

Multiple Molecular Dynamics Simulations Of A 28mer Oligopeptide Reveal Enhanced Sampling of Conformational Space

Sangeeta Sawant, A.S. Kolaskar

Bioinformatics Centre

University of Pune

Ganeshkhind, Pune 411007, India

sangeeta@bioinfo.ernet.in, kolaskar@bioinfo.ernet.in

Abstract

A strategy using multiple runs of MD simulations has been used for conformational searches of 28-mer peptide PSV which demonstrate the effectiveness of the strategy to sample the conformational space more efficiently than the single, long MD simulation. Further, the use of cluster analysis and $C\alpha$ distances has been made to detect native like folds of the peptide that are generated in the relatively short time scale simulations.

Keywords: oligopeptide, multiple MD simulations, conformational searches, cluster analysis.

1 Introduction

In Computational Structural Biology, the prediction of three-dimensional (3D) structures of oligopeptides, having amino acid chain lengths of 10-40, is an important area of research due to the biological significance of oligopeptides (Zornik et. al. 1999, Nichols et. al. 1999a, Nichols et. al. 1999b Morris et. al. 2000). The structural studies may also provide insights into the structure-function relationship of oligopeptides and to some extent, in the protein folding problem.

The 3D structures of oligopeptides with 10-40 amino acid residues have a higher flexibility as compared to the 3D structures of full protein domains and are stabilized primarily by short range and medium range interactions. Long range tertiary interactions may not be possible due to the smaller size of the oligopeptides. This inherent flexibility implies existence of multiple low energy structures of oligopeptides. It is necessary to identify these plausible structures of oligopeptides using computational methods, as structure determination using experimental techniques involves several problems due to instability and insolubility of oligopeptides.

The smaller size of oligopeptides and the structural aspects as described above, restrict the approaches to prediction of the 3D structures to *ab initio* approaches only, as homology modeling and fold recognition methods are tailored for full, larger protein domains. The *ab initio* approaches include conformational searches using techniques such as energy minimization, Molecular Dynamics (MD) simulations, simulated annealing, Monte

Carlo simulations etc. to generate several "decoy" structures followed by screening based on either statistical or empirical energy functions (Hardin et. al., 2002; Osguthorpe, 2000).

We have employed MD simulations using multiple runs approach to predict the plausible structures of a set of selected oligopeptides and have shown that the approach works better for exploring the conformational space of oligopeptides as compared to a single continuous MD simulation (Sawant 2002). Reasonable structures were obtained from this approach for oligopeptides that exist and function as independent units. In the present studies, we have applied the method to a 28-mer peptide sequence designed to adopt a specific β - β - α fold, the NMR structure of which is available in PDB (Code: 1PSV, Dahiyat et. al. 1997). Our results demonstrate that the multiple runs MD approach is able to sample the conformational space more effectively and in a reasonably short time scale of MD simulations, the native-like fold of the oligopeptide can be attained.

The MD simulation protocol developed by us involves multiple discrete MD runs; each of 1 nanosecond (ns) duration. A random, allowed initial structure was used in the first MD run of the peptide and for each successive MD run, the lowest energy allowed conformation from the previous run was used as the starting conformation. The ability and efficiency of this protocol to explore the conformational space of the peptides has been studied and compared with the results from a simulation using the traditional protocol of a continuous (single) MD run over an equivalent time scale. The structural features and characteristics of the low energy conformations have also been studied. Previously, multiple MD simulations have been shown to be useful for enhancing the conformational space searches in the terminally restrained loop fragment of tRNA^{Asp} (Auffinger et. al. 1995; Auffinger & Westhof, 1997) and also for conformational searches at room temperature in the vicinity of the native structure of crambin (Caves et. al. 1998). However, detailed studies using such an approach towards conformational searches of oligopeptides with lengths in the range of 20-40 amino acids have not been reported.

2 Methods and Materials

The amino acid sequence of the oligopeptide from the PDB entry 1PSV was used to build an initial random structure by assigning main chain dihedral angle ϕ, ψ values from the allowed regions of Ramachandran plot (Ramachandran and Sasisekharan, 1968) to each amino

acid residue. The omega (ω) angles of all the residues were assigned a value of 180° (corresponding to the trans configuration of peptide unit).

The initial structure was optimized to remove the short contacts and local strains if any, using steepest descent and conjugate gradient algorithms until a maximum derivative of $0.01 \text{ Kcal/mole } \text{Å}^\circ$ was reached. The MD simulations were carried out using two protocols: the multiple runs MD simulation and single run simulation. Simulation parameters used for the two protocols were identical. For all the simulations, Discover module of the InsightII software (InsightII, 2000) was used. The AMBER force field (Cornell et. al. 1995) was used with an implicit solvent effect, using a distance dependent dielectric constant of 4.0. A group-based method was used for calculation of non-bonded interactions with the cut-off distance of 14.0Å° so as to include maximum non-bonded interactions in the oligopeptide during the simulations. The non-bonded 1-4 interactions were scaled by a factor of 0.5 as is recommended for the AMBER force field. The temperature of 400K was used with a time step of 1 femtosecond (fs) during the equilibration as well as production runs. An equilibration run of 50 ps was followed by the MD production run.

2.1 Multiple Runs MD Simulation Strategy

Five MD simulations of 1 ns duration each were carried out (NS1, NS2, NS3, NS4 and NS5). For the first MD run, the optimized initial structure was used as starting structure while for each successive MD run, the lowest energy structure resulting from the previous run was used as the starting structure. The total duration of MD simulations carried out in this way was 5 ns.

2.2 Single Run MD Simulation Strategy

The routine MD protocol was used to run a single 5 ns long MD simulation (5NS_CONT). The initial optimised structure of the peptide mentioned above was used as a starting structure in both the strategies. All the other simulation parameters employed in the two strategies were also identical.

2.3 Data Capture and Analysis

In both the MD strategies, the intermediate frames were captured at 10 ps interval during the production run, thus generating a total of 500 structures of the peptide in each strategy. All these frames were minimized using steepest descent and conjugate gradient algorithms with the maximum derivative criterion of $0.001 \text{ Kcal/mole } \text{Å}^\circ$. The minimized structures from all the trajectories were analyzed using cluster analysis tool PDBCLUST (Momary et. al. 1975, Nemethy et. al. 1983 & 1992, Sippl et. al. 1984, Ripoll et. al. 1995) and Analysis module of the InsightII (InsightII, 2000) software. Backbone RMSD cut-off values of 3.0Å° , 2.5Å° , 2.0Å° , 1.5Å° were used for clustering. The allowed lowest energy structures with a long residence time in the most populated clusters were identified. For defining the allowed structures, the program PROCHECK (Laskowski et. al. 1993; Morris et.

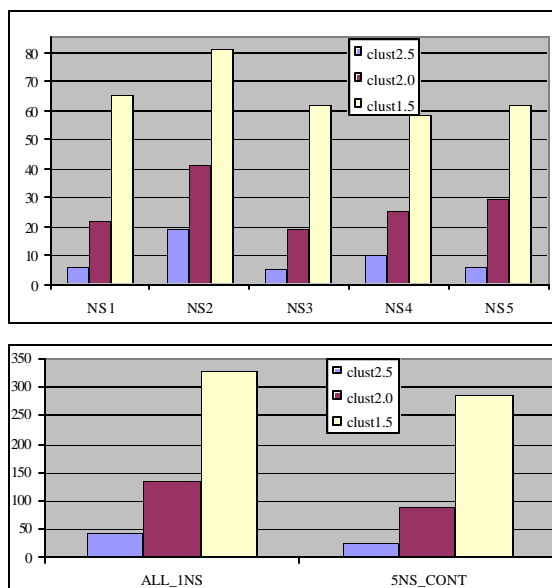


Fig 1: Histograms of the number of clusters obtained for the five 1 ns trajectories of PSV peptide (top) and for the ALL_1NS and 5NS_CONT data sets (bottom) at different cut-offs of backbone RMSD.

al. 1992) was used with a criterion that a conformation is allowed only if the values of (ϕ , ψ) angles of all the residues lie in the fully or partially allowed regions of the Ramachandran plot. The five 1 ns trajectories of the multiple runs MD strategy were clustered individually as well collectively, by combining the data from the 5 trajectories together (ALL_1NS data set). In order to estimate the efficiency of the multiple runs MD strategy in sampling the conformational space, clustering results of the ALL_1NS data set were compared with those of the 5NS_CONT. Additionally, two more analyses were carried out for the same purpose using the Analysis module of InsightII:

- Variation of the main chain dihedral angles ϕ and ψ of each amino acid residue as a function of simulation time was examined for each trajectory.
- Variation of the N-to-C terminal distances of the peptide in each trajectory, measured in terms of distance between C_1^α and C_{28}^α was examined.

3 Results

MD simulation is an effective tool for conformational searches as it provides detailed atomic data of temporal evolution of the molecular system. The parameters used in MD simulations are of key importance because a realistic simulation of the molecular system and its environment demands accurate representation and careful treatment of the various parameters. In the present studies, the AMBER force field (Cornell et. al. 1995) was used along with other parameters including temperature, time step of data integration and data capture, solvent effects etc. that were standardized for a set of oligopeptides (Sawant 2002). Usage of implicit solvent model by choosing an appropriate dielectric constant has been considered to be a reasonable approximation of the

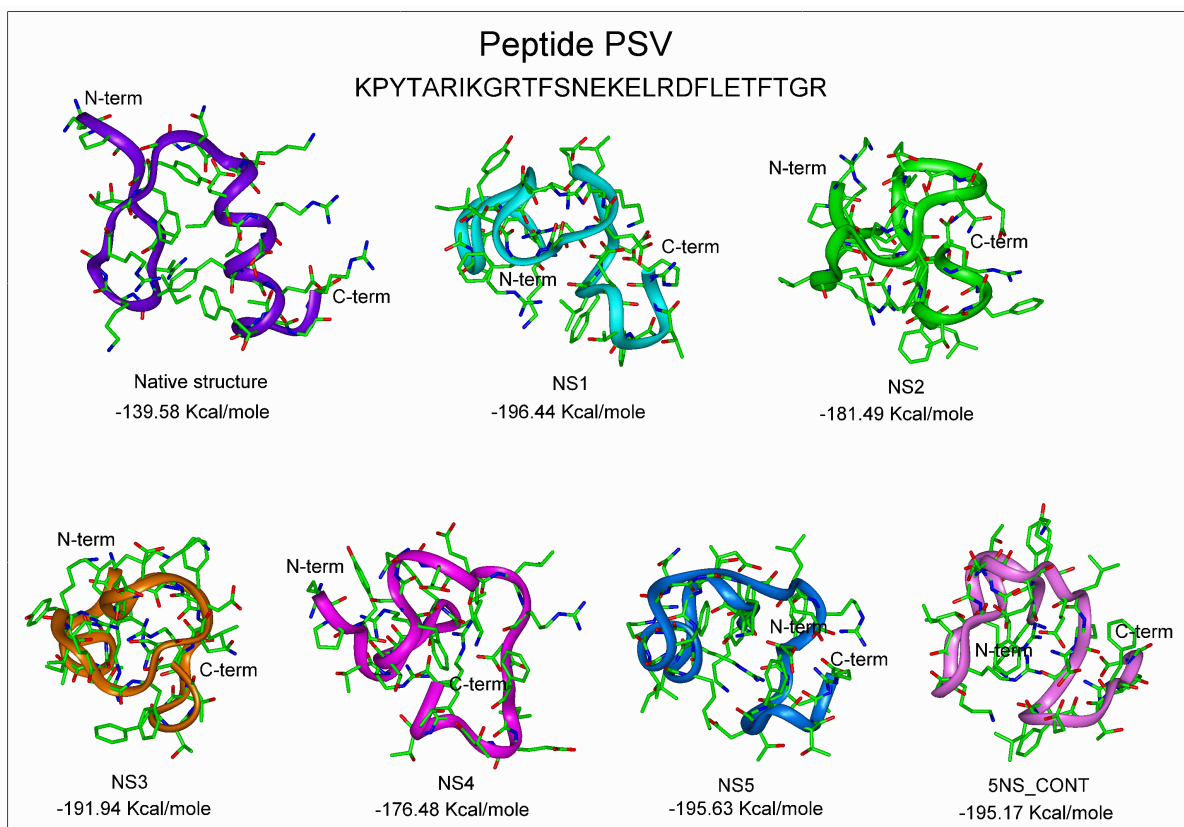


Fig. 2: The native structure of the PSV peptide (1PSV), the lowest energy allowed structures obtained in the five 1 ns trajectories and the 5NS_CONT trajectory. The backbones of all structures are shown as ribbons.

solvent effect (Daggett et. al. 1991) for conformational searches. The default NVT ensemble (canonical ensemble) was used where conditions of constant-temperature, constant-volume are maintained. This is the appropriate choice when conformational searches of molecules are carried out in vacuum without periodic boundary conditions. The temperature used for the simulations was optimum and folding-unfolding movements were observed in the peptide as can be seen in the plots showing the variation of N-to-C terminal distances of the PSV peptide (fig.4). MD simulations were not carried out at high temperature as the geometry of the molecules may get distorted during such a simulation e.g. the extent of bond stretching and angle bending may be high leading to structural changes rather than only the desired conformational changes. We have observed in simulations of smaller peptides at high temperature that the conformational sampling does not improve at high temperatures (Sawant, 2002).

The equilibration period of 50 ps was sufficient to bring the molecular system to a state of equilibrium wherein thermodynamic properties such as temperature, pressure, potential energy etc were uniform over the ensemble.

In the production run, intermediate frames were captured at 10 ps intervals. This has yielded a sufficiently large ensemble of the structures of PSV peptide representing significantly varying conformational states of the molecule. This can be seen from the clustering results (Fig. 1). A comparison of the data set captured at 10 ps

with a data set captured at 1 ps intervals have shown that no loss of structural data occurs by increasing the data capture time interval (data not shown). This indicates that the essential structural transitions can be captured at 10 ps interval with adequate efficiency.

3.1 Comparative analysis of MD trajectories obtained using two strategies

In the multiple runs MD strategy, the starting structure used for each MD run was a low energy structure representing a distinct local minimum on the energy landscape of the oligopeptide. The independent MD runs provided different initial velocities that guided each run to different paths in the conformational space. This entails the probability that the effective conformational space that could be sampled in such a strategy would be larger, as compared to the traditional MD simulation involving a continuous, single run. This hypothesis was examined by analysing the trajectories of the candidate peptide.

3.2 Clustering of MD Trajectories

Cluster analysis of MD trajectories provides useful data such as (a) the sets of conformationally similar frames from which the best conformation can be selected as a representative, (b) an estimate of the stability of various conformations as represented by the sizes of clusters and the time frames over which they occur; and (c) the conformational transitions from one state to another, that

Cluster no.*	No. of frames in cluster	Trajectory
RMSD cut-off 3.0Å^o		
1	052	NS4
2	100	NS5
3	043	NS4
4	211	NS1 (few), NS2
5	083	NS1
RMSD cut-off 2.5Å^o		
1	51	NS4
2	43	NS5
3	39	NS4
6	23	NS1
7	17	NS1
9	19	NS1
11	08	NS2
14	05	NS2
16	15	NS2
18	13	NS2
19	20	NS5

*: only those clusters that contain the frames with allowed structures are shown.

Table I: Clustering Data for ALL_1NS set at 3.0Å^o and 2.5Å^o

indicate the path of folding or unfolding. In the present studies, clustering was used for estimating the extent of conformational diversity resulting from the two strategies of MD simulations as well as for predicting the most plausible peptide structures on the basis of the cluster stability.

For cluster analysis the values of RMSD of backbone atoms were used as classification criterion (Fig. 1). At 3.0Å^o cut-off, very few clusters were obtained in which the distribution of frames was skewed, with most of the frames grouped in one cluster. On the other hand, the cut-off of 1.0Å^o resulted in generation of a large number of clusters, each one having very small number of individual frames, implying that at this value of RMSD, most of the conformations were unique.

3.3 Comparison of clustering data of multiple runs and single run MD trajectories

The results of combined clustering of five 1 ns trajectories (ALL_1NS) were compared with those of 5 ns single run MD trajectory (5NS_CONT). At each cut-off, the number of clusters obtained for the ALL_1NS data set was higher than the number of clusters for 5NS_CONT trajectory (Fig 1, lower panel). The energy-minimized frames of the trajectory that were clustered represent distinct minima on the potential energy surface. Thus, at a given resolution (cut-off), a cluster of such minima would represent a distinct conformational substate (Troyer and Cohen, 1995). Hence, the number of clusters obtained at that cut-off is a measure of conformational diversity. Larger the number of clusters in

Trajectory	Lowest energy (L)	Highest energy (H)	Difference in L & H
NS1	-189.17	-151.16	-38.01
NS2	-190.68	-153.96	-36.72
NS3	-191.94	-153.63	-38.31
NS4	-200.38	-151.62	-48.76
NS5	-195.63	-157.70	-37.93
5NS_CONT	-205.71	-130.82	-74.89

Table II: The energy values in Kcal/mole obtained in the five 1ns and 5NS_CONT trajectories of PSV peptide

a trajectory, larger is the conformational diversity, indicating that the extent of sampling of the conformational space is higher. A higher number of clusters in ALL_1NS clustering clearly indicate that the multiple runs MD strategy was able to generate a more diverse ensemble of peptide structures as compared to that from single run MD strategy. This is a significant finding which supports the hypothesis that multiple runs MD strategy can sample a larger conformational space as compared to single run MD.

The ALL_1NS and 5NS_CONT data clusters at various RMSD cut-offs were examined to identify the most populated clusters containing allowed structures of the peptide (definition of allowed frames in the methods section). The assumption made here was that the clusters with large groups of temporally consecutive frames having allowed structures would represent the highly plausible structures of the peptide. At all the cut-offs used for clustering the ALL_1NS data set, (ranging from 3.0Å^o to 1.5Å^o), the allowed structures of individual 1 ns trajectories were segregated into independent clusters (Table I). No convergence of frames from two or more trajectories into a single cluster was observed (except for cluster 4 at 3.0Å^o). This implies that the five 1 ns trajectories explored various distinct regions of the conformational space through rapid structural transitions. However, the data indicates that these transitions were not reversible (across distinct trajectories). This is in contrast with the results for shorter peptides (with ≤ 15 amino acids) where reversible structural transitions were observed in the 1 ns trajectories (Sawant, 2002).

On the other hand, in the 5NS_CONT trajectory, the frames from temporally distinct stretches were found to be clustered together especially at the RMSD cut-offs of 3.0Å^o and 2.5Å^o. At these two cut-offs, the largest clusters represent approximately 50% and 30% of the total MD duration respectively. It appears from these results that in the longer, single run MD simulation, the peptide conformations get trapped into local minima and the structural transitions are much slower.

3.4 Trends of lowest energies in the two strategies of MD

The lowest energy values of the successive 1 ns runs of the PSV peptide varied over a range of ~ 10 Kcal/mole (Table II). The average difference in the highest and

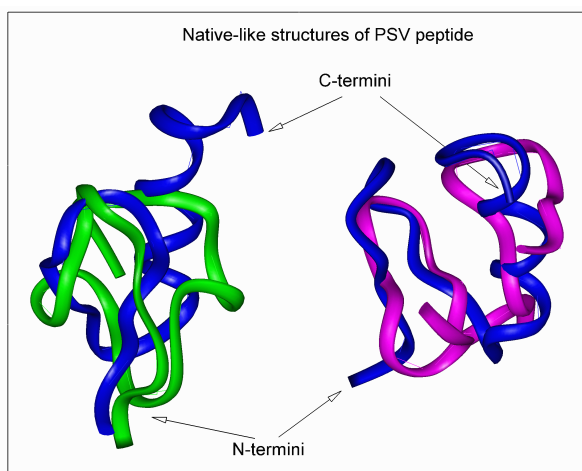


Fig. 3: Partially native-like structures from the NS2 (left, shown in green) and NS5 (right, shown in magenta) trajectories of the PSV peptide superimposed with the native structure (shown in dark blue in both panels). The N-terminal stretches in both the structures have an extended structure very similar to the native one. NS2 structure also shows a helical turn.

lowest energies of the individual 1 ns trajectories was of the order of ~ 40 Kcal/mole while the difference in the highest and lowest energies of the 5NS_CONT trajectory was ~ 75 Kcal/mole. The minimum-most energy of the 5NS_CONT trajectory was significantly lower than the minima of any of the 1 ns trajectories. The structure from 5NS_CONT corresponding to the lowest energy minimum (Fig. 2) had a high C^α RMSD value of 6.5\AA when superimposed with the native structure. It did not possess any well-formed secondary structural stretches either.

All the low energy conformations generated in various trajectories had negative non-bond energies and were stabilized by several hydrogen bonds between the atoms of the main-chain and side-chains.

3.5 Structural features of the low energy allowed structures of PSV

PSV is a 28-mer synthetic peptide with a sequence designed to adopt a β - β - α fold (Fig. 2). It has an extended conformation in the stretches Lys1-Arg6 and Arg10-Asn14 while the stretch Glu15-Glu23 is helical. The extended structures are not classified as β -strands by the definitions of DSSP (Kabsch and Sander, 1983) as they lack the necessary H-bonding pattern.

The lowest energy allowed structures of the peptide obtained in the NS1, NS2, NS3, NS4 and 5NS_CONT trajectories are shown in Fig. 2. It can be seen that none of these structures possess the perfect native fold. However, a closer look at the data revealed that many of the low energy structures obtained from the simulations had the fully or partially formed individual local structural elements of the native fold (extended and helical regions). Two of such representative structures are shown in Fig. 3, in superimposition with the native

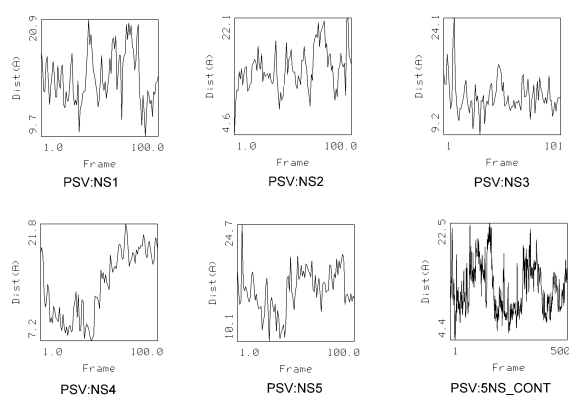


Fig. 4: The plots of the Ca1- Ca28 distances as a function of (consecutive) frame numbers in the five 1ns and 5NS_CONT trajectories. (The variations of Ca1- Ca6, Ca9- Ca13 and Ca15- Ca23 distances were also plotted as discussed in the Results section, but are not shown here).

structure. An analysis of the variation of the distances between C^α atoms of three pairs of amino acid residues, viz. Lys1-Arg28, Lys1-Arg6, Gly9-Ser13 and Glu15-Glu23 (which mark the boundaries of the secondary structures) in each trajectory was carried out (Fig. 4). These distances can be used to identify the type of local regular structure in the respective regions as local regular structures have characteristic values of end-to-end distances. A comparison of the native distances and the distances in MD-generated frames showed that several frames in all the five 1 ns trajectories had well-formed individual secondary structure elements. However, their orientations and packing were different from those in the native. Some of these structures had higher energy values (difference of >3.0 Kcal/mole) as compared to the minimum-most energy of the corresponding simulation. The average RMS deviation of the C^α trace of native-like structures with that of the native PSV structure was 5.5\AA (calculated using the superimposition method in InsightII).

3.6 Conformational variations of individual amino acid residues

The plots of ϕ and ψ values of individual amino acid residues as a function of frame numbers in each trajectory revealed that in the 1 ns multiple MD trajectories, most of the amino acid residues sample a wide range of the conformational space, adopting right handed and left handed helical conformations, extended structure, as well as some of the disallowed conformations, with varying residence time in each (Fig. 5, a, b). On the contrary, in the 5NS_CONT trajectory, majority of the amino acid residues tend to reside in one conformation for longer duration (fig. 5 b). It may be stated on the basis of these plots that a high diversity of conformations adopted by the amino acid residues is a marked feature of the multiple MD run strategy as against the low conformational diversity in the single run MD.

4 Discussion

The amino acid sequence of a protein is the most significant factor that influences its folding into a 3D structure (Anfinsen, 1973). Besides this, the environmental parameters such as temperature, solvent, pH, also affect the protein folding process and the 3D structure. The structural attributes (ϕ, ψ angles, secondary structure, solvent accessible surface etc.) of the local elements of a protein are a consequence of the context of the folded protein, in addition to all the above factors. For oligopeptides which exist as independent functional units, the same principles of folding as for the larger protein domains may apply. However, the size of the molecule is a limiting factor, which excludes the possibility of stabilizing, long-range interactions. The local interactions and to some extent, medium range interactions would dominate the structure formation of oligopeptides. The present studies have tried to investigate these assumptions using the conformational search approach involving MD simulations.

The low energy structures generated in the present studies of PSV peptide have shown that the folding process of the oligopeptides may follow a similar pattern as that for proteins where the local regular structures are formed initially which consequently undergo rearrangement to render compact packing. All types of secondary structures

i.e. helices, strands and turns were observed in the frames generated. The short extended structures of the PSV peptide were having a marginally longer lifetime and better energy values as compared to the extended stretches formed in oligopeptides with ≤ 15 amino acids, which were observed to be highly transient and high energy structures (Sawant, 2002). This may be due to the medium range tertiary interactions present in longer peptides (≥ 15 amino acids) that are not possible in the shorter ones. Another distinct feature of PSV simulation studies was that the formation of helical structure was much slower as compared to that in shorter peptides. In the MD simulation studies of 12-15-mer peptides helix formation was observed in few hundred ps simulation time (Sawant, 2002); while in the case of PSV, only 1 or 2 turns of helices could be observed over 1ns simulation.

The fact that short stretches of regular structures such as helices or strands are seen in some of the plausible structures irrespective of a higher energy value, implies that these could be the potential sites of nucleation for the protein folding process. In case of full, larger protein domains, these high energy structures may get stabilized by tertiary interactions.

Although the structural transitions of longer oligopeptides such as PSV are expected to be slower as compared to those of shorter oligopeptides (≤ 15 a.a.) the trajectories generated here over a total of 5 ns simulation have shown that the multiple runs MD strategy is more appropriate for observing the structural transitions than the single MD

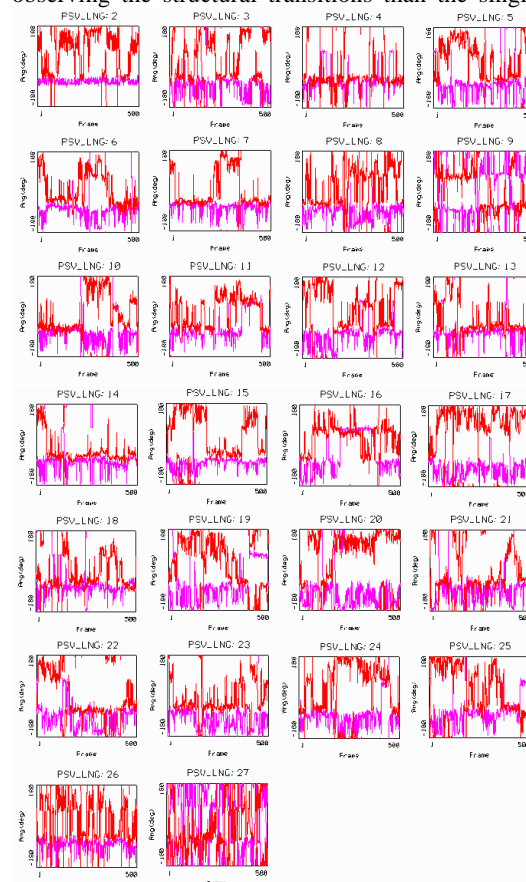
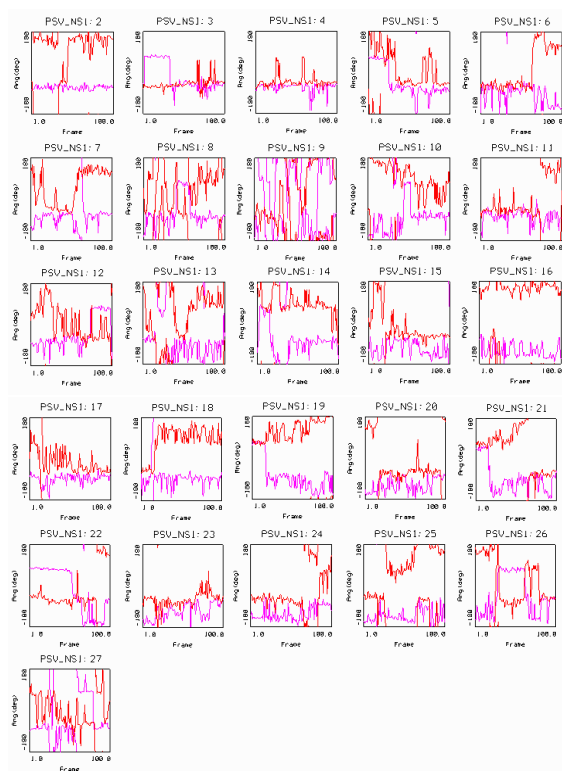


Fig. 5: The variations of the main chain dihedral angles (ϕ, ψ) of individual amino acid residues of the PSV peptide. (a) Plots for the NS1 trajectory (above) Similar plots were obtained for the other 1 ns trajectories. (b) Plots for 5NS_CONT trajectory. (next column) Plots of variation of ϕ are shown in magenta and those of ψ in red colour.

run. This is the effect of using distinct initial conditions in each MD run, viz. different starting structure and initial velocities assigned in the beginning of the run. These parameters may lead the path of the simulations in dissimilar directions along the potential energy landscape resulting into conformational diversity. Thus a relatively shorter MD simulation could generate peptide structures similar to the native fold. Further investigations may be necessary to verify whether additional MD runs would lead to structures more closely resembling the native structure. The kinetic and MD simulations of the folding-unfolding events of the 61-mer protein, Engrailed Homeodomain (from *D. melanogaster*), one of the fastest folding-unfolding proteins, have revealed that the half-life of unfolding at 100°C is of the order of 7.5 ns (Mayor et. al. 2000). The time scale of 5-7 ns is therefore sufficient to observe the folding unfolding events and adequate sampling of conformational space of the peptides considered in the present studies, which are much smaller than the Engrailed Homeodomain.

The MD simulation approach used by us does not allow one to study the macroscopic properties that may be desired from MD simulations. This is because a continuous time scale is not used here. However, our interest is to enhance the conformational sampling so as to observe formation of native-like structures of oligopeptides, which is well served by the strategy used. It may be argued that use of higher temperature with a continuous time scale or use of Simulated Annealing approach can be done for achieving these objectives. Our studies on small peptide MD simulations with higher temperature have convincingly proven that the degree of sampling at higher temperatures is very similar to that at the temperature used here (Sawant, 2002). On the other hand, the approach of Simulated Annealing, which is known to be useful, also necessitates the use of higher temperatures that may result in undesired geometric changes in the molecule through higher degree of bond stretching and angle bending events. These approaches therefore were not explored.

The interval of data collection used during the MD production runs corresponds to a coarse grid sampling of the conformational space. This data collection frequency does not lead to a loss of information due to its coarseness, as was observed in the comparison of the 10ps and 1ps data capture results (Sawant, 2002). In *ab initio* methods of protein structure prediction, coarse grid sampling is used to generate the low resolution structure decoys which are later refined using statistical or physical energy functions for optimization (Xia et. al. 2000; Betancourt & Skolnick, 2001). In the current approach, since an all-atom-based representation is used, the method directly yields 3D structure of the peptides.

Clustering procedures are commonly used for the analysis of MD trajectories and are an effective way of data reduction, to focus on the extraction of most important characteristics of the dynamics data. This technique, in combination with the basic assumption that two structures that are similar to each other at a given cut-off of RMSD, represent a conformational substate, which in turn represents a local minimum on the potential energy

landscape of the protein or peptide (Troyer and Cohen, 1995); can be effectively used to identify structural transitions as well as stable, low energy structural states. The definition of the RMSD cut-off depends on the expected resolution of the conformational differences/changes, ranging from small torsional changes to hinge-bending movements. For distinguishing structural differences in hexapeptides, a resolution of 1.5Å° was shown to be adequate (Cohen et. al. 1993). In the present studies, the RMSD cut-offs in the range of 1.5 to 2.5 Å° have yielded structural resolution at three levels. For PSV peptide RMSD of 2.0Å° was found to be optimum, which facilitated identification of distinct structural states.

It has been argued that the energy minima on the conformational space that are in close proximity do not necessarily represent kinetically accessible states of the proteins (Becker and Karplus, 1997). This may not apply in entirety to the oligopeptides of the type as studied here, as the pathways of folding available for these smaller molecules may be quite different than those of the larger native domains. The rapid structural transitions observed in the 1 ns MD trajectories of PSV peptide point at this view. However, confirmation of this would warrant more detailed studies of the nature of the potential energy surface of the oligopeptides.

5 Conclusions

The results presented here clearly indicate that in the MD simulations of the 28-mer oligopeptide using multiple runs approach, different energy basins are explored while the single continuous MD simulation may get trapped into a single local energy basin and explore only a small region of conformational space. The relatively short time scale simulations resulted in crude native-like models that could be identified using a combined approach of cluster analysis and C^α distance analysis. The approach of multiple runs MD simulation therefore is useful and may be successfully employed for the prediction of oligopeptide structures through conformational searches.

6 References

- ANFINSEN, C.B. (1973): Principles that govern the folding of protein chains. *Science* **181**:223-230.
- AUFFINGER, P., LOUISE-MAY, S., WESTHOF E. (1995): Multiple molecular-dynamics simulations of the anticodon loop of tRNA(Asp) in aqueous-solution with counterions. *J Am Chem Soc* **117**:6720-6726.
- AUFFINGER, P., WESTHOF, E. (1997): RNA hydration: Three ns of multiple molecular dynamics simulations of solvated tRNA(Asp) anticodon hairpin. *J Mol Biol* **269**:326-341.
- BECKER O.M., KARPLUS M. (1997): The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J Chem Phys* **106**:1495-1517.
- BETANCOURT, M.R., SKOLNICK, J. (2001): Finding the needle in a haystack: deducing native folds from

- ambiguous ab initio protein structure predictions. *J Comput Chem* **22**:339-353.
- CAVES, L.S.D., EVANSECK, J.D., KARPLUS, M. (1998): Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Sci* **7**:649-666.
- COHEN, B.I., PRESNELL, S.R., COHEN, F.E. (1993): Origins of structural diversity within sequentially identical hexapeptides. *Protein Sci* **2**:2134-2145.
- 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. (1995): A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J Am Chem Soc* **117**:5179-5197.
- DAGGETT, V., KOLLMAN, P.A., KUNTZ, I.D. (1991): Molecular dynamics simulations of small peptides: Dependence on dielectric model and pH. *Biopolymers* **31**:285-304.
- DAHIYAT, B.I., SARISKY, C.A., MAYO, S.L. (1997): De novo protein design: towards fully automated sequence selection. *J Mol Biol* **273**:789-96.
- HARDIN, C., POGORELOV, T.V., LUTHEY-SCHULTEN, Z. (2002): *Ab initio* protein structure prediction. *Curr Op in Str Biol* **12**:176-181.
- INSIGHTII (2000): Accelrys Inc, San Diego, CA, USA.
- KABSCH, W. and SANDER, C. (1983): Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**:2577-2637.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. and THORNTON, J. M. (1993): PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**:283-291.
- MAYOR, U., JOHNSON, C.M., DAGGETT, V., FERSHT, A.R. (2000): Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci* **97**:13518-13522.
- MOMANY, F. A., MCGUIRE, R. F., BURGESS, A. W., SCHERAGA, H. A. (1975): Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J Phys Chem* **79**: 2361-2381.
- MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. and THORNTON, J. M. (1992): Stereochemical quality of protein structure coordinates. *Proteins Struct funct Genet* **12**:345-364.
- MORRIS, M.C., CHALOIN, L., HEITZ, F., DIVITA, G. (2000): Translocating peptides and proteins and their use for gene delivery. *Current Opinion in Biotechnology* **5**:461-466.
- NEMETHY, G., GIBSON, K.D., PALMER, K.A., YOON, C.N., PATERLINI, G., ZAGARI, A., RUMSEY, S., SCHERAGA, H.A. (1992): Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J Phys Chem* **96**:6472-6484.
- NEMETHY, G., POTTLE, M. S., SCHERAGA, H. A. (1983): Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J Phys Chem* **87**:1883-1887.
- NICHOLS, R., LIM, I., MCCORMICK, J. (1999a): Antisera to multiple antigenic peptides detect neuropeptide processing. *Neuropeptides* **33**:35-40.
- NICHOLS, R., MCCORMICK, J., COHEN, M., HOWE, E., JEAN, C., PAISLEY, K., ROSARIO, C. (1999b): Differential processing of neuropeptides influences *Drosophila* heart rate. *J Neurogenet* **13**:89-104.
- OSGUTHORPE, D.J. (2000): Ab initio protein folding. *Cur Op in Str Biol* **10**:146-152.
- RAMACHANDRAN, G.N., SASISEKHARAN, V. (1968): Conformation of polypeptides and proteins. *Adv Protein Chem* **23**:283-438.
- RIPOLL, D.R., POTTLE, M.S., GIBSON, K.D., LIWO, A., SCHERAGA, H.A. (1995): Implementation of the ECEPP algorithm, the Monte Carlo Minimization method, and the Electrostatically Driven Monte Carlo method on the Kendall Square research KSR1 computer. *J Comput. Chem* **16**:1153-1163.
- SAWANT, S. (2002): Development of methods to predict the 3D structures of proteins and oligopeptides. Ph.D. Thesis. University of Pune, Pune.
- SIPPL, M.J., NEMETHY, G., SCHERAGA, H. A. (1984): Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O-H...O = C hydrogen bonds from packing configurations *J Phys Chem* **88**:6231-6233.
- TROYER, J.M., COHEN, F.E. (1995): Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory. *Proteins Struct funct Genet* **23**:97-110.
- XIA, Y., HUANG, E.S., LEVITT, M., SAMUDRALA, R. (2000): Ab initio construction of protein structure using ahierarchical approach. *J Mol Biol* **300**:171-185.
- ZORNIK, E., PAISLEY, K. and NICHOLS, R. (1999): Neural messengers and a peptide modulate *Drosophila* heart rate. *Peptides* **20**:45-51.