

The Structure of Ghrelin

Andreas Kukol

Running title: Ghrelin Structure

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK,
Tel. +44 (0)24 7652 8373, FAX +44 (0)24 7652 3701, Email:
a.kukol@warwick.ac.uk

Abstract – The structure of ghrelin, a 28-residue octanoylated peptide hormone, is only known up to the level of primary structure identifying an active core of residues 1-5 or 1-4 including octanoyl-Ser3 as necessary to elicit receptor response. This chapter reviews the results and limitations of experimental and computer modelling studies, which have appeared in the literature. The ¹H-NMR spectroscopy experimental studies revealed an unstructured and/or fast interconverting peptide at acidic pH, while molecular dynamics (MD) simulation studies at neutral pH pointed to a stable conformation over a time period of 25 ns in water and in the presence of a lipid bilayer. The significance of these findings is discussed with regards to the pH difference, the timescales accessible to simulation and NMR spectroscopy and the limitations of computational modelling. MD simulations of ghrelin in the presence of a lipid membrane revealed that the octanoyl side-chain did not insert into the lipid bilayer, but instead the peptide bound to the lipid headgroups with residues Arg15, Lys16, Glu17 and Ser18, which are located in a hairpin-like bend in the structure. The implications of these findings with regards to a recently obtained homology model of the ghrelin receptor are discussed.

I. Introduction

Ghrelin is a peptide hormone secreted from endocrine cells in the stomach acting as the natural agonist of the growth hormone secretagogue receptor (GHSR), a G-protein coupled receptor. It plays an important role in the control of growth hormone secretion and also in appetite regulation and food intake. Since its discovery in 1999 by Kojima et al. (1999) many reviews have appeared in the literature (Kojima and Kangawa, 2005; Korbonits et al., 2004) and most recent functional aspects are extensively discussed in this volume. Structural studies of ghrelin have concentrated on the primary structure, i.e. the sequence of amino acid residues, which consists of 28 residues with an octanoyl-group connected to Ser3 (fig. 1). The active core required for agonist potency of the human GHSR is the N-terminal tetrapeptide Gly-Ser-Ser(n-octanoyl)-Phe-COOH (Bednarek et al., 2000). The GHSR is a seven transmembrane helix G-protein coupled receptor, which has been discovered long before its natural ligand ghrelin was known (Guillemin et al., 1982).

An important goal in structural biology is to obtain the 3-dimensional tertiary structure of a biomolecule, which is the focus of the current chapter. In order to develop synthetic agonists and antagonists for the GHSR, which may be used as drugs for the treatment of pathophysiological conditions related to ghrelin and its receptor, it is advantageous to know the 3-dimensional structure of the natural agonist ghrelin. The structure may be helpful to understand pathophysiological conditions, binding to other proteins or lipid bilayers or it may be used as a pharmacophore model that defines essential structural features a synthetic agonist should possess, which is useful even in the absence of a structure for the ghrelin receptor. This chapter critically reviews experimental and theoretical studies relating to the 3-dimensional structure of

full-length human ghrelin with emphasis on molecular dynamics simulations. The structure of the GHSR is not known, although very recently a homology model has been presented (Pedretti et al., 2006), which will be discussed briefly below in relation to ghrelin structure.

II. Background

A. Experimental Methods

The main experimental methods to yield a high-resolution structure of proteins are x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. While x-ray crystallography is based on the analysis of diffraction patterns of crystals, NMR spectroscopy is the method of choice for structural analysis of smaller proteins (up to 25 kDa) in solution (Wuthrich, 1986). NMR spectroscopy relies on the absorption of radiowaves by nuclei (^1H , ^{13}C , ^{15}N , etc.) in a magnetic field. If the bonding structure of a molecule is known, as it is the case for proteins, 2D- and 3D-NMR spectroscopy is used to obtain torsion angles from measured coupling constants and distances from the nuclear overhauser effect (NOE). These data constrain the conformational space a molecule can adopt and with a sizeable number of such constraints, five or more per residue, an atomic 3-dimensional structure is calculated with the help of computers, usually involving in vacuo molecular dynamics calculations (see II.B below) and energy minimisation. While there is no lower size limit for NMR spectroscopy, the analysis of spectra becomes difficult, if the molecule exists in two or more conformations, in particular if these conformations interconvert between each other on a time scale, which is fast compared to the millisecond time resolution in analysing chemical shifts of ^1H -NMR spectroscopy. The possibility of temperature reduction in order to decrease the rate of interconversion is limited for biological samples by the

freezing point of water. Thus an NMR structural analysis of small flexible proteins/peptides is often not possible.

B. Computational Methods

Computational methods rely usually on a quantum mechanical or classical mechanical representation of a molecule. Given the size of biomolecules quantum mechanical calculations are too time consuming, thus a classical mechanical representation is often chosen, in which all molecules in the system are represented by ball-and-spring models. The forcefield describes the equations and parameters for the potential energy of the system in dependence of the atom coordinates. The forcefield takes into account bond lengths between atoms, bond angles, torsion angles, electrostatic and van der Waals interactions. The parameters of the forcefield are obtained from quantum mechanical calculations, vibration frequencies or chosen to reproduce thermodynamic properties using simulation; e.g. for the GROMOS96 forcefield used in the ghrelin study parameters have been chosen to reproduce the experimental heat of vaporization and density of aliphatic hydrocarbons (van Gunsteren et al., 1999). In order to speed up calculations GROMOS96 uses the united atom approach, which subsumes non-polar hydrogen atoms into their adjacent carbon atom.

In order to obtain stable, physically realistic conformations of molecules the procedures of energy minimisation (EM) and molecular dynamics (MD) simulation are applied. EM is always used after initial construction of a molecule in order to prepare the system for a subsequent MD simulation. EM tries to minimise the potential energy of the system as calculated from the forcefield by incrementally changing the atom coordinates according to a minimisation algorithm. EM is only able to find local minima and cannot overcome energy barriers. MD simulations

calculate the 'real' thermal fluctuations of the system at a given temperature and pressure over a defined period of time usually in the range of picoseconds or nanoseconds limited by the computational power or time the experimentator has available. A landmark MD simulation was the 1 μ s simulation of a 36 residue protein in explicit solvent starting from a fully extended structure, which folded into a marginally stable state that resembled the native experimentally known conformation (Duan and Kollman, 1998). In case the folding pathway is not of interest, but only the global energy minimum, simulated annealing (SA) MD is a powerful technique to find the most stable conformation. Rather than trying to mimic a natural process, SAMD is a computational technique to find the best solution to a problem, which has a large number of possible solutions (Kirkpatrick et al., 1983). SAMD simulations are started at a very high temperature and the system is progressively cooled down. At high temperature the system is able to effectively explore the conformational space overcoming high-energy barriers, while as the temperature is reduced lower energy states become more probable and eventually at zero temperature the system should reach the global energy minimum.

Inherent limitations of MD and SAMD simulations are the accuracy of the forcefield, the length of the simulation, e.g. picoseconds to nanoseconds, which may be short compared to the time scale of conformational changes of proteins (microseconds to hours) and the execution of only one or a few simulations compared to the statistical average over $6 \cdot 10^{14}$ molecules in an experiment using 1 mL of a 1 μ M solution. Thus the statistical significance of one MD simulation even carried out over several microseconds may be questioned. In the search for an energy minimum it is important to start the MD simulation from several different conformations as well as repeating the simulation of the same starting structure at different random initial atom velocities

as exemplified in the conformational search protocol for α -helical bundles in a lipid bilayer (Beevers and Kukol, 2006b; Beevers and Kukol, 2006c).

III. Results of Structural Studies

As mentioned in the introduction, structural studies of ghrelin have concentrated on the primary structure, i.e. the sequence of aminoacid residues and octanoylation of Ser3. So far only two studies relating to the 3-dimensional structure have appeared in the literature, one NMR study (Silva Elipe et al., 2001) and MD simulation studies in explicit water and in the presence of an explicit lipid bilayer (Beevers and Kukol, 2006a).

A. Nuclear Magnetic Resonance (NMR)

Using ^1H -NMR spectroscopy a structural analysis of full length human ghrelin and various truncated analogs including the octanoylated segment of the first five residues, which has shown binding to the human ghrelin receptor (GHSR), has been performed (Silva Elipe et al., 2001). NMR spectra were recorded at a temperature of 298 K in an $\text{H}_2\text{O}/\text{D}_2\text{O}$ mixture (90:10) with a small amount of trifluoro-acetic acid, which was necessary to increase solubility. This resulted in an acidic pH between 1.1 and 3.1 for the various samples. The low dispersion of chemical shifts indicated an undefined random coil structure in all cases. Furthermore, NOE connectivities were only detected between hydrogen atoms bonded to neighbouring atoms, e.g. $\text{HC}_\alpha\text{-NH}$ and within each individual amino acid residue. This indicates the absence of secondary structure elements like α -helices or β -sheets as well as a high mobility of the random coil structures. An exception was the five residue long truncated ghrelin peptide, which showed two NOEs between octanoyl-Ser3 and Phe4 indicating some

rigidity between the β -CH₂ group of Ser and the NH group of Phe4; unfortunately the derived distance has not been reported.

It can be concluded that under the experimental conditions chosen for the NMR experiment human ghrelin shows a random coil structure within the millisecond time frame accessible to standard ¹H-NMR experiments. A truncated ghrelin composed of residues 1-5 shows some rigidity between octanoyl-Ser3 and Phe4 but overall adopts a random coil conformation as well.

B. Molecular Dynamics (MD) Simulations

1. Structure in aqueous solution

The full length human ghrelin peptide in water was subjected to simulated annealing MD (SAMD) followed by 10 ns constant temperature MD at pH7 in order to investigate, if ghrelin is able to fold into a unique structure stable in the time frame accessible to MD simulations (Beevers and Kukol, 2006a). 20 random starting structures of an extended conformation were subjected to SAMD reducing the temperature from 450 K to 273 K over a period of 2 ns. The resulting structures were analysed for clusters of similar structures, because the independent folding of different random starting structures into a similar structure is indicative of a unique energy minimum, which may be prevalent in solution. One cluster of 13 structures was found, the members of which share a similar fold. The representative structure of this cluster (fig. 2A), which was most similar to all other members of the cluster, has no clearly defined fold apart from a bulge at residues His9 to Gln13 indicating the onset of α -helix formation and a short loop consisting of residues Ser18 to Lys20. A subsequent 10 ns MD simulation at constant temperature of 298K, showed that the

SAMD structure had not achieved its equilibrium fold, but underwent further structural changes. The endpoint structure of the 10 ns simulation after energy minimisation (fig. 2B) showed a clear formation of a short α -helix from Pro7 to Glu13 and the formation of a hairpin structure with Glu17 to Lys20 in the bending region. These structural elements formed after 6 ns in the constant temperature simulation and remained stable throughout the last 4 ns.

2. Lipid binding properties of ghrelin

Since the GHSR is a transmembrane protein ghrelin would naturally approach the vicinity of the cellular lipid membrane when binding to the GHSR. Furthermore, the hydrophobic octanoyl chain might either function as a lipid anchor increasing the local ghrelin concentration at the membrane surface or alternatively participate in direct binding to the GHSR. In order to investigate these questions and the stability of the structure obtained a 15 ns MD simulation of ghrelin in the presence of a lipid bilayer in water was performed (Beever and Kukol, 2006a). The system for simulation consisted of 128 dimyristoyl phosphatidylcholine (DMPC) molecules, 7184 water molecules and five chloride ions neutralising the +5 positive charge of the peptide. The structure in figure 2B was placed in the lipid bilayer system in such a position that the octanoyl chain points to the lipid bilayer in order to facilitate potential insertion in the lipid bilayer (figure 3). The progress of the simulation is shown in figure 4 in terms of the root mean square deviation (RMSD) of the backbone coordinates with respect to the starting structure. The RMSD is a measure of overall structural difference to a reference structure (chosen here at $t=0$) with values up to 0.5 nm backbone RMSD being indicative of normal structural fluctuations, while values above 1 nm would indicate a conformational change. The characteristic fold of the

structure did not change significantly throughout the simulation with an RMSD of around 0.25 nm after 15 ns. This is also apparent from the visual appearance of the snapshots shown at various times in figure 4. However, the orientation of the peptide with respect to the lipid bilayer changed during the simulation. During the first 3 ns the peptide approached the lipid bilayer and remained in its initial orientation. After about 5ns the peptide started to reorient such that the N- and C-terminus pointed away from the lipid bilayer towards the water phase, while the loop formed a close contact with the lipid bilayer headgroups. In particular the residues Arg14, Lys16, Glu17 and Ser18 were in close contact with the lipid headgroups. Most notably, the octanoyl side chain of Ser3 pointed into the water phase, although it was able to make contact with the lipid headgroups during the first 2 ns of the simulation.

In summary, the MD simulation studies revealed a stable fold of the ghrelin peptide in solution, which remained unchanged during a further 15 ns simulation in a lipid bilayer/water system. The octanoyl chain did not insert into the lipid bilayer, but a positively charged loop structure interacted with the zwitterionic lipid headgroups.

IV. Discussion

Contrary to the NMR experiments the MD simulations revealed a stable fold of the 28-residue ghrelin peptide, a loop structure with a short stretch of α -helix. Possible reasons for these differences are that the MD simulation has revealed only one of several conformations, which is stable over a time range of 25 ns but is able to adopt other conformations, which are, however, short-lived compared to the millisecond time resolution of standard ^1H -NMR spectroscopy. One should also take into account the pH difference between the NMR experiment at acidic pH and the MD simulation at neutral pH. At acidic pH the number of charges of ghrelin increases from +5 to +9

resulting in significant electrostatic repulsion between residues considering the small size of the peptide. The increased positive charge could very well prevent ghrelin from adopting a defined conformation. It is perhaps no surprise that some rigidity has been found in a truncated peptide composed of residues 1-5, which would only acquire a +1 charge at acidic conditions due to its protonated N-terminus. While ghrelin is secreted from the stomach under acidic conditions, physiologically more relevant is the conformation at neutral pH, when it approaches the lipid membrane bound receptors, which are expressed mainly in the hypothalamus and the pituitary (Howard et al., 1996).

The structural features of the octanoyl-serine residue in ghrelin are reminiscent of a detergent molecule composed of a hydrophilic headgroup and a hydrophobic tail. One might expect the hydrophobic octanoyl group to act as a lipid anchor attaching ghrelin to the lipid membrane thus facilitating binding to the receptor. However, the results of the MD simulation in the presence of a lipid bilayer lead to the hypothesis that the GHSR binding site consists of a hydrophobic pocket of a size, which can accommodate the octanoyl chain (Beevers and Kukol, 2006a). This postulate is in line with the minimal structural requirements of ghrelin necessary to elicit GHSR response, which consist of the first five residues including the octanoyl group (Bednarek et al., 2000). This hypothesis has been recently strengthened by the construction of a human GHSR model based on local homology modelling (Pedretti et al., 2006). This model has been subjected to computational docking studies using a tetrapeptide consisting of the first four residues of ghrelin including the octanoyl chain. The docking studies revealed that the octanoyl chain is accommodated in an hydrophobic pocket lined by extracellular loop 2 and shows interaction with residues Pro192, Trp193, Pro200, Ala204 and Val205 (Pedretti et al., 2006). For size

comparison the ghrelin model shown in figure 2B has been placed tentatively on a surface model of the GHSR coordinates obtained from Vistoli (Pedretti et al., 2006) in figure 5. Although no docking or MD simulation has been performed, the possibility of the interaction between the C-terminus of ghrelin and the receptor exists, facilitated by the hairpin-like structure.

V. Conclusions

The subject of the 3-dimensional structure of ghrelin is hampered by a lack of experimental and computational investigations. Based on computational modelling it may be concluded that ghrelin has a defined structure at neutral pH in aqueous solution and furthermore that the octanoyl side-chain does make direct contact to the GHSR upon binding rather than function as a mere lipid anchor. The emergence of GHSR homology models, which can be used for computational docking, is clearly an advantage for understanding of the receptor bound ghrelin structure. However, homology models need to be based on experimentally known structures, which are scarce in the area of G-protein coupled receptors and transmembrane proteins in general. Most homology models of G-protein coupled receptors are based on the crystal structure of bovine rhodopsin (Palczewski et al., 2000), although current approaches predict the structure of fragments based on local homology to various proteins (Ginalska et al., 2005). A recent docking model of a truncated ghrelin to GHSR has confirmed the postulated role of the octanoyl side chain (Pedretti et al., 2006). Further studies using all atom MD simulations of the GHSR/ghrelin in a lipid bilayer could expand the existing docking model and provide further insight into the GHSR-ghrelin interaction.

However, modelling results need to be backed up by experiments; NMR investigations taking into account the pH dependence of the charge state of ghrelin and measurements at lower temperatures in order to reduce structural fluctuations are clearly necessary in order to advance the field and ultimately contribute to development of novel drugs for the benefit of human health.

Acknowledgements

Figures of molecular structures have been prepared with VMD (Humphrey et al., 1996) using the molecular surface plug in (Varshney et al., 1994).

The author is grateful to Prof. G. Vistoli for providing the coordinates of the GHSR model, to Dr. A. Beevers for carefully reading the manuscript and to the Biotechnology and Biological Sciences Research Council, UK (grant no. 88/B19450) for ongoing support.

References:

Bednarek, M.A., S.D. Feighner, S.S. Pong, K.K. McKee, D.L. Hreniuk, M.V. Silva, V.A. Warren, A.D. Howard, L.H. Van Der Ploeg, and J.V. Heck (2000). Structure-function studies on the new growth hormone-releasing peptide, ghrelin: minimal sequence of ghrelin necessary for activation of growth hormone secretagogue receptor 1a. *J Med Chem* **43**(23), 4370-4376.

Beevers, A.J., and A. Kukol (2006a). Conformational flexibility of the peptide hormone ghrelin in solution and lipid membrane bound: a molecular dynamics study. *J Biomol Struct Dyn* **23**(4), 357-364.

Beevers, A.J., and A. Kukol (2006b). Systematic molecular dynamics searching in a lipid bilayer: Application to the glycoporphin A and oncogenic ErbB-2 transmembrane domains. *J Mol Graph Model* **in press**, available online.

Beevers, A.J., and A. Kukol (2006c). The transmembrane domain of the oncogenic mutant ErbB-2 receptor: A structure obtained from site-specific infrared dichroism and molecular dynamics. *J. Mol. Biol.* **in press**.

- Duan, Y., and P.A. Kollman (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**(5389), 740-744.
- Ginalski, K., N.V. Grishin, A. Godzik, and L. Rychlewski (2005). Practical lessons from protein structure prediction. *Nucleic Acids Res* **33**(6), 1874-1891.
- Guillemin, R., P. Brazean, P. Bohlen, F. Esch, N. Ling, and W.B. Wehrenberg (1982). Growth hormone-releasing factor from a human pancreatic tumor that caused acromegaly. *Science* **218**, 585-587.
- Howard, A.D., S.D. Feighner, D.F. Cully, J.P. Arena, P.A. Liberato, C.I. Rosenblum, M. Hamelin, D.L. Hreniuk, O.C. Palyha, J. Anderson, P.S. Pares, C. Diaz, M. Chou, K.K. Liu, K.K. McKee, S.S. Pong, L.Y. Chaung, A. Elbrecht, M. Dashkevich, R. Heavens, M. Rigby, D.J. Sirinathsinghji, D.C. Dean, D.G. Melillo, A.A. Patchett, R. Nargund, P.R. Griffin, J.A. DeMartino, S.K. Gupta, J.M. Schaeffer, R.G. Smith, and L.H. Van der Ploeg (1996). A receptor in pituitary and hypothalamus that functions in growth hormone release. *Science* **273**(5277), 974-977.
- Humphrey, W., A. Dalke, and K. Schulten (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**(1), 33-38, 27-38.
- Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671-680.
- Kojima, M., H. Hosoda, Y. Date, M. Nakazato, H. Matsuo, and K. Kangawa (1999). Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**(6762), 656-660.
- Kojima, M., and K. Kangawa (2005). Ghrelin: structure and function. *Physiol Rev* **85**(2), 495-522.
- Korbonits, M., A.P. Goldstone, M. Gueorguiev, and A.B. Grossman (2004). Ghrelin--a hormone with multiple functions. *Front Neuroendocrinol* **25**(1), 27-68.
- Palczewski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**(5480), 739-745.
- Pedretti, A., M. Villa, M. Pallavicini, E. Valoti, and G. Vistoli (2006). Construction of human ghrelin receptor (hGHS-R1a) model using a fragmental prediction approach and validation through docking analysis. *J Med Chem* **49**(11), 3077-3085.
- Silva Elipse, M.V., M.A. Bednarek, and Y.D. Gao (2001). ¹H NMR structural analysis of human ghrelin and its six truncated analogs. *Biopolymers* **59**(7), 489-501.
- van Gunsteren, W.F., X. Daura, and A.E. Mark (1999). GROMOS force field. *Encyclopedia Comput. Chem.* **2**, 1211-1216.

Varshney, A., F.P. Brooks, and M.V. Wright (1994). Linearly Scalable Computation of Smooth Molecular Surfaces. *IEEE Computer Graphics and Applications* **14**, 19-25.

Wuthrich, K. (1986). "NMR of proteins and nucleic acids". Wiley, New York.

Figure legends

Figure 1: The 28-residue aminoacid sequence of human ghrelin including the octanoyl-group attached to Ser3.

Figure 2: Peptide backbone structure of ghrelin with the octanoyl chain highlighted. (A) Representative structure of a cluster of similar structures obtained through 2 ns SAMD simulation reducing the temperature from 450 K to 273 K. (B) Structure after 10 ns MD simulation in water at 298 K followed by energy minimisation.

Figure 3: The system used for MD simulation of ghrelin in the vicinity of a lipid bilayer. The system shown consists of ghrelin, five chloride ions, 128 DMPC molecules and 7184 water molecules.

Figure 4: Root mean square deviation (RMSD) of the peptide backbone with respect to the starting structure at $t=0$. The arrows indicate the time points of snapshots displayed above the curve. Water molecules are not shown.

Figure 5: Superposition of the ghrelin model on the GHSR homology model. Residues of the GHSR, which have been proposed to interact with the octanoyl-chain are highlighted in light grey, while residues which have been proposed to interact with residues 1-5 of ghrelin are shown in dark grey. The interacting residues are based on a docking study with truncated ghrelin 1-4 (Pedretti et al., 2006).

Fig. 2:



Fig. 3:

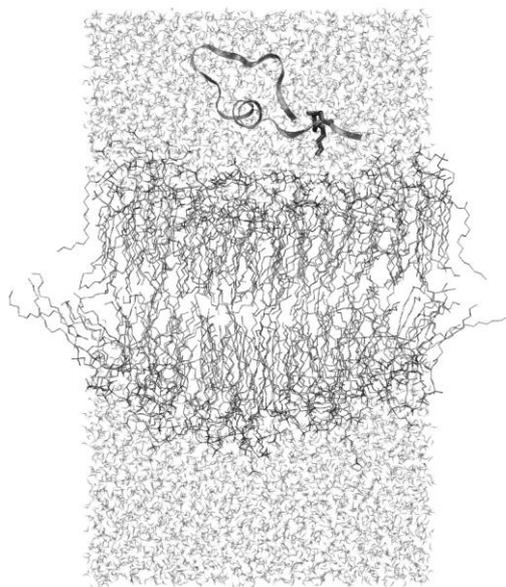


Figure 4:

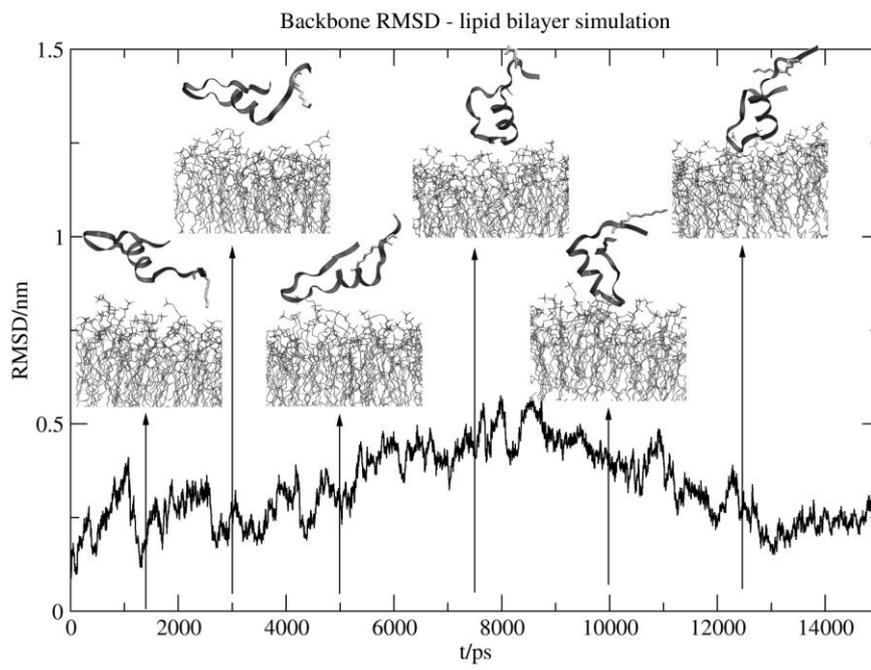


Figure 5:

