Theor Chim Acta (1993) 85: 61–67

Theoretica Chimica Acta

Theoretical studies of peptidic structures. Environmental effects*

Serafin Fraga[1] and Stephen E. Thornton[2]

[1] Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

[2] Protein Software Inc., Box 60499, University of Alberta Postal Outlet, Edmonton, Alberta, Canada T6G 2S7

Received August 26, 1991/Accepted December 13, 1991

Summary. The peptide Asp-Glu-Arg-Ser has been studied using the software package maPSI, with simulation of the pH conditions and the influence of the solvent. It is concluded, in agreement with the experimental observation, that in solution DERS is distorted from a true β-turn conformation, without H-bond but stabilized by a weak salt-bridge between the side chains of the Glu and Arg residues.

Key words: Theoretical simulation – Peptide – Tertiary structure – pH solvation – H-bond – Salt-bridge

1. Introduction

The understanding and interpretation, at the molecular level, of most biological processes involving proteins requires a knowledge of their tertiary structures and for that reason considerable efforts have been dedicated to their prediction.

A variety of software packages (such as AMBER, CHARMM, ECEPP, MMPEP, etc.) have been developed for that purpose [1]; at this laboratory, a package (ALTA:maPS, hereafter denoted as maPSI) has been assembled [2] from existing programs [3]. The general goal of any such package is the determination of the global minimum in the conformational energy hypersurface through variation of the dihedral angles, under a minimum energy criterion. The potential function used in the energy calculations is a characteristic and fundamental component of any such package.

The methodology implemented in a given software package is best tested with simulations of small peptides for which more detailed experimental information may be available, which is not always the case for large proteins. Recently, an NMR study has been completed for a series of tetrapeptides in order to investigate the influence of the sequence, the solvent, and the pH on the structure [4]. The success obtained [5] in the interpretation of earlier experimental results [6] has motivated this work, in order to further test the software package maPS, using a different approach.

* Dedicated to Professor Alberte Pullman

2. Methodology

The calculations have been performed using the software package maPSI, which consists of three main components: prediction of the secondary structure, determination of the tertiary structure, and simulation of the interaction with other systems.

The potential used in the energy calculation was developed originally for the study of molecular interactions and associations [7]. This potential was expressed as a summation of all the atom–atom pair potentials (AAPP) for the two interacting molecules, with the AAPP's given as $1/R$ expansions, with terms in $1/R$, $1/R^4$, $1/R^6$, and $1/R^{12}$. The coefficients of these terms were obtained by fitting, with some semiempirical considerations, of the accurate potentials obtained by Clementi and co-workers [8] from self-consistent field calculations for the interaction of amino acids, purines and pyrimidines with water. The AAPP's were later modified by inclusion of an appropriate dispersion contribution [9]. This potential was then adapted to the study of protein structures [3b, c] and its convergence characteristics studied in chosen systems [3d].

In its simplest form, the procedure used in the construction (from the amino acid sequence) and optimization of a protein structure is as follows. The second amino acid is added to the first, with formation of the corresponding peptidic bond in *trans* conformation. Then the ψ angle of the first residue and the ϕ and χ_1 angles of the second residue are optimized (fully or not, as desired). The construction of the complete chain is continued in this way, by successive addition of the remaining amino acids, with the option of optimizing at each step the chain already built. There is also the option of constructing the peptidic chain with given dihedral angles, say, for α-helices, β-chains or any of the β-turns. Once the peptidic chain has been constructed it may then be optimized, either in its totality or as required (given angles or regions), to whatever degree it is desired.

The optimization of the complete chain may be performed in either a single- or a double-pass modality. In the former the dihedral angles are optimized successively starting from the N-terminus, while in the latter the optimization proceeeds alternatively from both the N- and the C-termini; that is, one optimizes the angles associated with the first linkage, then the angles of the last linkage, next the angles of the second linkage, and so on, until all the linkages have been optimized. Either procedure is repeated for a given number of cycles or until convergence to a chosen threshold has been attained for all the dihedral angles.

Complete convergence to a given threshold is obtained in two steps. Partial convergence is reached when the maximum change for any dihedral angle, after a complete optimization cycle, is smaller than the threshold value. Because of the successive nature in which the optimization procedure is performed, it may happen however that restarting the optimization process at that point will produce in some of the dihedral angles changes greater than the threshold. That this may happen is easily understood from the following considerations. During a cycle each angle is optimized for the structure defined by the values of all the remaining dihedral angles at that moment, and therefore the last angle to be processed will be optimized to a better degree than all the others while the angle optimized at the beginning of the cycle will not be consistent with the structure obtained at the end of the cycle.

In order to reach total convergence, the optimization procedure must then be continued as follows. The values of the angles, when partial convergence has been reached, are adopted as reference. The optimization procedure is then continued until partial convergence, as defined above, is attained again. If at that point the change for each dihedral angle, with respect to the above reference values, is smaller than the threshold, total convergence is said to have been attained. If not, the process is repeated again.

In addition it is possible to perform the optimization of the peptidic chain in interaction with another system, which is allowed to move towards the first one until a stable association is formed. If the second interacting system is also a peptidic chain, it is also optimized (in the presence of the first chain), thus being possible to obtain the most stable association of the two interacting systems in their most stable conformation. The procedure may be summarized as follows. The first peptidic chain is placed with its geometric centre at the origin of coordinates and the second interacting molecule is displaced and rotated with respect to the first molecule as desired. Then the interaction process is allowed to start: the first chain is optimized in the presence of the second interacting system; next, the interacting system (if it is a peptidic chain) is optimized in a similar fashion, and finally it is allowed to move towards the first peptidic chain, along a path of extreme gradient. The process is repeated until convergence in all the dihedral angles and stabilization of the molecular association have been attained. The approach is controlled in order to ensure that the peptidic chain(s) have full opportunity to adapt itself (themselves) to the existence of the other molecule.

A particular case of the above option allows for the possibility of solvating a peptide chain, creating a solvation shell around it of up to 500 water molecules.

Optimization of the tertiary structure of a large protein may be rather time consuming and in such a case it may be worthwhile to proceed first with a prediction of its secondary structure. At this laboratory such a prediction is made through a procedure developed at this laboratory and based on the recognition, Lim, and Garnier–Osguthorpe–Robson methods [10]. Once this prediction has been made, the corresponding fragments are constructured as α-helices, β-chains, or joining loops and the complete chain constructed by coupling of those fragments. During the subsequent optimization of the resulting conformation, the structure of the α-helices and β-chains is maintained, with variation only of the dihedral angles of the joining loops. Thus a considerable saving of computing time is achieved while, it is believed, enhancing the possibility of arriving at a more accurate prediction of the tertiary structure. This assumption is based on the very satisfactory accuracy of the prediction for the secondary structure, with the corresponding reduction in the number of degrees of freedom.

3. Calculations

The calculations have been performed for the tetrapeptide Asp-Glu-Arg-Ser (or DERS, in short), on an IBM RISC System/6000 powerstation.

Crystallographic structures of the individual amino acids were used, with their dihedral angles ϕ, χ_1, and ψ rezeroed according to the IUPAC convention. The peptidic chain was built up as follows. First the peptide ER was constructed as a β-*I* turn [4, 5] and then the complete peptide was obtained by addition of the residues D and S, using only one cycle of optimization during the building

process. After construction, the N- and C-termini of the peptidic chain were then protected with the groups H_3C–CO– and –NH–CH_3, respectively, in order to simulate the experimental conditions. The protected chain was then fully optimized until all the dihedral angles ϕ, χ_1, and ψ had converged (with a convergence limit of ± 2 degrees).

In order to better simulate the conditions at neutral pH, the residues Asp, Glu, and Arg were used in charged form: that is, with ionized carboxyl groups in the side chains of the Asp and Glu residues and with the $N_{\eta 1}$ of the Arg residue protonated.

In the study of the conformation of the tetrapeptide DEKS, reported elsewhere [5], the protected peptidic chain was reoptimized again in the presence of water. The solvation process, with optimization of the peptidic chain at each solvation step, was continued until six water molecules were added. Such a procedure is not completely correct, because the water molecules were added successively instead of simultaneously, and therefore it may be considered only as an approximate simulation of the solvation process. Interestingly enough, however, it was observed that the presence of the water molecules had a strong effect on the structure of the peptide: the hydrogen bond present in the structure obtained in the simulation in vacuum was broken in the presence of the water molecules but the salt-bridge observed in the original structure was conserved.

The similarity between DEKS and the system (DERS) considered in this work leads us to believe that similar results would be obtained for the latter if the same procedure were to be followed. Consequently, it was decided to adopt a different approach, which would complement the original one. The new approach is as follows.

The peptide, with its tertiary structure fully optimized, is solvated with five water molecules, in an approximation to a first solvation shell. At the end of the solvation process, the water molecules are then incorporated into the peptide: that is, the water molecules are included in the listing of the atoms of the peptide, with the coordinates they had at the moment of stabilization in the solvation process; each water is incorporated as forming part of the side chain of the residue to which it was closer. The rationale for proceeding in this manner is that it may be assumed that those water molecules, constituting (part of) the first solvation shell, are rather strongly attached to the peptide and will follow the internal motions of the latter. The structure of this solvated peptide is then reoptimized, in order to simulate the effect of solvation on the tertiary structure.

The optimized structures of the non-solvated and solvated peptides, obtained above, were then used in the simulation of the corresponding molecular associations, as described above. In each of these simulations, only one attempt was made at obtaining one such molecular association, without trying to find the most stable one. The reason is that the purpose of the simulation is simply to determine whether and how much the tertiary structure of the peptide is affected by the presence of an interacting system.

4. Results

The interpretation of the experimental NMR results [4] for both DEKS and DERS, in 60/40 methanol/water solution at a pH in the range 5.5–6.6, leads to

Table 1. Dihedral angles (in degrees) for the various systems

Residue	Angle	System[a]					
		1[b]	2	3	4	5	6
D	ϕ	−61	−61	−72	−77	−55	−59
	χ_1	−34	−70	−35	−35	−35	−36
	ψ	171	142	−160	−134	150	150
E	ϕ	−53	−89	−66	−86	−89	−79
	χ_1	−54	58	−163	−63	58	65
	ψ	−18	0	−9	−24	−9	−27
R	ϕ	−71	−89	−51	−165	−64	−48
	χ_1	−40	68	−45	175	42	43
	ψ	−6	−3	−16	−36	−5	−20
S	ϕ	59	−103	−66	52	−66	−65
	χ_1	−7	−129	24	−23	−135	−130
	ψ	10	−4	−11	16	−6	−11

[a] See the text for details

[b] The original angles ϕ and ψ of E and R, in a β-turn, were −60, −30, −90 and 0°, respectively

similar conclusions. The results indicate that the same small proportion (estimated at less than 25% [11]) of molecules in a β-turn conformation coexist with a larger proportion of molecules in an extended conformation. The close-to-β-turn conformation may be stabilized by a H-bond (not always seen), involving the peptidic NH-group of the Ser residue and the peptidic CO-group of the Asp residue, as well as by a salt-bridge between the side chains of the Glu and Lys residues, in DEKS, and of the Glu and Arg residues in DERS.

The calculations for the non-solvated DERS peptide predict a slightly distorted β-turn conformation. The ϕ and ψ angles (see column for system 1 in Table 1) of the Glu and Arg residues are −53, −18, −71 and −6°, to be compared with the values −60, −30, −90 and 0° corresponding to a proper β-*I* turn. This conformation is stabilized by a H-bond between the peptidic O of the Asp residue and the H of the peptidic NH-group of the Ser residue. A weak salt bridge also exists between the side chains of the Glu and Arg residues; the distance between the geometric centre of the two oxygens of the side chain of the Glu residue and the geometric centre of the three nitrogens of the side chain of the Arg residue is 9.10 Å.

Solvation and interaction with other systems introduces a distortion, as observed from the values given in Table 1 for the dihedral angles of the various systems considered in this work. The notation used in this Table is as follows. System 1 denotes the non-solvated DERS peptide, fully optimized. System 2 is the peptide DERS solvated with five water molecules, added successively as described above. The initial conformation of the peptide, which was that of system 1, was maintained unchanged during the solvation process. After completion of the process, the structure of the DERS peptide (with the five water molecules incorporated as part of the side chains of the residues to which they were found to lie closer) was fully optimized again. Systems 3 and 4 are non-solvated DERS peptides involved in a molecular association (DERS)–(DERS), not necessarily the most stable one; no effect was made to find such a

conformation as the motivation for the calculation was to find out to what extent the molecular interaction would force the conformations of the monomers to change. Systems 5 and 6, in correspondence to systems 3 and 4, are solvated DERS peptides (system 2) involved in the molecular association (DERS-$5H_2O$)–(DERS-$5H_2O$). In this case, again for the same reasons, the resulting association may not necessarily be the most stable. In the two cases of molecular associations, the conformations of the interacting peptidic chains (systems 3 and 4 and systems 5 and 6, respectively) were fully optimized during the formation of the association.

Even if changes (with respect to system 1) are observed all across Table 1, it is worthwhile to mention that they all fall well within the allowed conformational space [3e]. The changes in the dihedral angles suggest changes in the main features of the structure of the peptidic chain and certainly the H-bond observed in the isolated peptide (system 1) has disappeared in all cases; that is, the solvation and the association have determined an opening of the chain, increasing the separation between its head and its tail. On the other hand, the change in the separation between the two charged groups involved in the salt bridge, although appreciable, is not so drastic: the separation observed for systems 2 through 6 is 11.38, 10.91, 11.16, 11.17 and 10.87 Å, respectively. Although weaker, it may be said that the salt bridge has subsisted.

Therefore it must be concluded, in agreement with the experimental observation, that in solution DERS is distorted from a true β-turn conformation, without H-bond but stabilized by a weak salt-bridge. It is possible, however, taking into account the results for the isolated peptide, that a molecular dynamics simulation of the system in solution might show the existence of a H-bonded conformer. On the basis of the present results, it would be expected that the proportion of such a H-bonded conformer would be smaller than that of the more extended form.

It must be emphasized that the simulations in this work, as well as those in the earlier work [5], represent only an approximation to reality, in an attempt to reproduce as much as possible the experimental conditions, such as charged residues (as appropriate for an approximately neutral pH) and the presence of interacting systems, as simulated by another peptidic molecule and some water molecules. The results, however, confirm the need for such simulations, taking into account the effect observed for the H-bond. [Future work will consider the effect of methanol.]

References

1. For details and references see, e.g., the following recent works: (a) Wolfe S, Weaver DF, Yang K (1988) Can J Chem 66:2687; (b) Wolfe S, Bruder S, Weaver DF, Yang K (1988) Can J Chem 66:2703; (c) Wolfe S, Khalil M, Weaver DF (1988) Can J Chem 66:2715; (d) Wolfe S, Yang K, Khalil M (1988) Can J Chem 66:2733; (e) Wolfe S, Bowers RJ, Shin HS, Sohn CK, Weaver DF, Yang K (1988) Can J Chem 66:2751; (f) Roterman IK, Gibson KD, Scheraga HA (1989) J Biomol Struct Dyn 7:391; (g) Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989) J Biomol Struct Dyn 7:421
2. Software package ALTA:maPS. Protein Software Inc., Box 60499, University of Alberta Postal Outlet, Edmonton, Alberta, Canada T6G 2S7
3. (a) Fraga S (1983) Comput Phys Commun 29:351; (b) Coghlan B, Fraga S (1985) Comput Phys Commun 36:391; (c) Seijo L, Coghlan B, Fraga S (1986) Comput Phys Commun 41:169; (d)

Fraga S, Singh B, Coghan B (1985) J Molec Struct (Theochem) 120:213; (e) Fraga S, Seijo L, Campillo M, Moscardo F (1988) J Molec Struct (Theochem) 179:27
4. Liu X, Scott PG, Otter A, Kotovych G (1992) Biopolymers 32:119
5. Thornton SE, Fraga S (1991) Can J Chem 69:1636
6. Otter A, Scott PG, Liu X, Kotovych G (1989) J Biomol Struct Dyn 7:455
7. Fraga S (1982) J Comput Chem 3:329
8. (a) Clementi E, Cavallone F, Scordamaglia R (1977) J Am Chem Soc 99:5531; (b) Scordamaglia R, Cavallone F, Clementi E (1977) J Am Chem Soc 99:5545; (c) Bolis B, Clementi E (1977) J Am Chem Soc 99:550
9. Hobza P (private communication)
10. (a) Thornton SE, San Fabian E, Fraga S, Parker JMR, Hodges RS (1991) J Molec Struct (Theochem) 226:87; (b) Fraga S, San Fabian E, Thornton SE, Singh B (1990) J Molec Recognition 3:65; (c) Thornton SE, San Fabrian E, Fraga S, Parker JMR, Hodges RS (1991) J Molec Struct (Theochem) 232:321
11. Kotovych G (private communication)