Chapter 4

Comparative Analysis of the Conformational Profile of Substance P using Simulated Annealing and Molecular Dynamics

4.1 Summary

The present study describes an extensive conformational search of substance P using two different computational methods. On the one hand, the iterative simulated annealing and on the other hand, molecular dynamics. Analysis of the results were carried out by applying an in-house methodology that classifies the different obtained structures through a cluster analysis method based on information theory. Furthermore, accessible values for the dihedral angles for the iterative simulated annealing conformational search and along the molecular dynamics have been obtained. Finally the conformational motifs that are characteristic of substance P in a hydrophilic environment are presented and compared with the experimental results available in the literature.

4.2 Introduction

Substance P (SP), Arg^{1}-Pro^{2}-Lys^{3}-Pro^{4}-Gln^{5}-Gln^{6}-Phe^{7}-Phe^{8}-Gly^{9}-Leu^{10}-Met^{11} is a neuropeptide of the tachykinin family, involved in smooth muscle contraction, hypotension, salivation and depression (Hökfelt et al., 2001). As for the rest of the bioactive peptides, knowledge of its conformational features is a necessary condition to rationally design new peptidomimetics. Spectroscopic methods alone, like NMR spectroscopy cannot provide all the structural details necessary to understand the conformational profile of the peptide in solution due to the flexibility of the molecule. Computational methods represent a complementary technique to extract structural information of the system. However, these methods require to be continuously contrasted with the experimental information available. In this context, SP is a large enough system to represent a challenge to conduct model simulations and on the other hand, an appealing system to contrast the results accumulated from the wealth of experimental information available. The goal of the present work falls within the scope of this latter category of studies.

Several reports both experimental and theoretical, devoted to discuss the conformational profile of SP have appeared in the past, however none of them can be considered definitive. Different spectroscopic studies were carried out in water and organic solvents using diverse techniques, including NMR studies (Chassaing et al.,1986, Sumner et al. 1990, Shukla et al., 1991, Patel et al., 2001), Circular Dichroism, Raman and IR Spectroscopy (Mehlis et al. 1975, Mehlis et al., 1980). Other studies were undertaken in model membranes and surfactants using NMR spectroscopy (Miyazawa, 1984, Schwyzer et al., 1986, Erne et al, 1986, Rolka et al., 1986, Williams et al., 1990, Young et al., 1994). Moreover, substance P was also studied by energy

calculations (Nikiforovich, et al. 1981, Manavalan et al., 1981, Manavalan et al., 1982) and more recently, by limited simulations using molecular dynamics (Wymore et al., 1999a, Wymore et al., 1999b, Coutinho et al, 1998).

This series of studies clearly point out the dependence of the conformation of the peptide with the environment. Thus, whereas in DMSO and pyridine the peptide exhibits an extended conformation in water, the NMR studies suggest *some degree of folding* in the region between residues 5 to 11 (Chassaing et al., 1986). Furthermore, in methanol, the same authors proposed either a 3_{10}-helix between residues Gln^5-Gln^6-Phe^7-Phe^8 or a α-helix at the segment Pro^4-Gln^5-Gln^6-Phe^7-Phe^8, being flexible the N-terminus region Arg^1-Pro^2-Lys^3. Moreover the C-terminal segment Gly-Leu-Met is predicted to be oriented towards the side chains of Gln^5 and Gln^6. In lipid membranes and sodium dodecylsulfate, SP adopts a α-helical structure (Schwyzer et al., 1986, Miyazawa, 1984, Rolka et al., 1986, Williams et al., 1990, Erne et al, 1986). Interestingly, a NMR study reported the comparison of the structures of a segment of the β-amyloid peptide (25-35) (Aβ25-35) and substance P showing a similar α-helical structure in their C-terminal regions in a trifluoroethanol/water solution (Lee et al., 1999).

In the present study, the conformational space of the free acid form of SP (SPOH) has been thoroughly investigated using two different computational methods. First, the structural features of SPOH were explored using simulated annealing (SA) in an iterative fashion as sampling technique (Filizola et al, 1997). The resulting low-energy conformations were then classified into clusters using a cluster analysis method based on information theory (Law, 1999). On the other hand, long molecular dynamics simulations were performed to sample the conformational space of a peptide in solution. This procedure was previously shown to be an adequate method to understand the conformational behavior of small peptides (Daura et al., 1998, Duan et al., 1998). Accordingly, we proceeded in a parallel effort by running two molecular dynamics simulations of the peptide soaked in a box of water molecules and using periodic boundary conditions at 300 and 400 K, respectively. The main goal of this study was to compare the performance of the simulated annealing procedure versus molecular dynamics simulations as tools to explore the conformational profile of medium size peptides. The results of these studies were compared with the results reported in a NMR study reported of the structure of SP in trifluoroethanol/water solution (Lee et al., 1999).

4.3 Methods

All the molecular mechanics calculations were carried out using the AMBER 5 package. For the SA calculations the Weiner all-atom force field (parm94) was used (Weiner et al.,1986),

whereas for the MD the Cornell all-atom force field (parm96) (Cornell et al., 1995) was used. Substance P was studied in its zwitterionic form (free acid) in both the SA and MD.

4.3.1 Simulated Annealing Calculations

The conformational space of SPOH was explored by using a SA protocol in an iterative fashion (Filizola et al., 1997). No explicit solvent was included in the calculations, although an effective dielectric constant of 80 was used to screen for the electrostatic interactions. An integration step of 2 fs was used together with the SHAKE algorithm to constrain bonds connecting hydrogen atoms. The initial extended structure, once minimized was quickly heated to 900 K, in order to force the molecule to jump to a different region of the conformational space. Subsequently, the 900 K structure was slowly cooled to 200 K at a rate of 7 K/ps and then minimized. This structure was stored on a library and used as the starting conformation for a new cycle of SA. In this way an energy rank ordered library of low energy conformations was generated. Low energy conformations were checked for uniqueness by comparing the dihedral angles of a new conformation with those already stored. A structure was stored if at least one of the backbone dihedral angles differed more than 60° in respect to any of the previous conformations already stored in the library. The conformational search was continued until 5000 structures were obtained.

4.3.2 Molecular Dynamics Calculations

The extended conformation of SPOH in its free acid form was generated using the AMBER program. The peptide was minimized with the module SANDER *in vacuo* using a dielectric constant of 80. 1500 cycles of steepest descent followed by the conjugated gradient method were applied until the RMS distance between two consecutive structures was smaller than 0.001 $kcal \cdot mol^{-1} \cdot A^{-1}$.

Subsequently, the peptide was soaked in a rectangular box containing 1150 TIP3P water molecules. The dimensions of the box were chosen in such a way that the minimum distance from the peptide in its extended conformation to the box wall was larger than 8 Å. The system was minimized fixing the coordinates of the peptide and allowing the water molecules to move. A dielectric constant $\varepsilon = 1$ and a cutoff of 10 Å were used.

A trajectory of 100 ps was carried out using periodic boundary conditions at constant temperature (300 and 400 K) and constant pressure (1 atm), with a cutoff of 10 Å. An integration step of 2 fs in conjunction with the use of the SHAKE algorithm to constrain the stretching of bonds

involving hydrogen atoms was applied. The translation and rotational motion was removed after 1000 steps and at the beginning of each 500 ps fragment of molecular dynamics. Then, both systems (at 300 and 400 K) were changed to Particle Mesh Ewald (PME) conditions with a grid spacing of 1 Å and a tolerance for the Ewald sum of 10^{-5}. Both trajectories at 300 and 400 K were computed for 40 ns and snapshots were recorded every picosecond for a later analysis.

4.3.3 Conformation Classification

Conformations were classified using, the computer program CLASICO, developed in this laboratory and designed to group tens of thousands of structures obtained from the SA and the MD trajectories into hundreds of patterns, easing the subsequent treatment of the information obtained. A secondary structure can then be automatically assigned to each of the structures left. Specifically, given a conformation, the procedure consists of assigning a letter to each of the residues of the peptide according to the values of the backbone dihedral angles, following the classification of Srinivasan and Rose (1999). In this scheme, the conformational space is partitioned into 36 boxes of the same size (60ºx60º), and are labeled with a one letter code as shown in Table 4.1. Letters assigned according to the previous step are then translated into conformation codes, by mapping them into secondary structure classes according to the following classification: S (sheet) = {A, F, G, L, M, R}; H (α-helix) = {O}; 310 (3_{10}-helix) = {P}; T = {J, O, P}; T' = {j, o, p}; U = {M, R}; U' = {m, r}. Finally, in a subsequent step, using a window of three consecutive residues the conformation codes are used to classify the residues into a conformational category, satisfying one of definitions listed in Table 4.2. Accordingly, each conformation can subsequently be represented by a combination of secondary structure motifs. This allows to account for the number of different patterns attained by the peptide and to assess the diversity of the conformational space sampled using different computational methods or by the different temperatures of the MD runs. Furthermore, the evolution of the patterns and by extension of the folding process can be followed. Patterns are numbered and plotted along the MD and the frequency of the different patterns can be graphically assessed by looking at the distribution of points.

Table 4.1. The conformational space is partitioned into coarse-grained bins defined by a letter code.

		ϕ						
		-180º	**-120º**	**-60º**	**0º**	**60º**	**120º**	**180º**
ψ	**180º**	A	G	M	S	m	g	a
	120º	F	L	R	X	n	h	b
	60º	E	K	Q	W	o	i	c
	0º	D	J	P	V	p	j	d
	-60º	C	I	O	U	q	k	e
	-120º	B	H	N	T	r	l	f
	-180º	A	G	M	S	m	g	a

Table 4.2. Conditions for secondary structure definition of three consecutive residues.

MOTIF	CONDITION	DEFINITION	CODE
3_{10}-helix	j and j+1 and j+2 $\in$ 3_{10}	j,j+1,j+2 = 3_{10}-helix	310
α-helix	j and j+1 and j+2 $\in$ H	j,j+1,j+2 = α -helix	H
β-strand	j and j+1 and j+2 $\in$ S	j,j+1,j+2 = β-strand	S
type I β-turn	j+1 $\in$ T and j+2 $\in$ T	j+1 = type I β-turn (residue i+1)	I1
	j $\in$ T and j+1 $\in$ T	j+1 = type I β-turn (residue i+2)	I2
type I' β-turn	j+1 $\in$ T' and j+2 $\in$ T'	j+1 = type I' β-turn (residue i+1)	i1
	j $\in$ T' and j+1 $\in$ T'	j+1 = type I' β-turn (residue i+2)	i2
type II β-turn	j+1 $\in$ U and j+2 $\in$ T'	j+1 = type II β-turn (residue i+1)	II1
	j $\in$ U and j+1 $\in$ T'	j+1 = type II β-turn (residue i+2)	II2
type II' β-turn	j+1 $\in$ U' and j+2 $\in$ T	j+1 = type II' β-turn (residue i+1)	ii1
	j $\in$ U' and j+1 $\in$ T	j+1 = type II' β-turn (residue i+2)	ii2
Coil	none of the above	j+1 = coil	-

4.3.4 Clustering of the structures

Patterns attained by the peptide are numerous and diverse. In order to tackle such diversity, we proceeded to group the conformations into clusters according to their similarity. Clusters were computed using an algorithm based on information theory. The algorithm considers that a pattern can be categorized unambiguously by the presence or absence of each of the secondary structure motifs. Therefore the pattern can be translated into a binary code and compute the number of bits or information contained in the pattern. This is done by using Shannon's equation :

$$H = -\sum P_i \cdot \log_2 P_i \text{ (bits per motif)} \tag{4.1}$$

where H is the entropy, or the amount of information contained in a pattern, and P_i is the probability of having a given secondary structure motif. We can also measure the change in the amount of information as R:

$$R=H_{after}-H_{before} \tag{4.2}$$

Therefore, the different patterns can be classified into several clusters eliminating those that do not affect the amount of information of the total system, i.e. grouping one pattern into a given cluster does not decrease R.

Accordingly, snapshots stored during the MD trajectories were translated into a categorization system. Since there are 11 different possible motifs (H, 310, S, I1, I2, i1, i2, II1, II2, ii1 and ii2) and considering d dihedrals pairs for any peptide, a pattern could be defined using a vector of $d\times11$ elements. The system containing all the patterns consists of a matrix with binary elements of $d\times11$ columns and r rows, being r the number of different patterns to classify. At each position P_{ij} , corresponding to the *i-th* pattern, the presence or absence for the pattern will be represented by 1 or 0, respectively. Since the method can only tackle with vectors containing at least one non zero component, a label was added to each vector to skip analysis of those vectors with all components equal to zero.

The information of the system is computed separately for rows and columns with Eq. 4.3 and 4.4 :

$$H_r = -\sum_i^r P_i \cdot \log_2 P_i \tag{4.3}$$

being *r* the number of patterns

$$H_c = -\sum_j^c P_j \cdot \log_2 P_j \tag{4.4}$$

being $c = 11 \times n + 1$ and *n* the number of dihedral pairs. The total system has information :

$$H_t = -\sum_i^r \sum_j^c P_{ij} \cdot \log_2 P_{ij} \tag{4.5}$$

If $Pij=P_i \times P_j$ then $H_t=H_r+H_c$.

If the second categorization is not independent of the first then the information capacity of the system will be less than the maximum given by the sum of H_r+H_c. The difference from the maximum amount of information can be defined as *J* by the equation :

$$J=H_r+H_c-H_t \tag{4.6}$$

being J a measure of the degree of dependence between two systems of categorization and the relative information as :

$$R = \frac{J}{H_r + H_c} \tag{4.7}$$

If the initial matrix is modified eliminating one row, we could define the two states : state 1 : r_1 rows, c_1 columns, information deficit J_1 and relative information R_1 and the state 2 : with r_2, c_2, J_2 and R_2, where $r_2 < r_1$, $c_2 = c_1$ and $J_2 \leq J_1$. Then the clustering consists in searching the second state that gives a lower value of ΔJ, and thus increasing or not decreasing R_2.

The algorithm was used to cluster conformations until all patterns were connected. The value of R along the clustering process was plotted for the simulated annealing and for the two MD simulations at 300 K and 400 K. Plots differ depending on the complexity of the system. In order to have a criterion common to all the systems studied to finalize the clustering process, we defined a normalized *R*, $\overline{R}$, dividing *R* by the value it takes in the first step R_1. Accordingly, all the plots begin with a $\overline{R}$ =1. Afterwards we plotted $\Delta\overline{R}$ versus the clustering step and defined a threshold of

0.01, i.e. when the difference between two consecutives $\overline{R}$ was bigger than 0.01, the clustering process was stopped and the clusters that were kept were considered to be the minimum number of clusters necessary to have a description of the conformational diversity of the total system.

For each cluster, a collection of patterns with several conformational motifs shared in common was obtained. In order to render the information in a useful and graphical way, the probability within a cluster for each motif at each residue was computed by adding the number of times that a pattern with such motif at such position had appeared in the trajectory. This plot is a *picture* of the average behavior of the cluster.

An overview of the methodology used to classify and group the structures obtained in the SA procedure and MD trajectories is shown in Figure 4.1. With these two techniques for the exploration of the conformational space of peptides 5000 to 50000 structures constituted the initial set. With the CLASICO program described above the structures were classified into patterns based in the presence of different secondary structure motifs. This reduced the initial set of structures to 200 to 500 patterns. Finally, the patterns were grouped with the program CLUSTERIT as described above in 10 to 30 clusters. This procedure allowed for the characterization of the complexity of the group of structures obtained in the SA procedure and the MD trajectories and for the interpretation of the folding process.

4.3.5 Transition Analysis

The sampling procedure carried out either with MD or SA proceeds by translating each structure into a pattern that is subsequently assigned to a cluster. It might be interesting to analyze the visiting sequence of the different clusters along the sampling process. This analysis may provide information about the intrinsic characteristics of the process. A *transition* is defined when the pattern of a structure is assigned to a different cluster of the previous one already classified. By extension, if the structure is kept in the same cluster, it can be considered as an *autotransition*. A cluster is considered *stable* when the probability of suffering an autotransition is higher than the probability of falling in that cluster by a random classification, according to the size of the cluster. It is also possible to study the *reversibility* of transitions. Indeed, it is possible to determine whether the transitions taking place between two clusters occur with the same probability in both directions. A ratio of 1 is obtained when transitions in both directions are equally probable. However, this ratio must be higher than 1 for clusters that are sources, and smaller than 1 for clusters that are sinks. The percentage of transitions to a given cluster in regard to the total number of transitions of the cluster is called *local transitions.* Similarly, the percentage of transitions in regard the total number of transitions is termed *global transitions* and provides a measure of the statistical relevance of such transition.

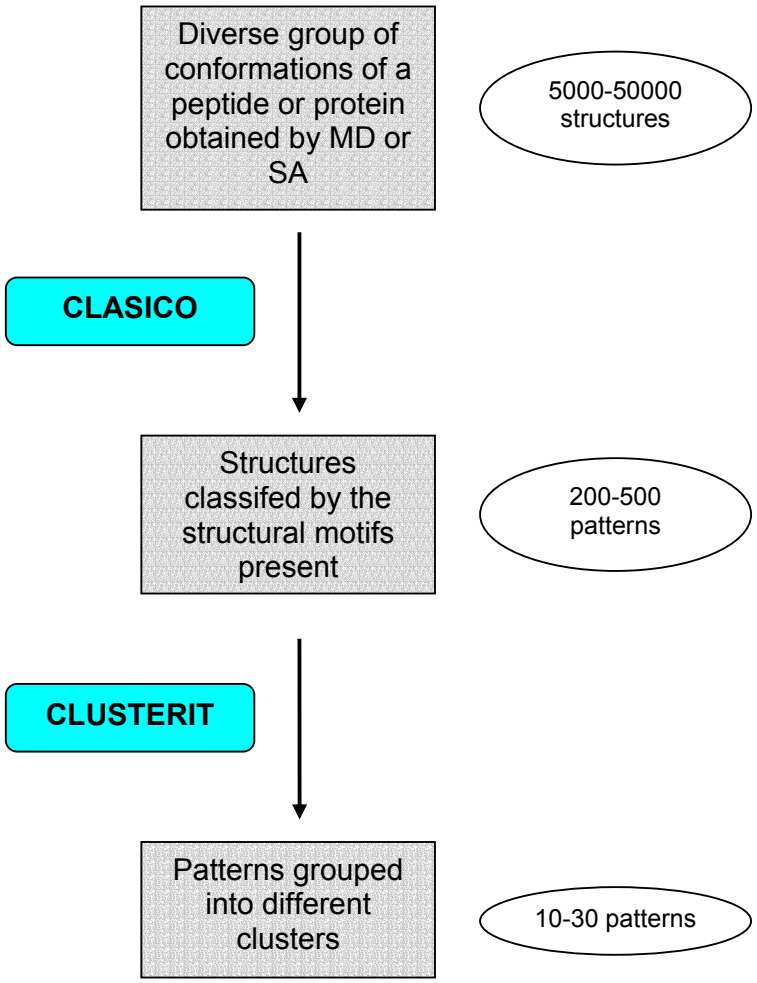

Figure 4.1. Overview of the methodology used for the characterization of the conformations obtained in the SA and MD trajectories at 300 and 400 K.

4.4 Results

4.4.1 Iterative simulated annealing of the peptide

In order to follow the evolution of the simulated annealing procedure the number of unique structures was calculated after each 100 new cycles. The number of unique conformations grows almost linearly with a slope of 0.96 (see Figure 4.2). However, the expected growth rate is logarithmic, reaching a plateau as we had previously seen in smaller systems (Corcho, et al., 1999). In the present case, after 5000 conformations the procedure had not yet reached the attenuation phase. This fact indicates that the conformational search should be extended beyond if the completeness criteria should be fulfilled. In order to check if the procedure had been trapped in a local region of the conformational space, the number of new structures after each 100 new conformations was monitored and found to oscillate around a mean value of 96.6 (see Figure 4.3), showing a good behavior in the simulated annealing process.

Energy distribution of unique conformations (density of states) after 5000 cycles of iterative SA is shown in Figure 4.4. Energies are referred to the global minimum obtained along the process. The bar width chose to construct the histogram was 0.8 kcal/mol with the maximum of the distribution located at 11.2 kcal/mol. As shown previously (Corcho et al., 2000), the rotational isomeric approximation can be used to predict the features of the density of states of a flexible molecule. Specifically, the maximum of the distribution U_0 can be computed from the effective number of torsional degrees of freedom, *f*; the mean number of rotamers of each rotor, *m+1*; and the window energy width, ε_0 (Flory, 1969):

$$U_0 = \frac{fm\varepsilon_0}{2} \tag{4.8}$$

In the present case, f = 22 (i.e. considering the 2 rotors of the backbone per residue (2x11 = 22)). Furthermore, considering m = 2 and $\tilde{\varepsilon}$ = 0.8; it can be obtained a value of 17.6 kcal/mol for $\tilde{U}$. However, the maximum of the distribution of our calculations is 11.2 kcal/mol. Therefore there is a difference of 6.4 kcal/mol between the predicted value from the rotational isomeric approximation and the actual value of the energy distribution computed from the SA calculations of the peptide. This discrepancy, as discussed previously (Corcho, et al. 2000) can be attributed to the presence of hydrogen bonds that stabilize energetically the conformations, enhancing the number of low energy conformations of the distribution to lower values.

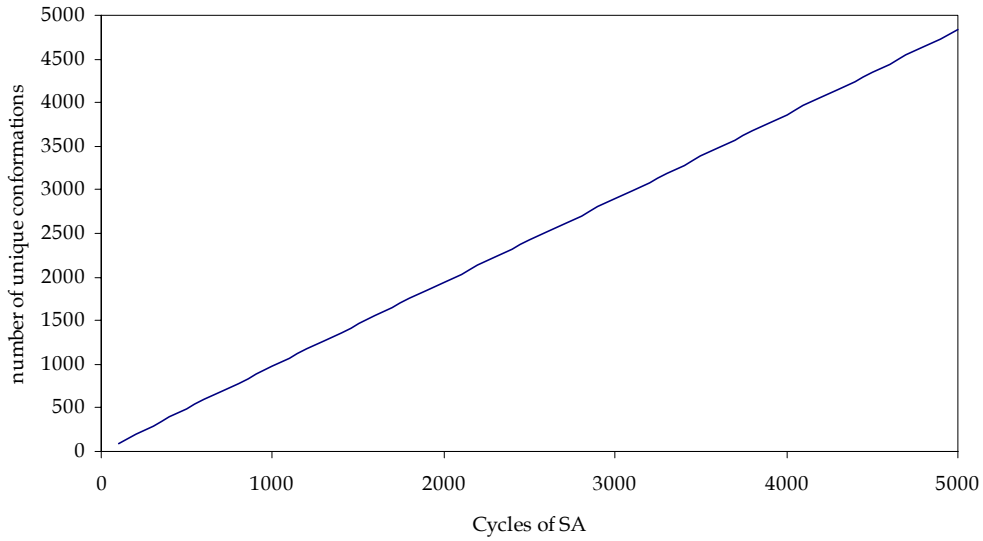

Figure 4.2. Evolution of the total number of unique conformations found along the SA procedure

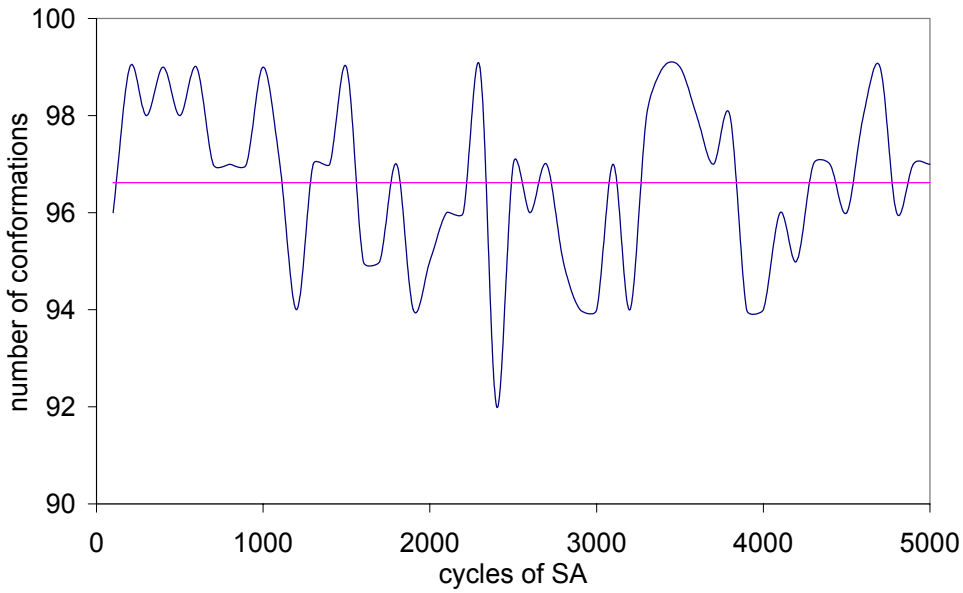

Figure 4.3. Evolution of new conformations for the SA procedure. The average value (96.6) is indicated as a straight line.

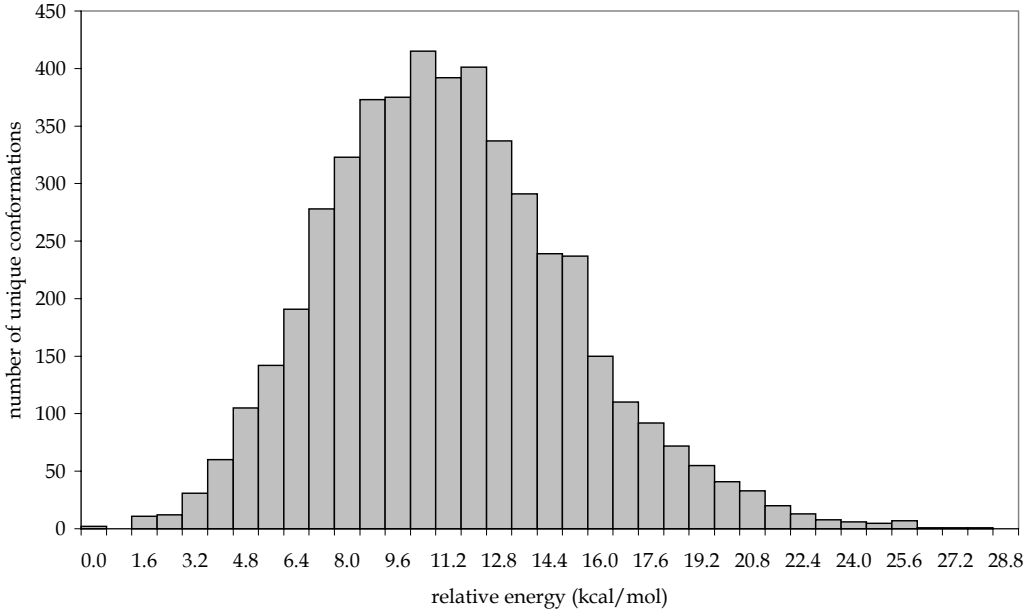

Figure 4.4. Histogram of the unique conformations of the SA procedure. The unique conformations have been previously rank-ordered by energy and referred to the lowest energy conformation found.

The conformational motifs of SPOH found in the SA exploration of the conformational space, were characterized using the *in house* algorithm CLASICO, described in the methods section. The statistics of all the motifs found in the peptide are shown in Figure 4.5 and Table 4.3. Each of the conformations found in the sampling process was labeled with an 11 character string, according to the motif present on each residue, providing 435 different patterns. The cumulative number of new patterns found during the exploration is shown in Figure 4.6. Similarly, to the plot of new unique structures (Figure 4.2), the plot shows a linear increase of the number of new patterns 5000 conformations. From both results it can be argued that the SA search has not reached convergence after 5000 cycles. From Figure 4.6 it can be deduced that all structures are not equally probable to find. Indeed, as it can be seen from this Figure, the structures most frequently sampled are contained within the 50 first patterns. Thus, the exploration proceeds by obtaining new conformations that are less probable. As a consequence the efficiency of the method decreases with the length of the exploration (Corcho et al., 1999), because the conformations that have not been already obtained are the less probable to find.

Table 4.3. Summary of secondary structure motifs present in the SA procedure and the MD trajectories at 300 and 400 K. (Percentages from the total structures of each experiment is shown in brackets).

DIHEDRAL PAIR NO.	SA	MD AT 300 K	MD AT 400 K
1	H(0.3), S(6.5)	S(69.3)	S(7.3)
2	H(0.8), S(12.4), I1(1.7), I2(1.3), II2(10.4), ii1(0.4)	S(69.3), II2(9.2)	S(7.3), II2(5.5)
3	H(4.4), S(13.1), I1(15.9), I2(2.1), II1(3.0), ii2(0.4)	S(69.4)	H(11.1), 3_{10}(8.2), S(7.3), I1(87.3)
4	H(5.2), S(10.0), I1(5.4), I2(16.6), II1(1.8), II2(3.0), ii1(0.9)	H(5.0), 3_{10}(5.1), S(40.3), I1(46.5)	H(15.2), 3_{10}(11.4), S(0.2), I1(9.5), I2(87.3)
5	H(6.5), S(5.1), I1(5.7), I2(11.0), i1(0.2), i2(0.2), II1(1.6), II2(1.8), ii1(0.6), ii2(0.9)	H(6.1), 3_{10}(9.3), S(0.1), I1(48.8), I2(46.5), ii1(0.2)	H(17.2), 3_{10}(13.1), I1(1.7), I2(87.7), ii1(0.2)
6	H(4.3), S(3.8), I1(4.4), I2(8.7), i1(0.6), i2(0.4), II1(3.8), II2(1.6), ii1(0.4), ii2(0.6)	H(6.1), 3_{10}(9.4), S(0.1), I1(1.1), I2(92.9), ii2(0.2)	H(14.5), 3_{10}(9.1), S(0.1), I1(7.7), I2(79.4), ii1(0.1)
7	H(3.4), S(1.9), I1(2.4), I2(7.2), i1(0.7), i2(0.9), II1(1.1), II2(3.8), ii1(2.0), ii2(0.4)	H(1.4), 3_{10}(6.0), S(0.1), I1(0.8), I2(86.8), i1(0.4), II1(0.1)	H(5.1), 3_{10}(9.3), S(0.1), I1(4.8), I2(81.6), II1(0.5)
8	H(1.4), S(0.7), I1(3.6), I2(3.6), i1(0.6), i2(1.0), II1(0.8), II2(1.1), ii1(3.8), ii2(2.0)	H(0.1), 3_{10}(1.3), I1(2.0), I2(13.3), i2(0.4), II2(0.1)	H(0.8), 3_{10}(6.7), S(0.1), I1(1.9), I2(44.8), II2(0.5), ii1(0.2)
9	H(0.8), S(0.2), I2(4.7), i2(0.9), II2(0.8), ii2(3.8)	3_{10} (1.1), I2(11.9)	H(0.2), 3_{10}(4.4), S(0.1), I2(38.2), ii2(0.2)

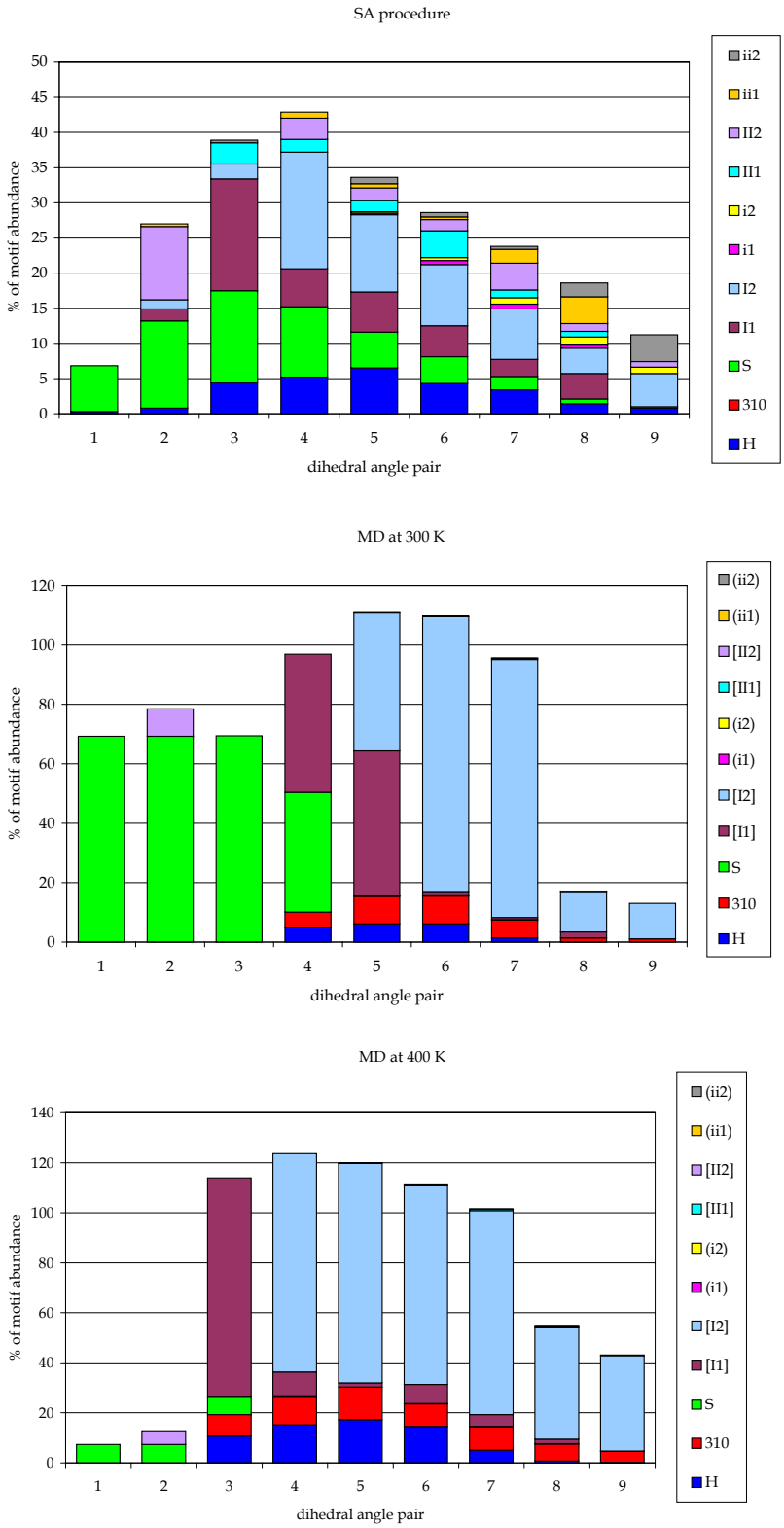

Figure 4.5. Cumulative statistics of conformational motifs for the SA procedure and the MD trajectories at 300 and 400 K.

Structures were classified into clusters using the CLUSTERIT algorithm, previously described in the methods section, using a convergence criterion for the clustering process of $\Delta\overline{R} \geq 0.01$ (Figure 4.7). The process yielded 18 clusters, shown in Figure 4.8. Clusters were rank-ordered according to the number of structures that were contained in each cluster. Table 4.4 summarizes the main structural features and the importance for each cluster in the SA procedure. Inspection of Table 4.4 reveals that, the 80% of the structures are described within the first 8 clusters, pointing out that structures are quite distributed amongst the different clusters. Cluster I is the most abundant cluster (32.1% of the total number of structures), containing a single pattern that exhibits no conformational motifs. The second most abundant cluster, cluster II (13.5%) exhibits as characteristic feature a α-helical turn running from residues 4 to 7. This cluster contains the second most abundant pattern found, with a 6.2% of the structures and corresponds to a conformation containing a type I β-turn between residues 4 and 5. The next most abundant cluster, cluster III (8.5%) corresponds to a group of structures exhibiting a II2 motif on the third residue (57%) and with a small percentage of the structures (30%) with a type II β-turn on residues 7 and 8. Cluster IV (5.6%) is characterized by a β-strand extending from 2 to 7 and in a smaller proportion (9 and 12%) two type II β-turns on residues 4 and 5 and 7 and 8, respectively. Cluster V (5.5%) exhibits three main features: two type I β-turns on residues 5 and 6 (46%) and 8 and 9 (38%), and one type II β-turn on residue 6 and 7 (24%). Cluster VI (4.9%) is characterized by a β-strand extending from 2 to 5 with varying percentages (29 to 90%). Cluster VII (4.8) groups patterns that show alternatively a β-strand in residues 4 to 8 or type II β-turns on residues 4 and 5 and 8 and 9, respectively. Cluster VIII is also characterized by a α-helical turn (35%) and type I β-turn (40-99%). The rest of clusters from IX to XVIII represent only 20% of the structures obtained in the SA procedure. These clusters show more rare events like type I' β-turn (residues 8 and 9 and 9 and 10), type II' β-turns (at residues 5 and 6 and 8 and 9) or an α-helical turn on residues 2 to 5, being 4 to 9 the most probable region for α-helices under these experimental conditions. Cluster XVIII only represents 0.8% of the structures obtained in the SA procedure and groups all the structures showing only β-strand motifs. The region extends from residues 2 to 7 at varying proportions (10-69).

Transitions between clusters were analyzed and the results are shown in Table 4.5 and Figure 4.9. Only 8 clusters, representing 80% of the structures and local transitions above 5%, are depicted for conciseness. From the transitions plot (Figure 4.9) it can be seen that the clusters are highly interconnected. Indeed, most of the clusters shown exhibit at least 5 arrows departing or arriving, indicating that each of these clusters are the source or the target in at least 5 instances of more than 5% of their respective local transitions. Furthermore, the connections do not seem to be preferred within a subset of clusters thus indicating that all structures belong to the same unique group of structures or that it does not exist a subset of clusters presenting transitions with a higher probability.

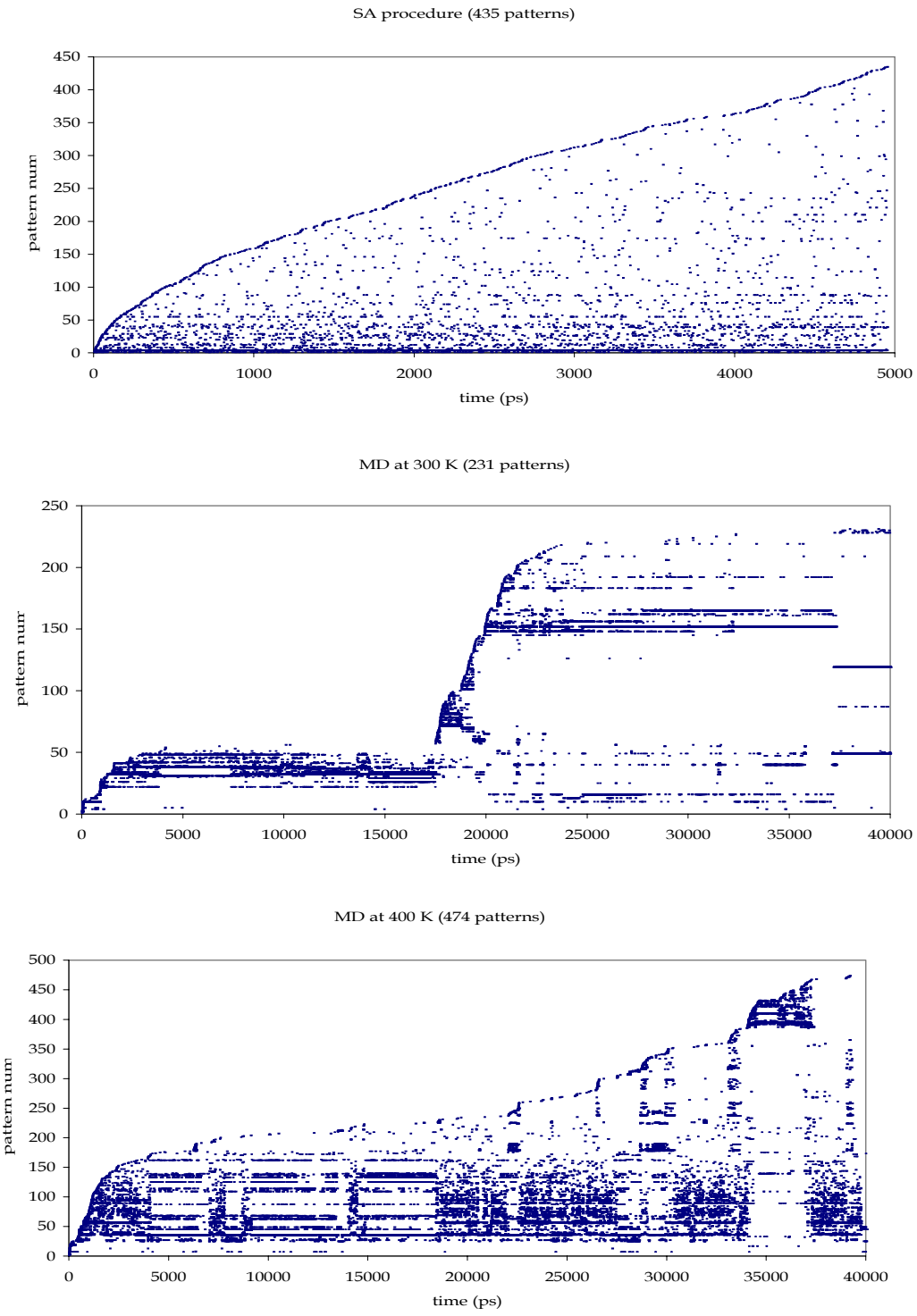

Figure 4.6. Evolution of patterns for the SA procedure and the MD at 300 and 400 K.

Table 4.4. Summary of the main structural features of the 18 clusters obtained using CLUSTERIT for the SA procedure. The second column shows the percentage of the total number of structures obtained in the SA that are contained within each cluster. The last column corresponds to the percentage of structures within each cluster that exhibit each motif. When the motif expands more than one residue with different percentages the interval is shown.

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
I	32.1	none	--	--
II	13.5	α-helical turn	4-7	10-30
		type I β-turn	4-7	12-95
III	8.5	II2	3	57
		type II β-turn	7-8	30
IV	5.6	β-strand	2-7	15-95
		type II β-turn	4-5	9
		type II β-turn	7-8	12
V	5.5	type I β-turn	5-6	46
		type I β-turn	8-9	38
		type II β-turn	6-7	24
VI	4.9	β-strand	2-5	29-90
VII	4.8	β-strand	4-8	14-55
		type II β-turn	4-5	34
		type II β-turn	8-9	16
		II motif	3	14
VIII	4.7	α-helical turn	6-8	35
		type I β-turn	6-8	40-99
IX	4.1	α-helical turn	7-9	14
		type I β-turn	7-9	20-93
		type I β-turn	4-5	18
X	3.1	type I β-turn	4-6	12-16
		type I' β-turn	9-10	15
		type II' β-turn	9-10	84

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
XI	2.9	type I β-turn	8-9	40
		type I β-turn	4-5	12
		type II β-turn	4-5	13
		type II' β-turn	8-9	51
		II motif	12	12
XII	2.7	α-helical turn	2-5	12-31
		type I β-turn	3-5	24-86
		type II β-turn	7-8	17
		II motif	2	9
XIII	2.0	type II β-turn	5-6	67
		type II' β-turn	5-6	33
		II motif	3	18
XIV	1.7	β-strand	2-5	38-99
		type I β-turn	6-7	47
		type I β-turn	9-10	21-26
		type I' β-turn	8-9	11
		type II β-turn	5-6	16
		type II' β-turn	8-9	13
XV	1.1	α-helical turn	4-8	25-66
		type I β-turn	4-8	32-100
XVI	1.0	β-strand	6-10	15-60
		type I β-turn	5-6	8-10
		type II β-turn	9-10	48
XVII	0.9	α-helical turn	8-10	53
		type I β-turn	8-10	53-71
		type I β-turn	5-6	18-20
		II motif	3	18
XVIII	0.8	β-strand	2-7	10-69
		type I β-turn	6-7	26-31
		type II β-turn	4-5	17
		type II' β-turn	9-10	90
		type II' β-turn	5-6	10

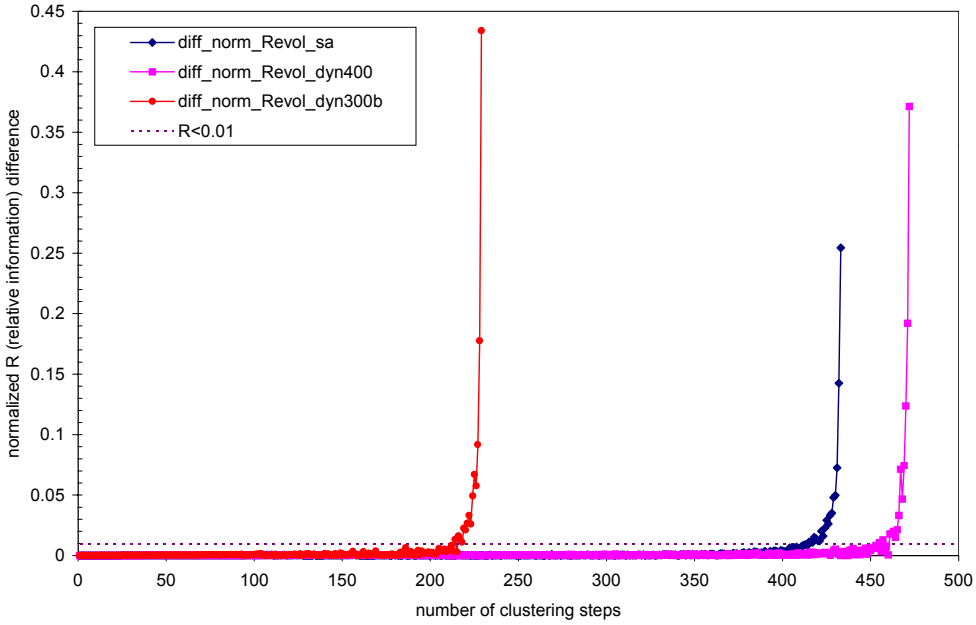

Figure 4.7. Evolution of the difference of normalized relative information $(\Delta\overline{R})$ between two consecutive clustering steps. The stopping condition of the clustering process, $\Delta\overline{R} \geq 0.01$, is also depicted.

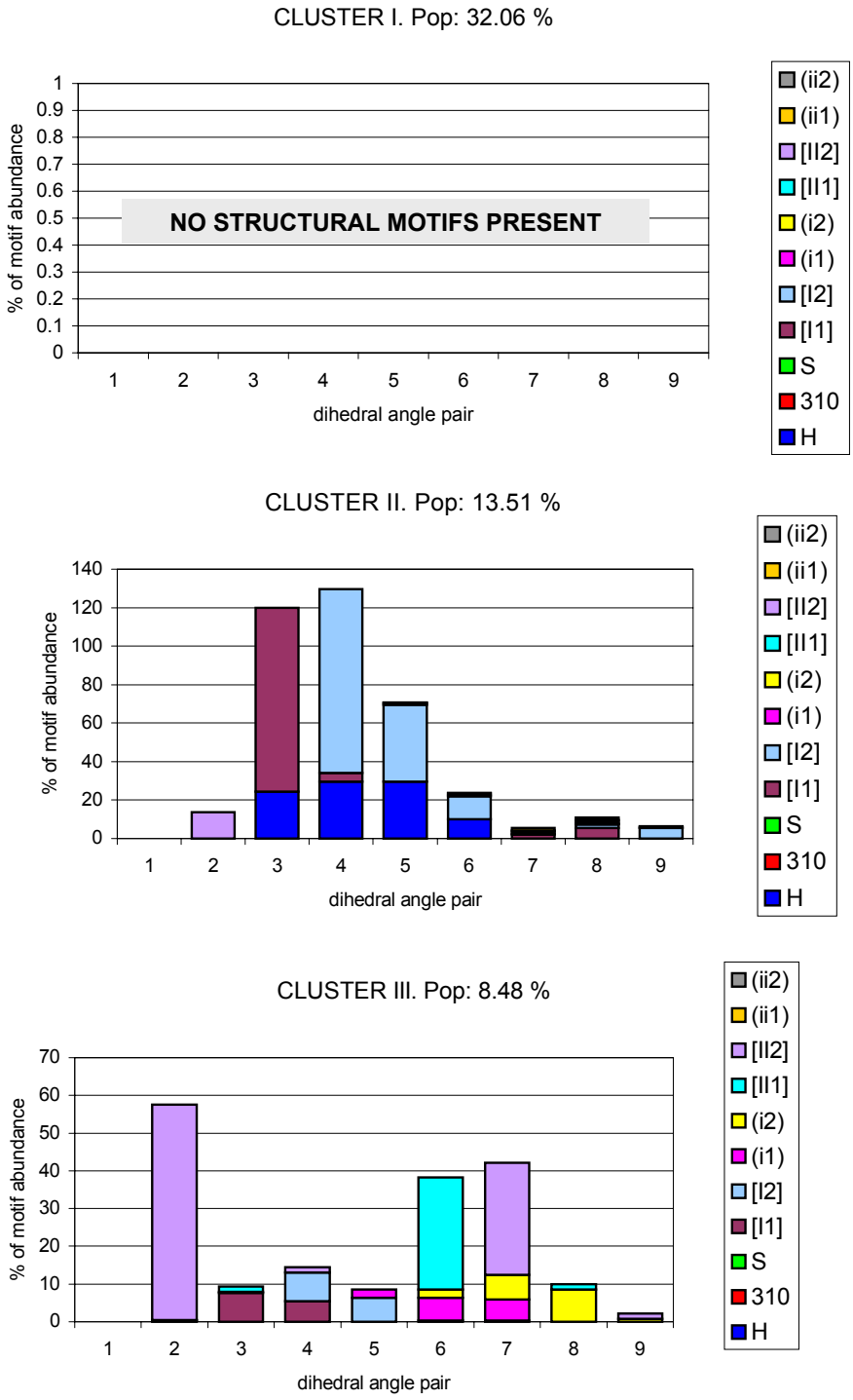

Figure 4.8. Clusters obtained for the SA procedure.

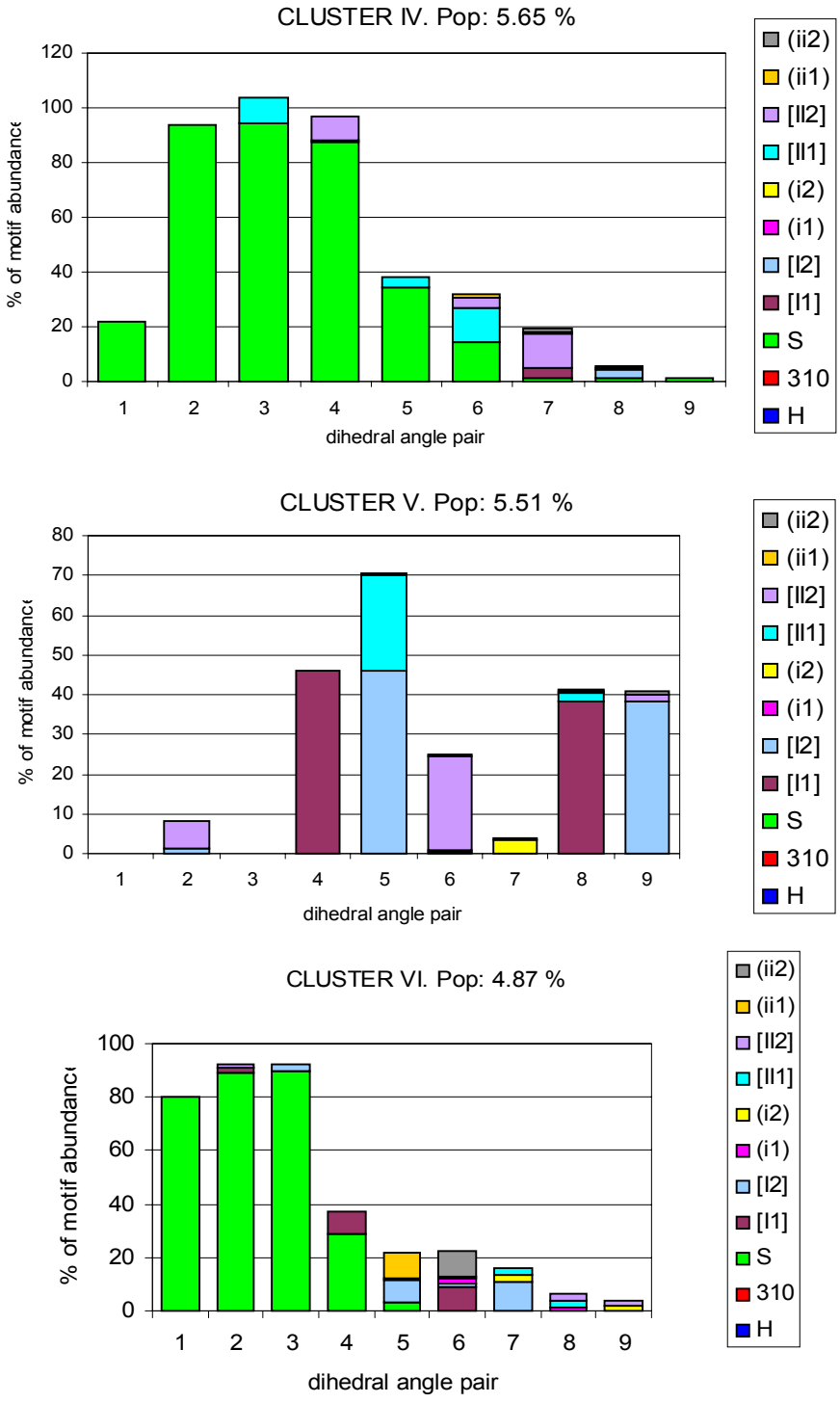

Figure 4.8. Clusters obtained for the SA procedure.

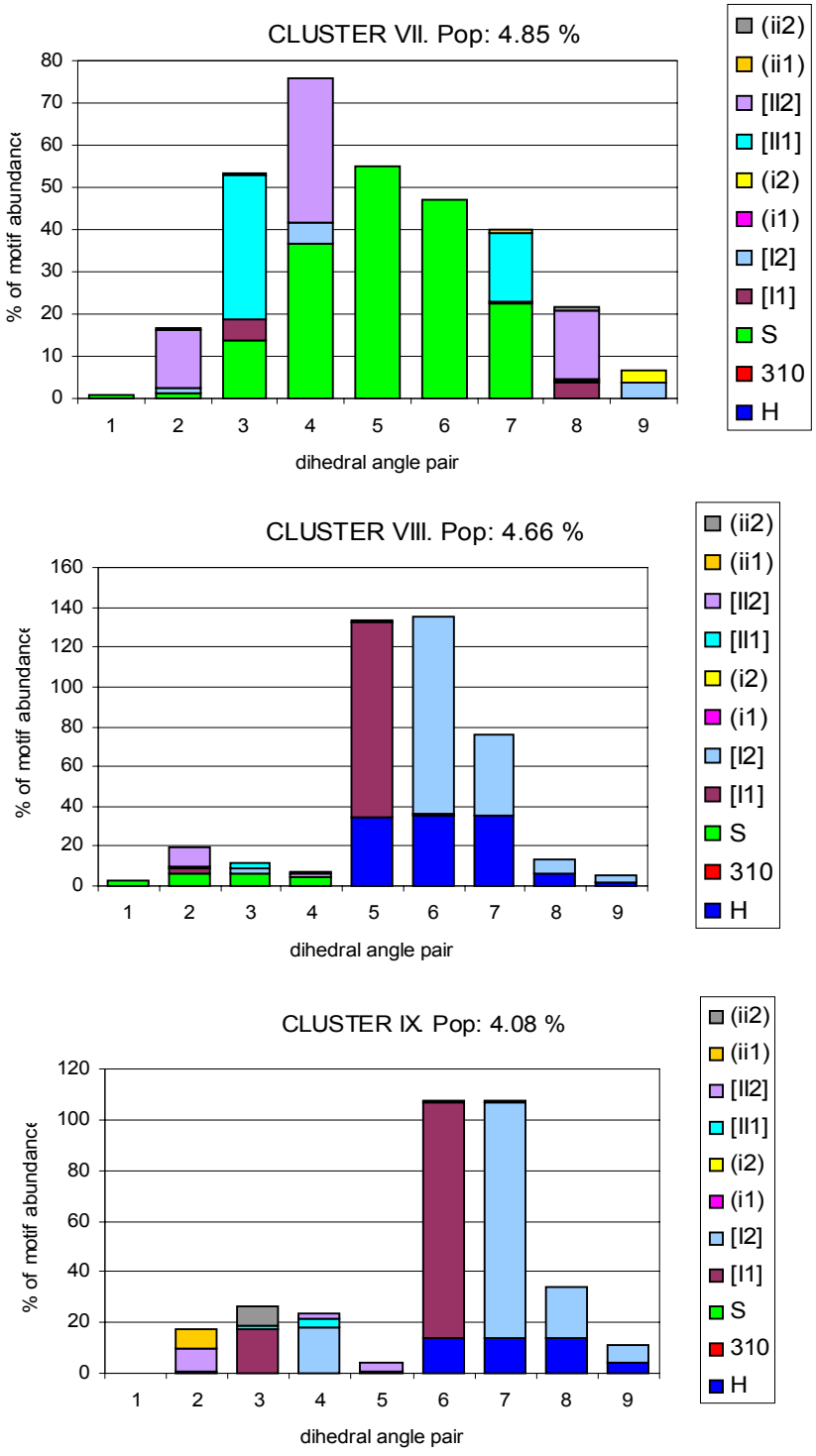

Figure 4.8. Clusters obtained for the SA procedure.

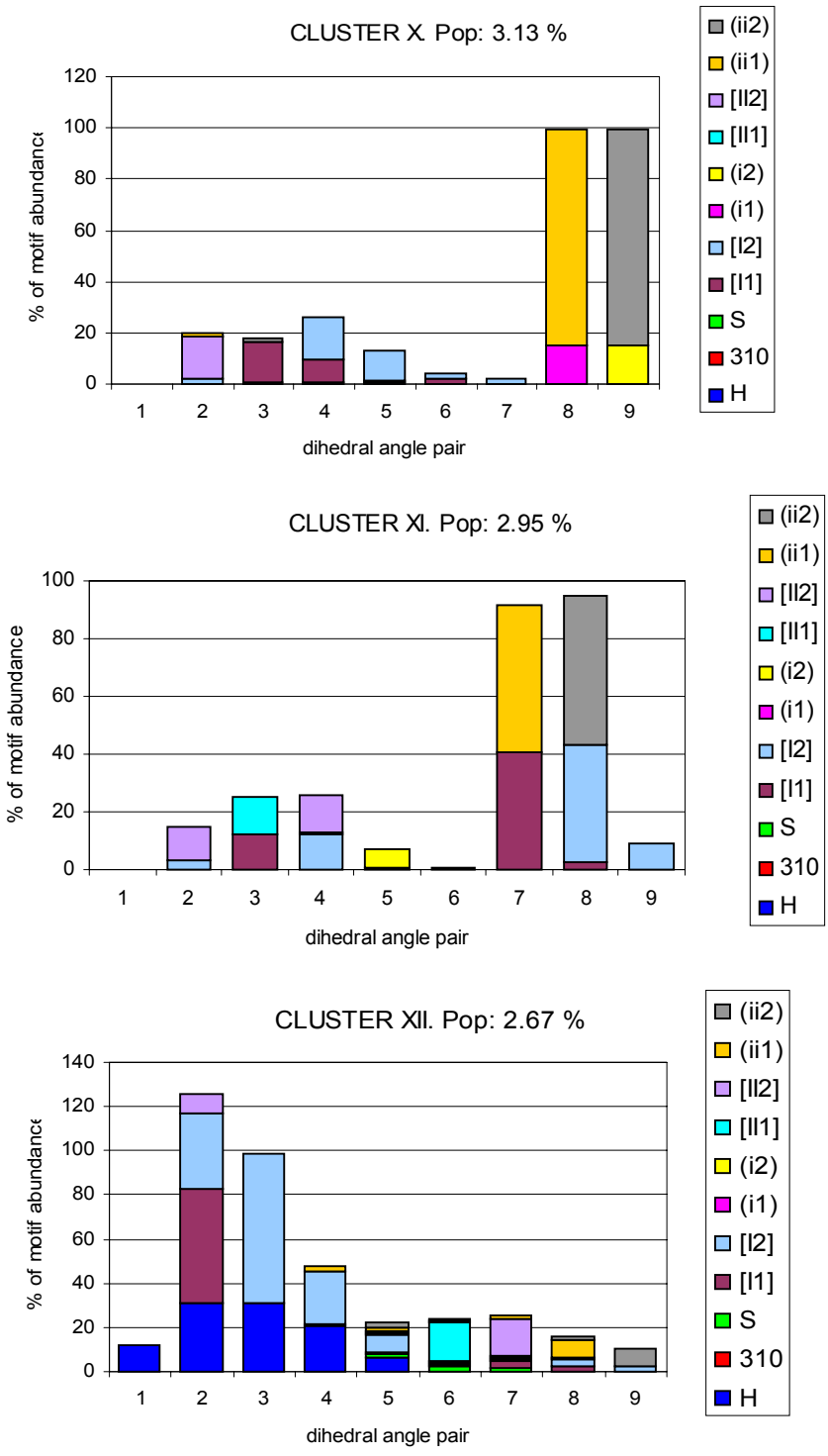

Figure 4.8. Clusters obtained for the SA procedure.

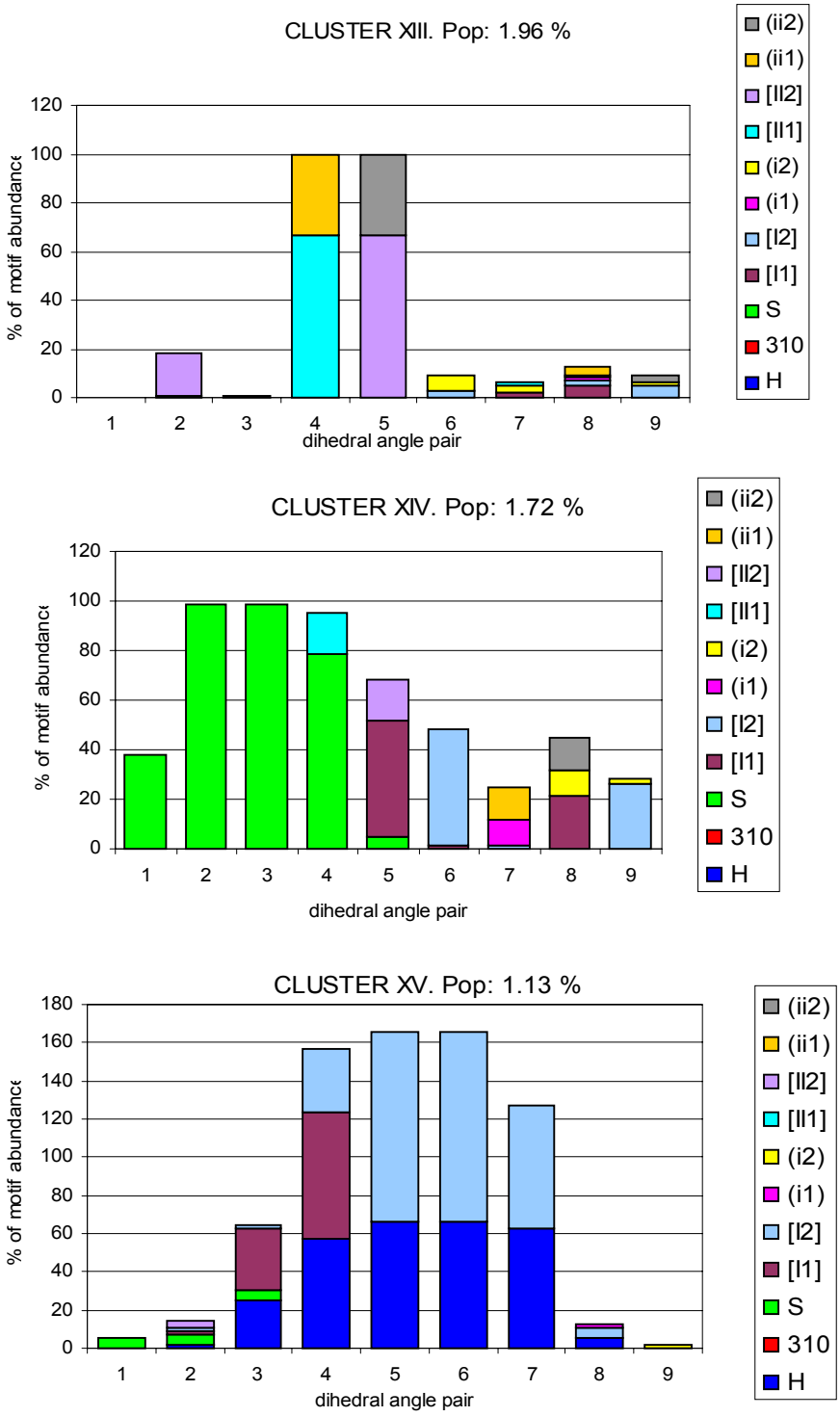

Figure 4.8. Clusters obtained for the SA procedure.

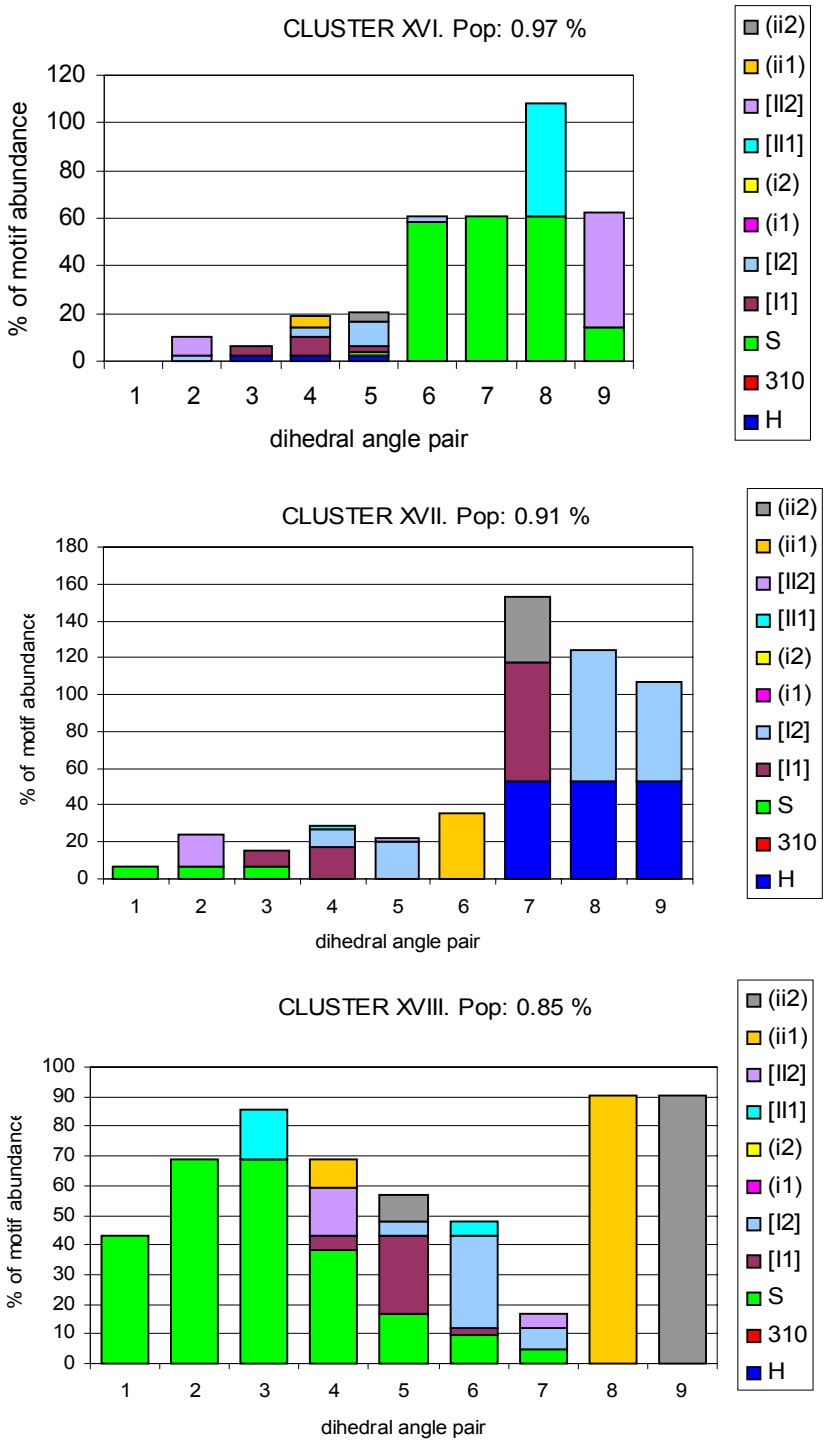

Figure 4.8. Clusters obtained for the SA procedure.

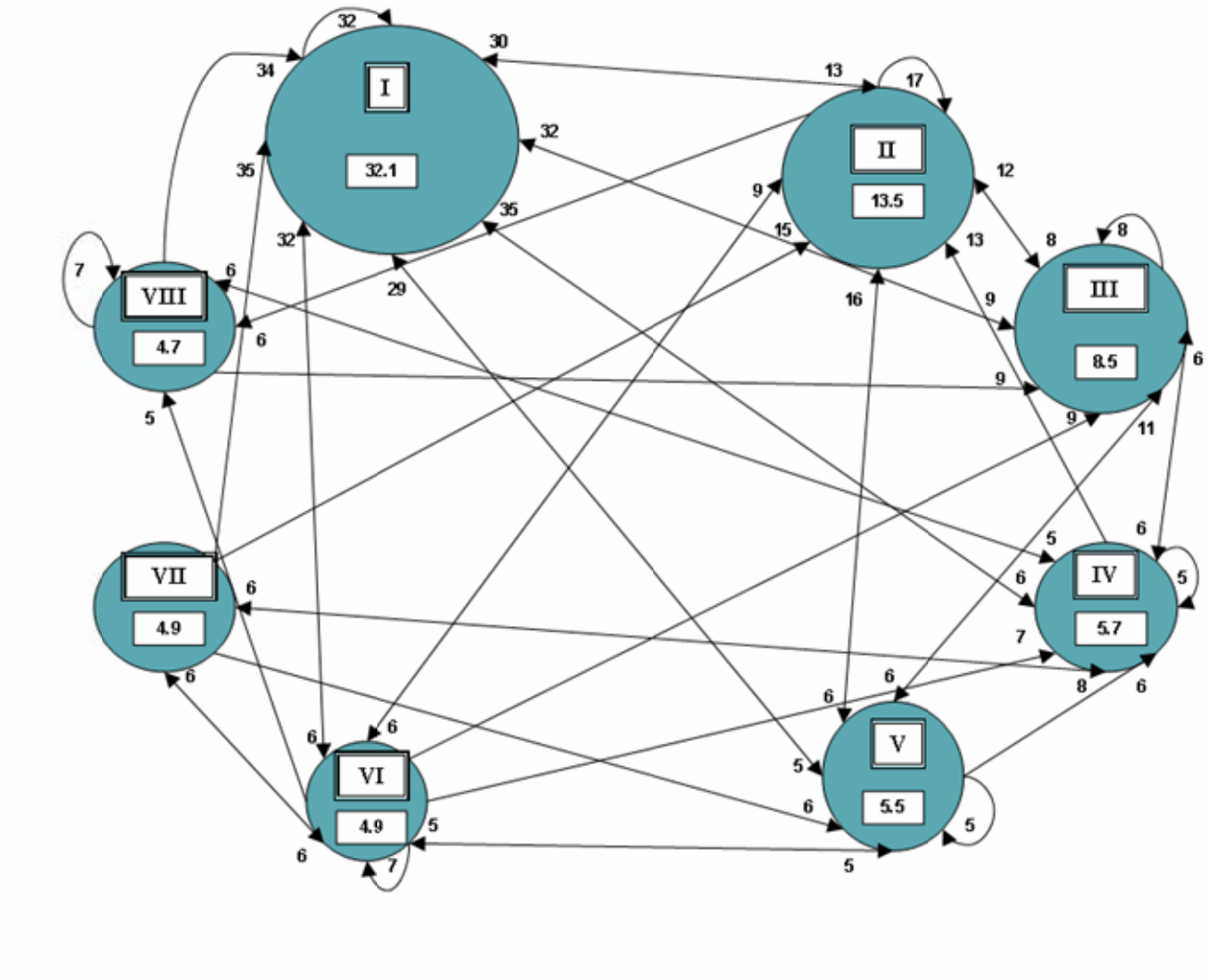

Figure 4.9. Transitions between the 8 largest clusters for the SA procedure

Table 4.5. Summary of the local transitions with probability larger than 5% obtained for the SA procedure.

CLUSTERS		TRANSITIONS			REVERSIBILITY	STRUCTURES	
Origin	End	Total	% Local	% Global		Total	%
I	V	87	5.48	3.35	1.1	1588	32.06
	II	204	12.85	8.12	1.03		
	IV	103	6.49	4.08	1.04		
	VI	89	5.6	3.33	1.17		
	III	137	8.63	5.45	1.03		
	I	510	32.12	10.3	1		
II	V	37	5.53	1.62	0.86	669	13.51
	VIII	43	6.43	1.49	1.39		
	II	113	16.89	2.28	1		
	III	51	7.62	2.08	0.98		
	I	198	29.6	8.12	0.97		
III	V	24	5.71	1.07	0.83	420	8.48
	II	52	12.38	2.08	1.02		
	IV	26	6.19	0.87	1.53		
	III	35	8.33	0.71	1		
	I	133	31.67	5.45	0.97		
IV	VIII	18	6.43	0.61	1.5	280	5.65
	II	37	13.21	1.35	1.23		
	VII	16	5.71	0.69	0.89		
	IV	14	5	0.28	1		
	III	17	6.07	0.87	0.65		
	I	99	35.36	4.08	0.96		
V	V	14	5.13	0.28	1	273	5.51
	II	43	15.75	1.62	1.16		
	IV	16	5.86	0.59	1.23		
	VI	14	5.13	0.53	1.17		
	IX	16	5.86	0.59	1.23		
	III	29	10.62	1.07	1.21		
	I	79	28.94	3.35	0.91		
VI	V	12	4.98	0.53	0.86	241	4.87
	VIII	12	4.98	0.44	1.2		
	II	22	9.13	0.91	0.96		

	VII	15	6.22	0.61	1		
	IV	16	6.64	0.59	1.23		
	VI	18	7.47	0.36	1		
	III	21	8.71	0.79	1.17		
	I	76	31.54	3.33	0.85		
VII	V	14	5.83	0.48	1.4	240	4.85
	II	37	15.42	1.37	1.19		
	IV	18	7.5	0.69	1.13		
	VI	15	6.25	0.61	1		
	III	18	7.5	0.77	0.9		
	I	83	34.58	3.25	1.06		
VIII	VIII	16	6.93	0.32	1	231	4.66
	II	31	13.42	1.49	0.72		
	IV	12	5.19	0.61	0.67		
	III	20	8.66	0.81	1		
	I	79	34.2	2.87	1.25		
IX	V	13	6.44	0.59	0.81	202	4.08
	VIII	12	5.94	0.44	1.2		
	II	18	8.91	1.01	0.56		
	VII	12	5.94	0.44	1.2		
	IV	16	7.92	0.53	1.6		
	IX	14	6.93	0.28	1		
	III	18	8.91	0.71	1.06		
	I	57	28.22	2.22	1.08		
X	V	11	7.1	0.4	1.22	155	3.13
	VIII	9	5.81	0.4	0.82		
	II	22	14.19	0.91	0.96		
	VII	10	6.45	0.38	1.11		
	III	12	7.74	0.4	1.5		
	I	55	35.48	2.08	1.15		
XI	II	23	15.75	0.81	1.35	146	2.95
	VII	9	6.16	0.28	1.8		
	IV	8	5.48	0.3	1.14		
	VI	8	5.48	0.16	0		
	III	15	10.27	0.69	0.79		
	I	44	30.14	2	0.8		

XII	V	7	5.3	0.32	0.78	132	2.67
	II	16	12.12	0.71	0.84		
	VII	7	5.3	0.2	2.33		
	VI	7	5.3	0.26	1.17		
	III	12	9.09	0.42	1.33		
	I	45	34.09	1.72	1.13		
XIII	V	7	7.22	0.22	1.75	97	1.96
	VIII	5	5.15	0.22	0.83		
	II	16	16.49	0.57	1.33		
	VII	5	5.15	0.16	1.67		
	III	9	9.28	0.34	1.13		
	I	40	41.24	1.45	1.25		
XIV	V	7	8.24	0.3	0.88	85	1.72
	II	10	11.76	0.48	0.71		
	VI	6	7.06	0.18	2		
	III	8	9.41	0.28	1.33		
	I	26	30.59	1.07	0.96		
XV	V	3	5.36	0.1	1.5	56	1.13
	II	11	19.64	0.46	0.92		
	VII	3	5.36	0.06	0		
	VI	3	5.36	0.12	1		
	III	6	10.71	0.3	0.67		
	I	17	30.36	0.65	1.13		
XVI	V	4	8.33	0.12	2	48	0.97
	II	5	10.42	0.16	1.67		
	VII	3	6.25	0.1	1.5		
	IV	3	6.25	0.1	1.5		
	VI	3	6.25	0.1	1.5		
	III	5	10.42	0.24	0.71		
	I	19	39.58	0.79	0.95		
XVII	XIII	4	8.89	0.08	0	45	0.91
	II	5	11.11	0.24	0.71		
	VII	3	6.67	0.08	3		
	IV	4	8.89	0.1	4		
	III	5	11.11	0.18	1.25		
	I	15	33.33	0.55	1.25		

XVIII	V	3	7.14	0.08	3	42	0.85
	VIII	3	7.14	0.06	0		
	II	4	9.52	0.18	0.8		
	IX	3	7.14	0.08	3		
	XI	4	9.52	0.14	1.33		
	I	13	30.95	0.57	0.87		

4.4.3. Molecular Dynamics in a water box at 300 K

A 40 ns MD trajectory of the peptide at 300 K soaked in a box of water molecules was undertaken starting from its extended conformation, using periodic boundary conditions and Particle Mesh Ewald summations as described in the methods section. Total energy and density were kept fairly constant during the simulation at -9102 kcal/mol and 0.945 g/cm^3, respectively.

The CLASICO algorithm was used to analyze the structures of the MD simulation at 300 K in order to determine the conformational motifs attained by the peptide that are summarized in Figure 4.5 and Table 4.3. The different conformations were labeled according to the conformational motifs exhibited and classified into 231 different patterns. Figure 4.6 shows the evolution of new patterns sampled along time. At the beginning of the trajectory the profile exhibits a plateau at 5 ns after an initial steady increase of the number of patterns. This behavior can be associated with the sampling of new secondary structures from the extended conformation. The number of new patterns is very low until at nanosecond 17 the peptide samples an II2 motif located on the third residue, generating a wealth of new patterns from this point, until a second plateau is reached after 25 ns.

The folding process of the peptide can be analyzed in view of the evolution of patterns sampled. Since the simulation started from the extended conformation, the first patterns sampled correspond to structures with no conformational motifs or adopting dihedral angles of a β-strand. Subsequently, SPOH adopts a type I β-turn between residues 9 and 10 (457 ps) that unfolds quickly (724 ps), falling into a type I β-turn between residues 6 and 7 after 785 ps. This conformation appears to act as a *nucleation site* of the region that extends from residue 5 to 9, that folds lately to form two consecutive turns that will last for the last 39 ns of the MD trajectory with a combination of different turn types. These central turns adopt sporadically the form of a α- or 3_{10}-helix between residues 5 to 7 or between residues 6 to 8. The increase in the number of patterns from 17 ns corresponds to the sampling of an II2 motif in residue number 3. This motif is present

together with the two central turns in the molecule mainly although sometimes appears with the 3_{10}-helix and more rarely with the α-helix.

In summary, the simulation points to the tendency of SPOH to attain a helical turn in the center of the molecule, however this turn is not stable and the molecule fluctuates around an ensemble of folded conformations characterized by the presence of two consecutive turns between residues from 5 to 9.

The structures of SPOH previously characterized were classified into clusters using the CLUSTERIT algorithm, described in the methods section. A criterion of $\Delta\overline{R} \geq 0.01$ (Figure 4.7) was adopted to stop the clustering process. Using this criterion 16 clusters remained (Figure 4.10). Table 4.6 summarizes the main structural features of the clusters found from the analysis of the MD trajectory at 300 K. The four most abundant clusters, represent 85% of the structures (shown in Figure 4.10, clusters I to IV) suggesting that the bulk of structures is grouped in a small number of clusters. The most abundant cluster (containing 35.5% of the total number of structures) exhibits β-strand dihedrals angle for residues 2 to 5 and a type I β-turn on residues 6 to 9. The second most abundant cluster (21.8%) exhibits β-strand values on residues 2 to 4 and two consecutives type I β-turns extending from residues 5 to 8. Furthermore, 17% of the structures exhibit a 3_{10}-helical turn extending from residues 5 to 7. The third most abundant cluster (19.9%) exhibits a α-helical turn extending from residues 5 to 7, present in 17% of the structures of this cluster. The type I β-turn expands from residues 5 to 9 in a proportion going from 13 to 100%. Cluster IV groups structures presenting an II2 motif on residue 3 (72%) and type I β-turns extending from residues 6 to 8. The rest of clusters representing 15% of the structures exhibit less frequent structural features, like type I β-turns in two regions 5 to 7 and 9 and 10, or the β-strand (residues 2 to 4) in conjunction with a α-helical turn (residues 5 to 8), or the II2 motif (residue 3) together with a 3_{10}-helical turn. The type I' β-turn is rarely present for residues 6 and 7 and 8 and 9. Cluster XV and XVI, both accounting for 0.1% of the structures, group the structures that exhibit only β-strand or do not present any conformational motifs. To sum up, the peptide at 300 K can be described as a folded molecule exhibiting consecutive type I β-turns in the region expanding from residues 5 to 9 and with residues 2 to 4 exhibiting a β-strand or an II2 motif on residue 3. For residues 5 to 7 a α-helical turn or a 3_{10}-helical turn are the most probable structural features observed.

Transitions between clusters for the MD trajectory at 300 K were analyzed as described in the methods section and the data is presented in Table 4.7 and Figure 4.11. Only 9 clusters representing 98% of the structures and local transitions above 5% have been depicted for conciseness. Although transitions between the two most abundant clusters with local transitions

less than 5%, are also shown for later discussion. Comparison of Figures 4.10 and 4.8 suggest that clusters are less interconnected in MD than in the SA procedure. Furthermore, the connections observed, suggest that clusters can be grouped into two different classes. On the one hand, clusters I, IV, V, VI and VIII and on the other, clusters II, III, VII and IX. Both classes exhibit connections that are contained within each of the subgroups having few transitions between cluster I and II (less than 1% of their respective local transitions).

Table 4.6. Summary of the main structural features of the 16 clusters obtained using CLUSTERIT for the MD trajectory at 300 K. The second column shows the percentage of the total number of structures obtained in this MD trajectory that are contained within each cluster. The last column corresponds to the percentage of structures within each cluster that exhibit each motif. When the motif expands more than one residue with different percentages the interval is shown.

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
I	35.5	β-strand type I β-turn	2-5 6-9	89-100 15-100
II	21.8	β-strand 3_{10}-helical turn type I β-turn	2-4 5-7 5-8	100 16-17 87-100
III	19.9	α-helical turn type I β-turn	5-7 5-9	16-17 13-100
IV	7.9	type I β-turn II2 motif	6-8 3	99-100 72
V	4.0	β-strand 3_{10}-helical turn type I β-turn	2-5 6-8 6-10	86-98 94-100 13-100
VI	2.9	β-strand type I β-turn	2-5 7-8	96-100 29-30
VII	2.1	β-strand α-helical turn type I β-turn	2-4 5-8 5-8	100 83-100 100

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
VIII	2.0	β-strand type I β-turn type I β-turn	2-5 6-7 9-10	94-98 73 95
IX	1.9	3_{10}-helical turn type I β-turn II2 motif	5-8 5-9 3	37-100 34-100 35
X	0.8	β-strand type I β-turn	2-5 8-10	89-93 91-100
XI	0.4	β-strand type I β-turn type I' β-turn	2-5 5-7 8-9	99 88 100
XII	0.4	type I β-turn type I' β-turn	7-10 6-7	59-100 38
XIII	0.2	α-helical turn 3_{10}-helical turn type I β-turn II motif	5-7 8-10 5-10 3	26 100 97-100 59
XIV	0.1	β-strand 3_{10}-helical turn type I β-turn type I β-turn II motif	2-4 5-7 5-7 9-10 3	35 45-49 92-100 92 53
XV	0.1	β-strand	2-9	12-97
XVI	0.1	none	--	--

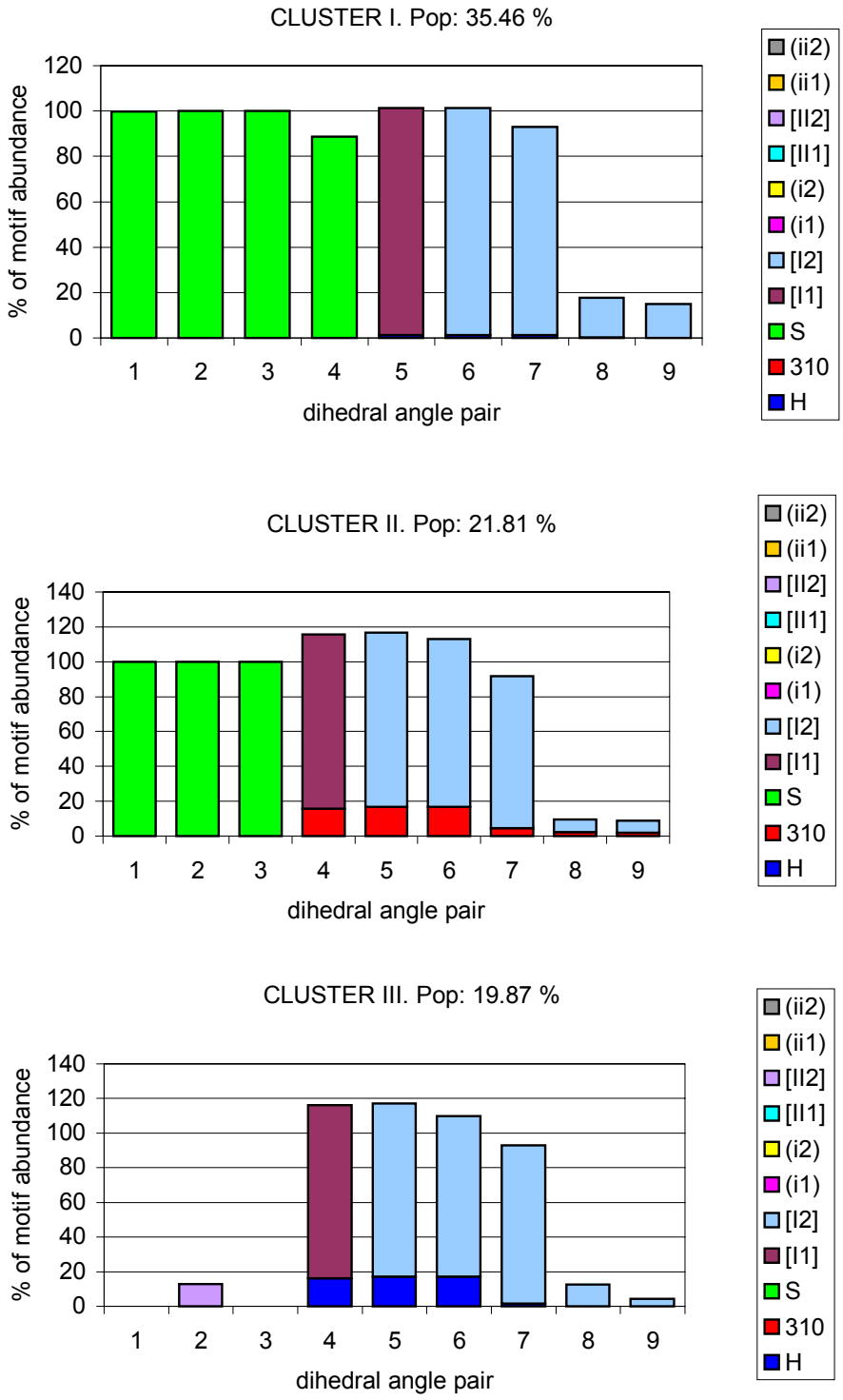

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

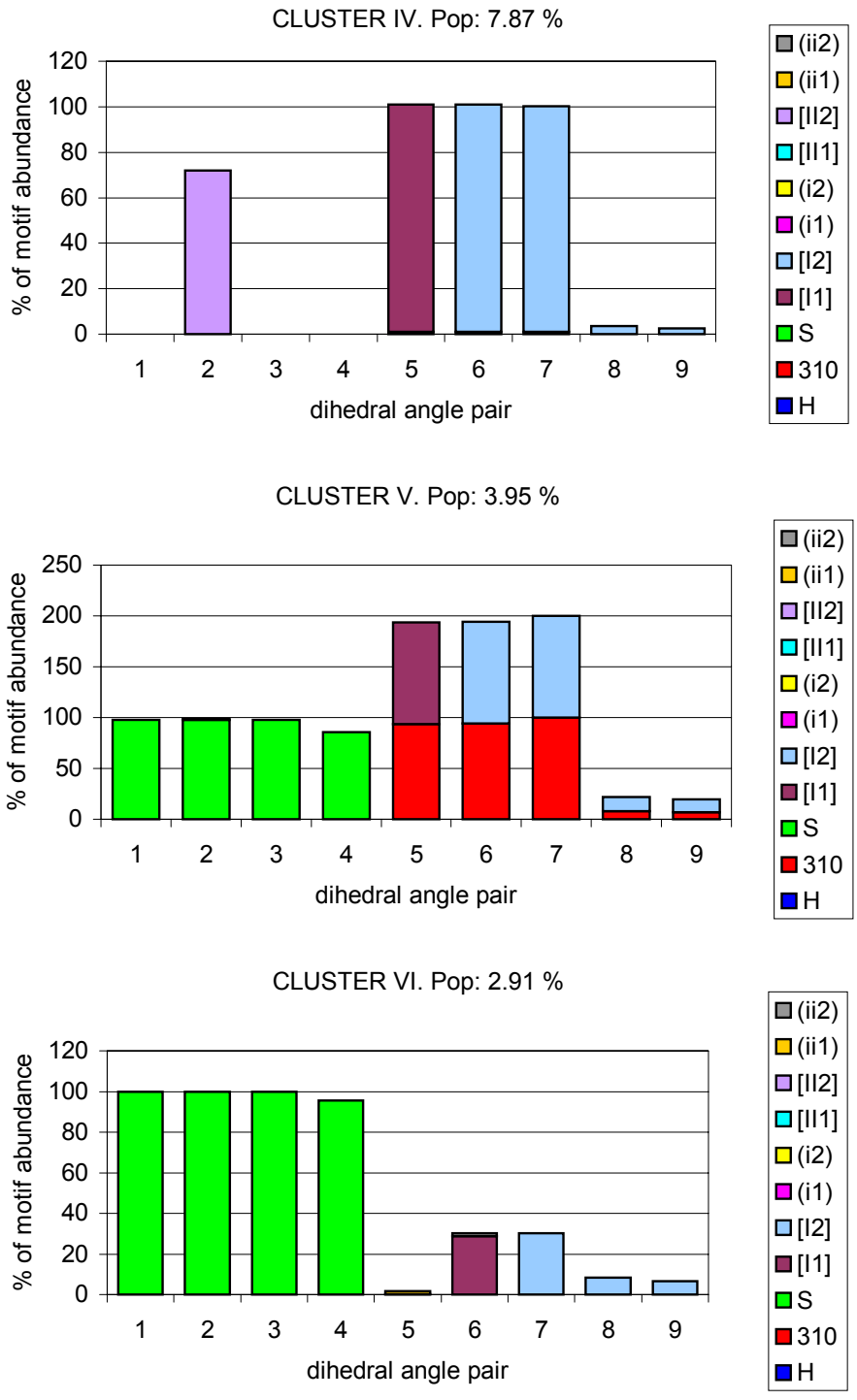

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

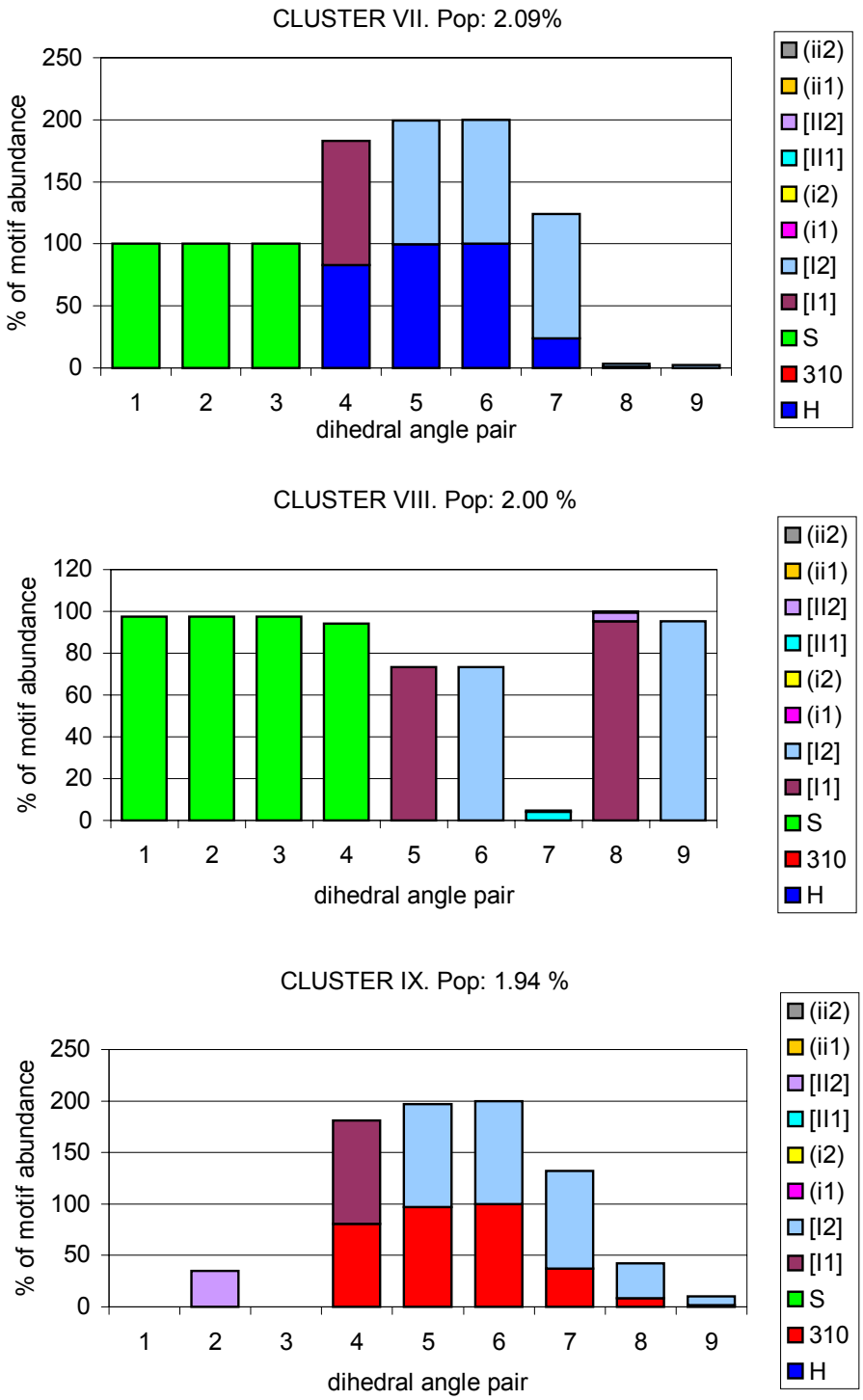

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

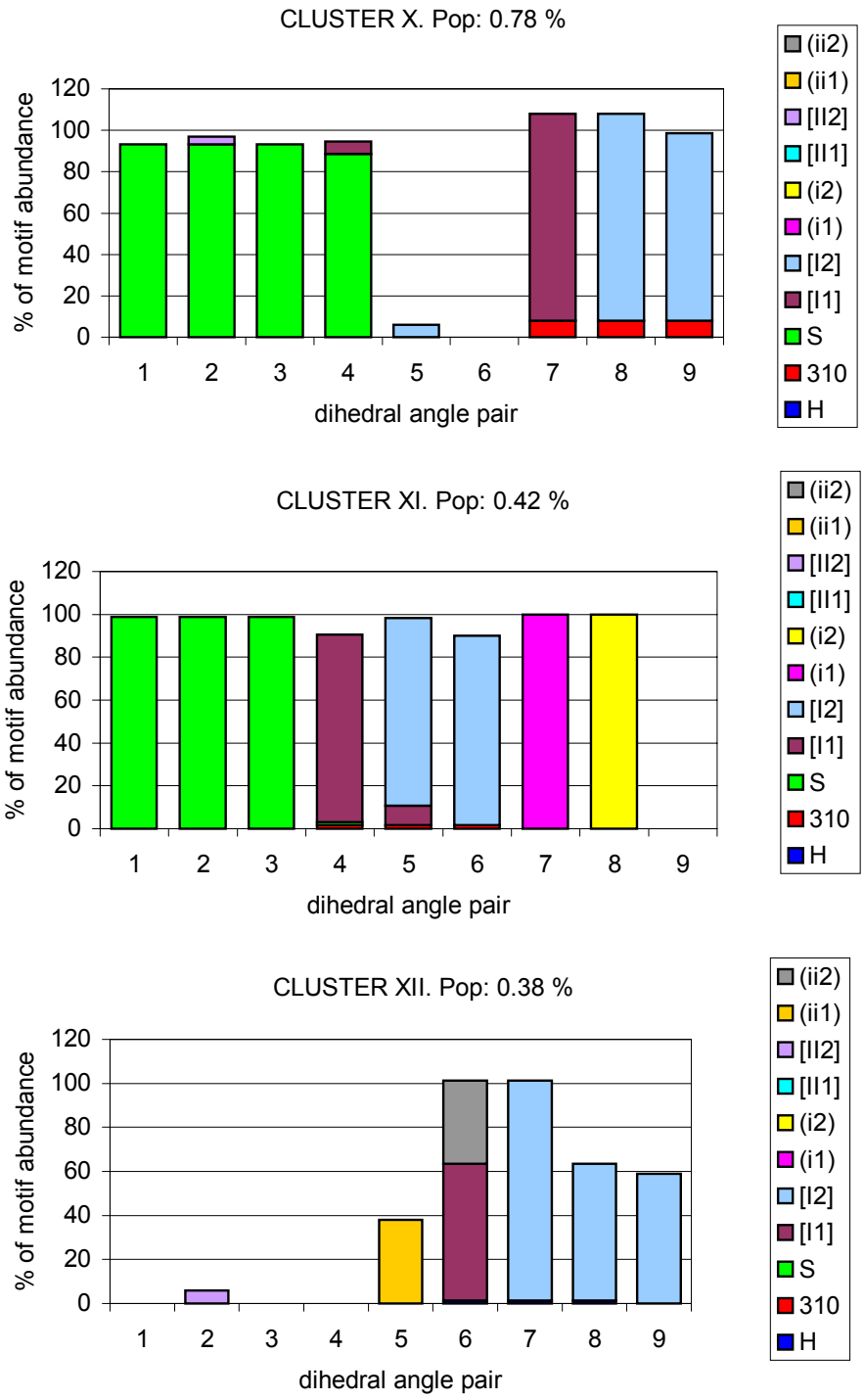

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

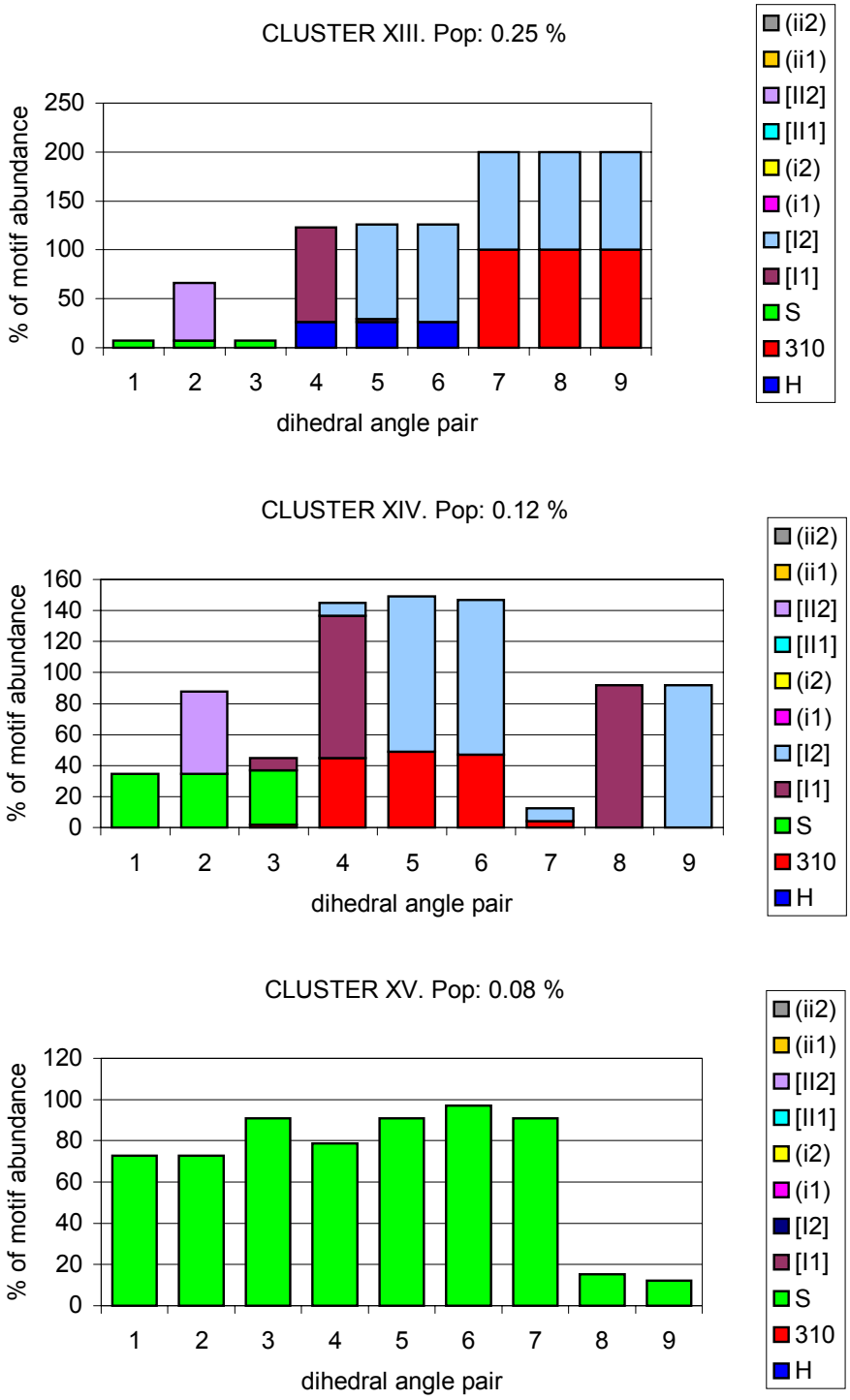

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

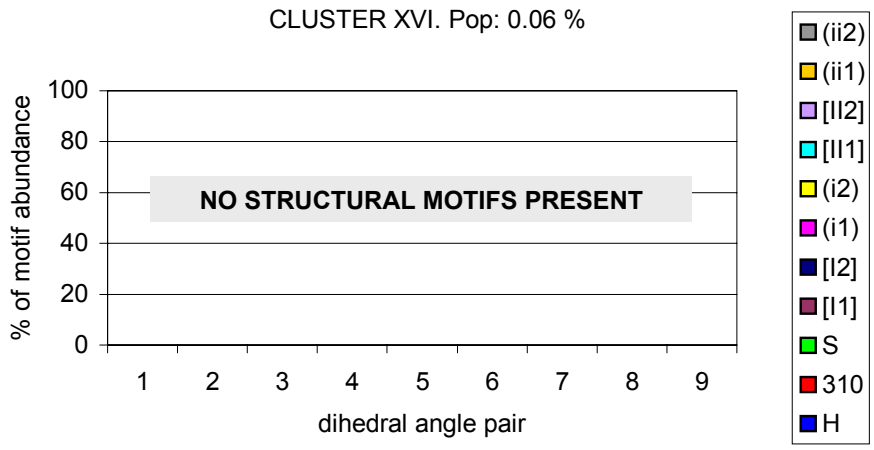

Figure 4.10. Clusters obtained for the MD trajectory at 300 K.

Table 4.7. Summary of the local transitions with probability larger than 5% obtained for the MD trajectory at 300 K.

CLUSTERS		TRANSITIONS			REVERSIBILITY	STRUCTURES	
Origin	End	Total	% Local	% Global		Total	% Local
I	V	999	7.02	4.96	1.01	14222	35.46
	I	12198	85.77	30.42	1		
II	III	658	7.52	3.28	1	8746	21.81
	II	7326	83.76	18.27	1		
	VII	511	5.84	2.56	0.99		
III	III	6516	81.77	16.25	1	7969	19.87
	II	656	8.23	3.28	1		
	IX	519	6.51	2.62	0.97		
IV	IV	2912	92.27	7.26	1	3156	7.87
V	V	532	33.61	1.33	1	1583	3.95
	I	991	62.6	4.96	0.99		
VI	I	371	31.82	1.85	1	1166	2.91
	VI	683	58.58	1.7	1		
VII	III	135	16.11	0.69	0.95	838	2.09
	II	514	61.34	2.56	1.01		

	VII	179	21.36	0.45	1		
VIII	VIII	396	49.44	0.99	1	801	2
	I	325	40.57	1.61	1.01		
IX	III	533	68.51	2.62	1.03	778	1.94
	II	71	9.13	0.36	0.95		
	IX	157	20.18	0.39	1		
X	VIII	30	9.55	0.16	0.86	314	0.78
	I	104	33.12	0.51	1.02		
	X	158	50.32	0.39	1		
XI	II	21	12.43	0.1	1	169	0.42
	XI	143	84.62	0.36	1		
XII	XII	108	70.59	0.27	1	153	0.38
	IV	12	7.84	0.06	1		
	III	11	7.19	0.06	0.92		
	VI	16	10.46	0.07	1.14		
XIII	III	38	38	0.19	1	100	0.25
	II	14	14	0.07	0.88		
	IX	10	10	0.04	1.25		
	XIII	34	34	0.08	1		
XIV	III	17	34.69	0.09	0.89	49	0.12
	II	14	28.57	0.07	1		
	IX	4	8.16	0.02	1		
	XIV	10	20.41	0.02	1		
XV	VI	6	18.75	0.03	1.2	32	0.08
	XV	21	65.63	0.05	1		
	XVI	6	18.75	0.03	1		
XVI	IV	2	8.7	0.01	0.67	23	0.06
	III	3	13.04	0.01	1		
	VI	9	39.13	0.04	1.13		
	XV	6	26.09	0.03	1		

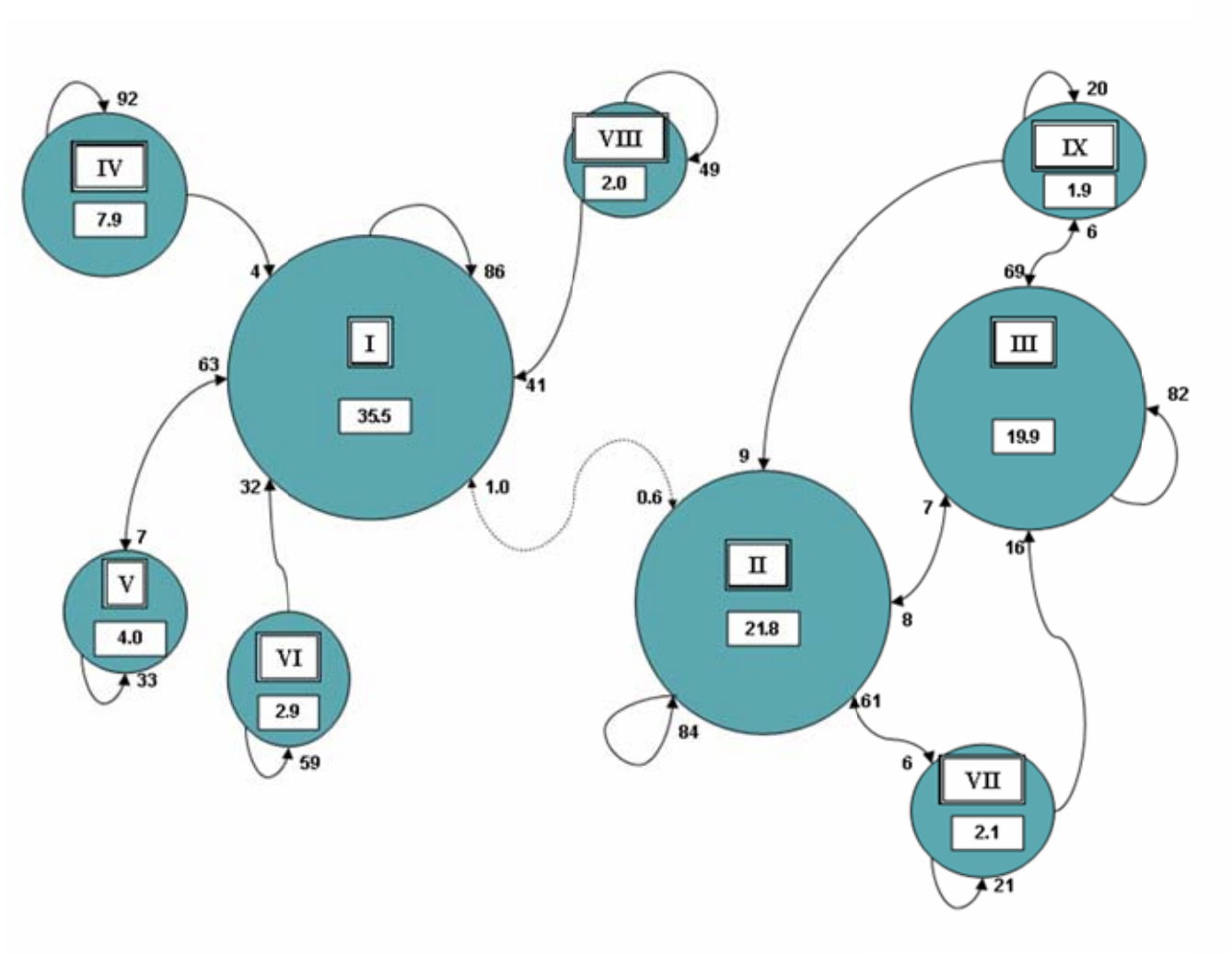

Figure 4.11. Transitions between the 9 largest clusters for the MD trajectory at 300 K.

4.4.4. Molecular Dynamics in a water box at 400 K

A 40 ns MD simulation of SPOH at 400 K was carried out starting from an extended conformation of the peptide soaked in a box of water molecules box using periodic boundary conditions and Particle Mesh Ewald summations as described in the methods section. The total energy and density kept fairly constant at -6771 kcal/mol and 0.870 g/cm^3.

The CLASICO algorithm was used to characterize the conformational motifs attained by the peptide during the MD simulation. All the motifs found are summarized in Figure 4.5 and Table 4.3. Conformations obtained were classified into 474 different patterns. The profile regarding new pattern sampling evolution is plotted in Figure 4.6. As can be seen, the number of patterns increases quickly and when around 200 patterns are sampled, after 4500 ps the sampling of new patterns slows down. In this case, a plateau is not reached, suggesting that at 400 K the exploration of new regions of the conformational space is a more continuous process than at 300 K.

As in the case of the 300K MD trajectory, starting from an extended conformation, the first patterns sampled correspond to structures that exhibit no secondary structure or exhibit some residues with dihedral angles corresponding to β-strands. Quickly, from picosecond 12 to 32 a type I β-turn appears between residues 9 and 10, unfolding subsequently and the turn does not appear until the 363 ps. At this point a central type I β-turn (residues 6 and 7) shows up becoming two consecutive turns at the 385 ps that extends from residues 4 to 7 and finally from residues 4 to 9 at picosecond 483. This behavior indicates that the peptide reaches a folded state in less than 500 ps at 400 K.

In residue 3 an II2 motif appears at picoseceond 9, although it disappears and appears along the trajectory. Until picosecond 6253 the motif does not appear together with a central turn. The central region exhibits alternatively α- or 3_{10}-helical turns extending from residues 3 to 5 and sporadically unfolds, the two most repeated patterns being a two consecutive (18.5%) or three consecutive type I β-turns (12.4%) extending from residues 4 to 8 and 4 to 10.

In summary, as found in the MD simulation at 300 K, the peptide fluctuates between α- and 3_{10}-helical turns for the central region (4 to 9 residues) being the folded state an ensemble of states containing two to three consecutive type I turns.

The structures sampled during the MD simulation at 400 K were classified using the CLUSTERIT algorithm described in the methods section. The convergence criterion for the

clustering process was set at $\Delta\overline{R} \geq 0.01$ (Figure 4.7). After following the clustering procedure only 14 clusters remained (Figure 4.12). The four most abundant clusters, accounting for 82% of the total number of structures are shown in Figure 4.12 (clusters I to IV). Accordingly, the majority of structures are grouped in a reduced number of clusters. Table 4.8 summarizes the main structural features exhibited by the peptide from the analysis of the MD trajectory at 400 K. The most abundant cluster, cluster I (containing 45.4% of the total number of structures) contains structures exhibiting consecutive type I β-turns running from residues 4 to 10, in different proportions (32-100%). Cluster II is the second most populated cluster (15.3%). This group is characterized by α-helical turns expanding from residues 4 to 8 (30-98) and type I β-turns running from residues 4 to 10 (37-100%). Cluster III (13.3%) groups structures showing 3_{10}-helical turns (16-79%) and type I β-turns (42-100%) expanding from residues 4 to 10. Cluster IV (7.5%) shows two type I β-turns, one located on residues 4 and 5 (96%) and another one, that in fact corresponds to two consecutive turns running from residues 7 to 10 (41-100%). The rest of clusters accounting for 18% of the total number of structures, show less frequent combinations of structural motifs. Thus, cluster V (6.4%) exhibits a β-strand going from residues 2 to 4 and for the C-terminal 3_{10}-helical turns (11-25%) and type I β-turns (37-100%) for residues 5 to 9 and 5 to 10, respectively. Cluster VIII (1.0%) and XIV (0.1%) exhibit a combination of α-helical turns and 3_{10}-helical turns for the region expanding residues 4 to 10. These are rare events where the peptide goes through when fluctuating between the α and 3_{10} conformations. Other non-abundant structural motifs found are type II β-turns on residues 8 and 9 and type II' β-turns on residues 7 and 8 and 9 and 10. Finally, two remarkable clusters are cluster IX (0.7%), a cluster that contains structures with no conformational motifs and cluster XII, that only contains a β-strand running from residues 2 to 10 (12-80%).

Transitions between clusters in the MD trajectory at 400 K were analyzed as described in the methods section and the data is presented in Table 4.9 and Figure 4.13. Only 8 clusters representing 98% of the structures and transitions with local transitions above 5% have been depicted for conciseness. From the transitions plot (Figure 4.13) it can be seen that the degree of interconnection lies in between the results obtained for the SA procedure and the MD at 300 K. Furthermore, the connections suggest the existence of two different areas that are formed by clusters I, V and VII and clusters I, II, III, IV, VI and VIII. Both subgroups have in common cluster I that facilitates transitions from one branch to the other.

Table 4.8. Summary of the main structural features of the 14 clusters obtained using CLUSTERIT for the MD trajectory at 400 K. The second column shows the percentage of the total number of structures obtained in the MD trajectory at 400 K that are contained within each cluster. The last column corresponds to the percentage of structures within each cluster that exhibit each motif. When the motif expands more than one residue with different percentages the interval is shown.

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
I	45.4	type I β-turn	4-10	32-100
II	15.3	α-helical turn type I β-turn	4-8 4-10	30-98 37-100
III	13.3	3_{10}-helical turn type I β-turn	4-10 4-10	16-79 42-100
IV	7.5	type I β-turn type I β-turn	4-5 7-10	96 41-100
V	6.4	β-strand 3_{10}-helical turn type I β-turn	2-4 5-9 5-10	93 11-25 37-100
VI	4.7	3_{10}-helical turn 3_{10}-helical turn type I β-turn type I β-turn	4-6 8-10 4-6 8-10	19 12 94-96 81-100
VII	3.9	β-strand α-helical turn type I β-turn	2-4 5-8 5-10	13 12-29 38-98
VIII	1.0	α-helical turn 3_{10}-helical turn type I β-turn	4-7 7-10 4-10	63-99 36-99 93-100
IX	0.7	none	--	--
X	0.6	β-strand 3_{10}-helical turn type I β-turn	2-4 6-10 6-10	78 15-38 61-97

CLUSTER	% OF STRUCTURES CONTAINED	MOTIFS PRESENT	RESIDUES INVOLVED	% OF STRUCTURES CONTAINING THE MOTIF
XI	0.5	3_{10}-helical turn type I β-turn type II β-turn type II' β-turn	4-6 4-7 8-9 7-8	8-10 67-100 83 15
XII	0.3	β-strand	2-10	12-80
XIII	0.3	β-strand 3_{10}-helical turn type I β-turn type II β-turn type II' β-turn	2-4 5-6 4-8 8-9 9-10	17 10-11 59-68 21 79
XIV	0.1	α-helical turn 3_{10}-helical turn type I β-turn	7-10 4-6 3-10	30-67 48-52 22-96

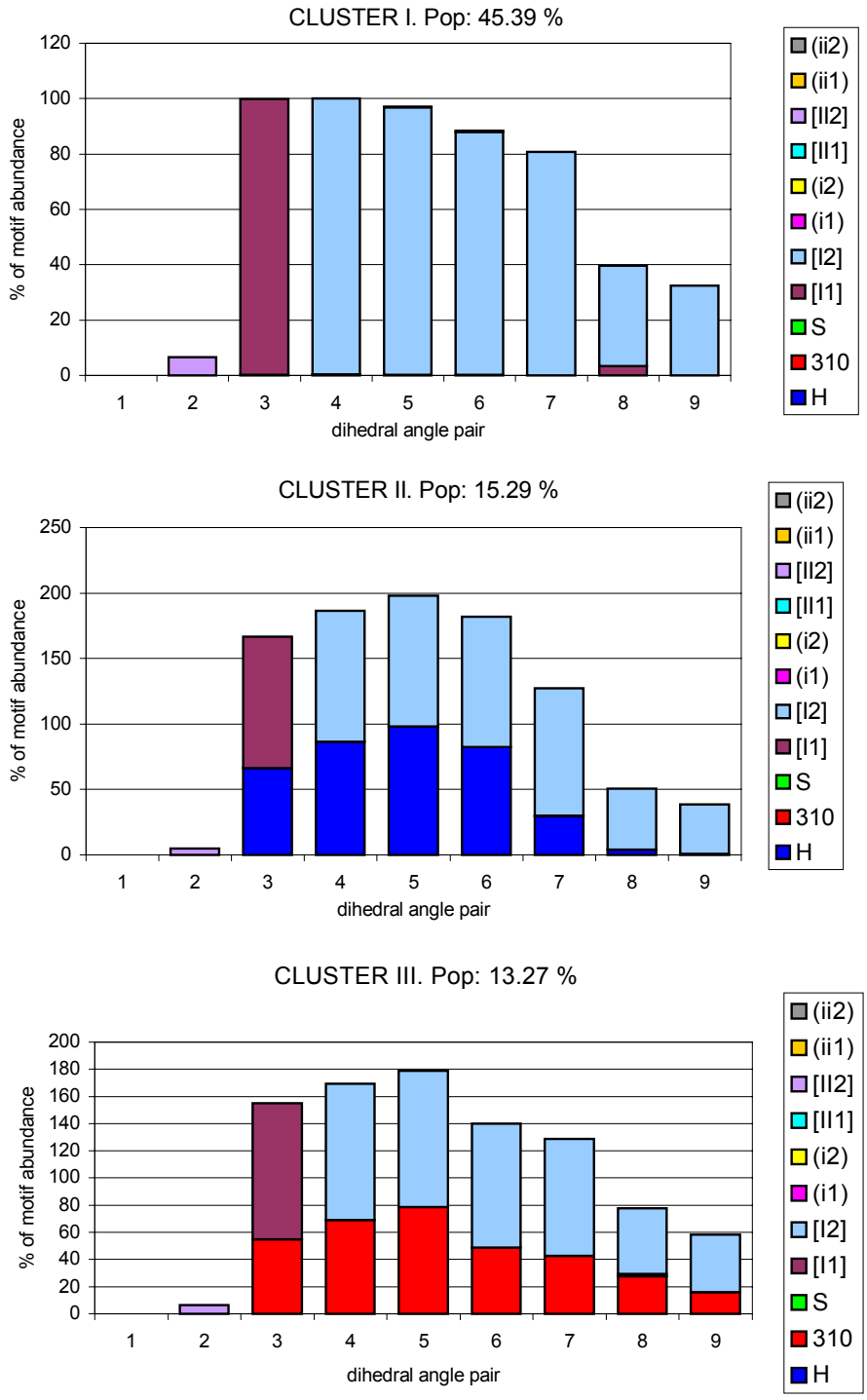

Figure 4.12. Clusters obtained for the MD trajectory at 400 K.

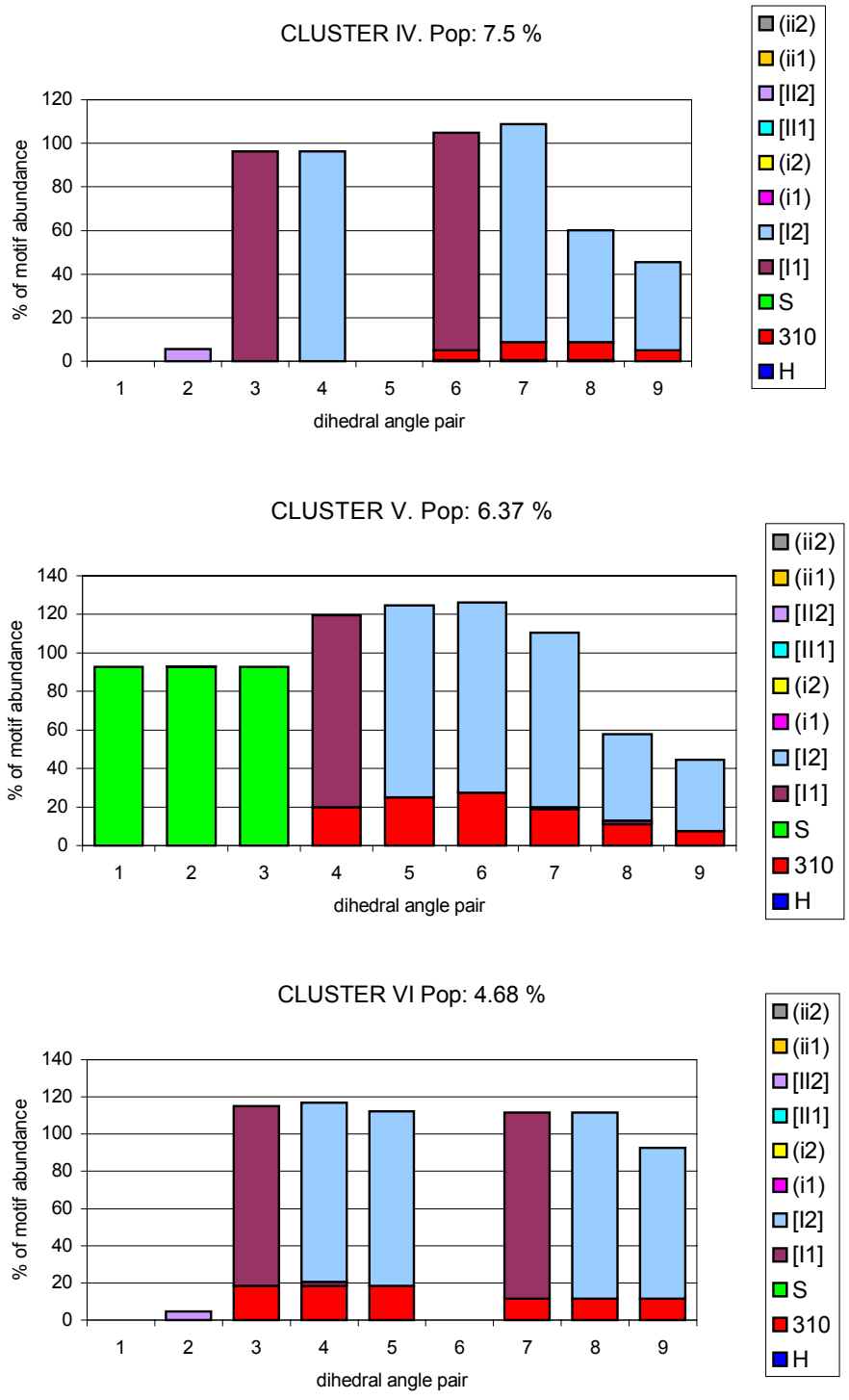

Figure 4.12. Clusters obtained for the MD trajectory at 400 K.

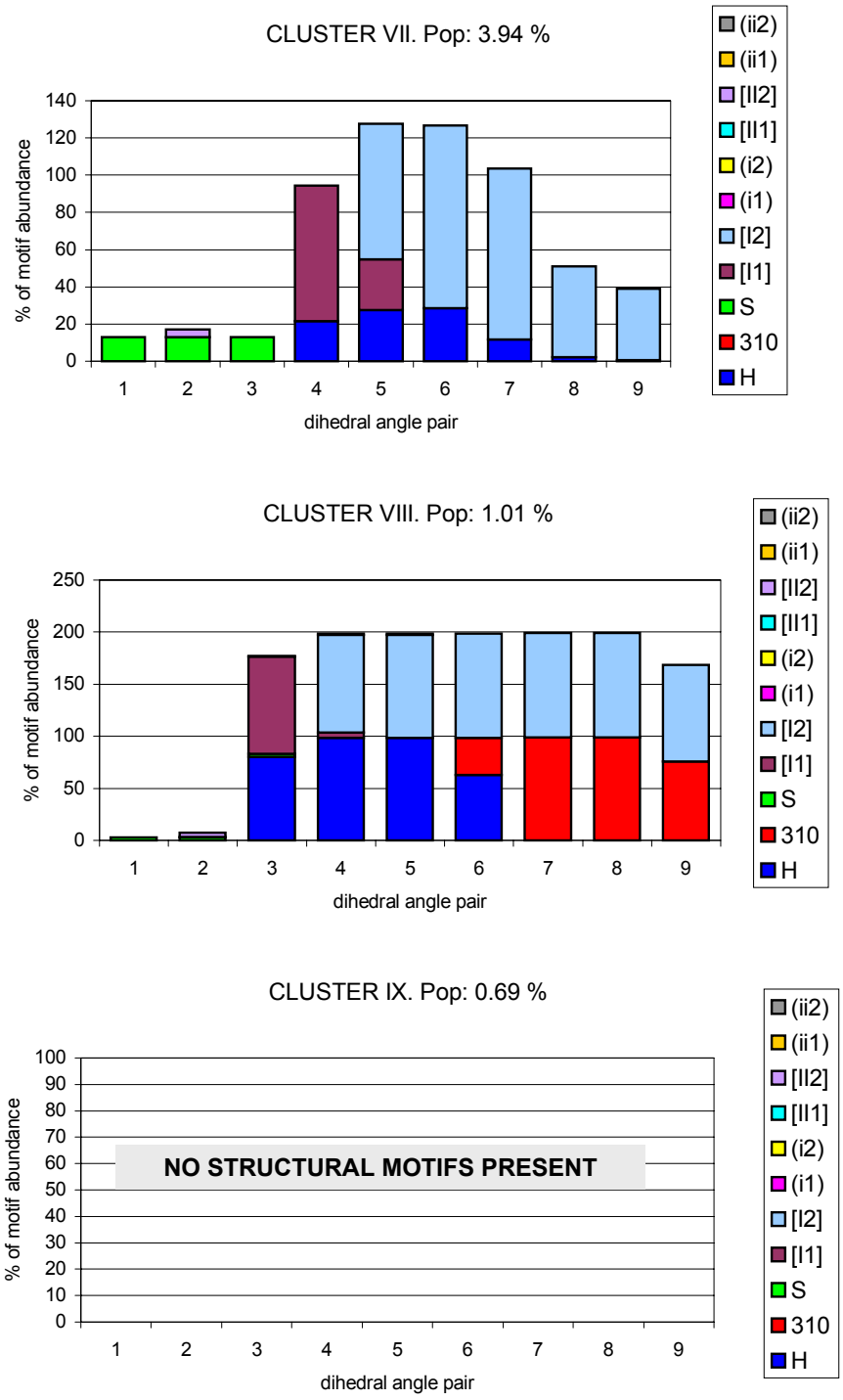

Figure 4.12. Clusters obtained for the MD trajectory at 400 K.

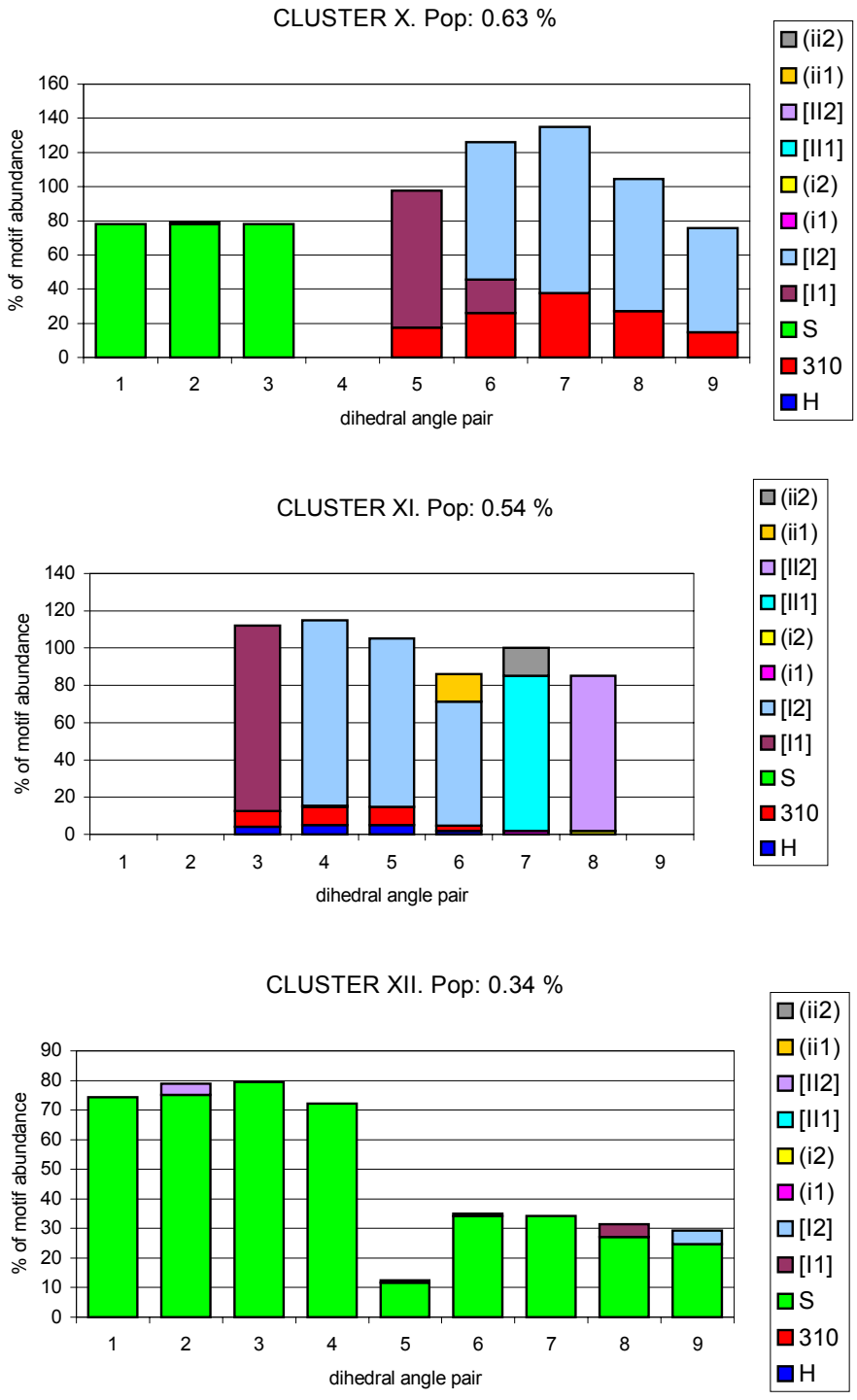

Figure 4.12. Clusters obtained for the MD trajectory at 400 K.

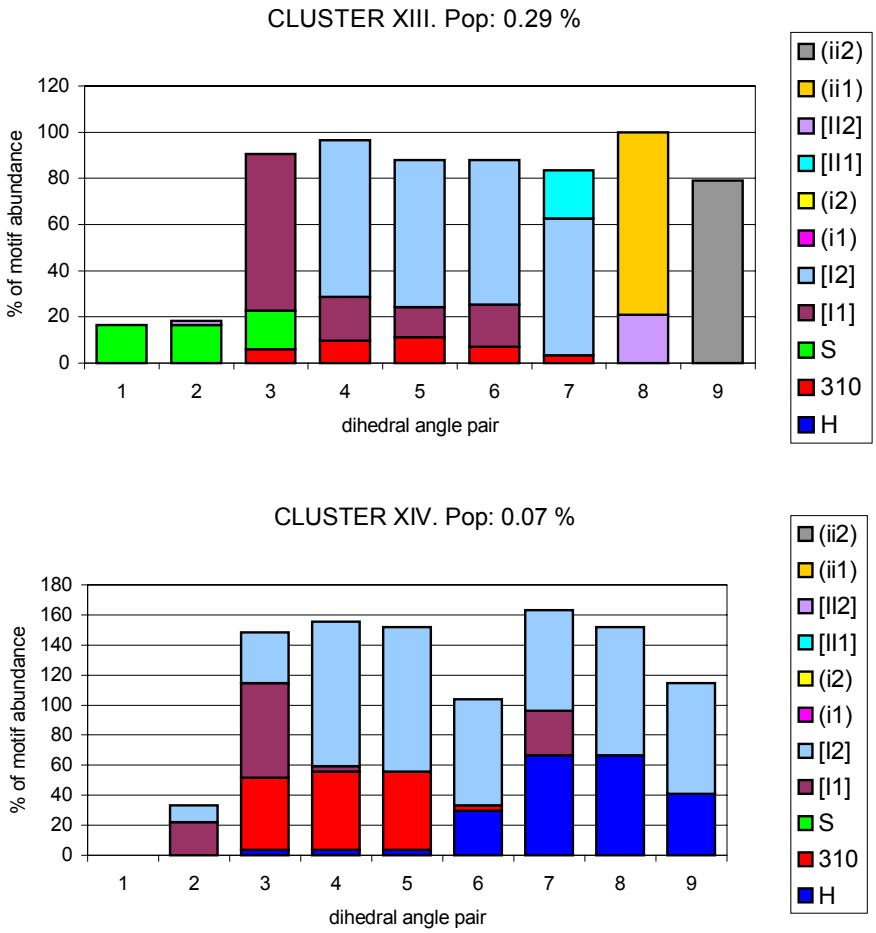

Figure 4.12. Clusters obtained for the MD trajectory at 400 K.

Table 4.9. Summary of the local transitions with probability larger than 5% obtained for the MD trajectory at 400 K.

CLUSTERS		TRANSITIONS			REVERSIBILITY	STRUCTURES	
Origin	End	Total	% Local	% Global		Total	% Local
I	IV	1163	6.39	5.81	1	18203	45.39
	III	2462	13.53	12.29	1		
	II	2757	15.15	13.67	1.01		
	I	10521	57.8	26.24	1		
II	IV	419	6.83	2.1	0.99	6132	15.29
	III	470	7.66	2.26	1.07		
	II	2050	33.43	5.11	1		
	I	2723	44.41	13.67	0.99		
III	III	1815	34.11	4.53	1	5321	13.27

	II	438	8.23	2.26	0.93		
	I	2468	46.38	12.29	1		
IV	IV	1068	35.51	2.66	1	3008	7.5
	III	167	5.55	0.84	0.99		
	II	425	14.13	2.1	1.01		
	I	1166	38.76	5.81	1		
V	VII	270	10.58	1.37	0.96	2553	6.37
	V	2055	80.49	5.12	1		
VI	VI	1010	53.87	2.52	1	1875	4.68
	III	161	8.59	0.8	1.01		
	II	125	6.67	0.63	0.97		
	I	481	25.65	2.37	1.02		
VII	VII	305	19.32	0.76	1	1579	3.94
	III	107	6.78	0.57	0.88		
	V	280	17.73	1.37	1.04		
	II	210	13.3	1.04	1		
	I	522	33.06	2.58	1.02		
VIII	IV	44	10.84	0.21	1.13	406	1.01
	III	68	16.75	0.36	0.91		
	II	95	23.4	0.49	0.95		
	VIII	36	8.87	0.09	1		
	I	112	27.59	0.58	0.93		
IX	XII	22	8	0.11	0.92	275	0.69
	I	38	13.82	0.18	1.12		
	IX	194	70.55	0.48	1		
X	VII	38	14.96	0.18	1.06	254	0.63
	V	99	38.98	0.5	0.97		
	X	75	29.53	0.19	1		
	I	22	8.66	0.08	1.83		
XI	XI	128	59.53	0.32	1	215	0.54
	I	68	31.63	0.33	1.03		
XII	XII	106	77.94	0.26	1	136	0.34
	IX	24	17.65	0.11	1.09		
XIII	XI	6	5.22	0.02	1.5	115	0.29
	IV	8	6.96	0.04	0.8		
	XIII	40	34.78	0.1	1		
	III	8	6.96	0.03	1.6		
	V	7	6.09	0.04	0.88		
	II	7	6.09	0.03	1.75		
	I	35	30.43	0.18	0.9		
XIV	VI	7	25.93	0.03	1.17	27	0.07
	III	6	22.22	0.03	0.86		
	I	11	40.74	0.04	1.57		
	XIV	2	7.41	0	1		

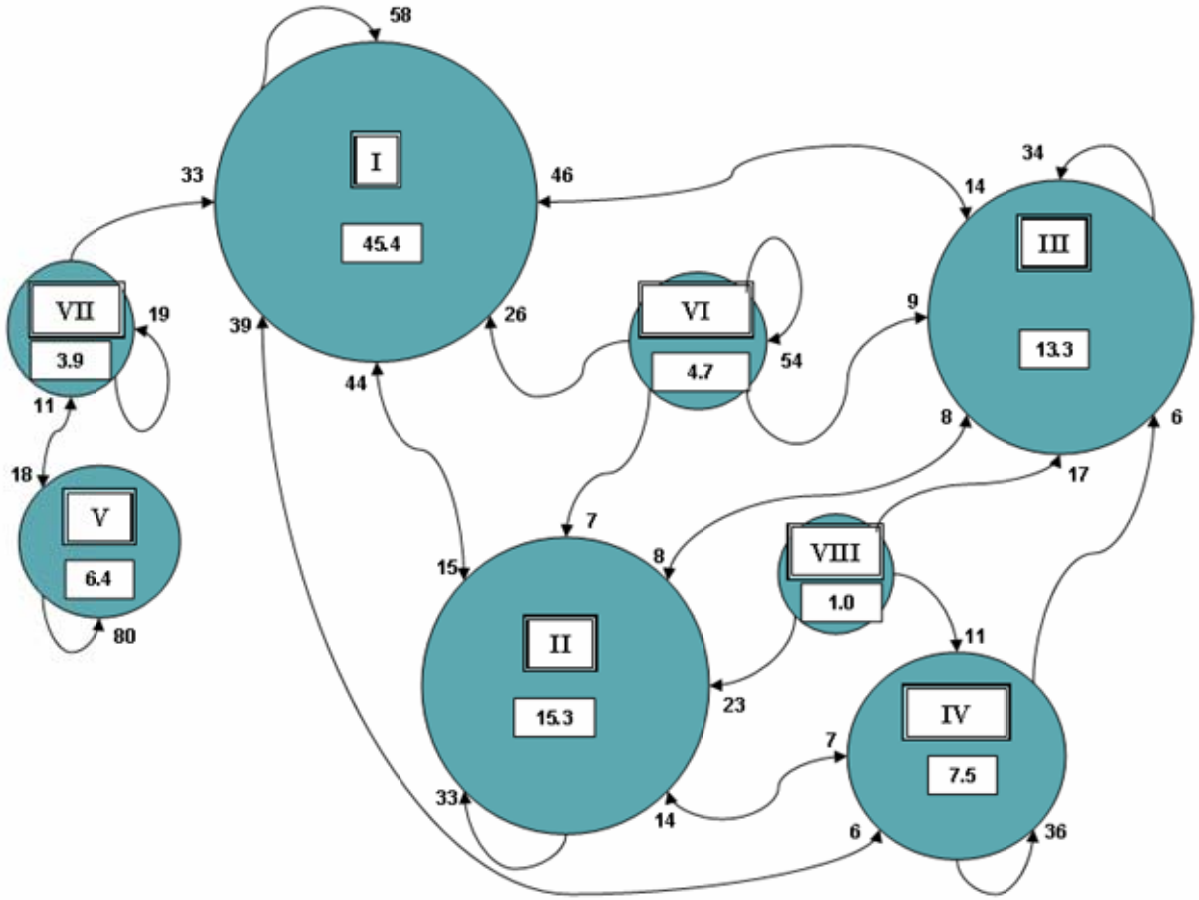

Figure 4.13. Transitions between the 8 largest clusters for the MD trajectory at 400 K.

4.5. Discussion

The MD simulation at 400 K yielded a larger number of patterns (474) than the SA procedure (435), and larger than the MD simulation at 300 K (231), suggesting that MD simulation at 300 K clearly samples a smaller portion of the conformational space. However, in qualitative terms the conformations sampled using the SA procedure and MD simulations are not identical. As can be seen in Table 4.3 and Figure 4.5, using the SA procedure each residue of the peptide exhibits a larger variety of motifs. In contrast, structures from the MD simulations show a fewer number of conformational motifs, although much more conserved (higher percentage of presence). Helical conformations represent a higher order of hierarchy in the structure, since they require having at least three consecutive residues with a turn like conformation. Although found in the SA procedure, they are much more abundant in the MD simulations. Comparison of patterns found with the SA procedure and MD simulations, was carried out using only information regarding turns and the β-strand motif, in order to include a broad similarity criterion between structures. Comparison yields that about 50% of the structures characterized using the SA procedure are also found in the MD simulations. Moreover, these structures represent a larger portion of the conformational space sampled by the MD simulation at 300 K (87% of the snapshots) than at 400 K (49% of the snapshots). This suggests that broadly speaking, all the structures sampled in the MD simulation at 300 K are included in the set of structures characterized in the SA procedure. The rest of structures sampled during the MD trajectory are probably regions that do not correspond to valleys of the conformational space and consequently, not sampled in the SA. This is also supported by the lower percentage of structures found in the MD simulation at 400 K. Indeed, at this temperature the system samples more often regions of the conformational space that are not valleys. Indeed, when the two MD simulations are compared only about 50% of the snapshots of the MD simulation at 400 K are recognized in the simulation at 300 K.

A more detailed comparison of the sampling performance between the SA procedure and the MD simulations, and between the two MD simulations at different temperatures can be done from the analysis of the dihedral angle distributions obtained in the sampling process. Let us first compare the distributions obtained from the SA procedure and those of the MD simulation at 300 K. Table 4.10 lists values of the maxima of the different dihedral angle distributions as well as their maximum probability value. The dihedral angle distributions have been plotted and are shown in the Annex section of the present work. In general terms, angle distributions obtained from the SA procedure are richer in the number of maxima. Moreover, angle distributions obtained from the MD simulations are contained within the regions covered in the SA procedure. However there are some differential features. Some of the maximum values of the angle distributions like for ψ_2, ϕ_3, ψ_5, ϕ_6, ψ_6, ϕ_7, ψ_7 and ϕ_{11} are shifted around 20°. In other cases, like for angles ψ_8, ψ_9, and ψ_{10} two

peaks around -60º and +60º exhibited in the SA distributions appear to be replaced by only one peak at 0º in distributions of the MD simulation. This is the reason that structures from the MD simulations exhibit a 3_{10} helical structure that cannot be attained in the SA procedure. Finally, omega dihedral angles only present one peaks at 180º, except for the angle ω_1, that was found to exhibit a *cis* conformation only in the SA sampling procedure, induced by the subsequent proline in the peptide chain.

The characteristic features of the distributions obtained from MD simulations at 300 K and 400 K are listed in Table 4.10. The dihedral angle distributions have been plotted and are shown in the Annex section of the present work. In most of the cases the distributions of both MD simulations are the same. The more significant differences correspond to ψ_4 and ψ_5 dihedral angles. For ψ_4 the maximum of the MD simulation at 400 K appears at -30º, whereas at 300 K appears at 170º. Interestingly, -30° is also a maximum in the SA distribution, suggesting that this is a low energy region, attainable by the residue. This value, together with ϕ_4 = –50°, the most probable value found in all the different sampling methods, corresponds to the $\alpha/3_{10}$ helix region. This provides an explanation to the differential nature of the helical structure sampled in the two MD simulations, shown in Table 4.3 and Figure 4.5. Thus, the region with propensity to form a helical structure at 400 K expands from residues 4 to 10, one residue longer than at 300 K. Dihedral angle ψ_5 at 400 K shows a single maximum around -20º, matching well with the main peak of the distribution at 300 K found around –30°. The lower peaks of the 300 K distribution, located around 90º and 160º are not present in the 400 K distribution. This is in accordance to Table 4.3, where at 300 K residue 5 exhibits similar chances of forming a helix or a strand, whereas at 400 K residue is involved in a helical structure. This differential behavior can be attributed to the secondary structure adopted by the previous residue in the peptide chain. Whereas residue 4 exhibits a high tendency to form a helical structure at 400 K, this tendency is small at 300 K as shown in the different ψ_4 distributions. Finally, angle ϕ_9 exhibits similar distributions in both MD simulations, although the maxima are interchanged, showing a higher propensity to form helical structures at 400 K.

In order to compare the results of the present study with the NMR structures in trifluoroethanol/water (Lee et al., 1999) and in SDS (Keire et al., 1986), patterns corresponding to these structures were calculated and compared with the average structure and the three more abundant patterns obtained in the MD trajectories at both temperatures (Table 4.11).

The NMR structure obtained in SDS presents a type II β-turn involving residues 4 and 5 and a type I β-turn involving residues 9 and 10. This pattern, although is not one of the most abundant patterns, it was identified in the SA procedure four times. However, the pattern was not sampled during the MD simulations.

Table 4.10. Maxima values for the probability distribution function for the dihedral angles of the SA procedure and the MD trajectories at 300 and 400 K.

ANGLE	SA		MD AT 300 K		MD AT 400 K	
	ANGLE VALUE	PROBAB.	ANGLE VALUE	PROBAB.	ANGLE VALUE	PROBAB.
ψ_1	90, 150	17.0, 23.6	130, 140	21.42, 20.92	130, 140	19.56, 19.01
ω_1	180, 0	47.7, 18.3	180	37.65	180	32.66
ϕ_2	-70	41.1	-60	33.11	-60	30.23
ψ_2	-40, 150	19.3, 6.0	-20, 170	1.92, 27.16	-20, 170	4.59, 18.62
ω_2	180	90.7	180	35.80	180	33.04
ϕ_3	-70, 60	17.2, 10.0	-60, 40, 50	21.83, 3.47, 3.63	-130, -120, -60,40, 50	7.63, 7.70, 15.90, 2.14, 2.11
ψ_3	-60, 80,150	2.0, 12.2, 19.8	70, 160	4.60, 21.02	80, 160	5.02, 21.57
ω_3	180	55.2	180	38.96	-170	34.08
ϕ_4	-70, 50	32.1	-70, -60	31.31,32.65	-50	30.32
ψ_4	-40, 70,80, 150	17.4, 4.7, 4.7, 4.7	90, 100, 170	5.82, 5.97, 24.33	-30, 170	20.99, 2.07
ω_4	180	89.5	180	32.12	180	41.65
ϕ_5	-160, -60, 60	6.4, 19.0, 4.3	-130, -60	3.32, 29.19	-60	29.41
ψ_5	-50, 70,160	13.3, 4.8,4.7	-30, -20, 90, 160	13.49, 13.08, 3.88, 6.63	-20	24.74
ω_5	180	87.6	180	35.56	180	43.00
ϕ_6	-160, -70, 60, 70	5.8, 17.8, 4.5, 4.5	-60, -50	31.77, 30.69	-120, -60	4.06, 21.52
ψ_6	-60, -50, 70, 80	13.0, 13.4, 5.2, 5.4	-20	26.65	-30	21.17
ω_6	180	86.5	180	46.61	180	37.21
ϕ_7	-160, -70, 60	7.1, 17.1, 4.4	-120, -60	2.03, 26.93	-120, -60	3.31, 23.92
ψ_7	-50, 70, 150	13.5, 6.8, 5.8	-30	20.41	-40, -30	19.23, 19.29
ω_7	180	88.7	180	41.54	180	38.88
ϕ_8	-150, -140, -70, 70, 150	4.5, 4.7, 11.7, 7.9, 5.0	-120, -70	7.48, 18.40	-120, -60	3.99, 22.39
ψ_8	-150, -50, 60, 150	3.3, 9.1, 7.2, 3.5	0	24.10	-10	20.68
ω_8	180	89.9	-180, -170	36.34, 34.63	180	35.86
ϕ_9	-80, 80, 170	8.7, 12.0, 6.7	-180, -70, 80	2.94, 2.98, 15.87	-70, 80	11.14, 7.30
ψ_9	-70, -60, 70	15.0, 15.1, 4.9	10, 20	17.98, 17.82	0	14.92

ANGLE	SA		MD AT 300 K		MD AT 400 K	
	ANGLE VALUE	PROBAB.	ANGLE VALUE	PROBAB.	ANGLE VALUE	PROBAB.
ω_9	180	89.0	-170	38.50	-170	32.67
ϕ_{10}	-160, -70,60	7.1, 20.6, 4.7	-120, -70	5.69, 19.52	-120, -70	7.56, 16.36
ψ_{10}	-60, -50, 70	16.6, 16.1, 6.9	-10, 0, 90, 170	16.61, 15.93, 2.30, 2.40	-10, 0	15.27, 14.96
ω_{10}	180	95.2	-180, -170	35.10, 33.85	-180, -170	30.72, 30.86
ϕ_{11}	-160, -70, 60	11.0, 19.6, 6.1	-130, -60	11.27, 13.75	-130, -60	7.89, 14.04

Table 4.11. Summary of secondary structure motifs present in the structures obtained by NMR experiments and the average structures from the MD trajectories at 300 and 400 K. The three most abundant patterns for the two MD trajectories are shown. Percentages from the total structures of each trajectory are shown in brackets for the most abundant patterns.

RESIDUE NUMBER		2	3	4	5	6	7	8	9	10
NMR SP in TFE/water					I1	I2	H/I2	H/I2	H/I2	
NMR SP in SDS				II1	II2				I1	I2
SA most abundant patterns	1st (32.06%)									
	2nd (6.18%)				I1	I2				
	3rd (4.34%)			II2						
MD at 300 K average structure		S	S	S	I1	I2	I2	I2		
MD at 300 K most abundant patterns	1st (22.78%)	S	S	S	S	I1	I2	I2		
	2nd (14.72%)	S	S	S	I1	I2	I2	I2		
	3rd (12.08%)				I1	I2	I2	I2		
MD at 400 K average structure				I1	I2	I2	I2	I2	I2	
MD at 400 K most abundant patterns	1st (18.49%)			I1	I2	I2	I2	I2		
	2nd (12.40%)			I1	I2	I2	I2	I2	I2	I2
	3rd (3.68%)			I1	I2	I2				

However, in the case of the structure derived from the NMR study in TFE/water, the conformation proposed agrees well with the structural motifs sampled in the MD simulations. Main differences consist in the number of consecutive type I β-turns present, running from residues 5 to 9 in the NMR structure and from residues 5 to 8 and 4 to 9 in the average structures of the 300 K and 400 K MD simulations, respectively. On the other hand, the structure was found in the SA procedure. Moreover, the type I β-turn between residues 5 and 6 corresponds to the second most populated fraction of structures.

Comparison of the cluster analysis obtained from the SA procedure and the MD trajectories at 300 and 400 K, suggests that the SA procedure explores a greater diversity of structures. Indeed, following the criterion to perform the clustering analysis, based on the loss of information during the process the initial R value is greater for the SA procedure (see Figure 4.7). In addition, ΔR gets over 0.01 when 18 clusters are characterized in the SA procedure, whereas this number is 16 and 14, respectively, for the MD trajectories at 300 and 400 K. When we look to the clusters, this diversity is translated into a greater variety of β-turns present at different regions of the peptide. Indeed, the SA procedure shows significant number of structures exhibiting type I, II, I' and II' β-turns. However, in the MD trajectories at 300 and 400 K, only type I β-turns are most abundant secondary motif, being type I' and II' β-turns exhibited scarcely. Another remarkable difference between the SA procedure and the MD is the fact that the most abundant cluster characterized in the SA procedure (32.1%) exhibits no conformational motifs. This cluster represents only 0.1 and 0.7% for the MD trajectories at 300 and 400 K, respectively. Accordingly, the SA procedure is enriched in unfolded structures, whereas these are rare events in the MD trajectories, under the experimental conditions used in the present study. However, several similarities between the clusters obtained from the SA procedure and the MD trajectories can be identified. Thus, in all three cases clusters containing a β-strand on residues 2 to 4 or 2 to 5 can be observed. Several clusters throughout the different simulations, exhibit a region with consecutive type I β-turns from residue 4 to 10, facilitating the formation of α and 3_{10}-helical structures in this region. However, whereas in the MD trajectories, clusters exhibiting any of the helix types, or a combination of both types can be identified, in the SA procedure only clusters exhibiting a α-helix can be characterized. Another interesting result derived from the inspection of the results is that the SA procedure exhibits a partition where the structures are more even distributed among the different clusters. Indeed, in a random clustering procedure, since the structures are evenly distributed among the different clusters, 80% of clusters account for 80% of the structures. In contrast, in a real situation, clustering obeys the 80/20 rule, i.e., 20% of the clusters account for 80% of the structures. The SA procedure lies in between these two extremes, it is necessary to consider 8 clusters out of 18 (44%) in order to account for 80% of the structures. On the other hand, both MD trajectories at 300 and 400 K fulfill the 80/20 rule. Specifically, 3 out of 16 (19%) of the clusters of the MD trajectory at 300 K account for 77% of the structures and 3 out of 14 (21%) of the clusters of the MD trajectory at 400 K account for 74% of the structures. This suggests that the SA procedure explores the conformations in a quasi-random way, whereas structures sampled during the MD are the subset of highly populated structures of the ensemble.

The study of transitions between consecutive conformations provides a detailed analysis of the relations of vicinity in the conformational space among the different groups of structures. Figure 4.9 indicates that the SA procedure within the protocol used is a method that explores the

conformational space in such a way that two consecutive conformations are not necessarily structurally related. If they were related it should be expected to find transitions onto the same cluster with a higher probability larger than a random move, which is not the case according to the results of Figure 4.9. Three clusters do slightly deviate from the standard behavior, jumping onto themselves with higher probability than expected II, VI and VIII. Cluster II exhibits a type I β-turn and α-helical turn on residues 4 to 7. Cluster VI exhibits a β-strand on residues 2 to 5. Cluster VIII exhibits a α-helical and a type I β-turn on residues 6 to 8. In addition cluster VI jumps to cluster II with a lower proportion, thus suggesting that this transition seems less probable in the SA procedure used.

From the study of the most frequent transitions between the different clusters in the MD trajectory at 300 K two different subgroups of clusters can be distinguish. On the one hand, clusters I, IV, V, VI and VIII and on the other, clusters II, III, VII and IX. Between both subgroups transitions only occur from cluster I to cluster II, the two most abundant and account for only a small proportion of their local transitions (0.6 and 1% respectively).

Structures belonging to cluster I are characterized by exhibiting a β-strand between residues 2 to 5 and consecutive type I β-turns in the segment 6 to 9. Clusters IV, V and VIII share in common type I β-turns on residue 6, but not on residue 5, as is the rest of clusters presenting such conformational motif. Structures that belong to cluster VI exhibit only a type I β-turn on residues 7 and 8, and accordingly corresponds to a partially unfolded group of structures. Clusters V, VI and VIII are unstable as there are frequent jumps back to the main cluster of the subgroup, cluster I. On the other hand, cluster IV seems to be quite stable as there are numerous jumps onto itself on 92% of the cases.

Clusters II and III are the most populated clusters of this subgroup (21.8 and 19.9%, respectively). Both cluster II and III are also the most stable clusters within this subgroup as they jump onto themselves on 84 and 82 of their transitions. Moreover, cluster II jumps with the same frequency to cluster VII and III. Transitions from cluster II to cluster III require both the unfolding of the β-strand of the N-terminal and passing from a 3_{10}- to a α-helical turn, whereas transitions from cluster II to cluster VII only represent the transition from a 3_{10}-helical turn to a α-helical turn.

From the study of the most frequent transitions between the different clusters on the MD trajectory at 400 K, we can establish the existence of only one group of structures given that all clusters are well interconnected although presenting a branch. Indeed, clusters I, V and VII form a separate path being cluster VII an intermediate state and cluster V the endpoint. Cluster I exhibits a type I β-turn between residues 4 to 10. Cluster VII is characterized by a β-strand on residues 2 to 4 and a α-helical turn on residues 5 to 8. This group of conformations is unstable and the

structures jump to cluster I or in some cases to cluster V, that exhibits the same β-strand motif at region 2-4 and a 3_{10}-helical turn on residues 5 to 9. This group of conformations although being poorly accessible along the MD trajectory and having only a population equal to 4% is the most stable of the eight clusters presented having 80% of autotransitions.

Clusters V and VII differ from the rest of clusters in the presence of the β-strand motif that induces the formation of type I β-turns at residue 5 instead of residues 4 as it is the case for the rest of clusters shown. Cluster II is the same cluster as cluster I but exhibits a α-helical turn on residues 4 to 8. Cluster III differs from cluster I in that exhibits a 3_{10}-helical turn on residues 4 to 10 and cluster IV is characterized by presenting two different regions, residues 4 and 5 and 7 to 10, that exhibit type I β-turns. Cluster VI exhibits two 3_{10}-helical-regions on residues 4 to 6 and 8 to 10. Therefore, cluster I is a great basin that sporadically visits cluster II, III and IV. Cluster II, III, IV and VI are relatively stable but in many cases they come back to cluster I (44, 46, 39 and 26%, respectively). Cluster VIII exhibits a α-helical turn on residues 4 to 7 and 3_{10}-helical turn on residues 7 to 10. It could be considered as a transition state between α-helix and 3_{10} conformations. As it can be seem from Figure 4.13 it lies in the middle between cluster II and III that correspond to the groups of conformations exhibiting α-helical and the 3_{10}-helical motifs, respectively. In other cases structures from cluster VIII will jump onto cluster IV presenting two different type I β-turns on regions 4 and 5 and 7 to 10.

4.6. Conclusions to chapter 4

The present work represents an attempt to compare the performance of two different methods to explore the conformational space of a peptide. On the one hand, the iterative simulated annealing procedure, a method intended to characterize local minima of the surface and on the other, molecular dynamics simulations, which provides information of the most populated conformations in solution.

For this purpose the conformational analysis of the substance P was subject to the two different methods. Conformations were classified according to their values of the backbone dihedral angles and clustered according to their similarity following an algorithm based on information theory.

Analysis of the structures obtained using both sampling methods suggest that those derived from the SA procedure are more extended. Furthermore, the SA procedure seems to sample the different dihedrals locally providing all the possible values attainable. This makes that structures are much more diverse than in the MD simulations. On the other hand, MD simulations sample

other conformations exhibiting larger secondary structure motifs that are not sampled by the SA procedure in the way is performed in the present work. These differences can be attributed to either cooperative effects between neighboring residues in the peptide chain or to an effect of the different treatment of the solvent.

In summary, the study reveals that the SA procedure is able to efficiently sample local motifs of the peptide structure. Furthermore, these motifs are relevant to those sampled in a MD simulation at 300 K, suggesting that it provides the conformational features of the peptide. However, the SA procedure does not sample all the breadth of conformations with a high content of secondary structure that is obtained by long MD trajectories.

Present work describes a novel method of clustering of structures based on information theory that efficiently groups structures upon their structural similarity. This method has proven to be a useful tool in the analysis of the evolution of a peptide through an extended MD. The use of this tool has also allowed for the study of the transitions that take place along the MD trajectories and the SA procedure. We have shown that the consecutive structures that are obtained from the SA procedure are not structurally related and therefore the method is not conditioned by the previous structures obtained and constitutes an efficient tool for obtaining a diverse set of structures. From the analysis of transitions for the MD trajectory at 300 K it has been shown that the structures are highly stable, they are kept in the same region of the conformational space. When the temperature is changed to 400 K the equilibrium between conformers become more fluid, thus obtaining a greater ratio of transitions between different cluster and a lower ratio of jumps onto the same cluster, thus suggesting that the lifetime of the clusters has decreased.

4.7 References to chapter 4

Case, D.A., Pearlman, D.A., Cadwell, J.W., Cheatham III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1997), *AMBER 5*, University of California, San Francisco.

Chassaing G., Convert O. and Lavielle S. *Eur. J. Biochem.*, 154, 77-86, (1986).

Corcho, F.J. Flilizola, M. and Perez, J.J. *J. Biomol. Struct. Dyn.*, 16, 1043-1052, (1999).

Corcho, F.J., Filizola, M. and Perez, J.J. *Chem. Phys. Letters,* 319, 65-70, (2000).

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman P.A. *J. Am. Chem. Soc.*, 117, 5179-5197, (1995).

Coutinho, E., Kamath, S., Saran, A. and Srivastava, S. *J. Biomol. Struct. Dyn.*, 16(3) 747-755, (1998).

Daura, X., Jaun, B., Seebach, D., van Gusteren, W.F. and Mark, A.E. *J. Mol Biol.*, 280, 925-932, (1998).

Duan, Y., Wang, L. and Kollman, P.A. *Proc. Natl. Acad. Sci. USA,* 95(17), 9897-902, (1998).

Erne, D., Rolka, K. and Schwyzer, R. *Helv. Chim. Acta.* 69, 1807-1816, (1986).

Filizola, M., Centeno, N.B. and Perez, J.J. *J. Pept. Sci.*, 3, 85, (1997).

Flory, P. J. *Statistical Mechanics of Chain Molecules*, Interscience Publishers, New York, (1969).

Keire, D.A. and Fletcher, T.G. *Biophys J.*, 70(4), 1716-27, (1996).

Law, G. *Clustering of Archaelogical Entities – An information Theory based Approach* (1999). Retrieved February 2000 from http://www.lawas.co.nz/PAPERS/CLUSTPAP/CLUSTPAP.HTM.

Lee, S., Suh, Y.H., Kim, S. and Kim, Y. *J. Biomol. Struct. Dyn.*, 17(2), 381-391, (1999).

Manavalan P. and Momany, F.A. *Int. J. Pept. Protein Res.*, 20, 351-365, (1982).

Manavalan, P. and Momany, F.A. *Pept. Synth. Struct. Proc. Am. Pept. Symp. 7th*, 713-716, (1981).

Mehlis B., Rueger, M., Becker, M., Bienert, N., Niedrich, H. and Oehme, P. *Int. J. Pept. Protein Res.* 15, 20-28, (1980).

Mehlis, B., Boehm, S., Bechker, M. and Bienert, M. *Biochem. Biophys. Res. Commun.*, 66, 1447-1453, (1975).

Miyazawa, T., *Kagaku Kyoiku*, 32, 18-21, (1984).

Nikiforovich, G.V., Balodis,I Yu and Cipens, G. *Bioorg. Khim.*, 7, 645-654, (1981).

Patel, A.B., Srivastava, S. , Phadke, R.S., *J. Biomol. Struct. Dyn.*, 19,129-138, (2001)

Rolka, K., Erne, D. and Schwyzer R. *Helv. Chim. Acta*, 69, 1807-1816, (1986).

Schwyzer, P., Erne D. and Rolka, K. *Helv. Chim. Acta*, 69, 1789-1797, (1986).

Shimohigashi, Y., Matsumoto, H., Takano, Y., Saito, R., Iwata, T., Kamiya, H. and Ohno, M. *Biochem. Biophys. Res. Commun.,* 193, 624-630, (1997).

Shukla, D.R. and Mahajan, S. *Proteins*, 85-95, (1991).

Srinivasan, R., Rose, G.D. *Proc. Natl. Acad. Sci. USA*, 96(25), 14258-14263, (1999).

Sumner S.C., Gallagher, K.S., Davis, D.G., Covell, D.G., Jernigan, R.L. and Ferretti, J.A. *J.Biomol. Struct.,* 8, 687-707, (1990).

Weiner, S.J., Kollman, P.A., Nguyen D.T. and Case, D.A. *J. Comput. Chem.,* 7, 230-252, (1986).

Wiliams, R.W. and Weawer, J.L. *J. Biol. Chem.,* 265, 2505-2523, (1990).

Wymore, T. and Wong, T.C., *Biophysical Journal*, 76, 1199-1212, (1999).

Yankner, B.A., Duffy, L.K., Kirschner, D.A., *Science*, 250, 279-282, (1990).

Young, J. K., Anklin, C., Hicks, R.P., *Biopolymers*, 34, 1449-1462, (1994)