Electronic Supplementary Material (ESI) for RSC Advances.

Detecting early stage structural changes in wild type, pathogenic and non-pathogenic prion variants using markov state model

Vinod Jani, Uddhavesh Sonavane, Rajendra Joshi*

High Performance Computing-Medical & Bioinformatics Applications Group,

Centre for Development of Advanced Computing (C-DAC),

Savitribai Phule Pune University Campus,

Pune 411007, India

rajendra@cdac.in

1. **Root Mean Square Deviation**

   The all atom RMSD of the WT, E200K and E219K has been calculated using gromacs tools.

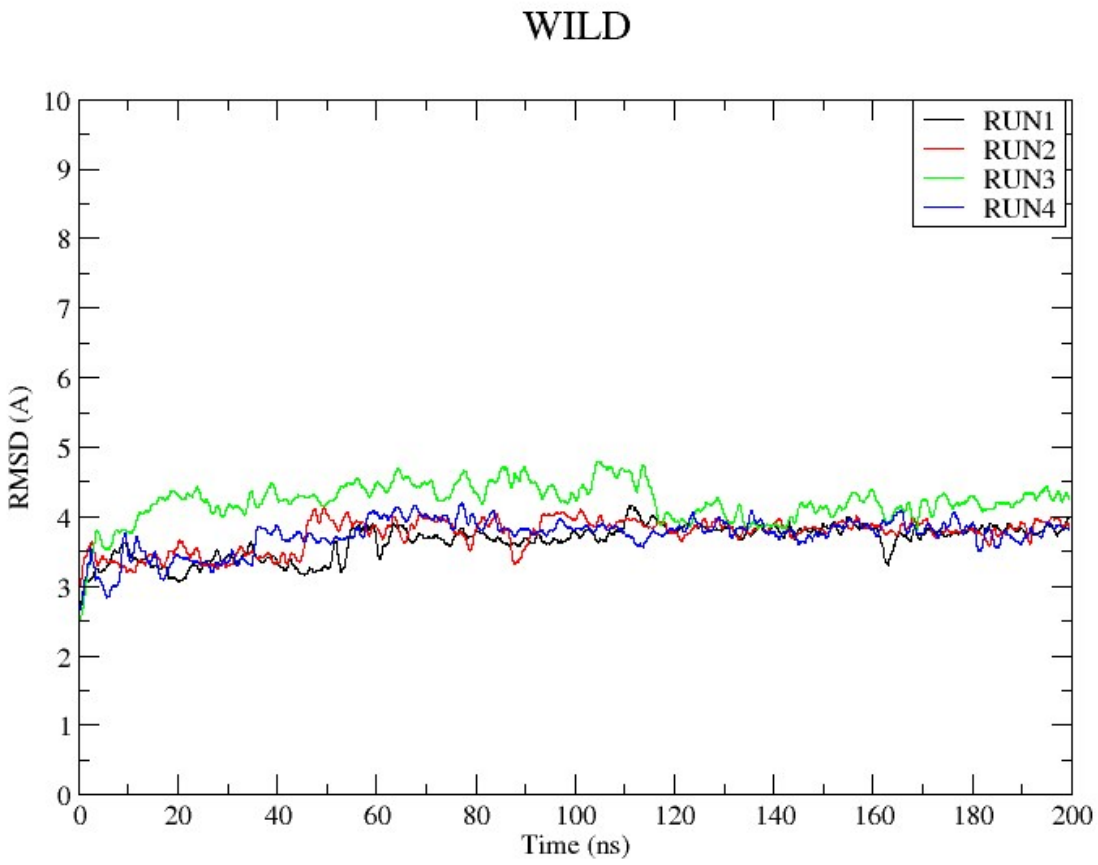

Figure S1 : RMSD plot for wild type variant for all four runs

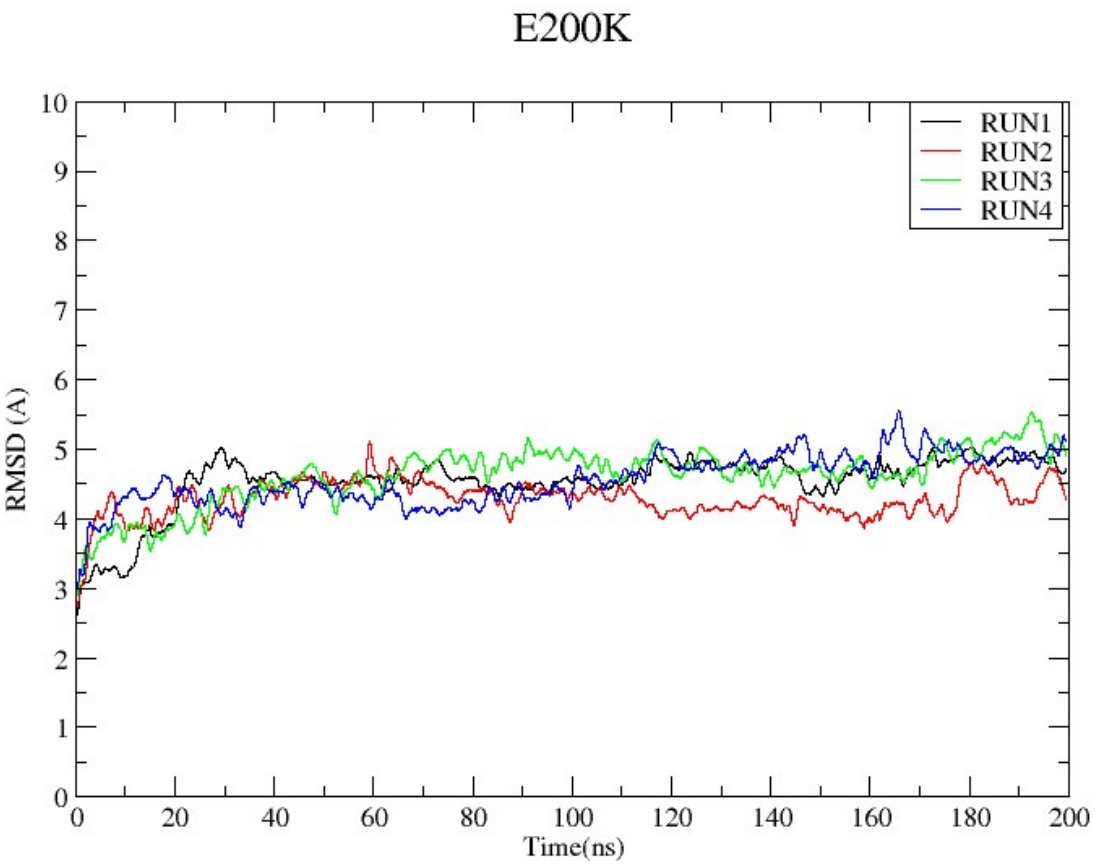

Figure S2: RMSD plot for E200K variant for all four runs

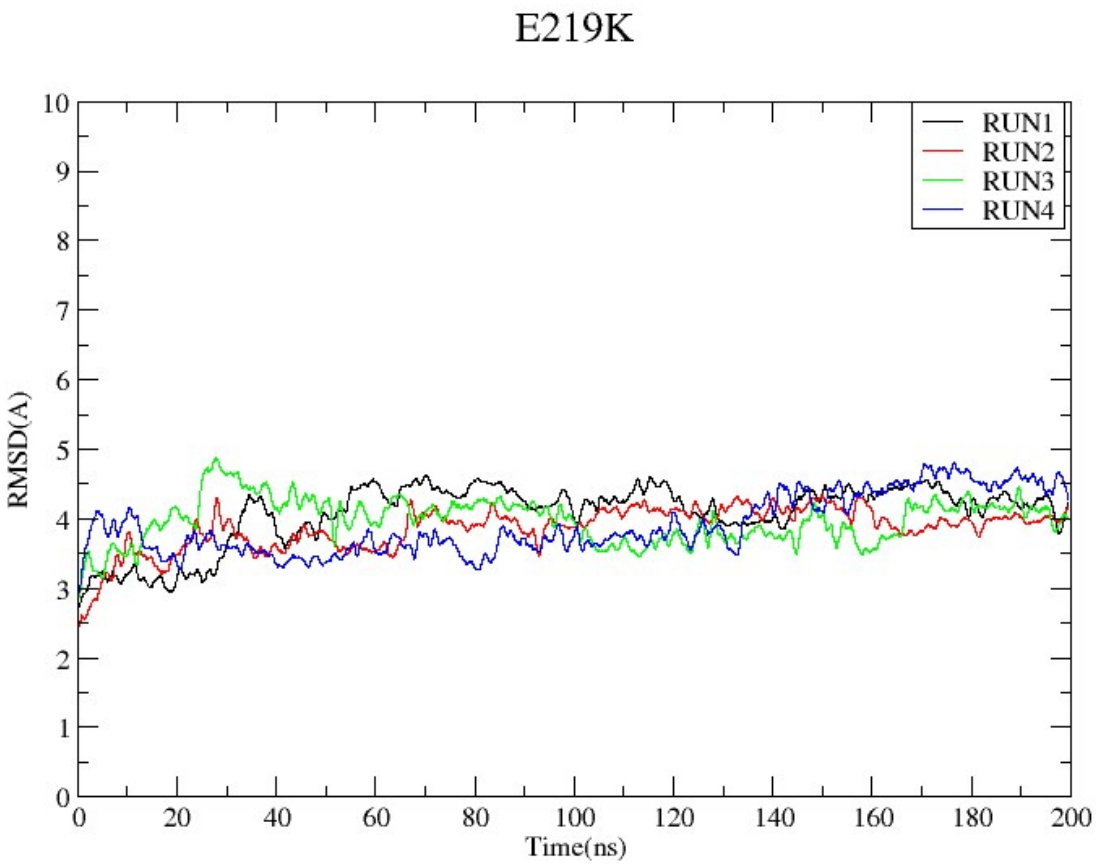

Figure S3: RMSD plot for E219K variant for all four runs

Supplementary Table 1: Average RMSD for each variants in different runs

	RUN1	RUN2	RUN3	RUN4
WILD	3.3 ±0.2	3.2 ±0.3	3.8 ±0.2	3.2 ±0.2
E200K	4.5 ±0.2	3.8 ±0.2	4.5 ± 0.2	4.3 ±0.2
E219K	3.9 ± 0.2	3.5 ± 0.3	3.7 ± 0.2	3.8 ± 0.2

2. Root Mean Square Fluctuation (RMSF)

The root mean square fluctuation (RMSF) gives an idea about residue wise fluctuation throughout the simulation. The average amplitude of RMSF fluctuation for all the three variants was around 3 Å as can been seen from figure S4. Wild type protein remains was stable throughout and showed slightly high fluctuation for residues 135 to 137 of around 3 Å. For E200K mutant there were few fluctuation peaks. For residues 138 to 142 the peak was around 2.8 Å. The peak for residues 167 to 170 was around 3 Å. Maximum fluctuations of around 4.5 Å was shown by residues in range of 187

to 194. For E219K maximum fluctuation was seen for residues 133 to 137 the same stretch where wild type protein had shown fluctuation. Thus WT and E219K had similar residue fluctuation and in both case fluctuations were observed in loop1. In E200K fluctuation was more as compared to WT and E219K.

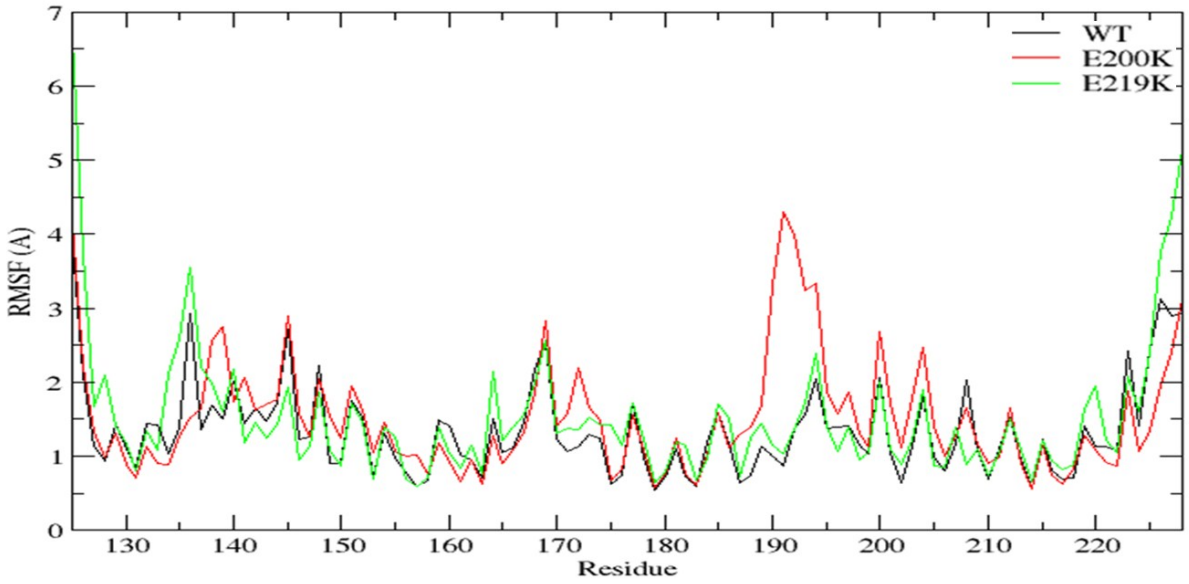

Figure S4: Root Mean Square Fluctuation for all the three variants. Black line is for WT, red for E200K variant and green for E219K variant.

3. Radius of Gyration (RGY)

The radius of gyration gives an idea about compactness of the structure. From Figure S5 it is clear that for wild type, structure remain stable throughout simulation with RGY varying in range of 0.2 Å i.e. from 14.8 to 15 Å. While for E200K it showed maximum fluctuation and it decreased to 14.3 from 15.1 Å i.e. the structure tends to get more and more compact. Also sudden dips in RGY value were clearly seen at regular intervals. The decrease in RGY values for E200K were seen at 20 to 40ns, then 130 to 140ns and from 170 to 180ns and 190 to 200ns. For E219K there was initial increase in RGY value up to 15.4 Å and then it decreased to 15 Å and for most of simulation it remained stable around 15 A.

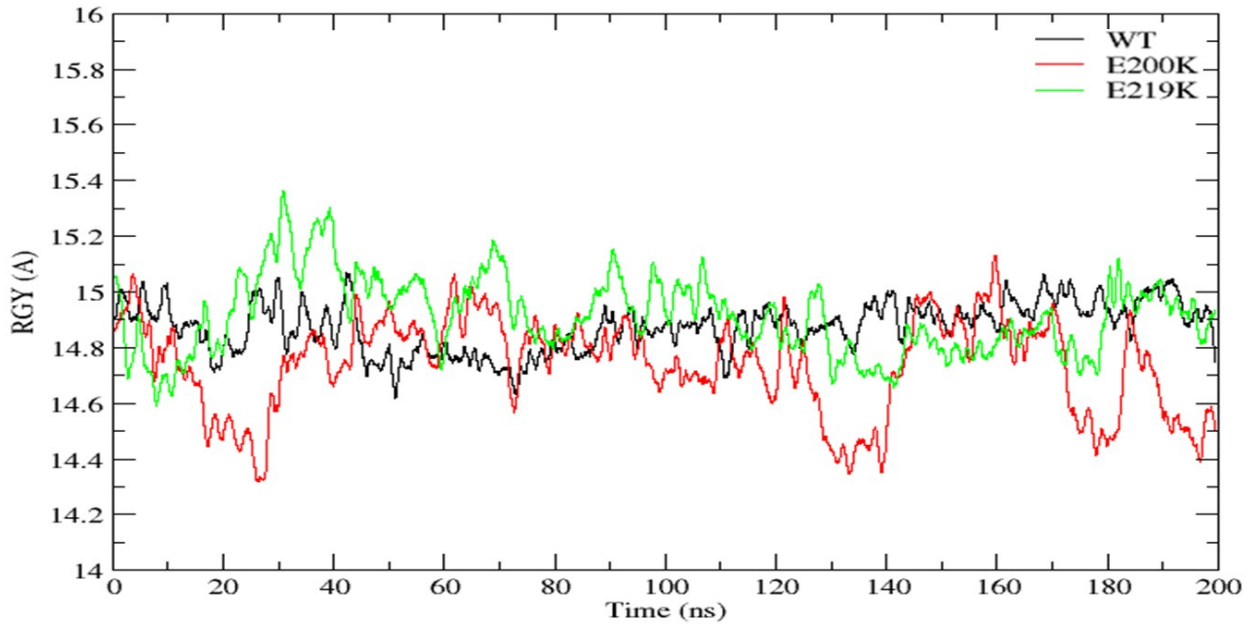

Figure S5: Radius of gyration plot for all the three variants. Black line is for WT, red for E200K variant and green for E219K variant.

4. Solvent Accessible Surface Area (SASA)

Solvent accessible surface area given an idea about amount of exposed surface of the protein. For all the three proteins initially SASA tends to decrease (Figure S6). For wild type after initial decrease in SASA value it remained constant throughout simulation. For E200K after initial decrease there was slight increase in SASA values and then sudden dip in SASA values at around 130ns followed by rise of 1000 $Å^2$ at around 140 to 160ns. For E219K much fluctuation was seen in SASA values these shows rearrangement of residues. For last 40ns the SASA values for E219K remained stable.

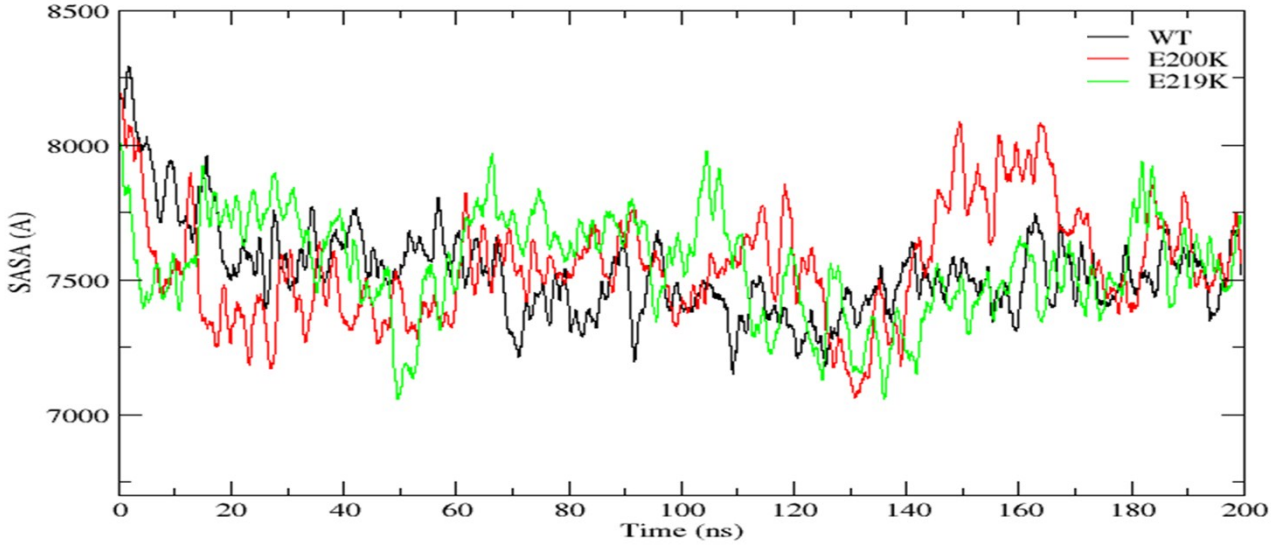

Figure S6: Solvent Accessible surface area plot for all the three variants. Black line is for WT, red for E200K variant and green for E219K variant.

5. Hydrogen Bonds

In order to maintain the stability of protein, various interactions plays important role among this hydrogen bond plays an important role in folding of α-helices and β-sheets. Hydrogen bond plays important role in maintaining local stability of protein hence hydrogen bonds were calculated and were classified into three types Mainchain-Mainchain (MS-MS), Mainchain-Sidechain (MS-SC), Sidechain-Sidechain (SC-SC). Figure S7 shows total number of hydrogen in the protein. Figure S7A shows hydrogen bond interaction in the entire protein in all three variants. From figure it is clear that number of hydrogen bond were least in E200K mutant. It has been observed that MC-MC hydrogen bonds were affected mainly in case of E200K mutant (figure S7B). The number of MC-MC hydrogen bond tends to decrease during course of simulation for E200K mutant as compared to wild type and E219K. In wild type and E219K mutant MC-SC hydrogen bond tend to remain stable throughout simulation (figure S7B). The MS-SC hydrogen bonds were not affected in all the three variants (figure S7C). The SC-SC hydrogen bond also tends to decrease for E200K mutant as compared to wild type and E219K (figure S7D). From this observation it can be concluded that secondary structure element were destabilize majorly in E200K relative to wild type and E219K.

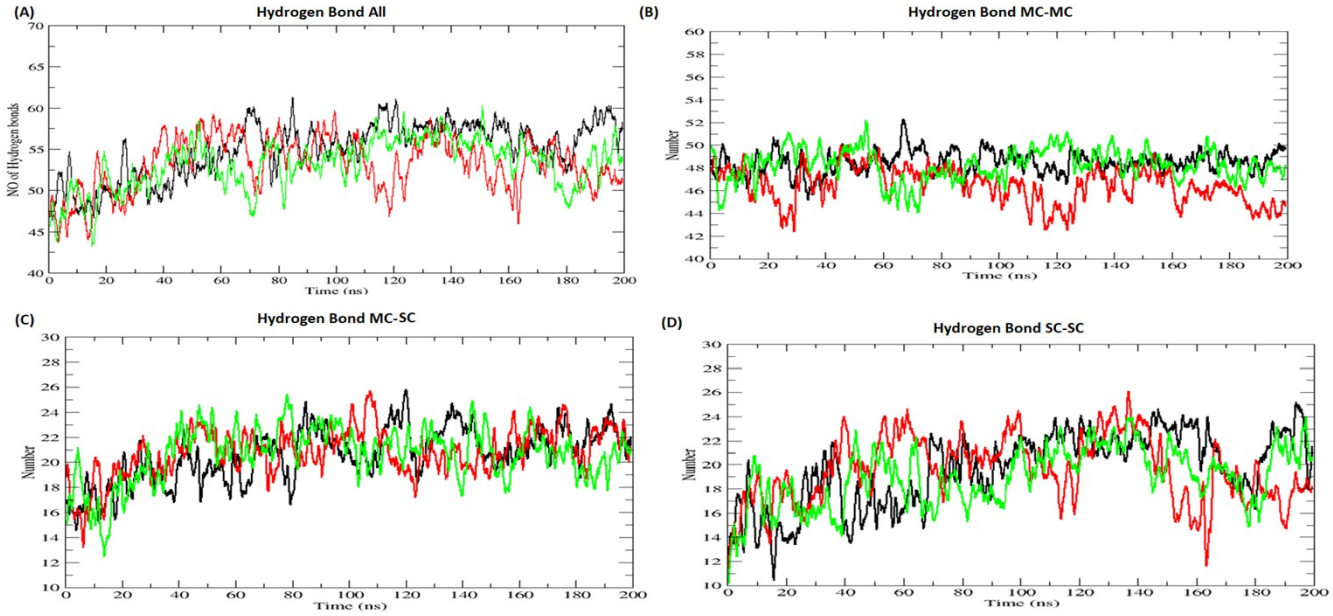

Figure S7: Hydrogen Bond for all the three protein variants

A) Hydrogen Bond for Entire Protein for all the three protein variants
B) MainChain-MainChain Hydrogen Bond for all the three protein variants
C) MainChain-SideChain Hydrogen Bond for all the three protein variants
D) SideChain-SideChain Hydrogen Bond for all the three protein variants

6. Markov State Model

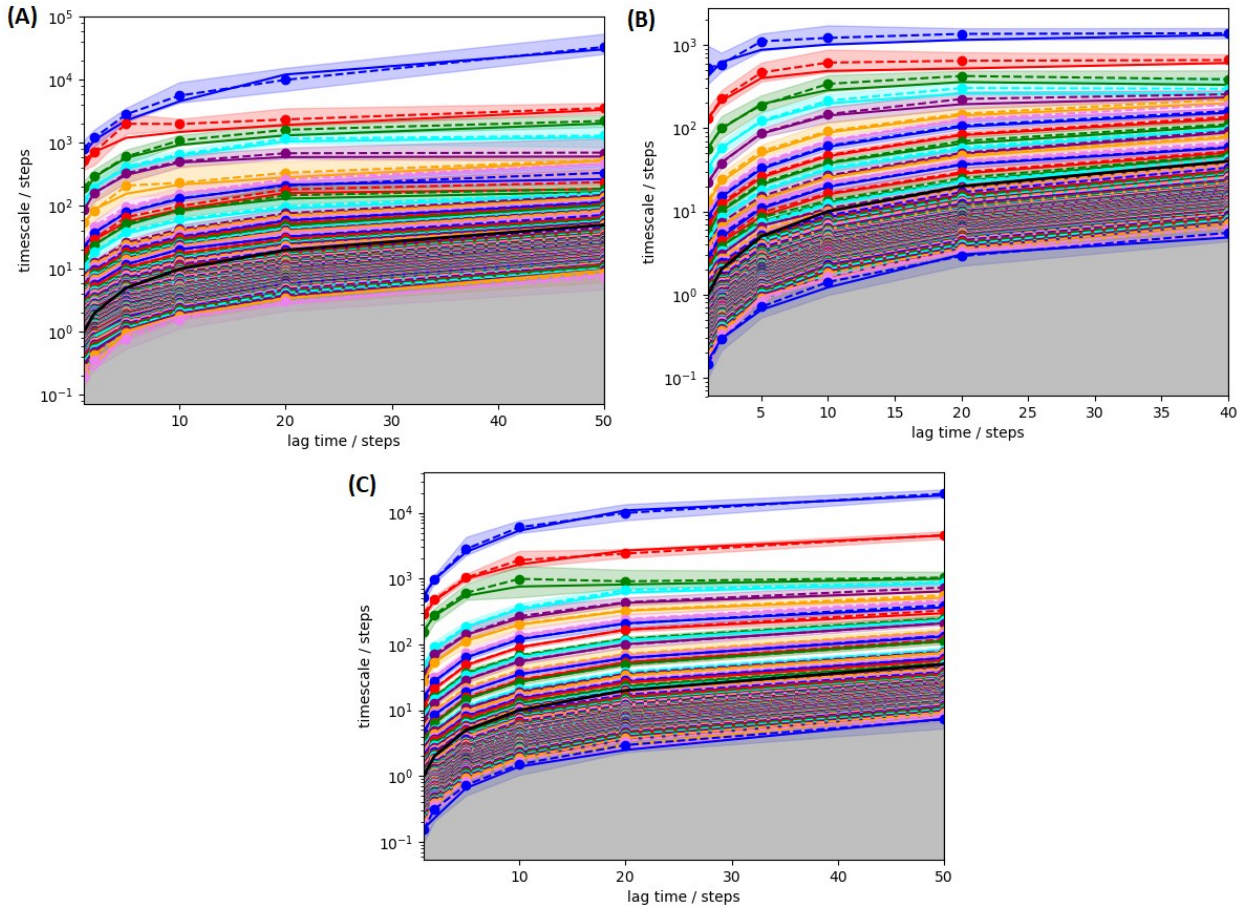

Figure S8a : Implied time scale of MSM

A) Wild Type
B) E200K
C) E219K

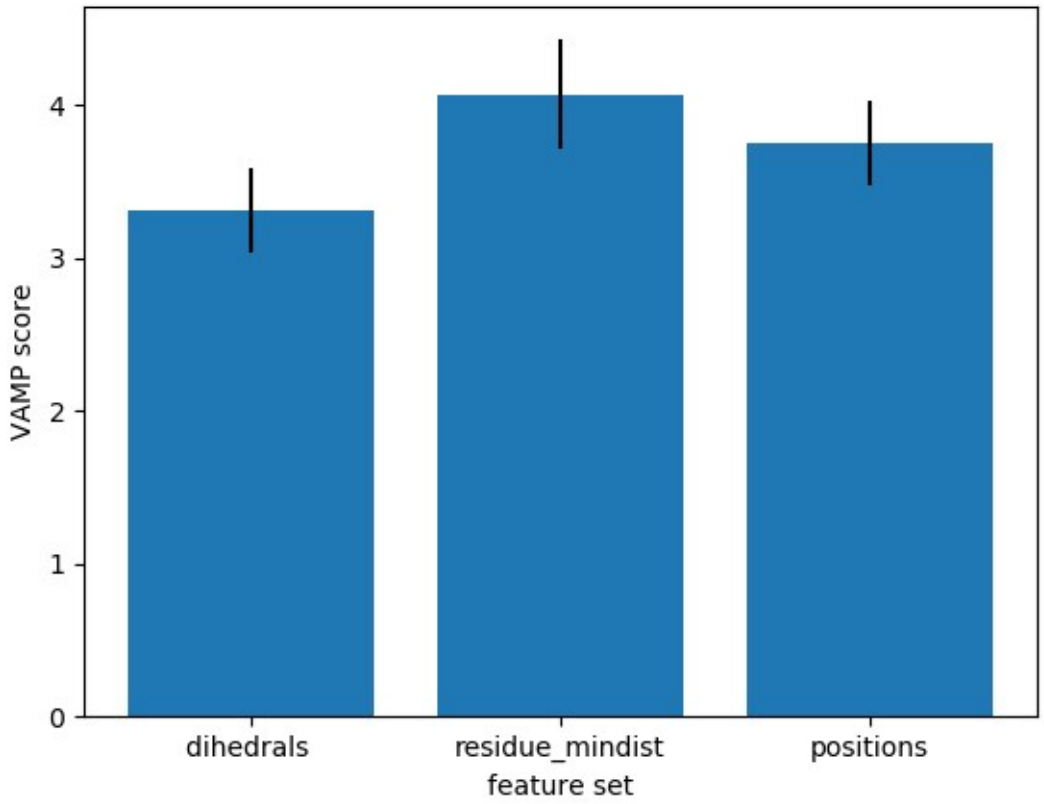

Figure 8b : Showing Vamp Score for different parameters i.e Dihedral angle, Residue_Min_dist and Coordinates(positions)

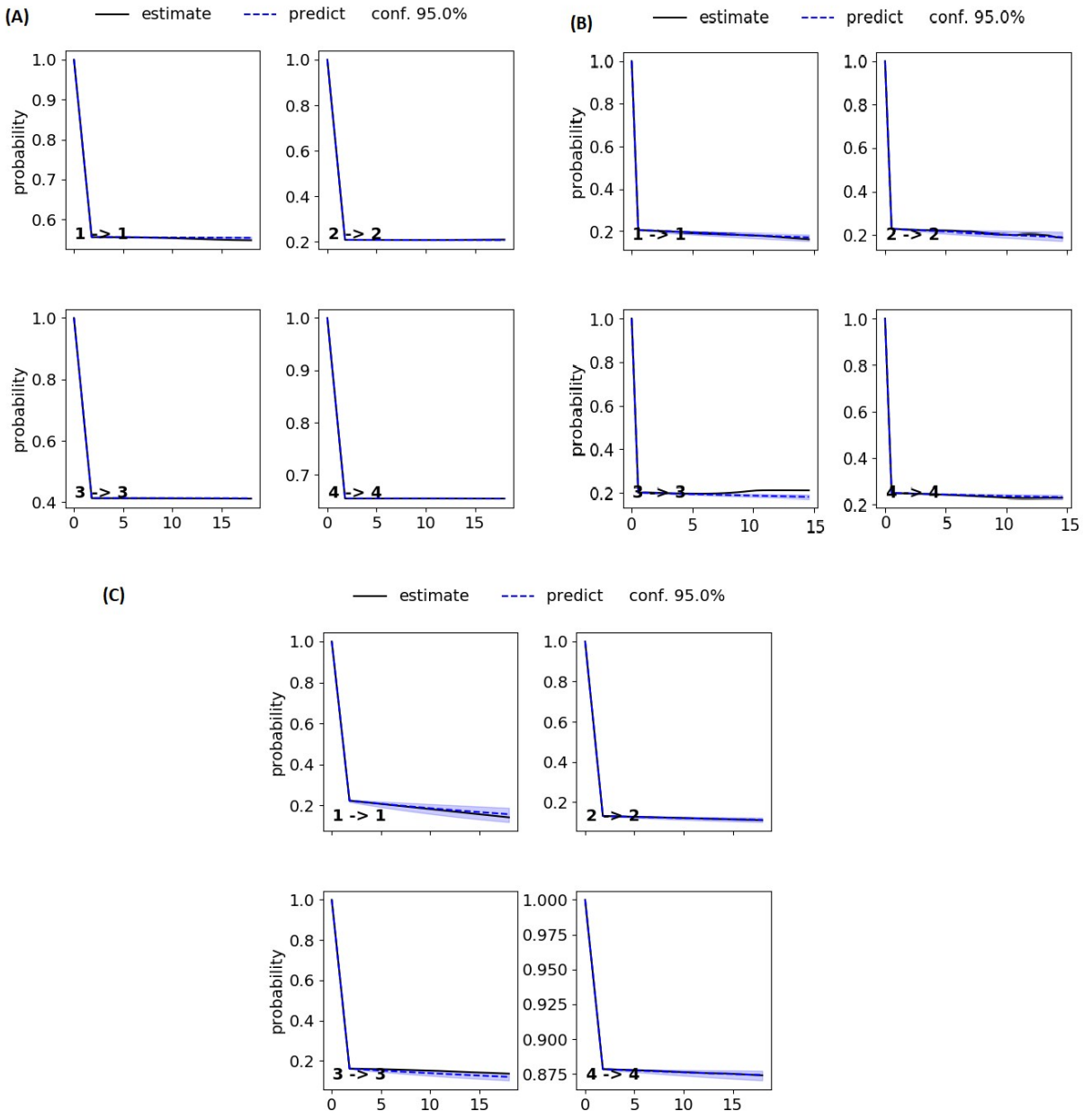

Figure S9: Chapman-kolmogorov tests of MSM estimates at lag time = 20ns.

A) Wild Type
B) E200K
C) E219K

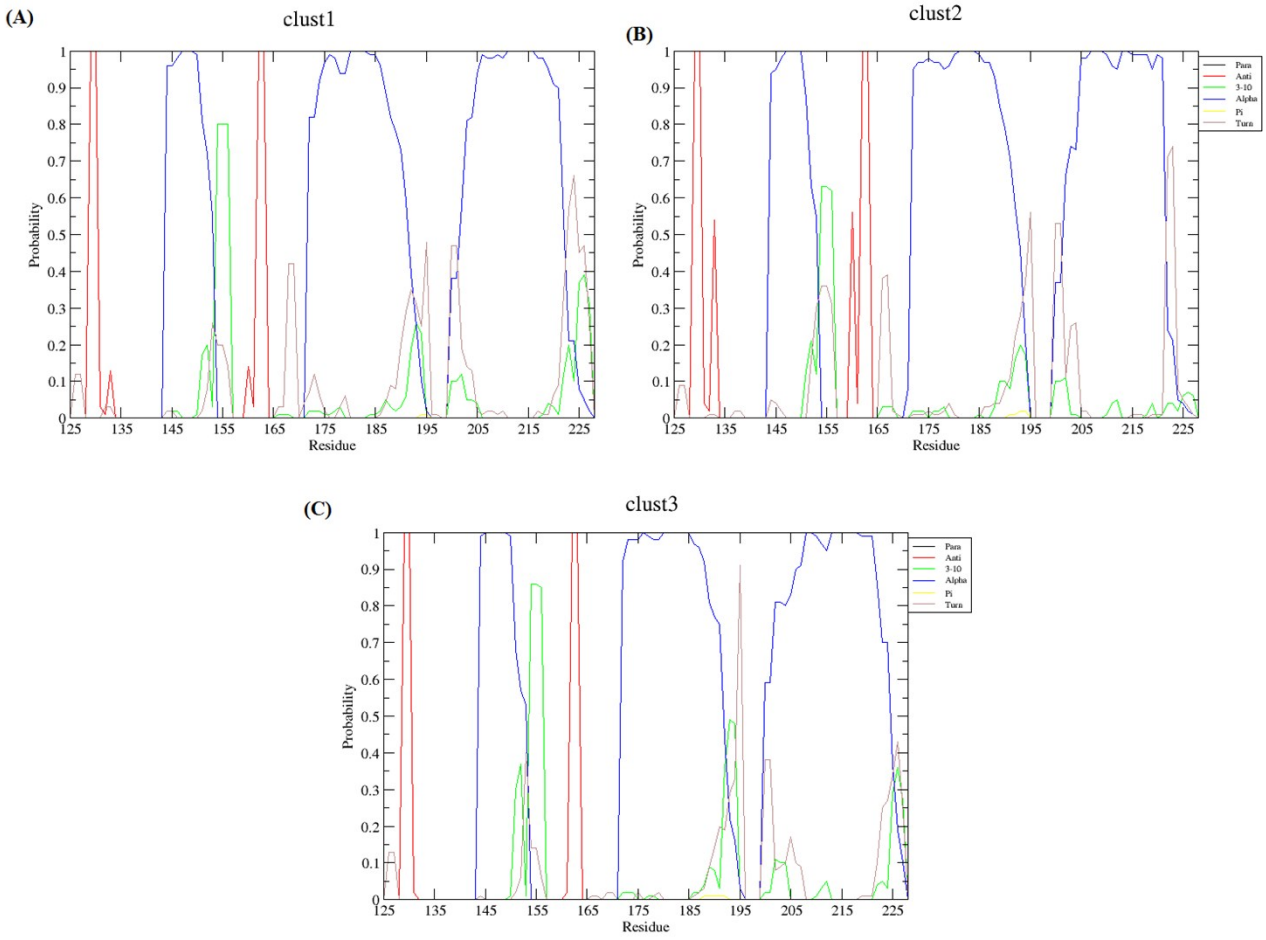

Figure S10 : Probability of Secondary structure for MSM clusters of wild type system

A) Cluster 1
B) Cluster 2
C) Cluster 3

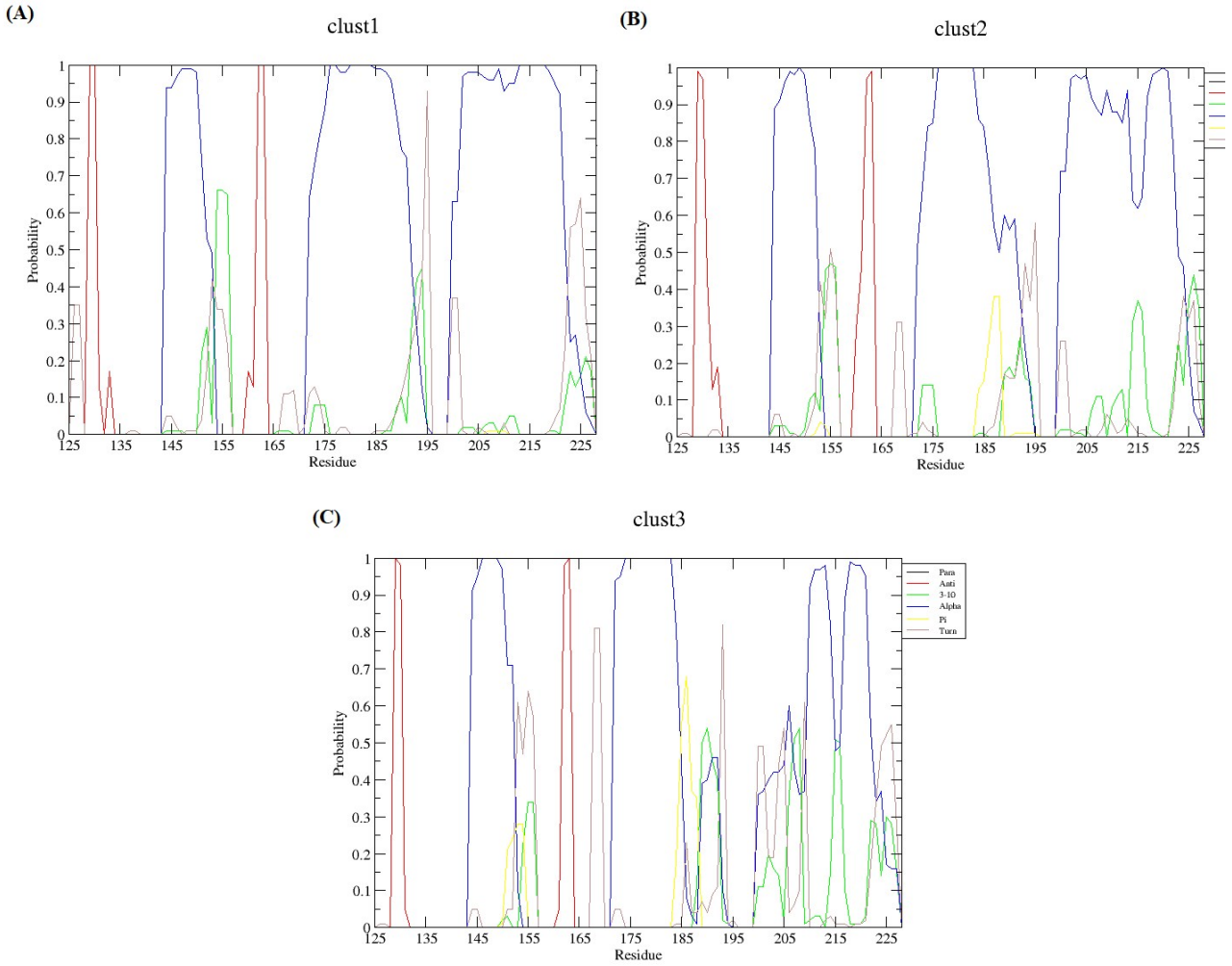

Figure S11 : Probability of Secondary structure for MSM clusters of E200K system

A) Cluster 1
B) Cluster 2
C) Cluster 3

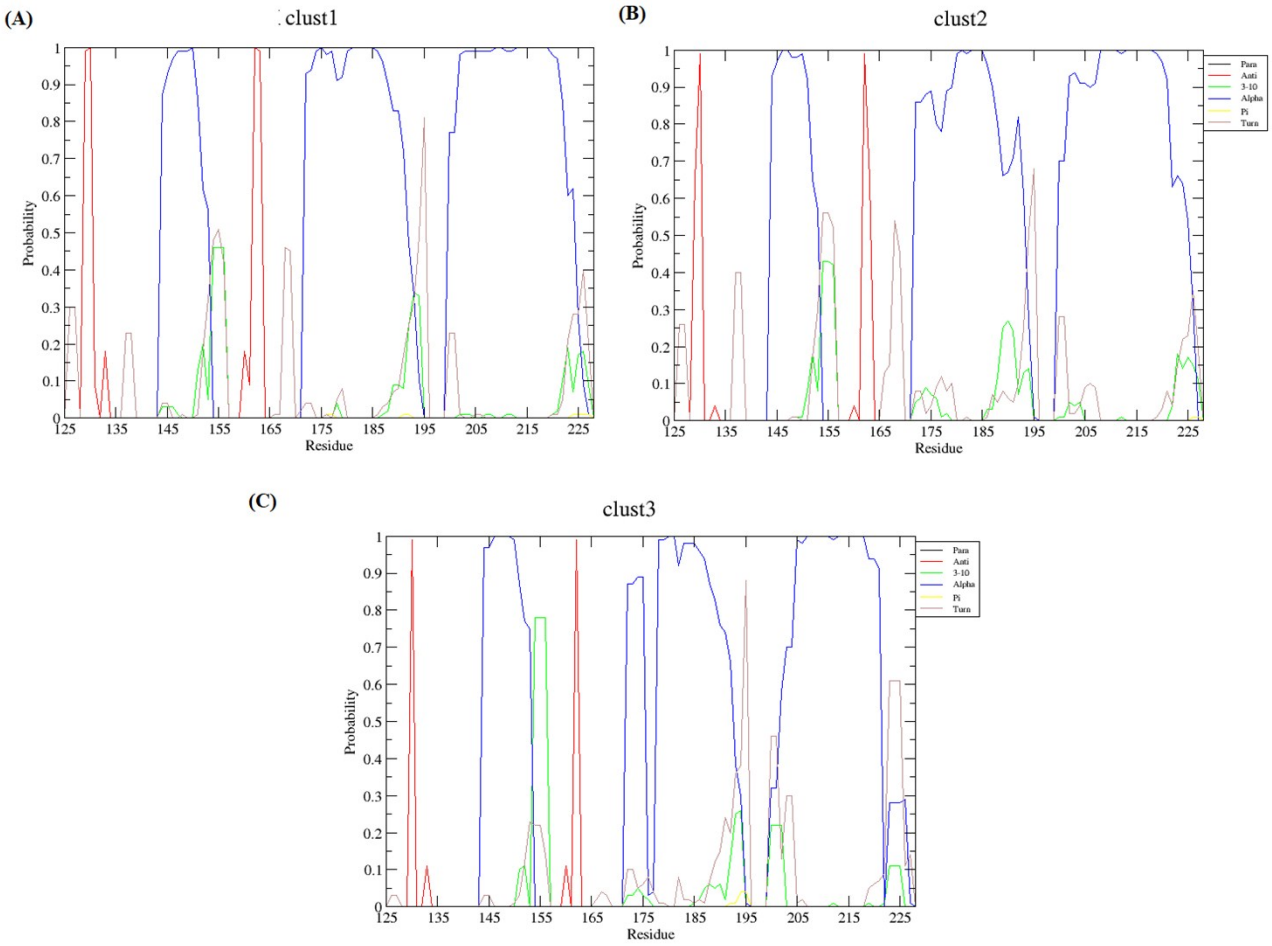

Figure S12 : Probability of Secondary structure for MSM clusters of E219K system

A) Cluster 1
B) Cluster 2
C) Cluster 3

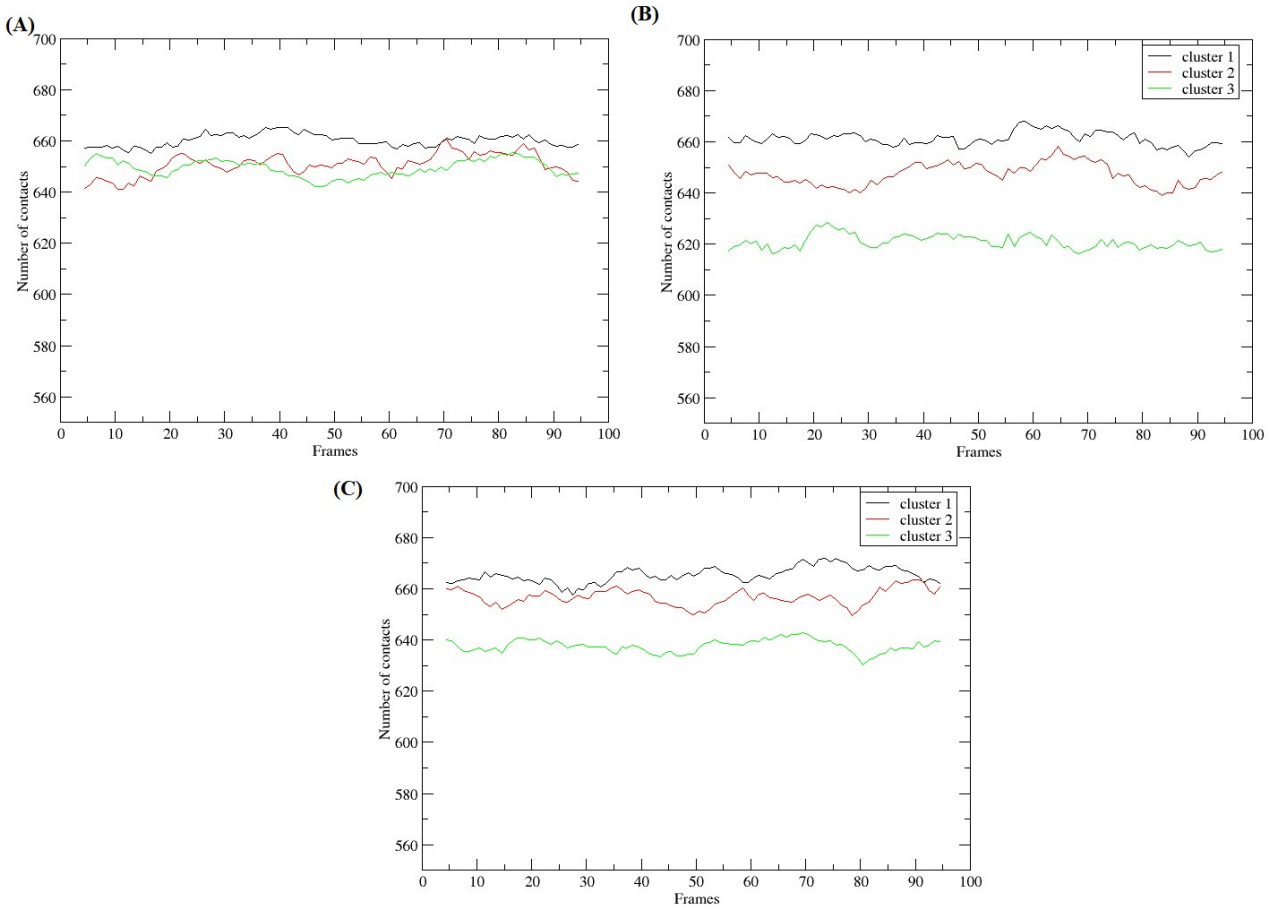

Figure S13 : Native contact plots for three clusters of all the three variants of MSM analysis

A) Wild type
B) E200K
C) E219K

Distance between Important Residues

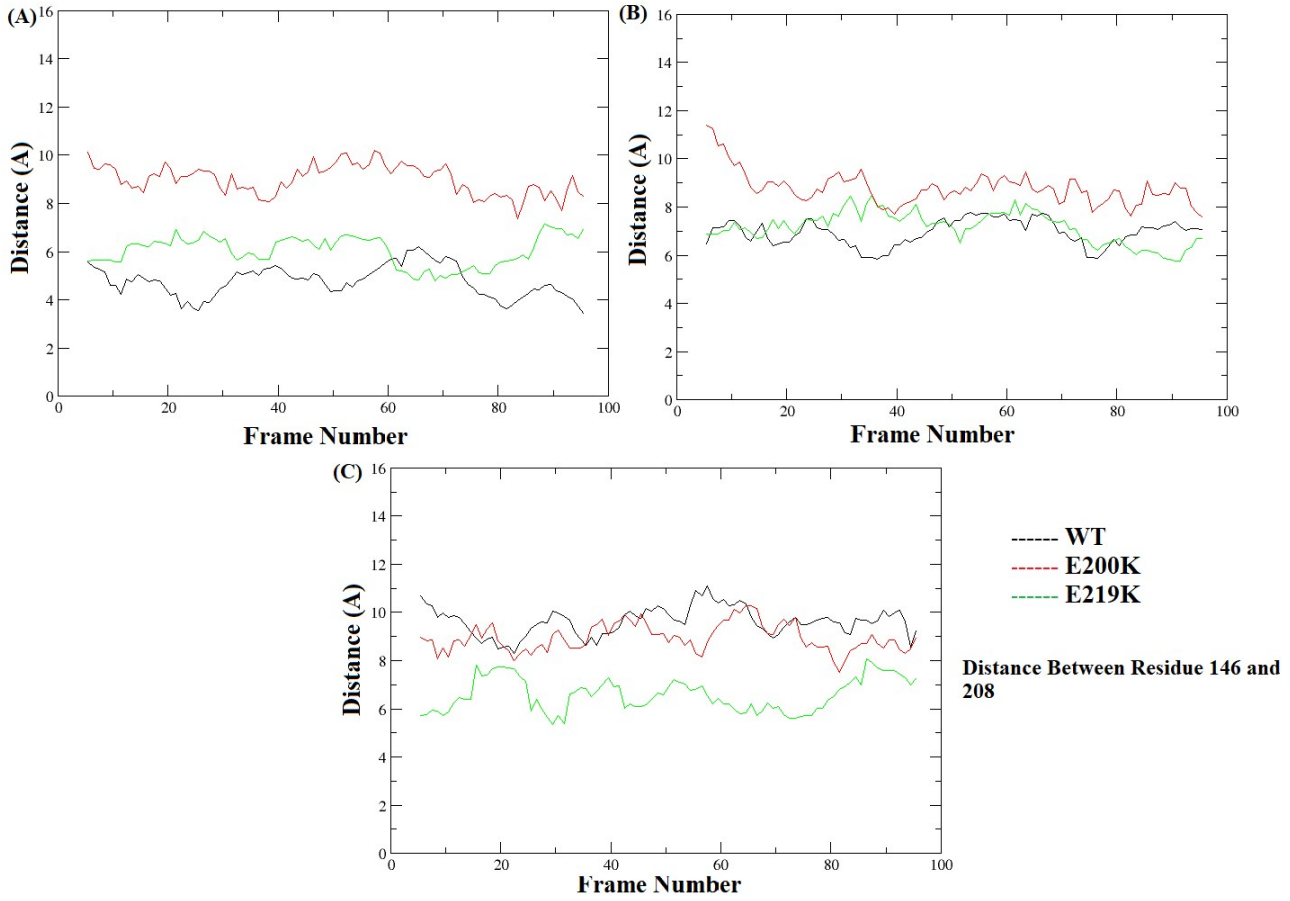

Figure S14 : Distance between Residues 146 and 208 for MSM states A) MSM state 1 B) MSM state 2 C) MSM state 3

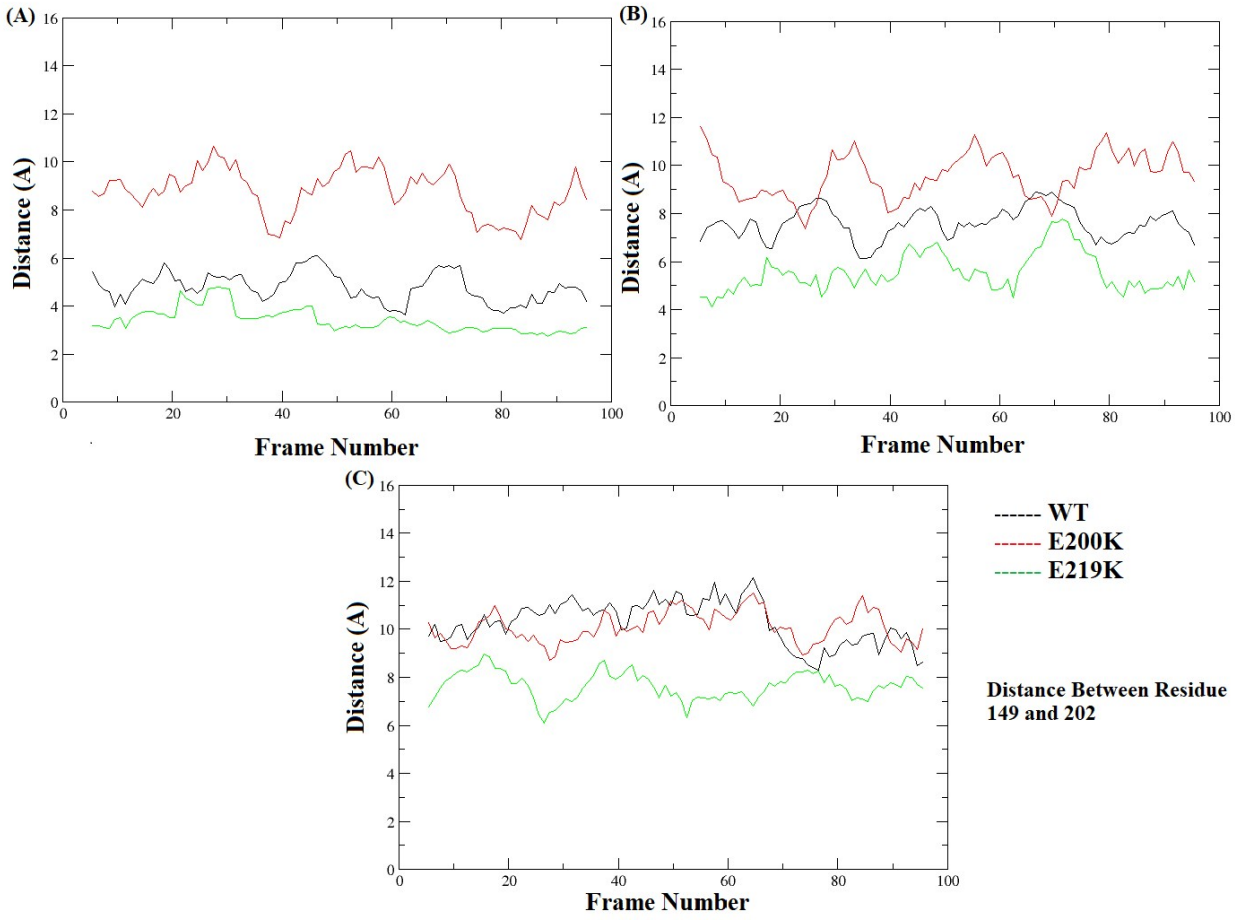

Figure S15: Distance between Residues 149 and 202 for MSM states A) MSM state 1 B) MSM state 2 C) MSM state 3

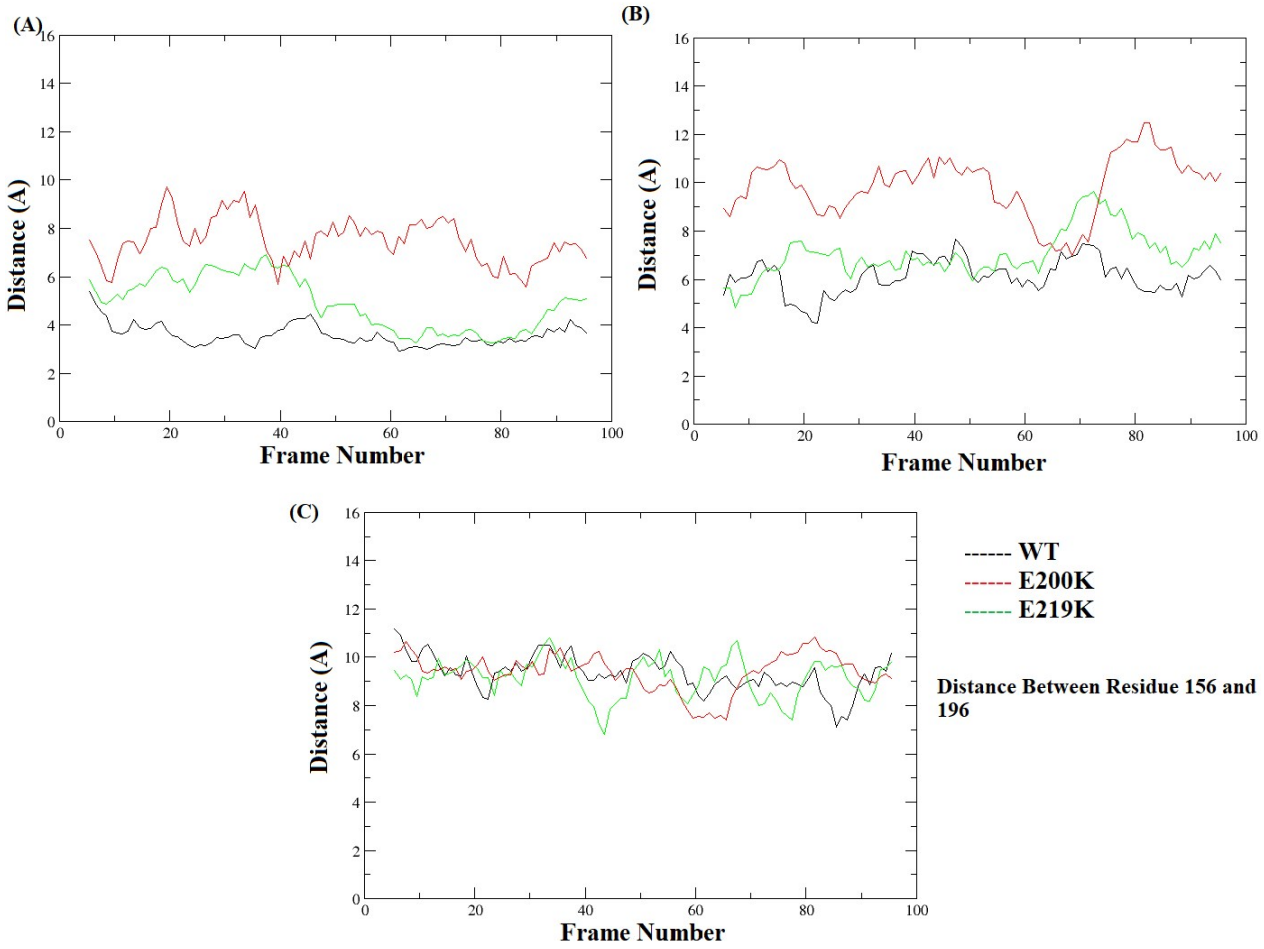

Figure S16: Distance between Residues 156 and 196 for MSM states A) MSM state 1 B) MSM state 2 C) MSM state 3

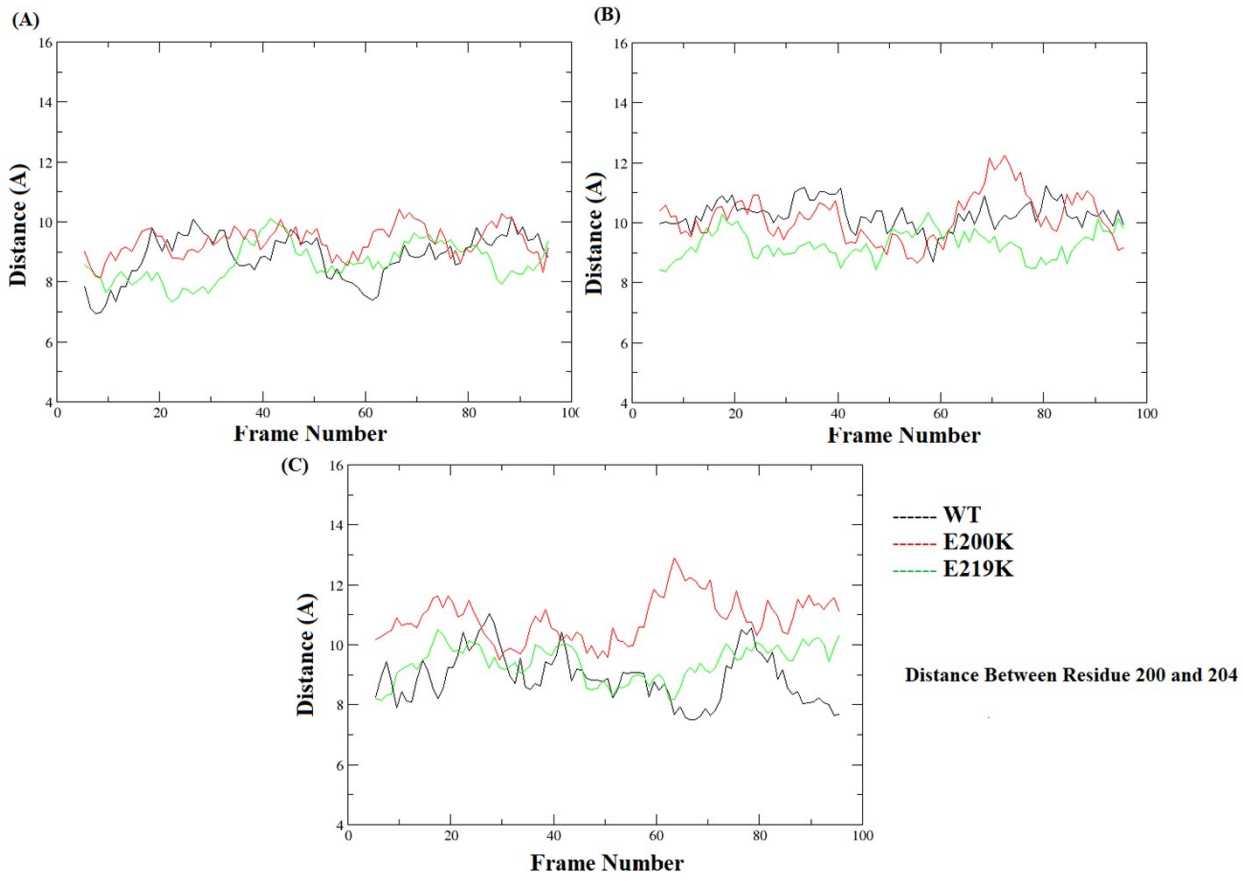

Figure S17: Distance between Residues 200 and 204 for MSM states A) MSM state 1 B) MSM state 2 C) MSM state 3