functions, such as stretching-stretching and stretching-bending terms. Also the functional form of some terms may be modified to improve PES description. Typical examples are replacement of the harmonic quadratic function in stretching terms by a Morse oscillator function, or the use of steric nonbonded terms other than L-J functions.

As regards the influence of the choice of model systems on calculated properties, the results of the present work recommend a thorough analysis of the functional groups thought mainly to contribute to the desired properties. This analysis will allow selection of model compounds containing those significant groups unaltered to the greatest possible extent, or, at least, the consideration of this principle together with others usually employed in such a choice.

REFERENCES

- 1 G. Naray-Szabo and P. Surjan, in G. Naray-Szabo (Ed.), Theoretical Chemistry of Biological Systems, Elsevier, 1986.
- 2 B.A. Hess, Jr., L.J. Schaad, P. Carsky and R. Zahradnik, Chem. Rev., 86 (1986) 709.
- 3 R. Arnaud, R. Subra, V. Barone, F. LeIj, S. OIivella, A. Sole and N. Russo, J. Chem. Sot. Faraday Soc. II, (1986) 1517.
- 4 M.J.S. Dewar, E. Zoebisch, E. Healy and J. Stewart, J. Am. Chem. Sot., 107 (1985) 3902.
- 5 W. Thiel, Tetrahedron, 44 (1988) 7393.
- 6 S. Schroder and W. Thiel, J. Am. Chem. Sot., 107 (1985) 4422; 108 (1986) 7985.
- 7 M.J.S. Dewar and K.M. Dieter, J. Am. Chem. Soc., 108 (1986) 8075.
- 8 U. Burkert and L.N. Allinger, Molecular Mechanics, ACS Monograph 177, Am. Chem. Soc., Washington D.C., 1982.
- 9 J. Caihet and P. Claverie, Acta Crystallogr., A-31 (1975) 448.
- 10 O. Ermer and S. Lifson, J. Am. Chem. Soc., 95 (1973) 4121.
- 11 E.M. Engler, J.D. Andose and P.v.R. Schleyer, J. Am. Chem. Soc., 95 (1973) 8005.
- 12 S. Lifson, A.T. Hagler and P. Dauber, J. Am. Chem. Soc., 101 (1979) 5111.
- 13 R.J. Abraham and R. Stolevik, Chem. Phys. Lett., 58 (1978) 4.
- 14 S.R. Cox and D.E. Williams, J. Comp. Chem., 2 (1981) 304.
- 15 P. Weiner and P.A. Kollman, J. Comp. Chem., 2 (1981) 287.
- 16 D.E. Williams and R.R. Weller, J. Am. Chem. Soc., 105 (1983) 4143.
- 17 G. Nemethy, M.S. Pottle and H.A. Scheraga, J. Phys. Chem., 87 (1983) 1883.
- 18 J. Hermans, H.J.C. Berendsen, W.F. van Gusteren and J.P.M. Postma, Biopolymers, 23 (1984) 1513.
- 19 P.A. Kollman, S.J. Weiner, D.A. Case, U. Chandra Sing, C. Ghio, G. Alagona, S. Profeta and P. Weiner, J. Am. Chem. Soc., 106 (1984) 765.
- 20 S.J. Wienr, P.A. Kollman, D.T. Nguyen and A. Case, J. Comp. Chem., 7 (1986) 230.
- 21 C. Jaime and E. Osawa, Tetrahedron, 39 (1983) 2769.
- 22 V. Barone, F. Fraternali, P.L. Cristinziano, F. Lelj and A. Rosa, Biopolymers, 27 (1988) 1673.
- 23 S. Lifson and P. Stem, J. Chem. Phys., 77 (1982) 4542.
- 24 S. Lifson, Gazz. Chim. Ital., 116 (1986) 687.
- 25 M.J.S. Dewar and G.P. Ford, J. Am. Chem. Soc., 99 (1977) 7822
- 26 G. Herzberg, Molecular Spectra and Molecular Structure, Vol. 2, Van Nostrand, Princeton, N.J., 1945.
- 27 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys., 21 (1953) 2087.
- 28 J.C. Owicki and H.A. Scheraga, J. Am. Chem. Soc., 99 (1977) 7403, 7413.
- 29 E. Clementi, Computational Aspects for Large Chemical Systems, Springer-Verlag, New York, 1980.
- 30 S. Romano, Int. J. Quantum Chem., 20 (1981) 921.
- 31 B. Bigot and W. Jorgensen, J. Chem. Phys., 75 (1981) 1944.
- 32 J. Kolofa, Mol. Phys., 63 (1988) 559.
- 33 A. Rahman, Phys. Rev., Al36 (1964) 405.
- 34 J.A. MC Cammon, B.R. Gehn and M. Kerplus, Nature, 267 (1977) 585.
- 35 M. Levitt, J. Mol. Biol., 170 (1983) 723.
- 36 M. Levitt, Ch. Sanders and P.S. Stem, J. Mol. Biol., 181 (1985) 423.
- 37 W.F. van Gusteren and H.J.C. Berendsen, in J. Hermans (Ed.), Proc. Workshop on Molecular Dynamics and Protein Structure, North Carolina University, 1984, pp. 5-13.
- 38 IUPAC-IUB Commission on Biochemical Nomenclature, Pure Appl. Chem., 40 (1974) 277.
- 39 E. Benedetti, in B. Weinstein (Ed.), Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, Marcel Dekken, New York, 1982 pp. 105-184.
- 40 R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1980.
- 41 J.C. Evans, J. Chem. Phys., 22 (1954) 1288; 31 (1959) 1435.
- 42 ST. King, J. Phys. Chem., 75 (1971) 405.
- 43 Y. Suagauwara, Y. Yamada, A.Y. Hirakawa, M. Tsuboi, S. Kato and K. Morokuma, Chem. Phys. Lett., 50 (1980) 105.
- 44 P. Amodeo, V. Barone and F. Fraternali, to be published.