Molecular Dynamics Simulations of Flaviviridae Family fusion peptides

Exploring conformational space of Flaviviridae family fusion peptides within MD simulations framework - force field and methodology assessment

A thesis report submitted towards the partial fulfillment of

BS-MS Dual Degree Programme

by

Sadhana Panzade

Under the guidance of Dr. Chandra Verma

&

Indian Institute of Science Education and Research Pune

Bioinformatics
Institute
A STAR

Dr. Chandra Verma
Division Head
Biomolecular Modelling & Design

Bioinformatics Institute
T +65 6478 8298 F +65 6478 9048
www.bii.a-star.edu.sg
Email: Chandra@bii.a-star.edu.sg
30 Biopolis Street
#07-01 Matrix Singapore 138671
Tel: +65 6478 8273
HP:+65 9479 7303

Certificate

This is to certify that this dissertation entitled “Molecular Dynamics Simulations of Flaviviridae family fusion peptides” towards partial fulfillment of BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune represents original research carried out by Sadhana Panzade at Bioinformatics Institute A*STAR Singapore under the supervision of Dr. Chandra Verma, Bioinformatics Institute A*STAR Singapore during the academic year 2014-2015.

Chandra Verma

Signature of the Supervisor Date: 24 March 2015

(Name of the Supervisor)

Chandra Verma

IISER PUNE

Declaration

I hereby declare that the matter embodied in the report entitled "Molecular Dynamics Simulations of Flaviviridae family fusion peptides" are the results of the investigations carried out by me at Bioinformatics Institute A*STAR Singapore under the supervision of Dr. Chandra Verma and the same has not been submitted elsewhere for any other degree.

Signature of the Student

Date: 25/03/2015

(Name of the Student)

Sadhana Panzade

Contents

Acknowledgements

I thank my supervisor Dr. Chandra Verma, Bioinformatics Institute A*STAR Singapore, who motivated me towards rational inquiry of the scientific study I have carried out at Bioinformatics Institute A*STAR Singapore as part of my Master's research project. His expertise and in depth understanding of the field constantly proved to be a great guidance for me in conducting my research. I am indebted to him for having the confidence in me for such an endeavor.

I am grateful for getting the opportunity to work with Dr. Jan Marzinek, Bioinformatics Institute A*STAR Singapore, who made immense efforts to help me understand the computational aspects of biology. I am thankful of him for training me in conducting research at such a level.

I am thankful to Professor Paul Matsudaira, Department of Biological Science, National University of Singapore, without whom I would not been able to learn as much as I did and get to work with some of the best researchers in the field of computational biology. I am grateful to him for making me realize the potential in me to pursue research as a career later in my life.

I thank Dr. Madhusudhan M.S Associate Professor; IISER Pune who kept me motivated towards my Master's thesis project with his constant feedback and played an important part in making this possible.

I thank MOM Singapore for the grant funding: MOE AcRF Tier 3 (MOE2012-T3-1-008) which played an important role in providing computational resources to conduct the study. I thank National University of Singapore, Bioinformatics Institute A*STAR Singapore, IISER Pune which is where it became all real and finally I express my ample gratitude towards everyone who supported me to make this research experience a success.

Abstract

This study comprises of conformational space explored within molecular dynamics simulations framework for flaviviridae family fusion peptides, Dengue virus serotype 2 fusion peptide (DENV2), Dengue virus serotype 3 fusion peptide (DENV3) and tick – borne encephalitis virus (TBEV) in explicit solvents at 300K with four different force fields, Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA using MD simulation techniques such as classical molecular dynamics simulations, replica exchange molecular dynamics simulations, simulated or temperature annealing that provide enhanced sampling. Experimental limitations such as failure to capture physiologically relevant time scales in order to study biological problems motivated this computational study where the protein/bimolecular dynamics can be feasibly accounted at such smaller time scales. The total simulation time was 33.6 μs. Different force fields were seen to have propensities towards different secondary structures. Amber99SB*-ILDN-Q simulations formed structures close of Dengue virus serotype 2 fusion peptide crystal structure in case of replica exchange molecular dynamics simulations method, whereas in case of methods, classic molecular dynamics simulations and temperature annealing/simulated annealing, not effective enough sampling Amber99SB*-ILDN-Q did not predict structures close to the crystal structure. Charmm22/CMAP, Gromos54a7 and OPLS-AA on the other hand were biased towards secondary structures. This study comprehends the use of sampling methods in combination with force fields in predicting structures for a small unstructured peptide that is highly dynamic in the solvent studied.

Introduction

Flaviviridae family viruses cause Dengue fever, Dengue hemorrahagic fever and dengue shock syndrome. A study conducted by World Health Organization (WHO) estimated upto 3.6 billion people at the risk of this disease (WHO, 2012; N. E. A Murray 2013). Flaviviruses are transmitted via mosquitoes. Some prominent flaviviruses are: yellow fever virus (YFV), Japanese encephalitis virus (JEV), West Nile virus (WNV), and four dengue viruses (DENVs: DENV1, DENV2, DENV3, DENV-4), tick-borne encephalitis viruses (TBEVs) are transmitted by ticks. The structural assembly of a flavivirus: flavivirus genome comprises of a positive-stranded RNA which encodes proteins, capsid, precursors of membrane (prM) proteins, envelope proteins (E) and non- structural proteins (NS)- (seven non-structural proteins-NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5). E proteins are involved in binding the receptors and fusion with the host cell membrane hence antibodies are induced against E proteins to study the neutralizing effects on the viruses (Lindenbach BD et al 2013).E proteins are involved in fusion process during viral infection process. These proteins undergo conformational changes during viral maturation and fusion stages of the virus life (Y.Modis et al 2004; Y Zhang et al 2004). E proteins are formed by three domains. Domain I makes the insertion loop (Y Modis et al 2004; Y Modis et al 2003; F.A. Rey et al 1995). Domain II makes the hydrophobic sequences that are recognized as fusion peptides and are responsible for insertion of the E protein trimers after rearrangement (dimer to trimer transition). These fusion peptides of the flaviviruses are conserved (Y Modis et al 2004; S. Bressanelli et al 2004; Y Modis et al 2003; S.L. Allison et al 2001; S.L.Allison et al 1999). Domain III connects to domain I initiating cellular receptor binding (J Hung et al 2004). E proteins lay flat on the virus surface as dimers. They experience rearrangement during low pH conditions in late endosome during fusion process exposing the fusion peptide. The fusion peptide exposed comprises of a tip containing highly conserved residues Trp101 and Phe108. E protein is most important in receptor recognition and attachment to the host cell membrane. It is seen interacting with various molecules (S. Bressanelli et al 2004; M. Liao et al 2005; D. Kato et al 2010; C.O.

Nicholson et al 2011). It can serve as a potential drug target. “Fusion peptides” of flaviviruses are motifs that differ in terms of single residues in a 15 residue sequence. For example: DENV2 and DENV4 (^{98}DRGWGNGCGLFGKGG^{112}) PDB entry 1OAN, DENV-1, DENV-3, as well as West Nile Virus (^{98}DRGWGNGCGLFGKGS^{112}) PDB entry 1UZG, Tick borne-encephalitis virus (^{98}DRGWGNHCGLFGKGS^{112}) PDB entry 1URZ. Previously it has been shown by nuclear magnetic resonance study (NMR) that a single mutation from tryptophan to alanine (W101 to A) can abolish the membrane fusion process (Melo MN et al 2009). Experimental findings have revealed that anti-peptide molecules recognize the fusion peptide motif of the flaviviruses from non-flaviviruses (Pattnaik P et al 2006). Fusion peptide recognition has lead to effective flavivirus vaccines against yellow fever virus, Japanese encephalitis virus and tick-born encephalitis virus infections have been developed, but WNV and DENV vaccines are still being worked on for human use (Baratte AD et al 2001; Patel JR et al 2009). Dengue viruses have four serotypes: (DENV1, DENV2, DENV3, DENV4) Infection from individual virus can be prevented (Sabin AB 1952). Studies revealed that cross-protection against dengue virus serotypes occurs over a short period of time causing dengue fever but the risk is still higher in causing dengue hemorrhagic fever even after primary infection (Halstead SB 1982). Hence it is a worth the effort to develop new vaccines for DENV. As it is experimentally established fusion peptide plays a crucial role during infection it was of high interest to study the dynamics of it using computational tools as experimental techniques limit the scope of studying bimolecular activities at atomic resolution and at physiologically relevant time scales. To capture the dynamic details of fusion peptides of flaviviruses were studied (DENV2, DENV3, TBEV) using molecular dynamics simulations (MD), a technique to study the evolution of fusion peptide conformations in time. Molecular dynamics simulation (MD) is a computational technique wherein the basic idea is to assign initial conditions (positions and velocities) to particles of the system (in this case the atoms of fusion peptide) and interaction of potentials (represented by mathematical formulations called “force fields”) in order to calculate forces among these particles. The evolution of the system in time is solved within classical mechanics framework specifically Newton's second law of motion:

$F = ma$ where F is the force and m and a are the mass and acceleration of the particle. Within MD frame work techniques such as simulated/temperature annealing and replica exchange molecular dynamics simulations are widely used to sample the conformational space of the protein of interest revealing the secondary structures that are attained during molecular dynamics simulations. Simulated/temperature annealing (TA/SA) is an enhanced sampling molecular dynamics simulation method to study conformational space for protein folding and unfolding (Galzitskaya et al 2000) where the system is linearly heated to a high temperature in order to overcome energy barriers followed by slow cooling to acquire a low energy states in order to reach minimum energy state confomations (Kirkpatrick et al 1983). Crystallographic and NMR studies widely use simulated annealing molecular dynamics simulations method. It is also used in protein modeling studies (Brunger et al 1999; Brunger and Adams, 2002; Li et al 1997; Möglich et al 2005). A main limitation of simulated annealing is that conformations of peptides or proteins may come across minimum energy traps and this is highly affected by the cooling process during annealing. Unless the simulations are longer the method does not assure any global minimum energy states. (Larrhoven and Aarts 1987). Replica exchange molecular dynamics simulation (REMD) is another enhanced conformational sampling MD method that provides sampling of conformation (Fukunishi et al, 2002; Mitsutake et al 2001; Nymeyer and Garcia, 2003; Okabe et al, 2001; Sanbonmatsuand Garcia, 2002, Sugita and Okamoto, 1999; Yoshida et al, 2005, Zhou 2003, Kannan and Zacharias, 2007a). REMD simulates several copies of the same system evolving independently at different temperatures. Various values of temperature are assigned for each replica so that the replicas at different temperatures exchange coordinates among each other in order to explore conformational space. REMD and TA/SA explore conformations from low to high energy states crossing energy barriers. They are being employed to study protein folding and unfolding. This report present a study of conformational space explored within molecular dynamics simulations framework for flaviviridae family fusion peptides, Dengue virus serotype 2 fusion peptide (DENV2), Dengue virus serotype 3 fusion peptide (DENV3) and tick –borne encephalitis virus (TBEV) in explicit solvents at 300K with four different widely used force field.

Amber ff99SB*-ILDN-Q (Sorin, E. J. et al 2005; Best, R. B et al 2009; Hornak, V et al 2006; Lindorff-Larsen, K et al 2010, Lindorff Larsen, K et al 2012), Charmm22/CMAP **(**MacKerell et al 1998; MacKerell, 2001; MacKerell, A. D. et al 2004), Gromos54a7 (Schmid N et al 2011) and OPLS-AA(Jorgensen, W. L et al 1996) using MD simulation techniques classical molecular dynamics simulations, replica exchange molecular dynamics simulations, simulated or temperature annealing. The total simulation time was 33.6 μs. Force fields Amber99SB*-ILDNQ, Charmm22/CMAP, Gromos54a7, OPLS-AA were employed to study the DENV2 conformational space using three methods MD, SA/TA and REMD. Further it was used as a platform to study fusion peptides DENV3 and TBEV and the influence of force fields and simulation methods.

Simulation Methodology

The starting structure for molecular dynamics simulations was a 15 amino acid sequence of the fusion peptide (FP) corresponding to DENV-2 & DENV-4 (protein data bank* entry: 1OAN, with the sequence: $^{(98}$ DRGWGGGCGLFGKGG1^{12}) DENV-1, DENV-3, WNV (^{98}DRGWGGGCGLFGKGS^{112}) (PDB entry 1UZG) and Tick-borne encephalitis virus (^{98} DRGWGGHCGLFGKGS^{112}) (PDB entry 1URZ)

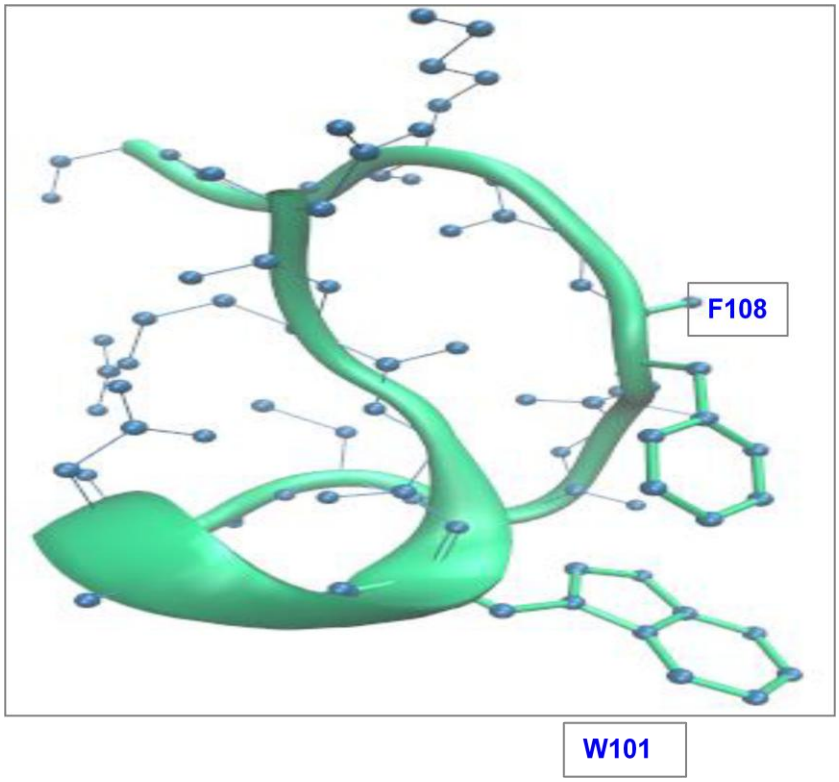


Figure:1a) Dengue virus serotype 2 fusion peptide (atomic (blue) & cartoon (green) representation). Aromatic residues exposed at viral infection (W101-Tryptophan & F108-Phenylalanine)

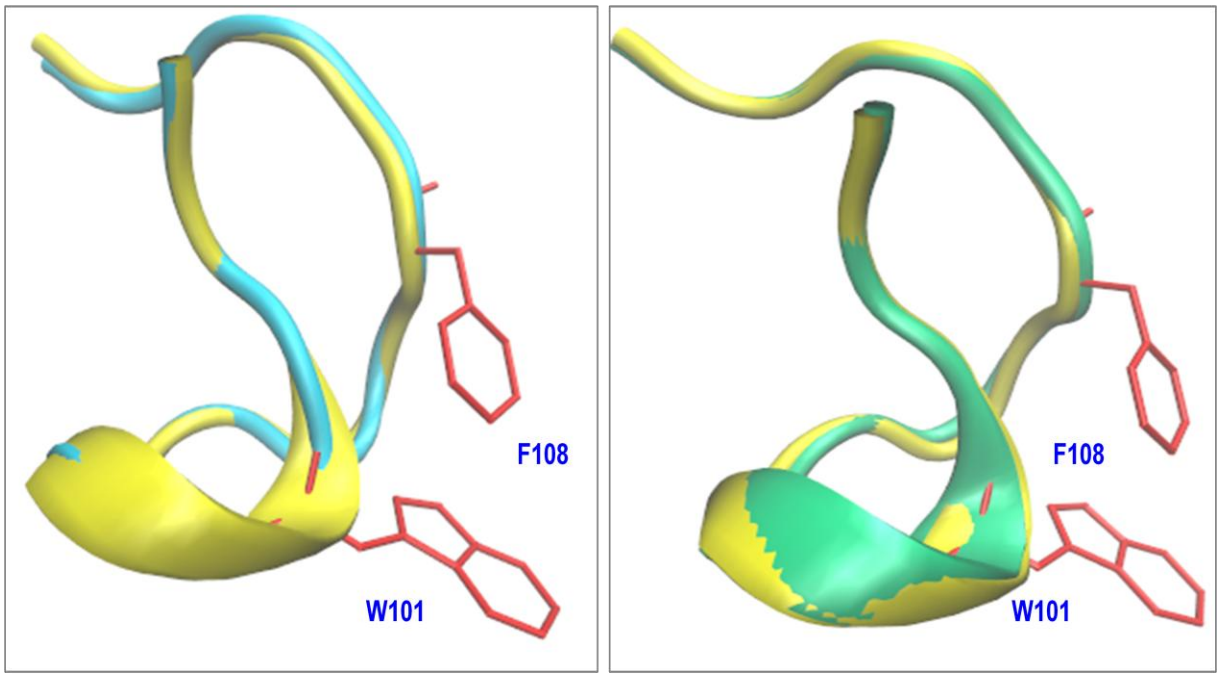


Figure:1b) Dengue virus serotype 3 fusion peptide and tick borne encephalitis virus structural alignment with respect to dengue virus serotype 2. Aromatic residues exposed (W101-Tryptophan & F108-Phenylalanine)

The FP was placed in the centre of the cubic box (4.5x4.5x4.5 nm) and filled with approximately 3000 water molecules. In order to neutralize the overall system charge, one Cl^- ion was added. CHARMM22/CMAP and Amber99SB*-ILDN-Q, Gromos54a7, OPLS-AA force fields were employed in four sets of simulations. The TIP3P water model was used in Amber99SB*-ILDN-Q and Charmm22/CMAP while SPC in case Gromos54A7 and TIP4P for OPLS-AA were employed. All MD simulations were performed using GROMACS 4.6 (Hess, B et al 2008) package. With a time step of 2 fs, equations of motion were integrated through Verlet leapfrog algorithm. Bond lengths were constrained with the LINCS (Darden T et al 1993) algorithm. The cutoff distance was 1.4 nm for the short-range neighbor list and van der Waals interactions. The Particle Mesh Ewald method (Essman U et al 1995) was applied for the long-range electrostatic interactions with a 1.4-m real space cutoff. The velocity rescale thermostat with additional stochastic term (Bussi G et al 2007) and Berendsen barostat were used to maintain the temperature and pressure (1 bar). Initial velocities were set according to the Maxwell distribution. Periodic boundaries were used in all directions. All simulations were performed using Bioinformatics Institute (BII) cluster. REMD is a method that simulates several copies or replicas of the same system evolving independently at different temperatures (Sugita Y, Okamoto Y 1999).The replicas exchange coordinates at an interval based on Metropolis criterion (Metropolis N et al 1953):

$$P(i \rightarrow j) = min\ (1, exp\ [\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right) * (U_i - U_j)]$$

Where $(i \rightarrow j)$ is the exchange probability, k_B is Boltzmann's constant Ui and Uj are the potential energies of individual configurations of the replicas at the temperatures T_i and T_j. By exchanging coordinates between simulations replicas at different temperatures REMD method allows the improved sampling of the conformational space visiting simultaneously many local minima. To ensure a uniform exchange probability, the temperatures were chosen according to the relation: $T_i = T_0 \exp ic$ proposed by (Sugita Y, Okamoto Y 1999) where T_0 and c can be changed to provide the desired exchange ratio. Application of each force field employed 36 replicas at the following temperatures:

297, 297.00, 300.22, 303.47, 306.74, 310.04, 313.38, 316.73,320.12, 323.54, 326.98, 330.45, 333.96, 337.49, 341.05, 344.66, 348.23, 351.88, 355.57, 359.28, 362.91, 366.69, 370.53, 374.38, 378.25, 382.17, 386.11, 390.09, 394.10, 398.15, 402.23, 406.35, 410.50, 414.72, 418.94, 423.20, 427.51 K; where the temperature increments ranged from 3 to 4K. Initial configuration was minimized using steepest descent (SD) algorithm. Each replica prior to the production run was equilibrated using 1ns in NVT and 1ns in NPT ensemble respectively (where N is the number of atoms, V is volume, P is pressure and T is temperature). Coordinates between replicas were exchanged every 2 ps and the production run (NPT) was run corresponding to 100 ns at given temperature. In total each REMD run employed 3.6 μs. Simulated annealing simulations in the NPT ensemble were run for 1000 ns involving 100 cycles of 10 ns each. The system was linearly heated up from 300 to 450 K for 1ns followed by 4 ns of the simulation at 450 K. Subsequently, the3 ns of cooling down followed by equilibration of 2 ns at 300K were employed. The data analyzed corresponded to last 2 ns of each cycle at 300K.

Results

In order to assess the convergence of each simulation the following properties were averaged over five equal simulation time intervals for DENV-2 fusion peptide: Gyration radius (R_g) (Figure 2), end to end distance ($E_{to}E$) (Figure 3) and solvent accessible surface area (SASA) (Figure 4). The error bars in each figure corresponds to the standard deviation and was compared with the crystal structure values. Gyration radius is defined as the compactness of the system as following:

$$Radius\ of\ gyration = \sqrt{\frac{1}{N_i} * \sum_i (r_i - r_{cm})^2}$$

Where r_i - r_{cm} is the distance between atom *i* and the center of mass of the molecule. The SASA is calculated based on the algorithm proposed by (Eisenhaber F et al 1995). It can be observed that in case of REMD method for all force fields the calculated R_g average was the closest to the crystal structure value (0.59 nm). However in case of TA/SA all force fields overestimated the experimental R_g value whereas MD showed a varying trend. In case of the end to end distance for REMD Charmm22/CMAP it is observed that the value is higher than the crystal structure value.

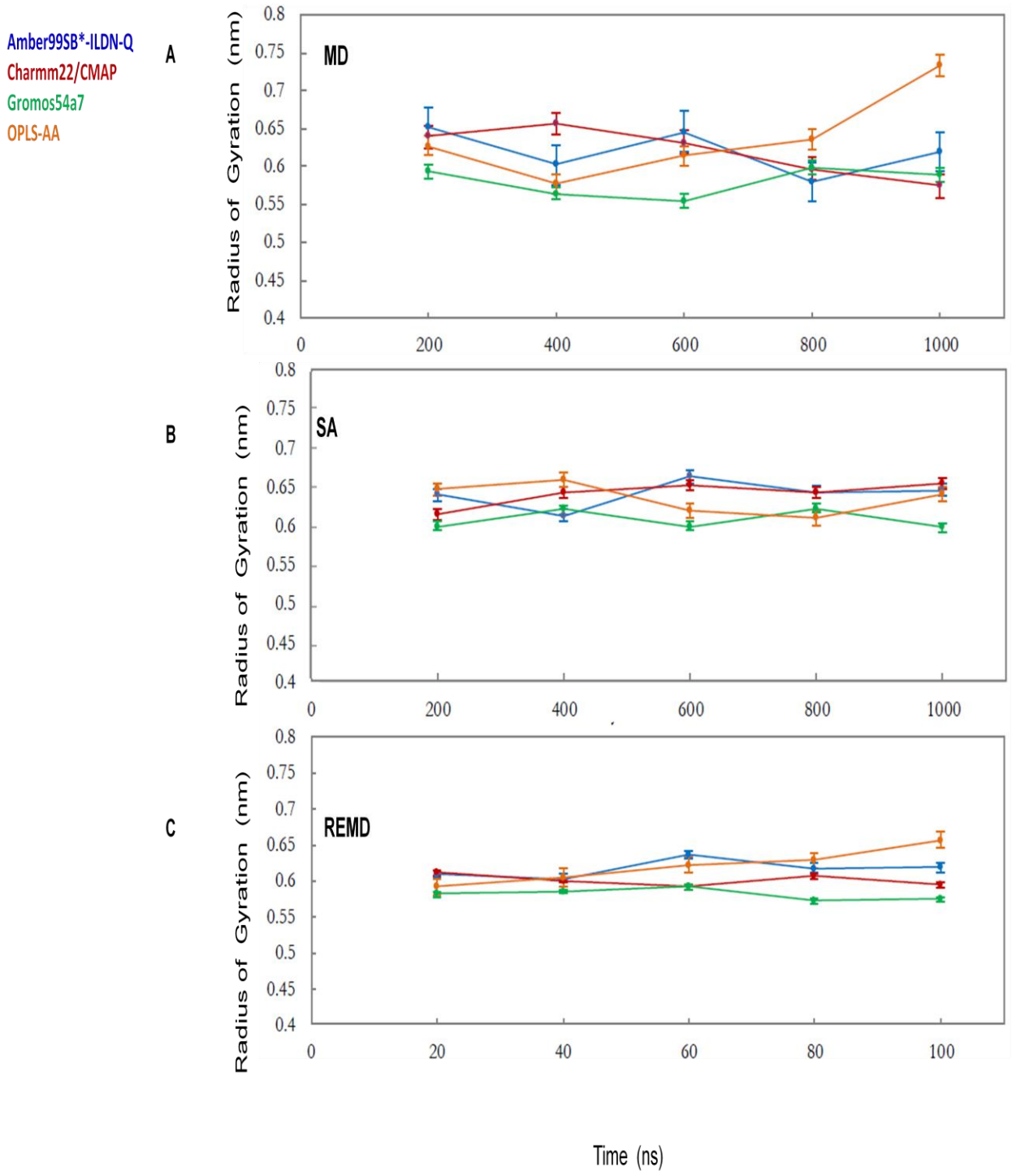


Figure: 2) Convergence analysis. The gyration radius obtained as an average obtained for MD, SA, REMD using force fields, Amber99SB*-ILDN-Q (blue), Charmm22/CMAP (red), Gromos54a7 (green) and OPLS-AA (orange).The gyration radius of DENV2 FP crystal structure value (0.59 nm).

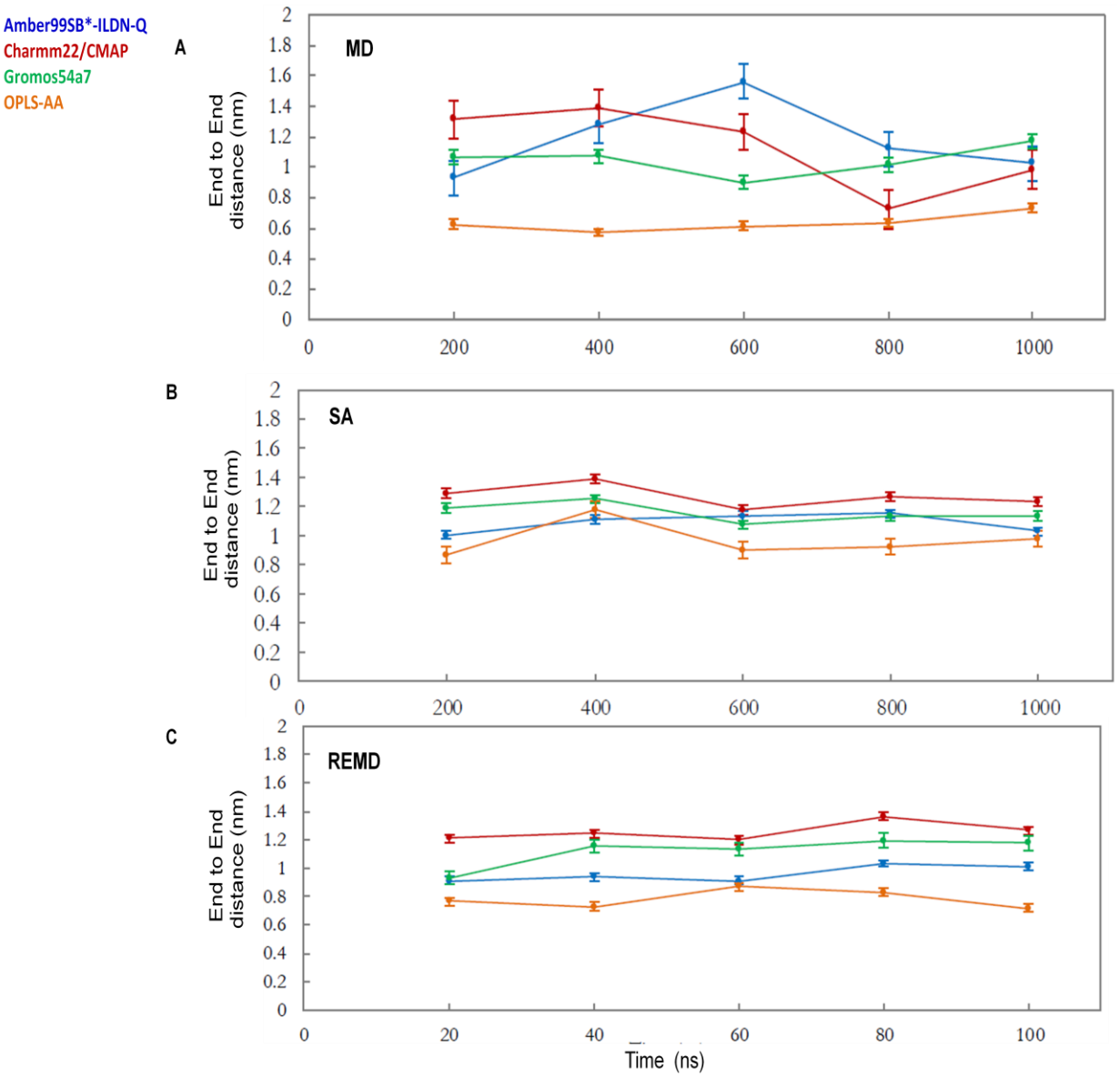


Figure: 3) Convergence analysis: End to End distance calculations is an average obtained from MD, SA, REMD for force fields Amber99SB*-ILDN-Q (blue), Charmm22/CMAP (red), Gromos54a7 (green) and OPLS-AA (orange). The end to end distance of DENV2 FP crystal structure value (0.83 nm).

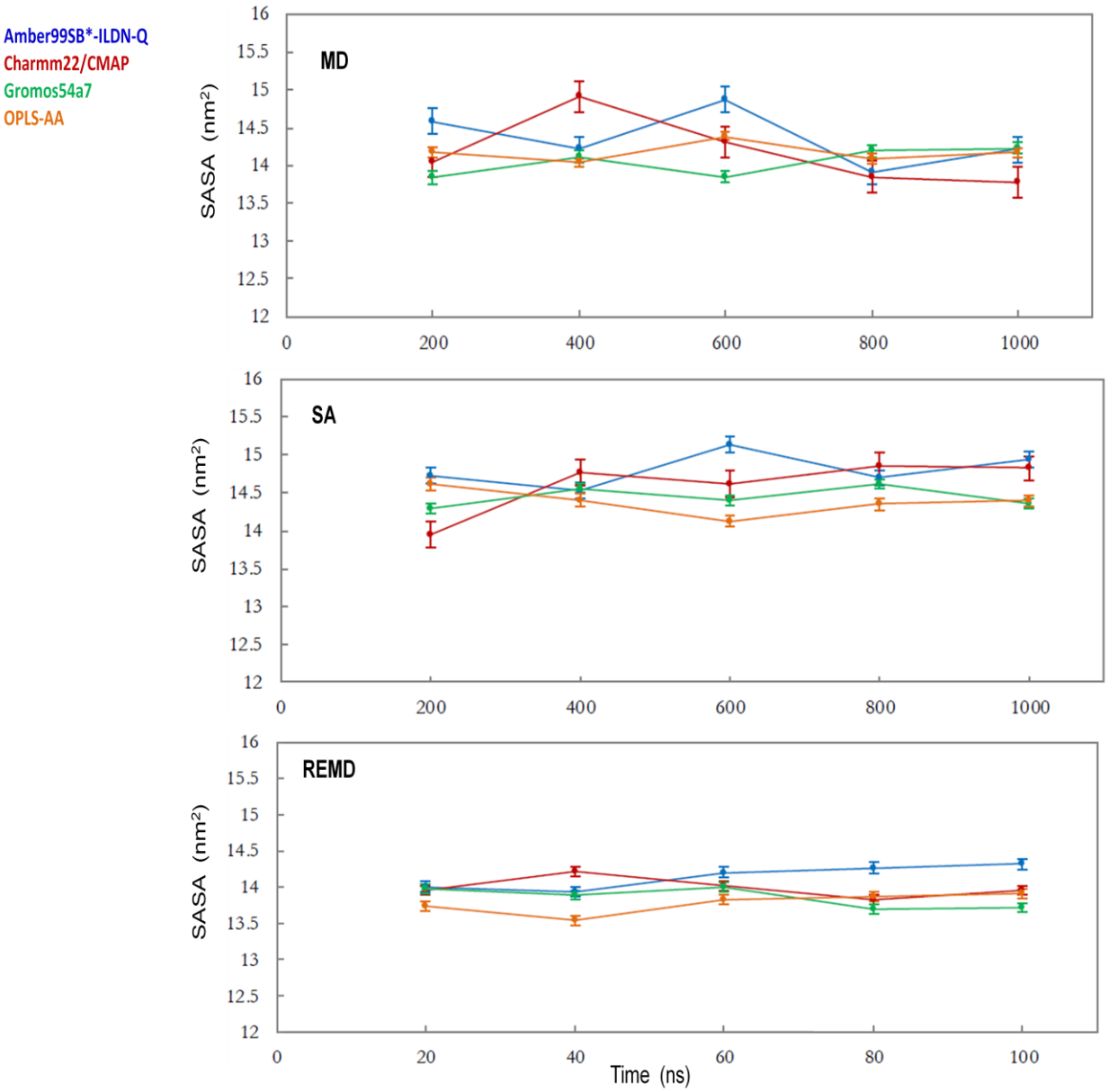
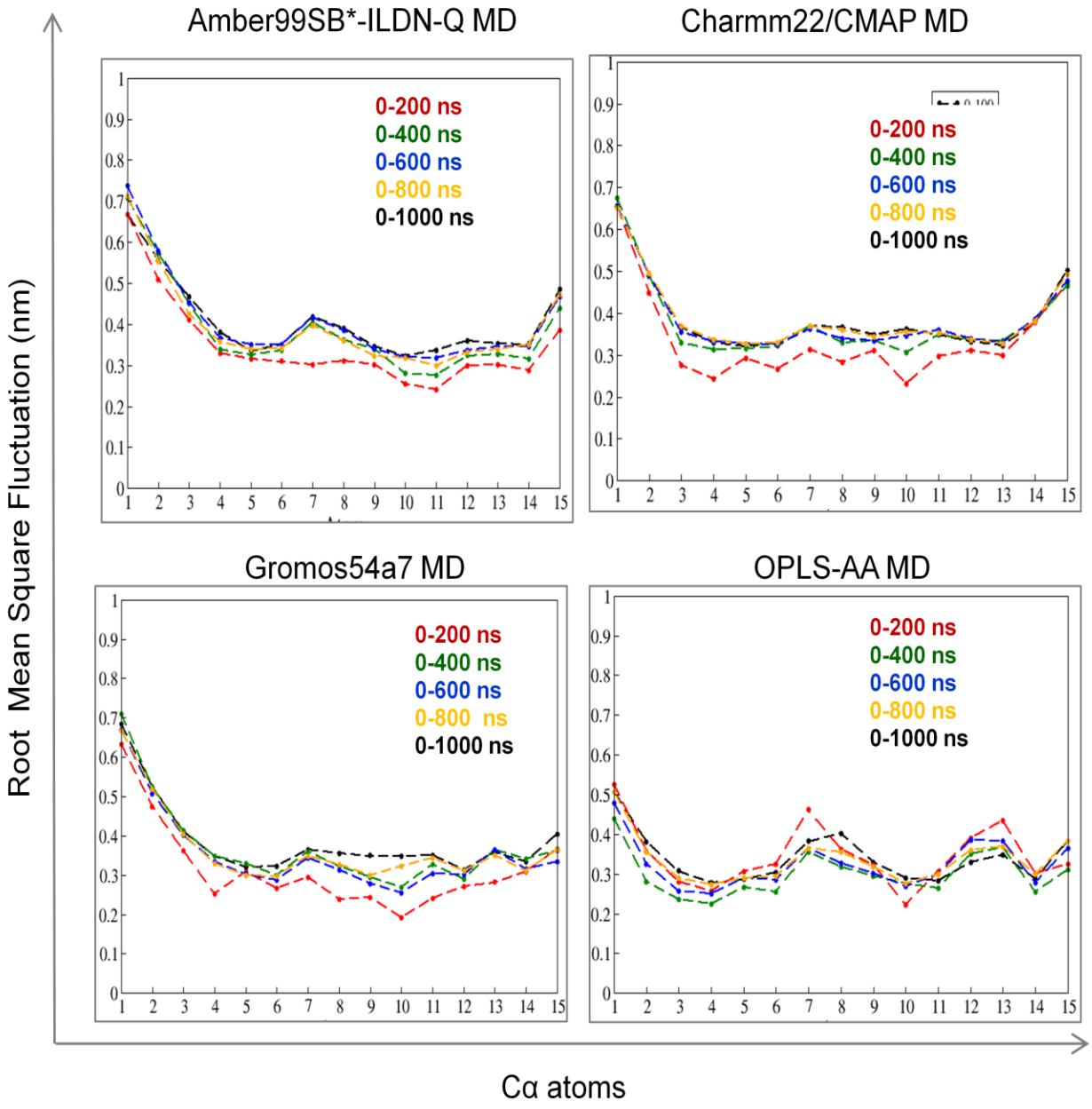


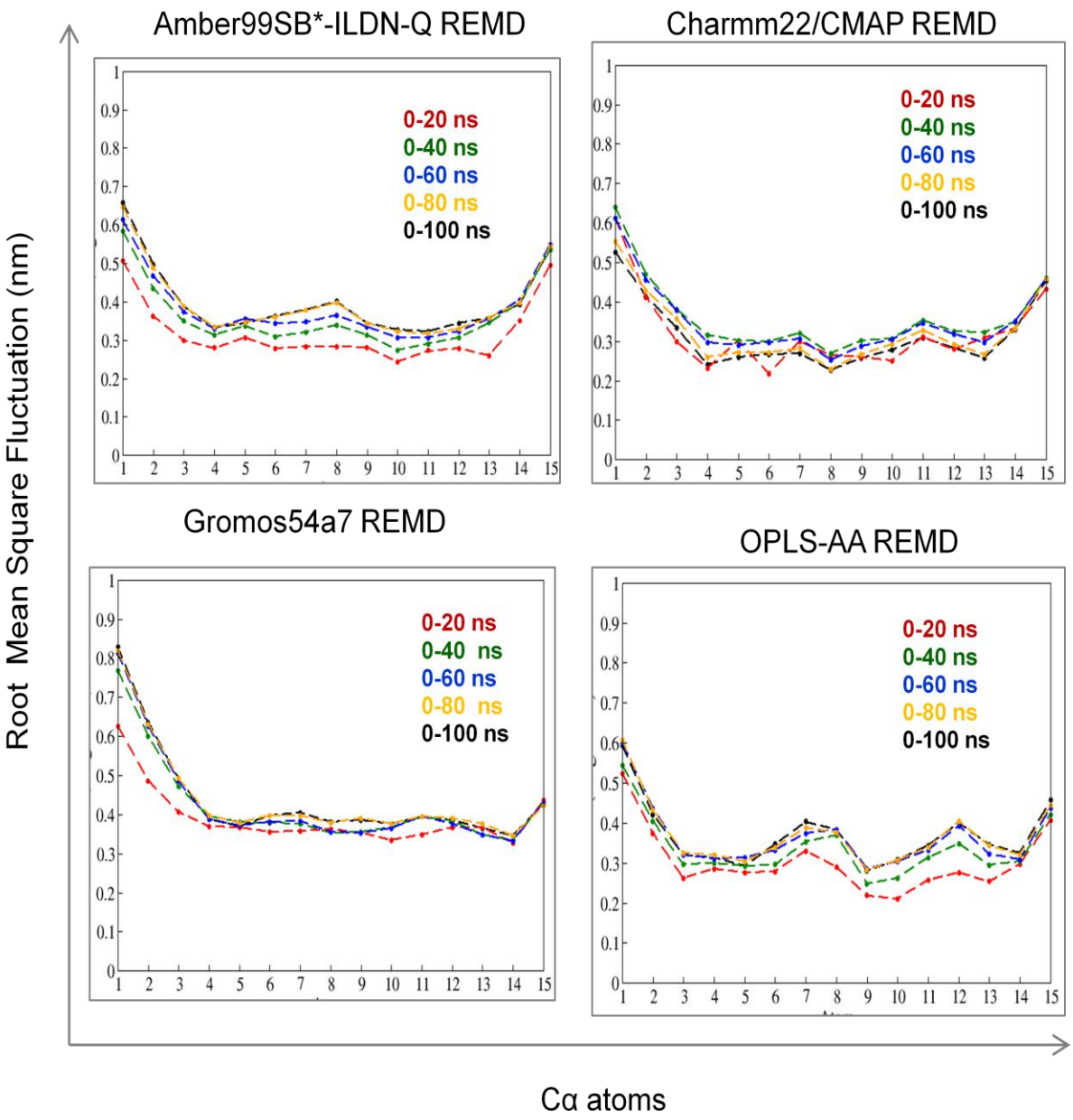
Figure: 4) Convergence analysis: Solvent accessible surface area calculations is an average obtained from MD, SA, REMD for force fields Amber99SB*-ILDN-Q (blue), Charmm22/CMAP (red), Gromos54a7 (green) and OPLS-AA (orange).The solvent accessible surface area for DENV2 FP crystal structure value (13.8 nm^2).

Capturing fluctuations due to each residue of the fusion peptide about its average position provided an insight into the flexibility of regions of the protein. The RMSF (Figure: 5) for C_α atoms was calculated using MD, TA/SA GROMACS tool (rmsf). RMSF assessment showed that the terminal regions are highly flexible, the reason being fusion peptide is dynamically changing conformations in time. MD showed increasing fluctuation as the simulation progressed for all force fields. Similarly in case of REMD, but the trend was quite different in TA/SA where the fluctuation was not pronounced at all. Root mean square fluctuation is calculated based on given equation between the position of C_α of a residue i and a reference position of residue.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j\,=\,1}^{T}\left(r_i(t_j) - ro_i\right)^2}$$

Where T is the time over which the average is calculated, ro_i is the reference position of the C_α of the residue *i*. This reference will be a time averaged position of the same residue *i*.


Amber99SB*-ILDN-Q MD
Charmm22/CMAP MD
Gromos54a7 MD
OPLS-AA MD
0-200 ns
0-400 ns
0-600 ns
0-800 ns
0-1000 ns
Root Mean Square Fluctuation (nm)
Cα atoms


Amber99SB*-ILDN-Q REMD
Charmm22/CMAP REMD
Gromos54a7 REMD
OPLS-AA REMD
0-20 ns
0-40 ns
0-60 ns
0-80 ns
0-100 ns
Root Mean Square Fluctuation (nm)
Cα atoms

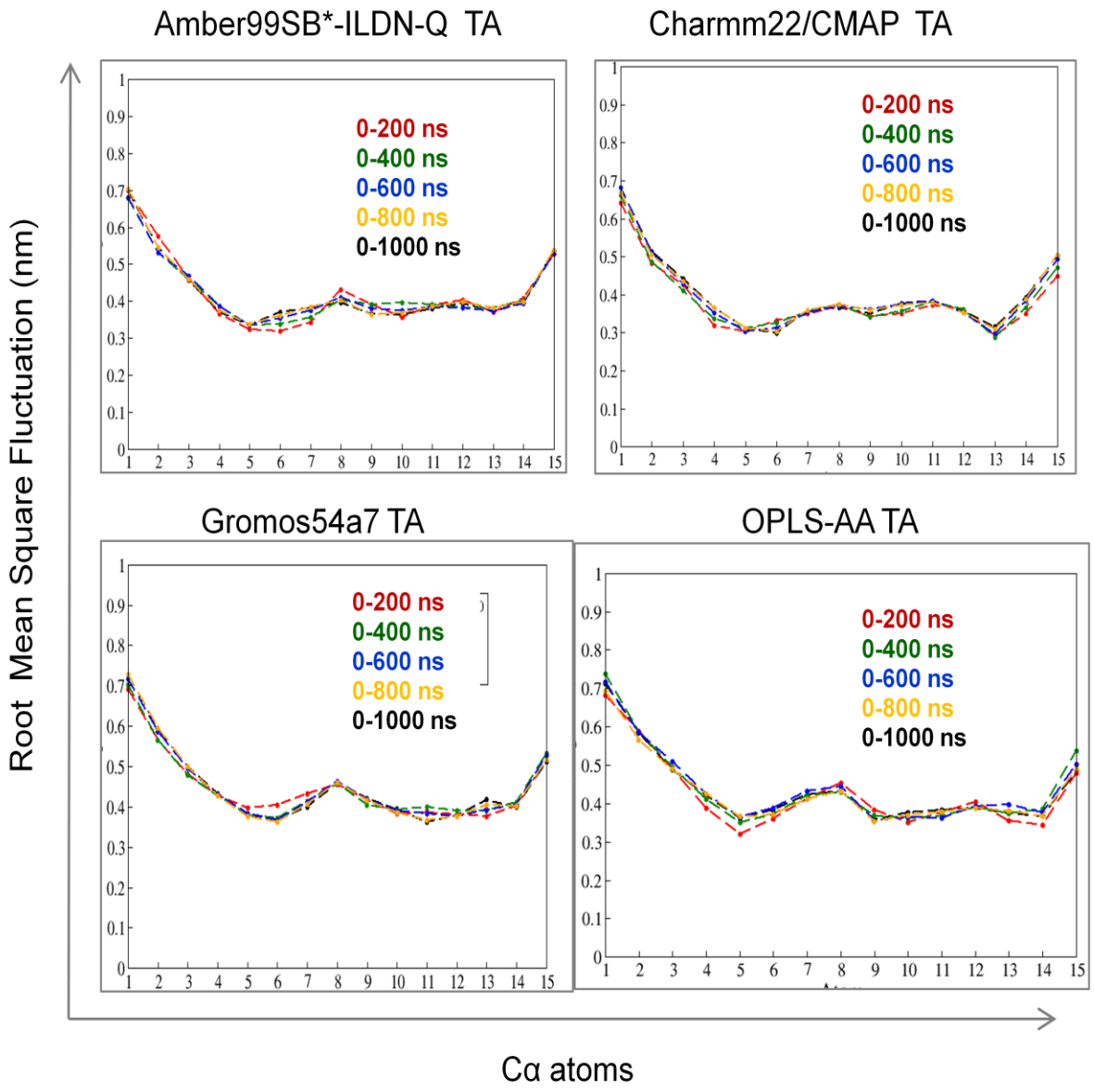


Figure: 5) Residue root mean square fluctuations for Cα atom of each residue. RMSF for MD, TA/SA, REMD for Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA force fields at different time points by dividing the trajectory into five parts in each simulation method.

Secondary structures predicted by Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA using simulation methods, MD, REMD and TA/SA to assess propensity towards certain structure, GROMACS tool (do_dssp) was employed. Secondary structure maps for dengue virus serotype 2 fusion peptide. Figure: 6 (a) & (b) shows the crystal structure fusion peptide dengue virus serotype 2 configurations.

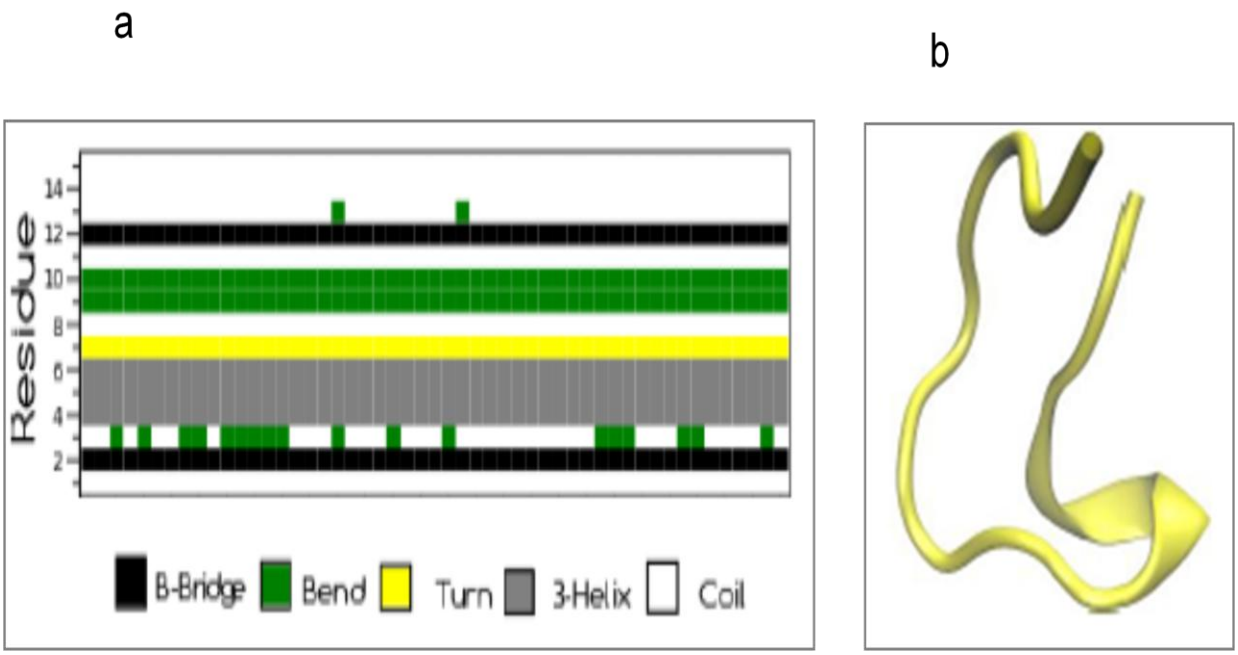


Figure: 6 (a) Secondary structure map for DENV2 FP (b) DENV2 FP (cartoon representation) (300K)

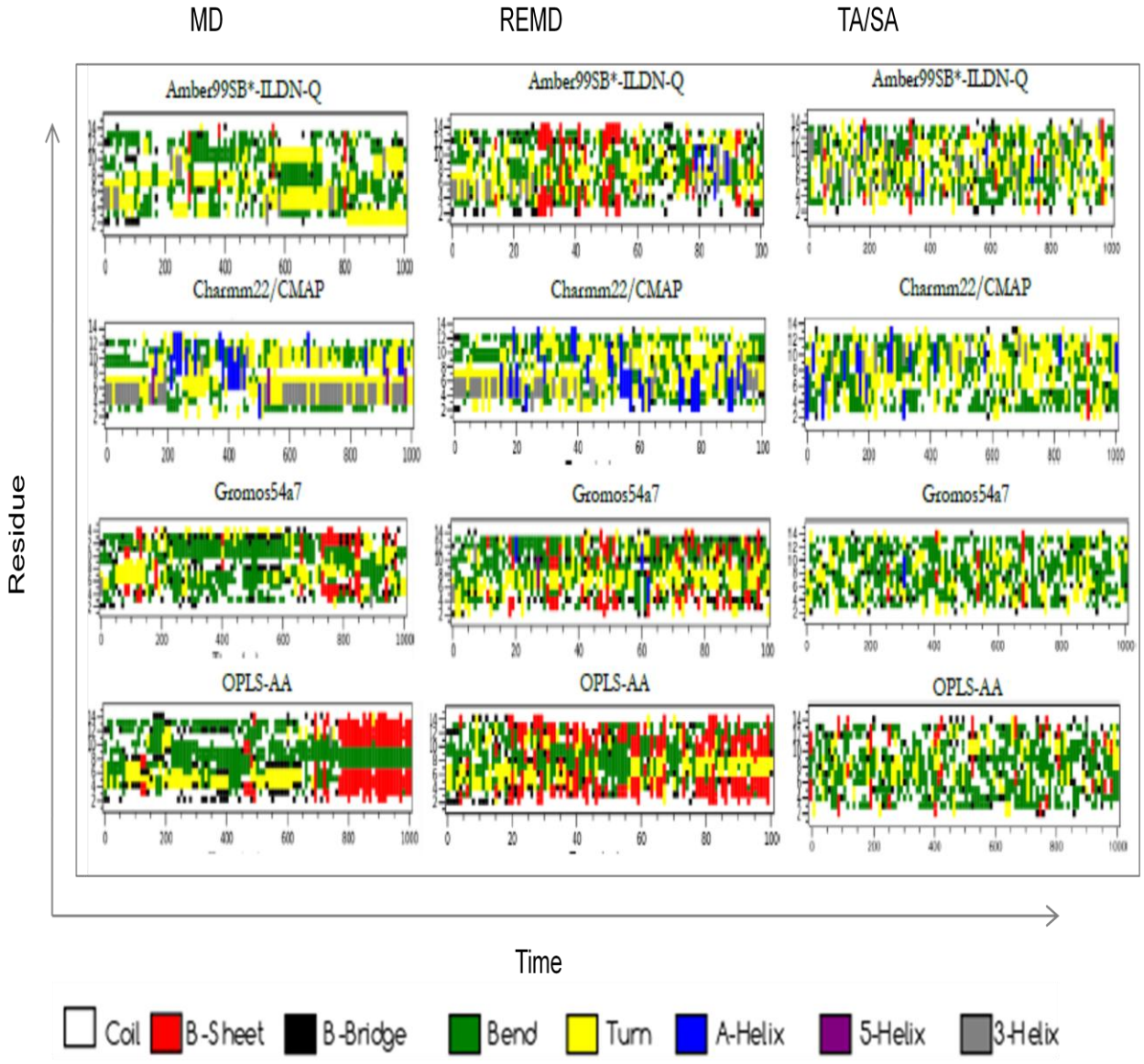


Figure: 7) Secondary structure analysis of DENV2 fusion peptide for MD, TA/SA and REMD for force fields Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA.

The secondary structure maps show clear differences in structure prediction by force fields. Amber99SB*-ILDN-Q predicts overall coil, bend, turn, β-bridges, β-sheets, α-helices, 310 -helix. REMD clearly showed efficient sampling compared to MD wherein only at four time points β-sheets appear and no α-helices at all in case of Amber99SB*-ILDN-Q. In case of TA/SA the sampling showed β-sheets and α-helices appear less frequently but these conformations were explored during the simulation run for force field Amber99SB*-ILDN-Q. Charmm22/CMAP showed significant bias towards α-helices in all three methods MD, REMD, TA/SA predict majorly bends, turns, 310 -helices. Gromos54a7 shows bias towards β-sheets in REMD and MD. In case of REMD Gromos54a7 exhibits a frequent appearance of β-sheets than MD where it appears only in the beginning and towards the end of the simulation run. The method TA/SA shows β-sheets appearing four times during entire simulations. OPLS-AA shows strikingly β-sheets bias during entire simulation run for all three methods. In case of MD OPLS-AA assumes β-sheets for a significant period of the simulations suggesting its overestimation of that structure and lack of sampling efficiency of MD method. The analysis provided a result suggesting Amber99SB*-ILDN-Q doing well in predicting the structure in all three methods MD,REMD,TA/SA for DENV2 FP hence for DENV3 and TBEV simulations were done using Amber99SB*-ILDN-Q for methods MD, REMD,TA/SA. The difference in sampling efficiency of MD, REMD and TA/SA is very well accounted in (Figure: 8). MD predicts prominently α-helical conformations of TBEV amongst coil, bend, turn, β-bridges, β-sheets,310 –helix conformations whereas DENV3 exhibits coil, bend, turn, β-bridges, β-sheets,310 –helix , α-helices over entire time. REMD sampled conformations much efficiently compared to TA/SA where all these structures appear throughout the simulation.

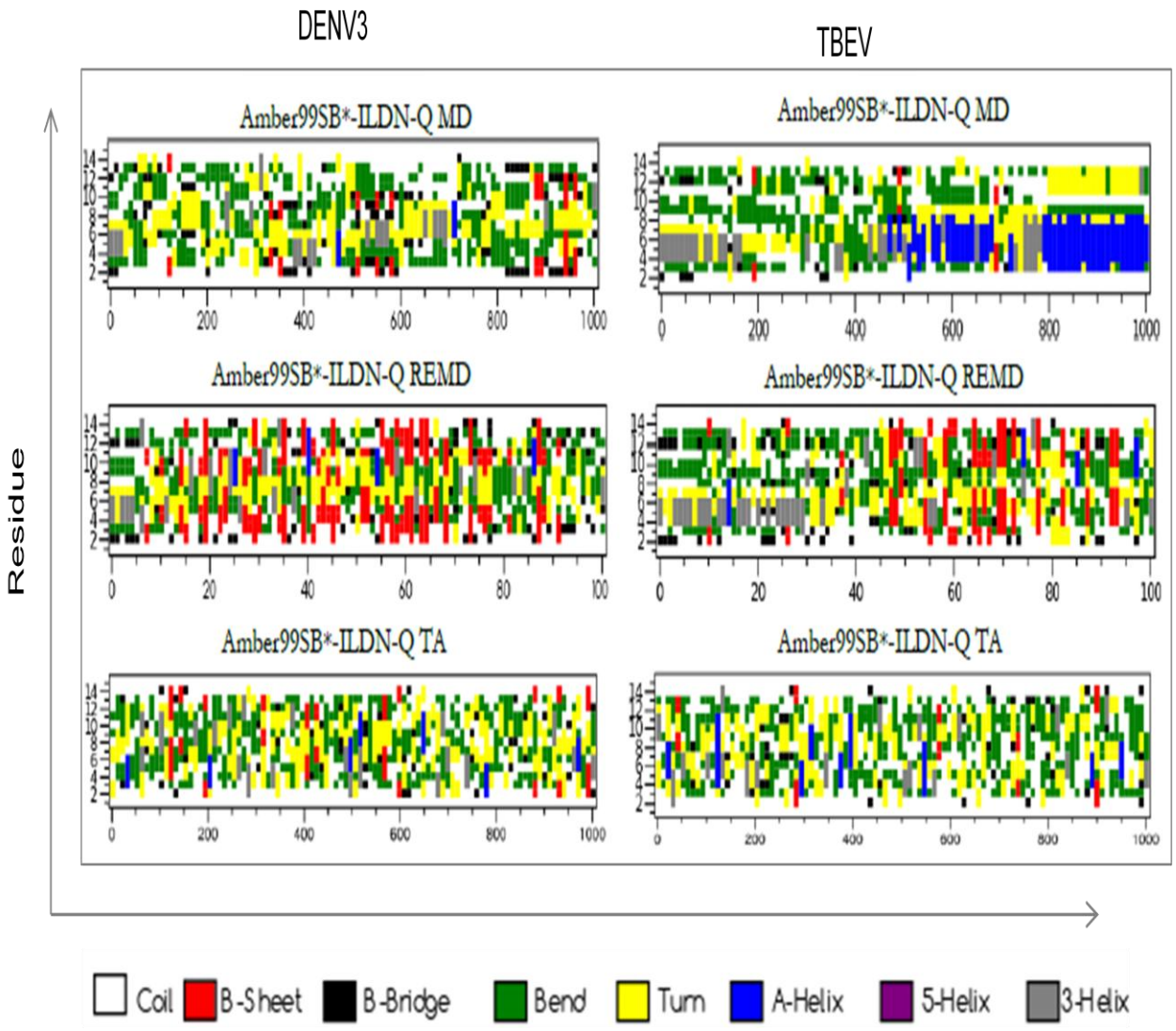


Figure: 8) Secondary structure analysis of DENV3 and TEBV fusion peptide for MD, TA/SA and REMD for force fields for Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA.

Next to identify the most prominent conformation of the entire conformational space based on hierarchical clustering method gromos (Daura et al. 1999)where it counts number of neighbors based on a cut-off (cut-off used was 0.45). To assess the structure with largest number of neighbors GROMACS tool (cluster) was used to find the representative conformation cluster for DENV2 FP. In (Figure:9) it clearly showed that MD does not sample as much as REMD and TA/SA do indicating these methods provide enhanced sampling exploring many conformations given fusion peptide is dynamic. It was seen (Figure: 10) that the largest cluster representing a stable minimum energy conformation in case of each force field and each method revealed that Amber99SB*-ILDN-Q force field has predicted a structure close to crystal structure in case of REMD. Charmm22/CMAP favors α-helical conformation over any other conformation and Gromos54a7 and OPLS-AA are seen to perform not so well in any of the sampling method.

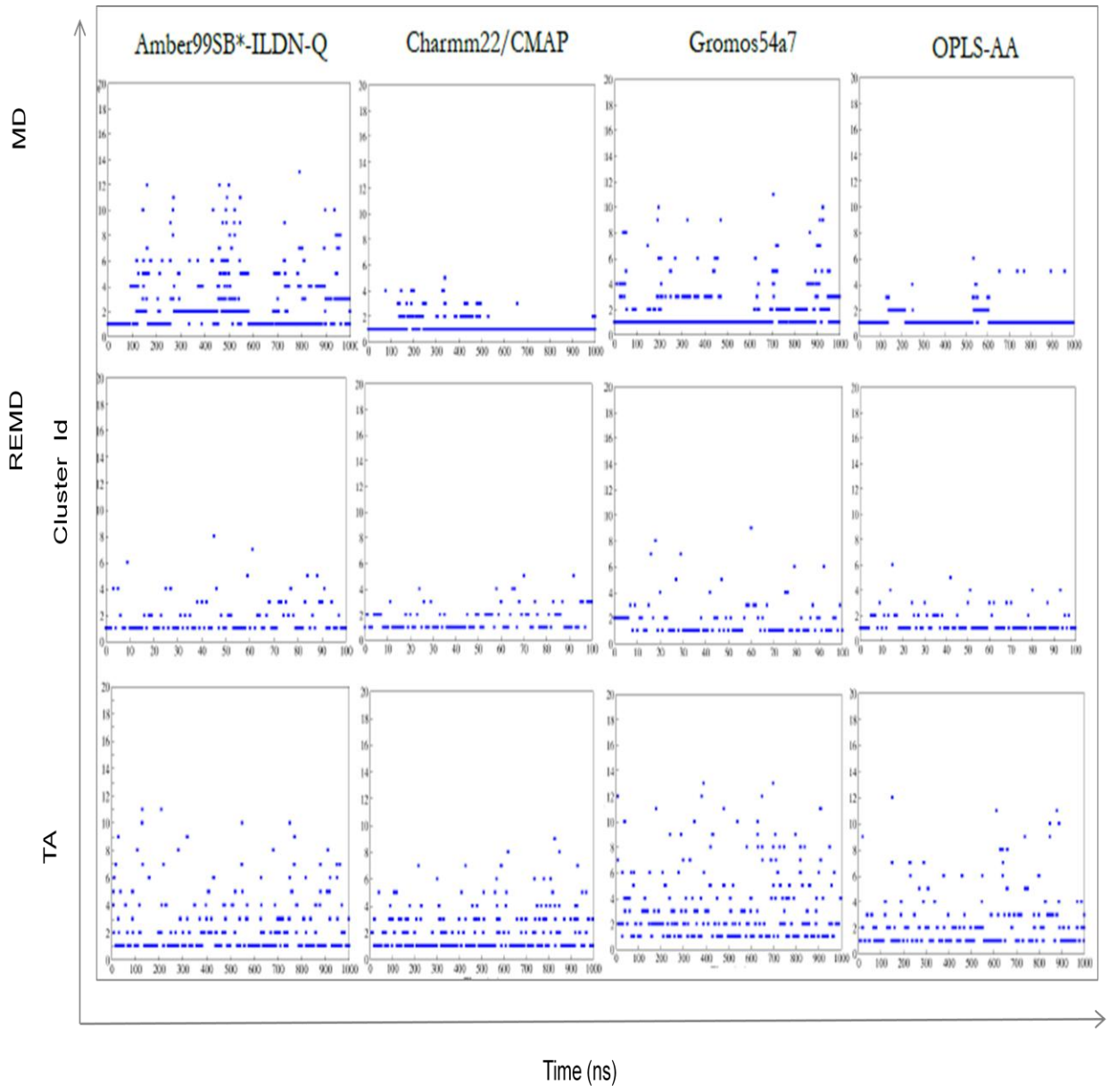


Figure: 9) Clustering: Number of clusters formed for entire simulation time for MD, TA/SA and REMD for force fields Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA.

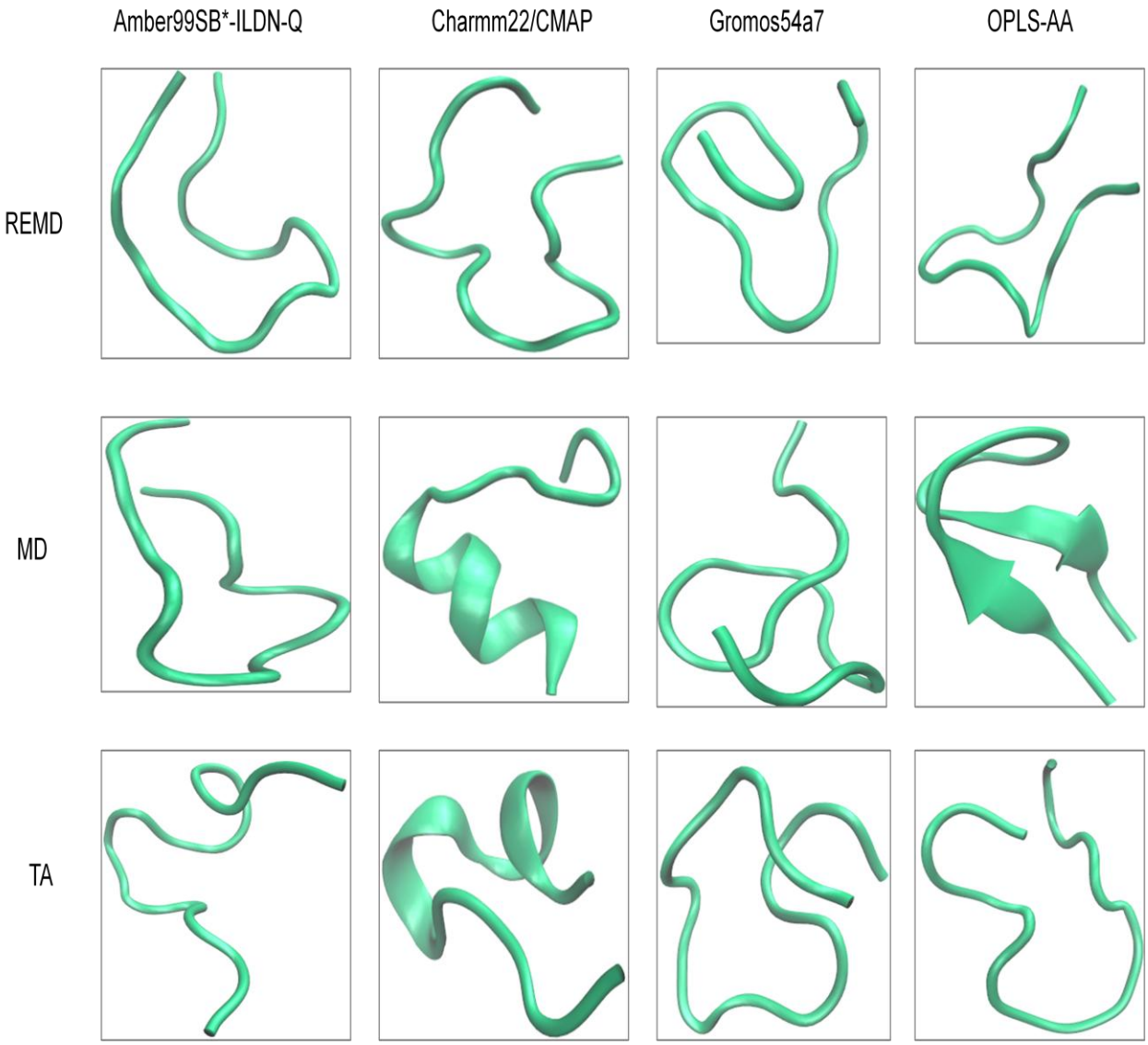


Figure: 10) A single representative cluster (Cluster1) that was prominent throughout the simulation time for MD, TA, REMD for force fields Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA

Principal Components Analysis (PCA) is a way of identifying patterns in data to highlight similarities and differences based on mathematical expressions. PCA is a powerful tool for data analysis since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available. The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data by reducing the number of dimensions, without much loss of information. Principal component analysis was done in order to identify functionally relevant motions. Eigen values against the index of eigenvectors were obtained from PCA on a combined trajectory for REMD, TA, MD for force fields (Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7, OPLS-AA. The analysis was done using GROMACS tool (covar, anaeig). The first two eigenvectors accounted nearly 40% of the conformations sampled.

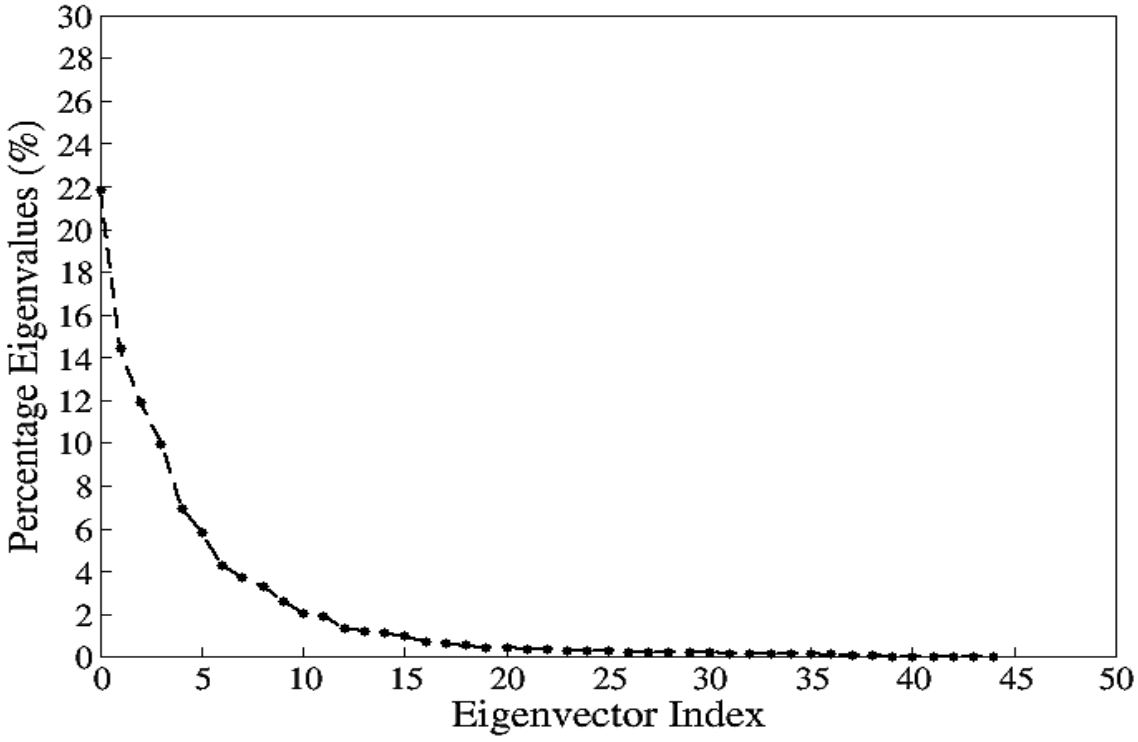


Figure: 11) First two principal components account for the major motions of the fusion peptide ~40% of all conformations.

In protein folding energy landscape maps all possible conformations of a molecular structure and spatial positions of the interacting molecules in a system and the corresponding gibbs free energy. It represents all possible conformations and also provides a qualitative idea about the low energy secondary and tertiary conformations.

Gibbs free energy landscape assessment was done using GROMACS tool (sham) to study the conformational space explored by MD, REMD and TA/SA. It is clearly seen that MD explored very few minimum energy conformations of flavivirus fusion peptides studied (Fig: 12 & 13) but REMD and TA/SA have sampled much more.

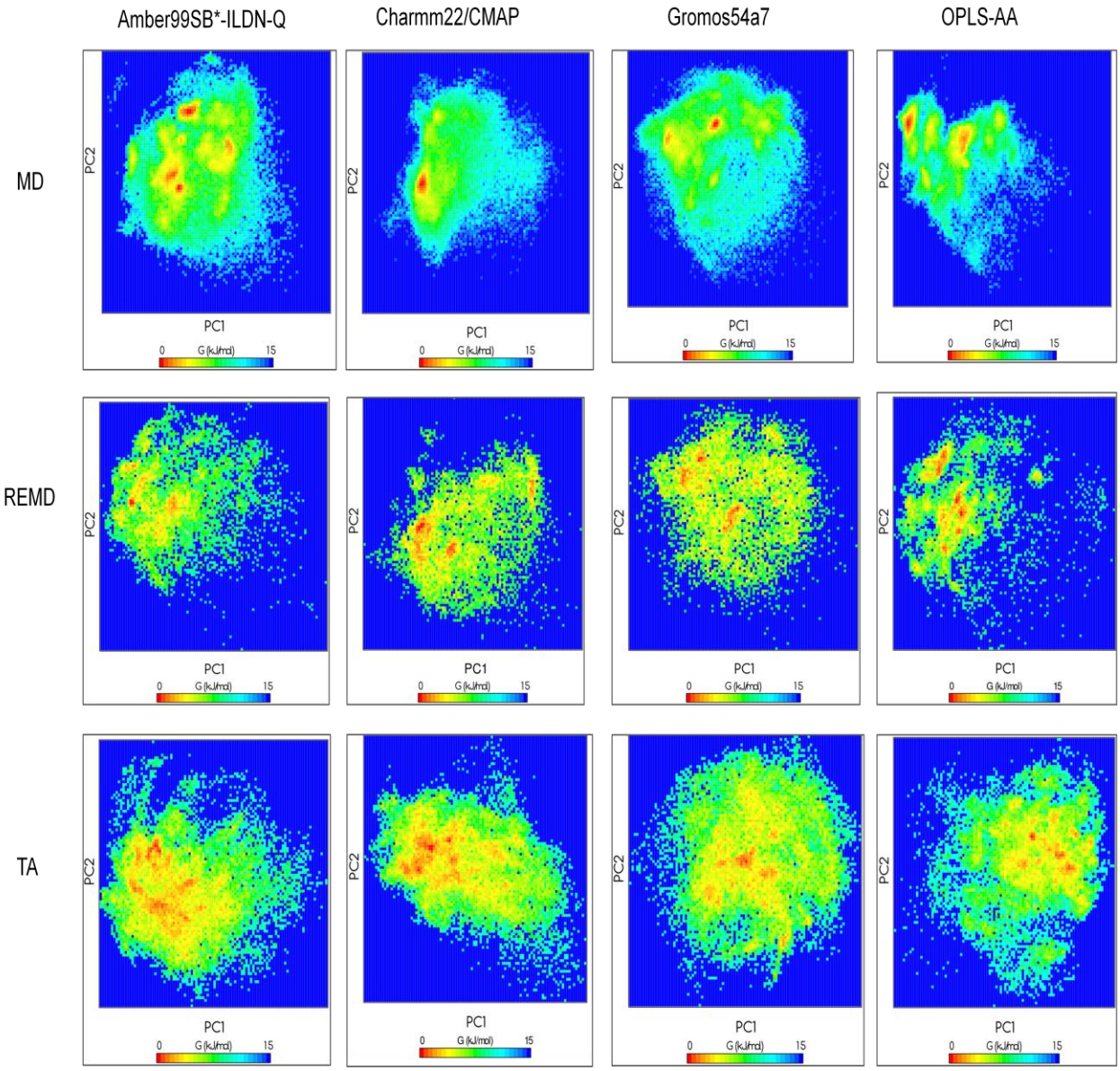


Figure: 12) Gibbs free energy landscapes describing conformational sampling by MD, REMD and TA for Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA force fields for fusion peptide dengue virus serotype 2

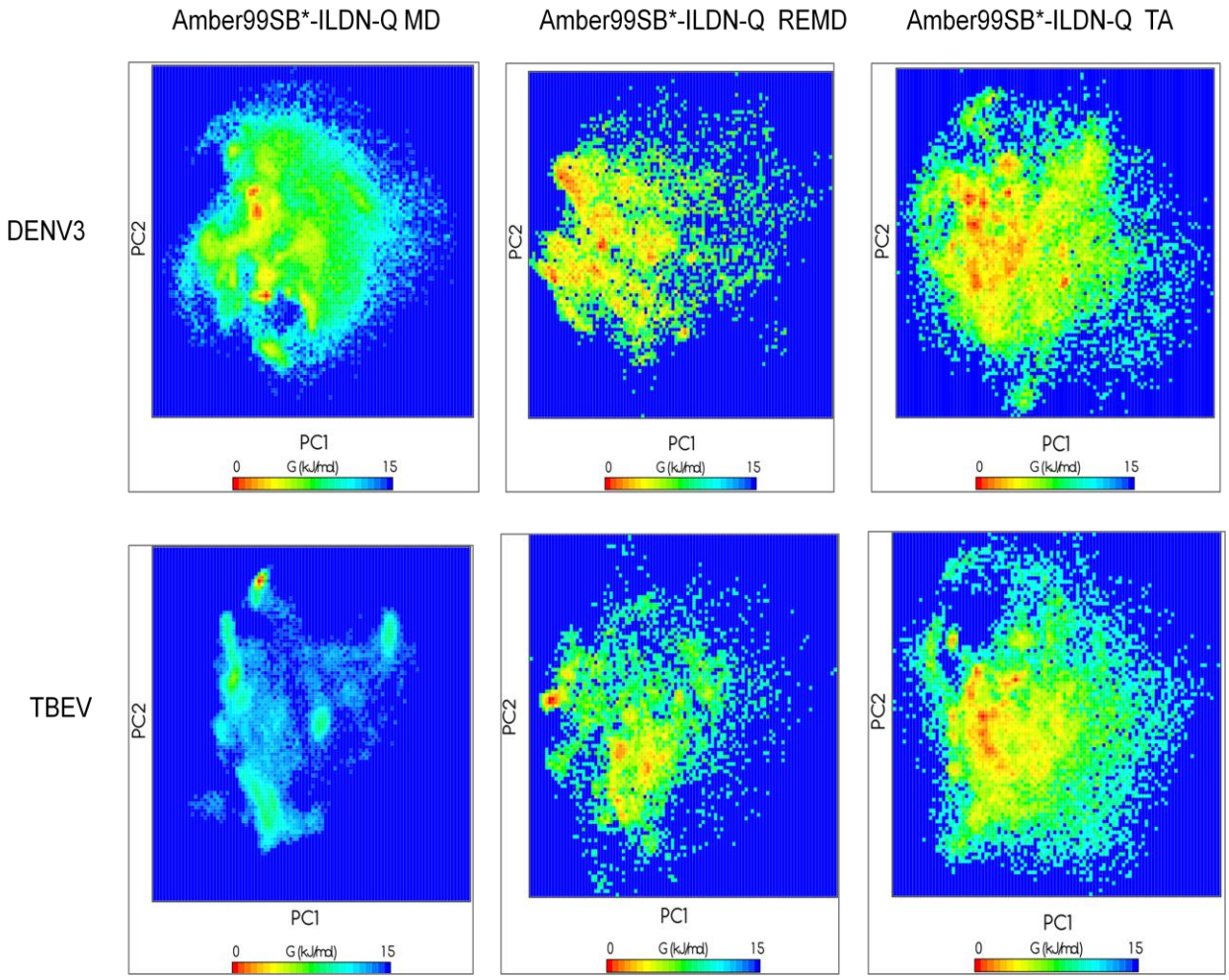


Figure: 13) Gibbs free energy landscape describing the conformational sampling by MD, REMD and TA for Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA force fields for DENV3 and TBEV fusion peptides

Conclusions

The conformational sampling was examined by simulation methods, classic molecular dynamics simulations, replica exchange molecular dynamics simulations and simulated/temperature annealing for flaviviruses dengue virus serotype 2, dengue virus serotype 3 and tick-borne encephalitis virus fusion peptides with four different force fields Amber99SB*-ILDN-Q, Charmm22/CMAP, Gromos54a7 and OPLS-AA force fields in explicit solvents. The starting structures were DENV2, DENV3 and TBEV simulations significant differences were noted between each force field for each simulation method in explicit solvent studied that has heavily influenced the conformations studied. Amber99SB*-ILDN-Q did well in predicting structure of fusion peptide dengue virus serotype 2 close to the crystal structure. It has performed well in predicting structures represented by major clusters as shown in cluster analysis, secondary structure analysis, configurationally it showed values of properties, gyration radius, end to end distance, solvent accessible surface area of the fusion peptide are very close to that of the values for these properties of the crystal structure. It showed effective sampling in case of REMD as compared to MD and TA/SA. Force fields Charmm22/CMAP, Gromos54a7 and OPLS-AA showed strong bias towards secondary structures reason being force field parameterization. Even though simulated/temperature annealing is an improved sampling technique exploring the conformational space significantly wider than MD or REMD it appears that it is not sampling configurations that are structurally similar to the crystal structure available. In comparison, REMD is known to overcome those issues. The analysis also revealed that fusion peptide dengue virus serotype 2 is an unstructured peptide and the MD study carried out in combination with force fields have influenced the structure prediction hence such a study can be of use for further investigations of small peptides that do not assume any specific conformations. The combination of a force field and method can serve as an input for further simulations studies. In addition to that it provides an idea to design molecules that can interact with the fusion peptide affecting the viral infection in a robust way possible in spite of the ongoing findings on blocking virus entry leading to diseased conditions.

References

World Health Organization, “Treatment, prevention and control global strategy for dengue prevention and control,” 2012

N. E. A.Murray, M. B. Quam, and A.Wilder-Smith, “Epidemiology of dengue: past, present and future prospects,” Clinical Epidemiology, vol. 5, no. 1, pp. 299–309, 2013.

Lindenbach BD, Murray C, Thiel HJ, Rice CM. Flaviviridae. In: Knipe DM, Howley PM, editors. Fields virology. Sixth ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 712–46

Y.Modis, S. Ogata, D. Clements, and S. C. Harrison, “Structure of the dengue virus envelope protein after membrane fusion,” Nature, vol. 427, no. 6972, pp. 313–319, 2004

Y. Zhang,W. Zhang, S. Ogata et al., “Conformational changes of the flavivirus E glycoprotein,” Structure, vol. 12, no.9, pp. 1607– 1618, 2004.

Y.Modis, S. Ogata, D. Clements, and S. C. Harrison, “Structure of the dengue virus envelope protein after membrane fusion,” Nature, vol. 427, no. 6972, pp. 313–319, 2004

F. A. Rey, F. X. Heinz, C. Mandl, C. Kunz, and S. C. Harrison,“The envelope glycoprotein from tick-borne encephalitis virus at 2 ˚A resolution,” Nature, vol. 375, no. 6529, pp. 291–298, 1995.

Y. Modis, S. Ogata, D. Clements, and S. C. Harrison, “A ligand binding pocket in the dengue virus envelope glycoprotein,”Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 12, pp. 6986–6991, 2003.

S. Bressanelli, K. Stiasny, S. L. Allison et al., “Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation,” The EMBO Journal, vol. 23, no. 4, pp. 728–738, 2004.

Y. Modis, S. Ogata, D. Clements, and S. C. Harrison, “A ligand binding pocket in the dengue virus envelope glycoprotein,” Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 12, pp. 6986–6991, 2003.

S.L.Allison, J.Schalich, K. Stiasny, C.W.Mandl, andF.X.Heinz, “Mutational evidence for an internal fusion peptide in flavivirus envelope protein E,” Journal of Virology, vol. 75, no. 9, pp. 4268–4275, 2001.

S. L. Allison, K. Stiasny, K. Stadler, C.W.Mandl, and F. X.Heinz, “Mapping of functional elementsin the stem-anchor region of tick-borne encephalitis virus envelope protein E,” Journal of Virology, vol. 73, no. 7, pp. 5605–5612, 1999.

J. Hung, M. Hsieh, M. Young, C. Kao, C. King, and W. Chang, “An external loop region of domain III of dengue virus type 2 envelope protein is involved in serotype-specific binding to mosquito but not mammalian cells,” Journal of Virology, vol. 78, no. 1, pp. 378–388, 2004

M. Liao and M. Kielian, “Domain III from class II fusion proteins functions as a dominant-negative inhibitor of virus membrane fusion,” Journal of Cell Biology, vol. 171, no. 1, pp. 111–120, 2005.

D. Kato, S. Era, I. Watanabe et al., “Antiviral activity of chondroitin sulphate E targeting dengue virus envelope protein,” Antiviral Research, vol. 88, no. 2, pp. 236–243, 2010.

C. O. Nicholson, J. M. Costin, D. K. Rowe et al., “Viral entry inhibitors block dengue antibody-dependent enhancement in vitro,” Antiviral Research, vol. 89, no. 1, pp. 71–74, 2011.

Barrett AD. Current status of flavivirus vaccines. Ann N Y Acad Sci 2001; 951:262–71. Patel JR, Heldens JG. Immunoprophylaxis against important virus disease ofhorses, farm animals and birds. Vaccine 2009;27: 1797–810.

Sabin AB. Research on dengue during World War II. Am J Trop Med Hyg 1952;1:30–50. Halstead SB. Immune enhancement of viral infection. Prog Allergy1982; 31:301–64. Eisenhaber F, Lijnzaad P, Argos P, Sander C, & Scharf M (1995) J. Comput. Chem. 16, 273-284.

Best, R.B., and Hummer, G. (2009). Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. Journal of Physical Chemistry B *113*, 9004–9015.

Freddolino, P.L., Liu, F., Gruebele, M., and Schulten, K. (2008). Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophysical Journal *94*, L75–L77.

Freddolino, P.L., Park, S., Roux, B., and Schulten, K. (2009). Force field bias in protein folding simulations. Biophysical Journal *96*, 3772–3780.

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GRGMACS 4: Algorithms for highly efficient, load balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation*4*, 435–447.

H. J. C. Berendsen, J. P. M. Postma, W.F.V.G., and Hermans, J. (1981). Interaction models fo rwater in relation to protein hydration. Intermolecular Forces 331.

Huang, W., Eichenberger, A.P., and van Gunsteren, W.F. (2012). Molecular dynamics simulation of thionated hen egg white lysozyme. Protein Science : A Publication of the Protein Society *21*, 1153–1161.

Ishikawa, T., Yamanaka, A., and Konishi, E. (2014). A review of successful flavivirus vaccines and the problems with those flaviviruses for which vaccines are not yet available.

Jorgensen, W.L., Jorgensen, W.L., Maxwell, D.S., Maxwell, D.S., Tirado-Rives, J., and Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc. *118*, 11225–11236.

Mackerell, A.D. (2004). Empirical force fields for biological macromolecules: Overview and issues. Journal of Computational Chemistry *25*, 1584–1604.

Mittal, J., and Best, R.B. (2010). Tackling force-field bias in protein folding simulations: Folding of Villin HP35 and Pin WW domains in explicit water. Biophysical Journal *99*, L26–L28.

Piana, S., Lindorff-Larsen, K., and Shaw, D.E. (2011). How robust are protein folding simulations with respect to force field parameterization? Biophysical Journal *100*, L47–L49.

Ponder, J.W., and Case, D. a. (2003). Force fields for protein simulations. Advances in Protein Chemistry *66*, 27–85.

Sorin, E.J., and Pande, V.S. (2005). Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *88*, 2472–2493.

Zhou, R. (2007). Replica exchange molecular dynamics method for protein folding simulation. Methods in Molecular Biology (Clifton, N.J.) *350*, 205–223.