PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

C-Terminal Tail Insertion of Bcl-x_L in Membrane Occurs via Partial Unfolding and Refolding Cycle Associating Microsolvation

Atanu Maity[1,¥], Souvik Sinha[1,¥], Debabani Ganguly[2], Shubhra Ghosh Dastidar[1,*]

[1]Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VII M, Kolkata 700054

[2]Department of Chemistry, Indian Institute of Engineering Science and Technology, Shibpur, Howrah - 711103

*Address for correspondence

sgd@jcbose.ac.in

[¥]Authors contributed equally

Keywords: Bcl-xL, membrane, PMF, Molecular dynamics

Physical Chemistry Chemical Physics Accepted Manuscript

Abstract: Bcl-x_L, a member of Bcl-2 family of proteins remains distributed over cytosol and mitochondrial membrane maintaining a balance between apoptosis and survival of the cell. Passage to the membrane is essential for its biological functions (e.g. to antagonize pro-apoptotic proteins of Bcl2 family), which is known to be initiated by the insertion of the C-terminal segment into the membrane. This tail, comprised of ~24 residues, was reported to act as a pseudo-inhibitor of the protein itself, adapting helical conformation. It gets released from the confinement when the Bcl-x_L approaches to the membrane. This article, hereby reports the events associated with the insertion of a helical tail into an explicitly modeled all-atom membrane, which reveals a partial unfolding to refolding cycle of the peptide, correlating with the early insertion to a fully inserted state. The polar interactions have been found to have dominating role to steer the peptide towards the membrane in desired orientation. The landscape of the potential of mean force (PMF) is consistent with the proposed mechanism. Molecular dynamics further brings the insight that the peptide insertion associates an encapsulation of thin water layer around the peptide throughout the course of insertion, which motivates the protein to refold once the insertion is complete.

Physical Chemistry Chemical Physics Accepted Manuscript

Introduction:

Bcl-2 family proteins play one of the salient role in the process of programmed cell death[1,2]. Based on the function and presence of different homology domains, this family of proteins is sub-divided to two classes; one class has the anti-apoptotic function (for example Bcl-2, Bcl-x_L etc.), whereas the other class is of pro-apoptotic proteins which directly or indirectly promotes the cell death in response to different cell stresses, e.g. Bax, Bak (having BH1, BH2 and BH3) or Bid, Bad (having BH3 domain only). Balance between the expression of these proteins tipped off the favor towards or against the cell death process. Most of the Bcl-2 family proteins are generally tail-anchored to membranes of different organelles by the help of amphipathic C-terminal, whereas Bcl-x_L specifically targets mitochondrial outer membrane (MOM) [3] during its course of action. In a healthy cell, Bcl-x_L usually maintains a dynamic equilibrium between cytosolic and membrane embedded population[4,5], whereas such equilibration is influenced towards membrane associated population for the purpose of retro-translocating the pro-apoptotic partner (e.g. Bax) by recognizing and binding to its BH3 domain while apoptosis signaling flagged off.[6] So for a functional understanding of Bcl-x_L, it is important to understand its membrane-associated topology. For this purpose, Bcl-x_L has been studied in different membrane-mimicking environment like micelle[7], liposome[8], nanodiscs[9] etc. These have identified the existence of several helical segments, including a stretch of residues at the C-terminal of Bcl-xl to be transmembrane. Although different studies have identified different helices to be transmembrane, but C-terminal stretch (residue index 210-233) has responded as transmembrane in almost all the studies.[4]

In the reported structure of Bcl-x_L (PDB code 1BXL), [10] the residues 210-217 are mutated, and the residues beyond 217 are not reported. Reconciling available experimental evidences,

Physical Chemistry Chemical Physics Accepted Manuscript

using computation it has been reported to occupy the BH3-binding pocket as pseudosubstrate in apo Bcl-x_L[11]. The conformation of C-terminal has been characterized both in water-soluble form and membrane-inserted form by Yong Yao et al.[9] It has been found that the C-terminal forms a transmembrane helix inside the membrane bilayer and adopts one favorable tilt angle of ~25° with the bilayer normal and rest of the protein remains in the cytosolic part as a globular domain. Such structural data is in healthy agreement with other studies regarding the narrower orientation of transmembrane helix in apolar environment than in aqueous solution[12,13]. The sequence of the 24 residues (F210 - R233) stretch shows that its two ends are enriched with residues having titrable side chain, and when seen in the context of their secondary structure, these residues are strategically positioned by nature to interact with the charged surfaces of the membrane. Kaufman et al. have also proposed that such polar amino acids also provide significant assistance in mitochondrial targeting[3]. But before such a situation could be achieved, some of the charged residues have to pass through the thick hydrophobic slab for which the driving force is non-trivial to guess. From the available reports, it is understood that, when required, the C-terminal domain peptide (CTPEP) is released from the binding groove of Bcl-x_L and inserts itself into the MOM where it forms a structured segment. But the events, connecting a step where the CTPEP is released from the binding pocket and the step of attaining a stable transmembrane topology, conceive the mechanism of this insertion process. In this context of transition from one environment to another, there is a virgin set of questions hanging around just to be touched. How does the C-terminal of anti-apoptotic Bcl-x_L protein get inserted into the membrane? Does it preserve its secondary structure during the insertion? Is there any sequence of conformational changes occurring during the insertion and the post-insertion stages? *In-silico* approaches using molecular modeling and MD simulation can provide useful insight regarding such concerns.

Physical Chemistry Chemical Physics Accepted Manuscript

The membrane insertion of protein/peptide is well appreciated as a many-body complex process and there may be many different pathways of insertion carrying different probabilistic weightage. In general, a typical membrane can be robustly subdivided into two regions [14]; the hydrophobic region consisting of large aliphatic chains and the interfacial region, consisting of polar head groups. Starting from the bulk of the solvent up to the center of the membrane, it is 25 Å to 30 Å wide region depending on the lipid composition and this has a polarity gradient. Henceforth, profiling the mechanism of insertion of a peptide into a membrane bilayer has correlation with the bilayer depth [15]. Now, issues like, how the conformation of C-terminal is changing during the insertion process in response to variable environment and how does it affect the energy profile of insertion, become important and there lies the worth of this study. All–atom molecular dynamics simulation of the C-terminal stretch (F210-R233) of Bcl-x_L has been performed to obtain the energy profile of membrane insertion and its exclusive relationship with the structural dynamics. The experiments have reported the peptide to be helical in membrane [9] and it is unlikely that it folds after insertion, which has been also reported to be thermodynamically infeasible [16]. Therefore the peptide plausibly folds first in cytosol and then gets inserted; or in other words, out of the pool of all possible conformations present in cytosol, only the helical conformations are thermodynamically eligible to enter in the membrane. Eliminating the unfolding↔folding simulation of the peptides in cytosol due to infeasibility within a reasonable timescale, in this work, only the helical peptides were considered as the starting conformation to study the insertion [9].

There are many available methods for studying membrane insertion like Steered molecular Dynamics (SMD) [17], Umbrella Sampling (US)[18], Adaptive Biasing Force (ABF)[19] etc. The general principle is to allow the substrate for sampling at different distances along a specific

Physical Chemistry Chemical Physics Accepted Manuscript

user-defined reaction coordinate (RC) and then to calculate the potential of mean force along that coordinate. This sampling can be obtained by either forcing the substrate along the RC or by simply positioning it at different RC value and allowing an equilibrium simulation. SMD follows the first and unfortunately that is very unlikely for the present system of interest as the unrealistic 'pulling' would possibly disturb the secondary structure. On the other hand, keeping the dihedral angles of the peptide rigid would not serve the purpose of the present investigation. So, in this case, it is an optimal choice to allow the sampling of the peptide at different RC value in its predefined secondary structure from which the free energy profile can be obtained as a function of depth of the membrane where the peptide is allowed for a reasonable time to structurally relax and equilibrate. Here, ABF protocol has been chosen for the present investigation. ABF helps the system to avoid some kinetic traps and ensures efficient sampling in some of the higher and obscure valleys of the energy surface.

Method:

Modeling:

The structural coordinate of C-terminal peptide of Bcl-x_L (F210-R233) is not described in available structures (1BXL)[10]. But the similar domain of Bax (pro-apoptotic member of Bcl-2 family) with significant sequence similarity has reported structure (figure 1a) (1F16)[20]. Using Modeller[21], structure of the C-terminal domain of Bcl-x_L was modeled with respect to that of Bax. The modeled structure was capped at C and N termini using n-methyl and acetyl group respectively to prevent conformational trapping due to over-stabilization of the polar contacts formed by uncapped (charged) terminals. The peptide was then placed at different positions along the membrane normal in a pre-equilibrated DOPC bilayer containing 80 lipid molecules in

Physical Chemistry Chemical Physics Accepted Manuscript

each leaflet, obtained from CHARMM-GUI membrane builder[22]. The choice of DOPC was motivated by the abundance (54%) of PC type lipids in outer MOM [23] and to minimize the equilibration run (MD) time by choosing only one lipid type. For one-dimensional (1D) PMF calculation, peptide membrane assembly was solvated in a water box of size 96 Å × 96 Å × 150 Å; a larger box size of dimension 96 Å × 96 Å × 160 Å was used for two-dimensional (2D) PMF calculation. Required numbers of potassium and chloride ions were added to maintain an ion concentration of 0.15 (M) after neutralizing the head-groups and amino acid charges.

Molecular dynamics simulation:

For initial preparation and minimization, CHARMM bio-molecular simulation program [24] was used and simulations were run using NAMD 2.10[25]. CHARMM all36 parameter set was used for both protein[26] and lipid[27]. Initially, each system was minimized gradually using ABNR (Adoptive Biasing Newton-Raphson) and SD (Steepest Descent) method. After minimization, systems were equilibrated for 5 ns first using NVT and then NPT condition. The temperature was kept at 300 K using Langevin Dynamics with damping coefficient of 1/ps. Pressure was kept constant at 1 atm using Langevin piston method [28] with a coupling constant of τ_p = 0.5ps. All the bonds involving H-atom were constrained using SHAKE algorithm[29] and 2 fs integration time step was used consequently. Short-range non-bonded interactions were truncated at 14 Å. Particle Mesh Ewald method [30] was used to compute the long-range electrostatic interactions.

PMF calculation:

From the molecular dynamics simulation, PMF calculations were done using Adaptive Biasing Force algorithm implemented in the Collective Variable module[31] of the NAMD. During the production dynamics, a history dependent biasing force was applied to cancel out the running

Physical Chemistry Chemical Physics Accepted Manuscript

free energy gradient. The gradient of this force was stored along chosen collective variable (reaction coordinate), which was then integrated to get the free energy landscape

$$e^{-\beta W(Z)} = \int e^{-\beta V(X)} \delta_{\xi(X)-Z}(dX) \qquad (1)$$

where W(Z) is the free energy along z-coordinate, Vis potential which changes as V(X) – $W_t[\xi(X)]$ and W_t is an update that eventually converge with free energy W The Dirac delta function $\delta_{\xi(X)-z}$ (dX) is defined by the subset {X,ξ(X) = z} where ξ is the given reaction coordinate. Simply, it is just flattening the energetic barrier between different conformational states without much affecting the characteristics of the dynamics [19]. During 1D PMF calculation, the distance between the center of mass (COM) of the protein and the same of the membrane was chosen as the 'translocation' reaction coordinate (Z) (figure-1b). The total length of the reaction coordinate was 74 Å. Different systems, each for separate MD run, was prepared by placing the CPTEP as different Z values, starting from -56 Å (i.e. above the membrane) upto 18 Å (i.e. below of the membrane), with a distance increment of 2 Å. Each of such MD trajectory is referred as a simulation window and have been identified as w1, w2, w3, etc. upto w37 for the range Z=-56 Å to Z=18 Å. As the time for convergence is proportional to the square of the length of the transition path reaction coordinate was stratified with windows of size 2 Å with a force constant of 100kcal/ $Å^2$mol. [32] Each window was simulated for 10 ns using ABF module after the initial minimization and subsequent equilibration as mentioned earlier. So, a total of 370 ns of sampling using ABF were obtained to compute 1D PMF. All these windows have used 1000 MD steps prior to applying the bias. For exploring a specific region of conformational space (will be discussed in proper section later) more efficiently, another collective variable was

Physical Chemistry Chemical Physics Accepted Manuscript

chosen in addition to the previous one and then 2D PMF was calculated[33]. An angle between the helical axis of the peptide, the axis connecting center-of-mass of the lipid and a point on the helical axis was introduced as the new collective variable (figure-1c). The angle was varied from 90° to 180° where 90° inclination means horizontal orientation atop the membrane surface. Along the previous reaction coordinate, nine windows were considered using 2Å stratification and along the newly introduced reaction coordinate, three windows were considered using 30° stratification. Total twenty seven windows were simulated for 10ns using ABF module as before. An additional restraint was applied in these windows just to restrict the peptide from moving along xy-plane. All 2D windows have used 500 MD steps of sampling prior to applying the bias. Details of all windows simulated are provided in table S1 and S2. To get the 1D PMF along any one of the reaction coordinates from 2D PMF, following equation is used[34] –

$$e^{-\beta w(x)} = e^{-\beta w(x_c, y_c)} \frac{\int e^{-\beta w(x,y)} dy}{\int e^{-\beta w(x_c, y)} dy} \qquad (2)$$

where $\beta = (1/k_B T)$; k_B is Boltzmann Constant , T is absolute temperature, $W(x,y)$ is the 2D PMF for the range of (x,y), $W(x_c,y_c)$ is value of the PMF at any arbitrary point (x_c,y_c), $W(x_c,y)$ is value of the PMF for an arbitrary x-coordinate but throughout the range of y-axis.

Results and discussion:

1D PMF of the peptide insertion:

The plot of PMF, presented in Figure 2, appears to be minimal at the 'origin' of the reaction coordinate (Z = 0), i.e. at the center of the bilayer (in between window w28 and w29). Instead of a sharp barrier, a relatively flat region has been observed around ~38kcal/mol above the surface

Physical Chemistry Chemical Physics Accepted Manuscript

of the bilayer and while crossing this region, it seems to be a spontaneous insertion (gradual downfall in the plot). PMFs calculated from different time windows (i.e. different length of simulations, e.g. 1ns, 2ns etc.) have shown that the difference in energy values between two consecutive windows has decreased gradually and for the final two windows (8-9 ns and 9-10 ns) it is less than 2kcal/mol (figure S1). This proves the convergence of PMFs. The energy plot can be classified into three different regions up to the minima based on the nature of the plot: (a) 'association' region with almost a flat surface in the energy landscape where the peptide is predominantly present in bulk water, and interact with the membrane only through its most terminal two residues (reaction coordinate ranging from -56 Å to -38 Å) (b) 'absorption' region with a relatively flat slope in the energy plot where the peptide gets closer to the lipid surface and almost all residues are interacting with the head-groups (reaction coordinate ranging from -38 Å to -24 Å) and (c) 'insertion' region with a relatively steeper slope where the peptide crosses the head–group region and penetrate into hydrophobic core of the lipid (reaction coordinate ranging from -24 Å to 0 Å) (figure 2). The insertion profile also depicts that the slopes in the upper and lower leaflets are not symmetric; rather in the upper leaflet it is steeper than that is in the lower leaflet. This is because during the passage through lower leaflet, the peptide is already stabilized by interaction with both the upper and lower head-group regions using its polar terminal residues, whereas during insertion process through the upper leaflet, stabilizing interactions changes abruptly, particularly when it crosses the upper head group region and gets exposed to the hydrophobic core of the membrane.

The flat plateau region at ~38 kcal/mol corresponds to the 'association' of the peptide upon bilayer head and here distance of the peptide from the membrane is such that the peptide is mostly exposed in water. From experiment [9], it is known that the CTPEP(C-terminal peptide of

Physical Chemistry Chemical Physics Accepted Manuscript

Bcl-xL) can experience varying degrees of flexibility in water. To check that possibility, a relatively unstructured helix (a helical hairpin, obtained at the end of a 100 ns simulation of the free CTPEP), was set as a new starting structure to rerun the simulations in the "association" and "absorption" regions to rescan the PMF in that range (details available in Supporting Information). As shown in the Figure S2, this exercise has only smoothened the PMF landscape, and have slightly changed the slope of the PMF in the 'absorption' range, without elucidating any significant energy barrier. Therefore, to have a consistent reaction coordinate of membrane insertion across all windows, the helical CTPEP was used as the starting structure everywhere.

In the absorption range, the vertical approach of the peptide towards the membrane has not been maintained during the simulation, i.e. the intermolecular interactions did not prefer to maintain vertical approach. In such a situation, a wide range of relative orientation of the peptide and membrane plausibly appeared energetically equivalent (approximately), which is consistent with the horizontal nature of 1D PMF in this range. Therefore it is insufficient to obtain an insight of the driving forces from a one-dimensional plot of PMF. To explore such possibility the calculations in that region (38 Å to 56 Å above the membrane center) were repeated adding another reaction coordinate in the PMF, which is the angle (θ) between the peptide axis and the membrane z-axis (an axis connecting upper to lower leaflet) (figure 1c). The peptide orientation with respect to membrane was stratified into three separate angle ranges (θ = 90˚ to 120˚, >120˚ to 150˚ and >150˚ to 180˚) in between horizontal (θ = 90˚) and vertical (θ = 180˚) orientations.

2D PMF of the peptide insertion:

Physical Chemistry Chemical Physics Accepted Manuscript

The change in free energy profile with respect to the two coordinates considered for 2D PMF calculation was presented as a histogram in figure 3. The consideration of the second collective variable has revealed the importance of sampling the possibility of different angular orientations. When the peptide is closer to the membrane (Z ~ -38Å to -42Å), continuously distributed closely-spaced low-energy conformations along wide range of inclination angle (more precisely θ ~120˚ to 160˚), have been sampled. But conformations far away from the membrane (Z ~ -50Å to -56Å) are showing minima more towards slightly tilted horizontal orientation (θ ~ 110˚ to 130˚). Moreover, these distant low-energy conformations are not-at-all distributed continuously but as separated islands in the energy landscape. In between these closer and distant low-energy regions, horizontal orientations (θ ~ 90˚ to 110˚) are favored around Z ~ -42Å to -50Å. The difference between the low-energy states lies within the range of 2-3kcal/mol. When the peptide is far away from the membrane it prefers a nearly horizontal orientation (for reference, membrane surface is set as horizontal) (figure 3a(i)), plausibly because at these distances only electrostatic interaction (being long ranged) is the driving force and the charged residues at the C and N terminus both get attracted. There are previous studies on ligand-protein complexes where the role of electrostatics has been proposed to be significantly strong for large separation that steer the moieties to come closer to each other and leading to complex formation[35]. To get the insight of forces driving the insertion events in the system, the energy of interaction between the peptide and membrane has been calculated and electrostatic and van der Waals' components were distilled out (figure 3b and 3c). The electrostatic interactions were observed to be the dominating component not only to attract the peptide towards the membrane but also to influence their relative orientations. At a larger distance of separation, the horizontal orientation of peptide (with respect to membrane surface) is largely favored. Being in agreement with the 2D PMF

Physical Chemistry Chemical Physics Accepted Manuscript

plot, electrostatic component widens the preference zone of conformation (θ = 120˚ - 160˚) (figure 3a (iii)) near the membrane, i.e. in the 'absorption' region. These angular orientations of the peptide are suitable for getting inserted into the membrane, plausibly by disturbing least number of lipids. For more confirmation, 2D PMF was integrated according to equation 2 with respect to orientation angle and obtained a 1D PMF along translocation coordinate (figure 3d). Though the energy difference is in 2-3 kcal/mol range, this 1D PMF is now showing troughs and ridges in a zig-zag landscape. It confirms the presence of closely spaced conformations with similar population distribution in the association region that causes the absence of clear hills and valleys in 1D PMF calculation.

Energy barrier: Overall, the PMF calculations did not reveal any energy barrier as such, which is contrary to the traditional reports and concepts of a process. In general, energy barrier separates two stable states. But the absence of a barrier in this plot is plausibly because the event of membrane insertion is only a fragmented part of the entire biological event that involves two stable states; one is the membrane inserted state, the other one is the complexed state where CTPEP is stable as a pseudo-substrate of the Bcl-x_L. In a previous report by Maity et al. [11], it has been already shown that CTPEP bound to the pocket of Bcl-x_L is a stable state. Therefore, it is rational to hypothesize that the release of the CTPEP from the binding pocket of Bcl-x_L refers to an energy barrier and once this barrier is crossed, the insertion to the membrane is a spontaneous process.

Effect of membrane environment on secondary structure:

As appeared from the initial model, the secondary structure of the 24-residue peptide was helical. As expected, the secondary structure got disrupted once it started to get inserted into the membrane and the extent of change in structure is correlated with the depth of insertion. These

Physical Chemistry Chemical Physics Accepted Manuscript

changes in the structure took place within the first 2 ns of unbiased equilibration run and it got optimized within this interval, therefore the ABF calculation on the rest of the MD simulation was expected to be free from positional biasness and artifacts.

In the '*association'* region, as the direct interaction of the peptide is very less or only through some terminal residues, the helical integrity is maintained more or less. Residues F214 to F230 are in α-helical or 3_{10} helical conformation and the residues in two terminals are unstructured. The small changes like conversion of helix to turn or helix to coil during the simulation are mainly due to interaction within the peptide (figure 4). Both horizontal (figure 4a) and vertical (figure 4c) orientation of the peptide are found in this region and in some windows the helices also bend to form a hairpin of helix (figure 4b). The variation in orientation emphasizes the non-specific protein-lipid interaction at a distant position only through the terminal residues.

In the *absorption* region, although the peptide is closer to the membrane head-group and more residues are now interacting with the charged head group, the helical assembly is maintained in the initial windows by introducing a kink between two helical parts. One of these two parts is engraved into head group region and the other is associated with water just above the membrane or the first hydration shell as described by Marrink et al. [36]. The position of the kink depends on the reaction coordinate i.e. distance between the center of mass of protein and membrane. Due to the competition between maximizing interaction with membrane head-group and maintaining helical integrity, helices are stabilized in a 'chair-like' orientation on the membrane surface. Only in the last part of absorption (window w17), since significant extent of the peptide had already been entered into head-group region, fourteen residues from C-terminal became unstructured and the N-terminal (W213 to V220), which is outside the head-group region, remains in α-helical conformation (figure 4d). This leads to a 'coil-helix-coil' motif with the

Physical Chemistry Chemical Physics Accepted Manuscript

second coil embedded into membrane head. Thus during its initial approach to membrane, CTPEP gets unfolded and its C-terminal get stabilized at the charged head-group region. The *insertion* region is associated with the most interesting change/switching of secondary structure at different extent of insertion. The very first window (w18) has similar kind of story as in window w17, only the helical segment has now been shifted towards the C-terminal (T219-Y223) (figure 4e). In the next window (w19), as the peptide moves downwards by 2 Å, the previously helical part overlaps with the charged head group and as a result that turned into a coil to form two shorter helices at its two ends (F210-F214 and V224-F230) (figure 4f). So starting from a 'coil-helix-coil' structural motif in the association region, it started inserting as a 'helix-coil-helix' motif. This was also maintained in next two windows (w20 and w21) with only exception of residues in first helix. In window w22, the extent of insertion forced the N-terminal helix to overlap with head-group and helicity was lost. At the same time, the C-terminal helix propagates more towards the N-terminal (T219-F230) as more residues are now embedded into the hydrophobic core of membrane. In the next window (w23), as the N-terminal comes closer to the upper layer head-group, the terminal unstructured region now forms a small helix and gets oriented horizontally on top of head group surface, that introduced a two residue coil between two helical segments (figure 4g). But that helix also dissolves just in the following window (w24) and the residues of the small helix and some residues of the C-terminal helix becomes unstructured (figure 4h). While moving to the next window (w25), it was advantageous for the peptide to accommodate itself in the membrane in such a way that the two helical segments meet together and form a larger helix (L215-F230) (figure 4i). Thus the 'coil-helix-coil' motif was restored but this time with the helix inside the bilayer. But even this largely helical conformation is not the most stable state rather the energy plot moves downward with a characteristic

Physical Chemistry Chemical Physics Accepted Manuscript

secondary structure that has two unfolded regions connected by an alpha-helical segment. In the window (w28) before the minima, residues (L225-F230) towards C-terminal become 3_{10} helix / turn from alpha helix conformation (figure 4j**)**. In the window (w29) after minima, secondary structure/helicity towards C-terminal was gained whereas same was lost from N-terminal residues (L215-T219) (figure 4k). The maximum stability is a combined effect of the hydrophobic interaction of the helical portion within the hydrophobic core and electrostatic interaction of anchored charged residues with the charged head groups. In the windows near minima, the tilt angle with the membrane normal is found to be ~30˚ which is close to the experimental reports of Yao et al.[9] So, as the peptide goes deeper into the membrane, helix propagates from N-terminal to C-terminal and consequently, the position of the coil gets shifted towards N-terminal. Continuation of this process leads to a structure where the helices from two ends merge and attain maximum helicity within the hydrophobic region of the lipid. Now this large helix gets anchored to the head groups by the terminal charged residues. So one hypothesis could be built from here that during the insertion process, the peptide will get unfolded from C-terminal to N-terminal sequentially and finally inside the bilayer, the helical structure will be reconstituted again.

Water carried out into hydrophobic core:

Though macromolecules need to have some charged assistance to cross through the hydrophobic core, small molecules like water can easily permeate through membrane hydrophobic core in some basal pathways [37]. This permeation process has a contribution to the membrane physiology and in some cases, it has keen biological relevance[38,39]. To assess this in case of CTPEP insertion, the fractions of total number of water molecules that are present at different distances along the perpendicular direction with respect to membrane surface were calculated. In figure 5a,

Physical Chemistry Chemical Physics Accepted Manuscript

the blue and orange dots are describing the position of phosphorus and nitrogen atoms respectively at two head group, and the region in between them is the hydrophobic core of lipid tail. The fraction of water in the lower leaflet is more in the case of window w19, whereas window w37 is showing greater fraction of water in the upper leaflet of the membrane (figure 5a). But, in window w27, where the helix is completely inserted with maximum helical content, no clear preference is observed in any of the leaflets in terms of confined water fraction. This observation will be more precise from the figure S3 (a)-(c). Window w27 (figure S3b) is clearly showing lesser number of water around 4 Å of the protein embedded within the hydrophobic core (hydrophobic core is defined to be the region in-between the carbonyls of lipid molecules in the upper and lower leaflet) throughout the range of simulation in comparison to window w19 and w37 (figure S3a and S3c respectively). The lower leaflet water number is increasing towards the end of the simulation for window w19, whereas the upper leaflet is showing less significant changes (figure S3a). But, both have driven the increased water population towards the end of the simulation. This same thing happened in case of window w37, but more contribution is led by upper leaflet water count (figure S3c). It was more insightful to look into the radial distribution of water molecules around amino acids in the three windows (figure 5b-d). In window w19, the maximum water density at ~2.8 Å was concentrated around C-terminal R232 and K233 in lower leaflet. The density was intermediate around the residue at the center (G222 and V223) as their backbones are involved in water mediated helix formation inside the upper leaflet. The lower density around N-terminal residues (F210 and N211) was due to high exchange rate of water around them with bulk water (figure 5b). In w27, due to a compact transmembrane placement of the peptide, both N and C-terminal residues (R232, K233, F210 and N211) are positioned near the lipid the head-group region resulting, similar density (3.5-5.5)

Physical Chemistry Chemical Physics Accepted Manuscript

of water around them whereas the central residues (G222 and V223) are inaccessible to water (figure 5c). The scenario for w37 is just opposite to w19 and the water distributions correlates with the amino acid distribution in membrane: N-terminal (at upper leaflet) > center (at lower leaflet) > C-terminal (exposed to bulk water) (figure 5d). The specific role of water inside lipid bilayer is either to solvate the sidechain / backbone or to fill the blank or vacuum created due to conformational change of the protein.

The decay of residence time correlation function of the water solvating the peptide inside the lipid bilayer has been presented in the Figure 6. The correlation function C(t) which measures the probability of water to remain within some distance criteria for time t, was calculated using the following expression [40, 41] –

$$C(t) = \frac{1}{N}\sum_{j=1}^{N_w}\frac{\left\langle P_j(0)P_j(t)\right\rangle}{\left\langle P_j(0)^2\right\rangle} \qquad (3)$$

Where the $P_j(t)$ is 1 if the *j*'th water molecule is within the distance criteria at time t and $P_j(t)$ is 0 it is outside that distance range. The rate of the decay of the correlation differs in different windows; data of three representative windows w19, w27 and w37 have been shown to represent three different degree of insertion of the peptide as well as different extent of secondary structure. Among these three windows, w27 had the maximum secondary structure content (and closest to the PMF minimum) and therefore the formation of intra-peptide H-bonds releases the bound waters from the peptide surfaces faster in this case compared to other windows. This is why the residence time correlation decays much faster in w27, whereas formation of protein-water Hbond with higher probability (as the intra-peptide H bonds are less) in w19 and w37, decays the same much slowly.

Physical Chemistry Chemical Physics Accepted Manuscript

.

Thus water molecule is bridging the gap between peptide terminal and lipid head group inside the bilayer. Johansson et al. have shown the effect of solvation to amino acids during membrane insertion and specifically characterized the side chains of a fully inserted transmembrane helix accordingly [42]. Our observation is not only in agreement with that but also reveals mechanistic insights that explain the need of micro-solvation for membrane integration of peptide.

Role of charged residues in insertion:

The C-terminal charged residues arginine and lysine had a significant role to help the insertion starting from its initiation at one side of the bilayer to its exclusion on the other end of the bilayer. At the very initial step of insertion, the positively charged nitrogen rich side chains are found to stretch itself towards the lower leaflet head-groups and soak in some water inside the membrane to get stabilized via micro-solvation (figure 7a). As the peptide penetrates into the membrane, the extent of stretching required for this kind of stabilization gets reduced. Once the peptide is inserted to maximum extent, both the side chains can directly interact with oppositely charged groups in the lipid surface (figure 7c). This gradual change in the requirement of stretching for stabilization is reflected in the RMSD (root mean square deviation) of R232 and K233 in the insertion windows (figure 7b). In the region away from the membrane, these charged residues face both positively charged acetylcholine group and negatively charged phosphate group. As a combined effect of opposing electrostatic interactions, C-terminal in this region experiences random fluctuations rather than having any characteristic trend in compaction reflected in radius of gyration measure (figure S5).

Water molecule helps in helix propagation:

Physical Chemistry Chemical Physics Accepted Manuscript

If we look at the helical content of the peptide it varies not only in quantity but also in the specific residues, which were part of the helix. Depending on the depth of the bilayer, peptide is found to lose its secondary structure during initial equilibration stages, which were regained later to some extent depending on the membrane environment around protein. The reconstitution of the helical segment was guided by the interaction between amino acids as well as by some intermediating water molecule caged within the peptide backbone. This kind of structural reconstruction was found in windows where the peptide is hanging around the interface of water and bilayer. One such observation is for window w19, where the rigid stretch gets unstructured during equilibration (figure 8a). But after a while, around 4 ns, some water molecules invaded this uncoiled region by the help of lipid polar head group and were intermediating S228 side chain and backbone carbonyls of V224 and L225 for formation a 3_{10} helix (figure 8b and 8c). This continues up to 8.5 ns and gradually the amino acids get enough time to attain favorable orientation for helix formation. At this point, S228 can directly interact with the backbone of V224 with its side chain hydroxyl group and form an alpha-helical stretch (figure 8d and 8e).

Conclusion:

A series of events accomplishing the insertion of C-terminal tail of Bcl-x_L (CTPEP) in membrane has been presented in this report, which complements the biochemical evidences. Starting from the situation where the helical CTPEP is freely available in solution (i.e. released from the binding pocket of Bcl-x_L in a real biological context) to approach towards the membrane surface, the computed PMF reveals it to be a downhill movement on the energy landscape, i.e. indicates it to be spontaneous. The 1D and 2D PMF plots together have revealed how and to what extent the translation of the peptide associates a preference for the orientation to approach towards the membrane surface, which has a relation with the natural design of its sequence (i.e. how the

Physical Chemistry Chemical Physics Accepted Manuscript

charged residues have been distributed on its both ends); the electrostatic interactions dominates to steer the peptide towards the membrane. While passing through the membrane, the peptide carries a thin layer of water jacket, which stabilizes the polar functional groups. Such micro-solvation has been found to have a role to re-stabilize the intra-protein polar interactions which gets disrupted as the peptide experiences strong perturbation of environment-change; first, it crosses the charged head groups of the membrane and then immediately gets exposed to the hydrophobic core. So far the authors are aware, this work for the first time, presents the micro-solvation mediated peptide insertion mechanism into membrane with a quantitative estimation of the thermodynamic feasibility of the mechanism.

Acknowledgement:

The work is funded by the project no. BT/PR793/BID/7/370/2011 sanctioned by the Department of Biotechnology (DBT), Government of India. For funding, AM thanks CSIR (Council of Scientific and Industrial Research), Government of India and SS thanks UGC (University Grant Commission), Government of India. DG thanks DBT, Government of India for funding through Ramalingaswami fellowship scheme.

Physical Chemistry Chemical Physics Accepted Manuscript

Figures:

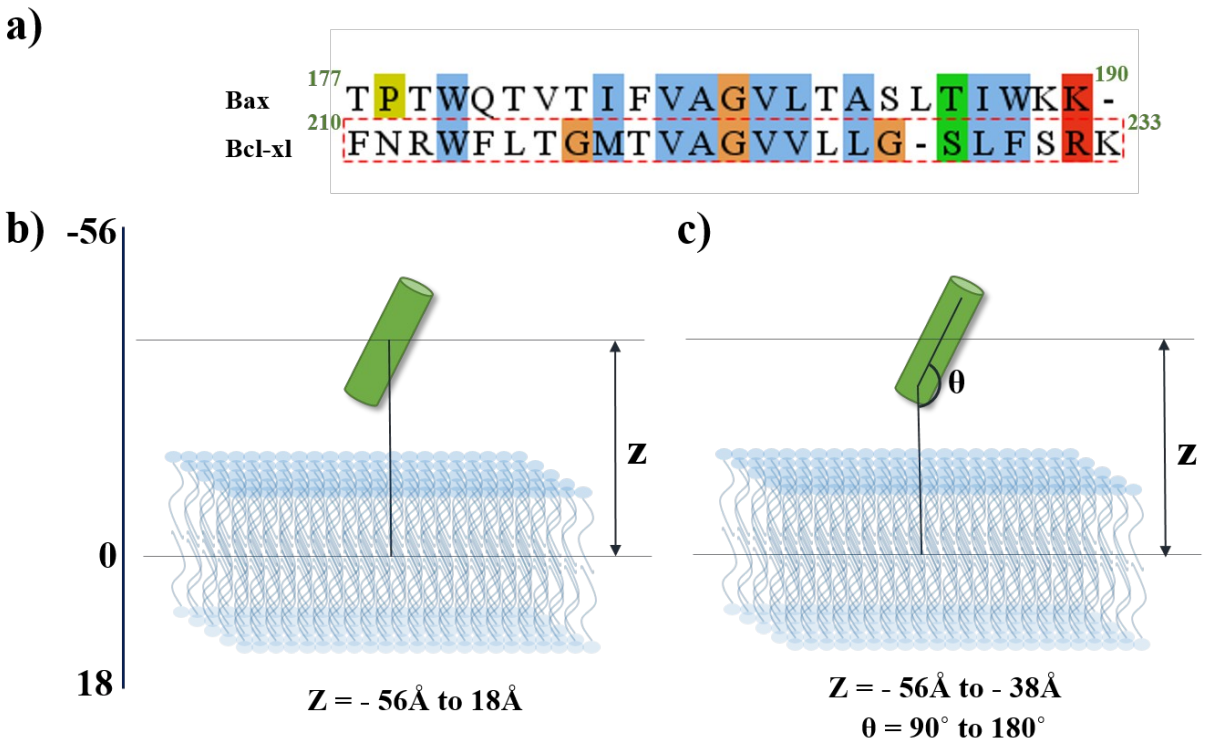


Figure 1 : (a) Sequence alignment of the C-terminal residues of Bcl-x_L (Uniprot id Q07817) and Bax (Uniprot id Q07812). **(b)** Definition of reaction coordinate 'Z' is the distance between center of mass (COM) of membrane and COM of peptide along the membrane normal. Z = -56 Å to 18 Å has been scanned to construct the 1D PMF. **(c)** 'θ' is the angle between the helical axis of the peptide and the axis connecting COM of the lipid and a point on the helix axis. θ= 90° to 180° has been scanned to compute the 2D PMF in the range of Z = -56Å to -38Å.

Physical Chemistry Chemical Physics Accepted Manuscript

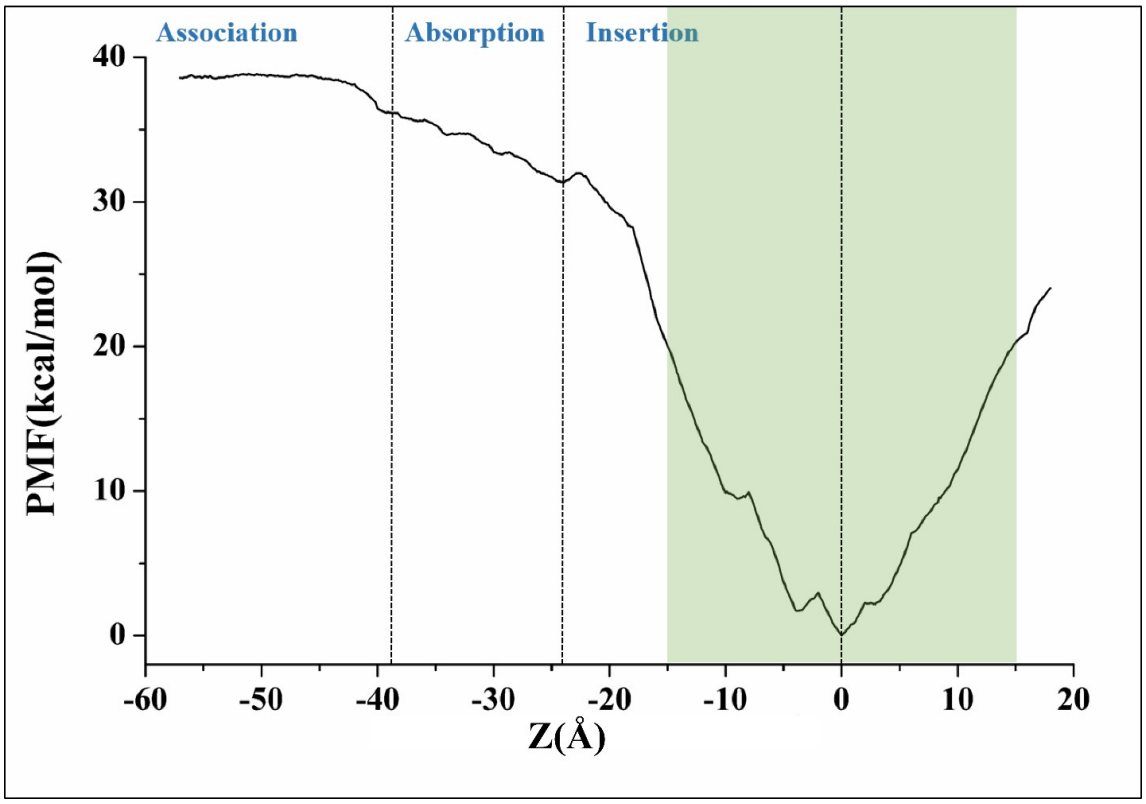


Figure 2 : 1D PMF profile of the insertion of peptide in the membrane, computed over 9-10 ns of simulation of each window (w1 to w37). The horizontal axis, 'Z', has been defined in figure 1. The ranges of Z corresponding to 'association', 'absorption' and 'insertion' events have been indicated by dashed lines. The light green shade describes the region of lipid bilayer. Further details are available in supporting information Figure S1.

Physical Chemistry Chemical Physics Accepted Manuscript

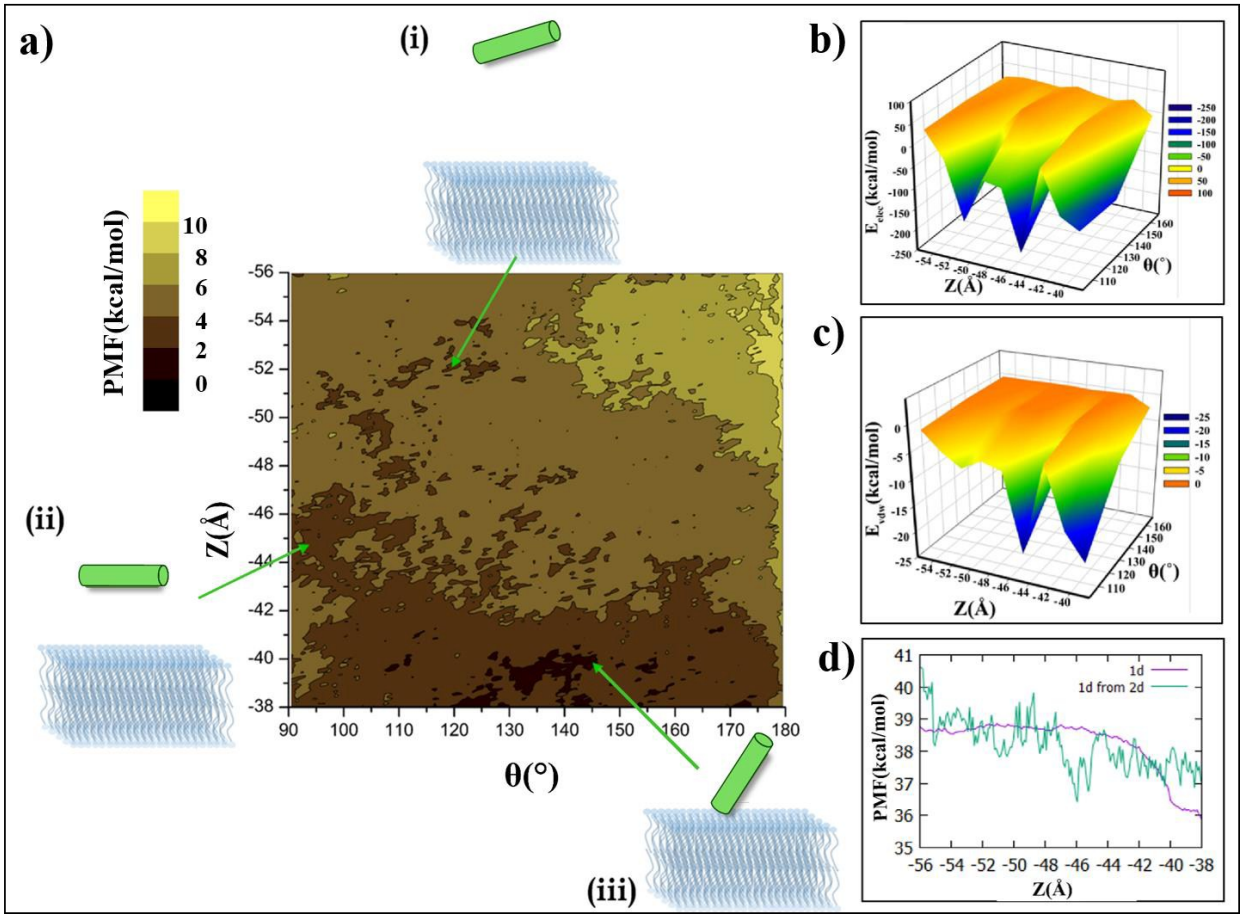


Figure 3: (a) 2D PMF plot of peptide entering into the membrane, shown using a black to yellow color scale. The axes, Z and θ are defined in Figure 1; a few situations for different θ have been shown in (i), (ii) and (iii) panels. Energy surfaces of (b) electrostatic interactions and (c) van der Walls interaction between the peptide and lipid bilayer as a function of Z and θ. The color scale blue to red corresponds to the lower to higher values of interaction energy. (d) Overlay of 1D PMF and the same extracted from the 2D PMF data points.

Physical Chemistry Chemical Physics Accepted Manuscript

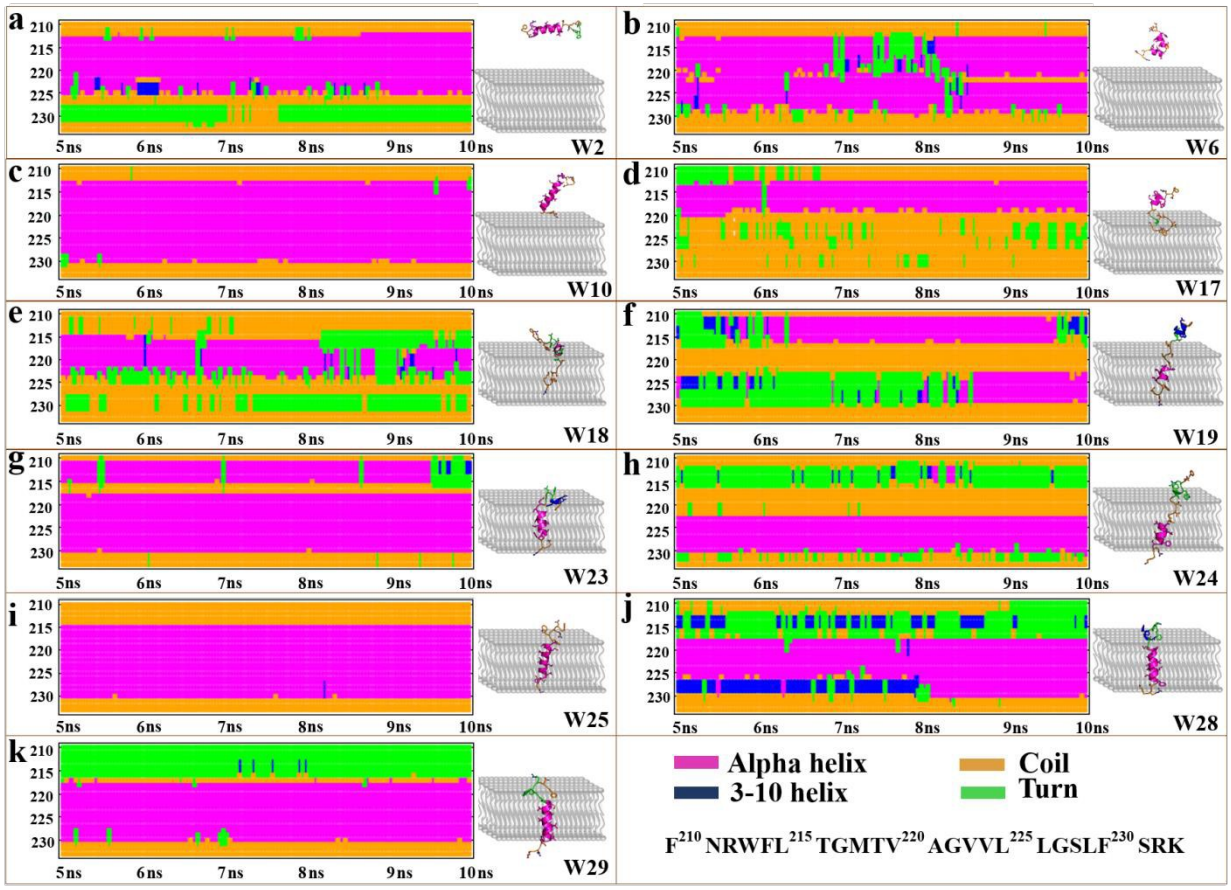


Figure 4 : Evolution of secondary structural elements of the peptide dueing 5-10ns in representative windows: (a) w2, (b) w6, (c) w10, (d) w17, (e) w18, (f) w19, (g) w23, (h) w24, (i) w25, (j) w28 and (k) w29. Structures of the peptide at the end of the 10 ns simulation of each window are presented at right side of the corresponding secondary structure plot and have been overlayed on a cartoon representation of the membrane to show the degree of insertion. The secondary structure color code and sequence of the peptide have been provided at the right-bottom.

Physical Chemistry Chemical Physics Accepted Manuscript

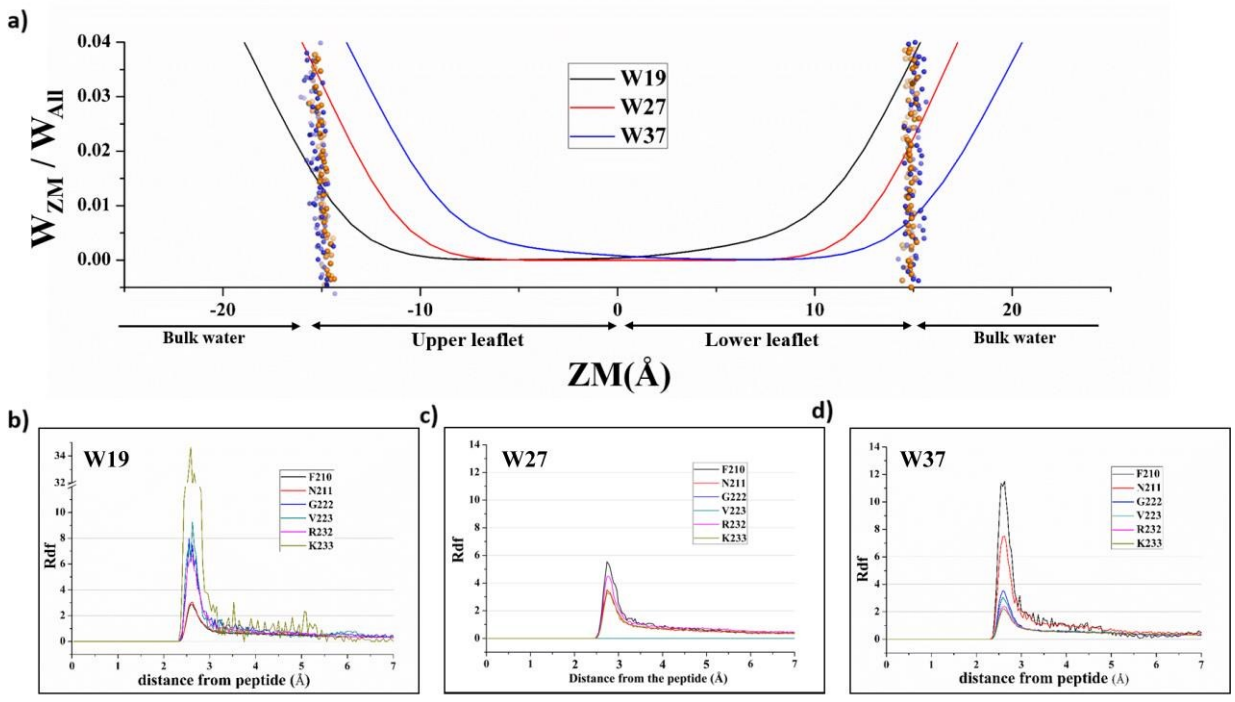


Figure 5: Plots showing distribution of water: **(a)** Fraction of total water, W_{ZM} / W_{All} (where W_{ZM} is the number of water molecules within the distance range (ZM - 2.5) Å to (ZM + 2.5) Å along the membrane normal and W_{All} is total number of water) is plotted against the distance (ZM) along the membrane normal, averaged over last 5ns simulation in three representative windows. Atoms of the lipid head-groups have been shown in Orange (Phosphorous) and Blue (Nitrogen). (b), (c), (d) Radial Distribution Function (Rdf) of waters around different residues of the peptide, averaged over last 5ns simulation in three representative windows w19, w27 and w37 respectively.

Physical Chemistry Chemical Physics Accepted Manuscript

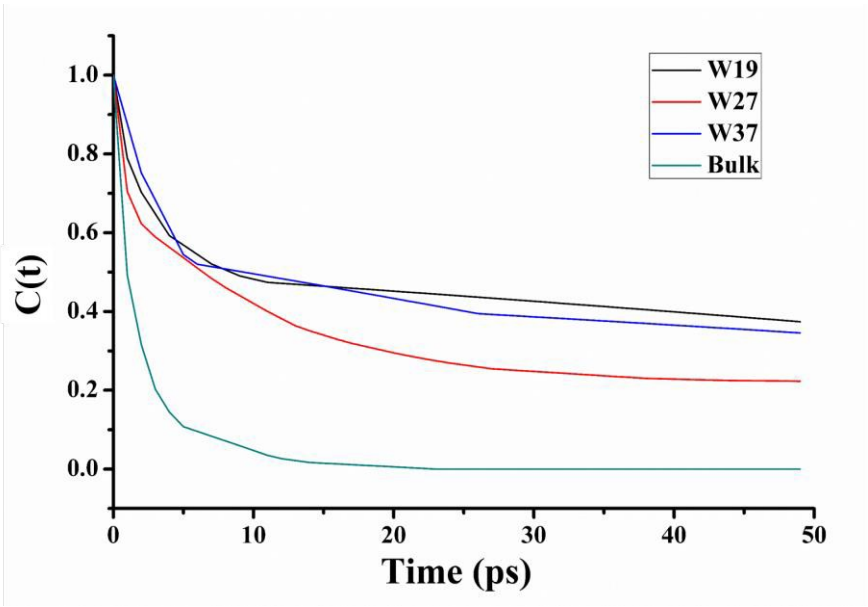
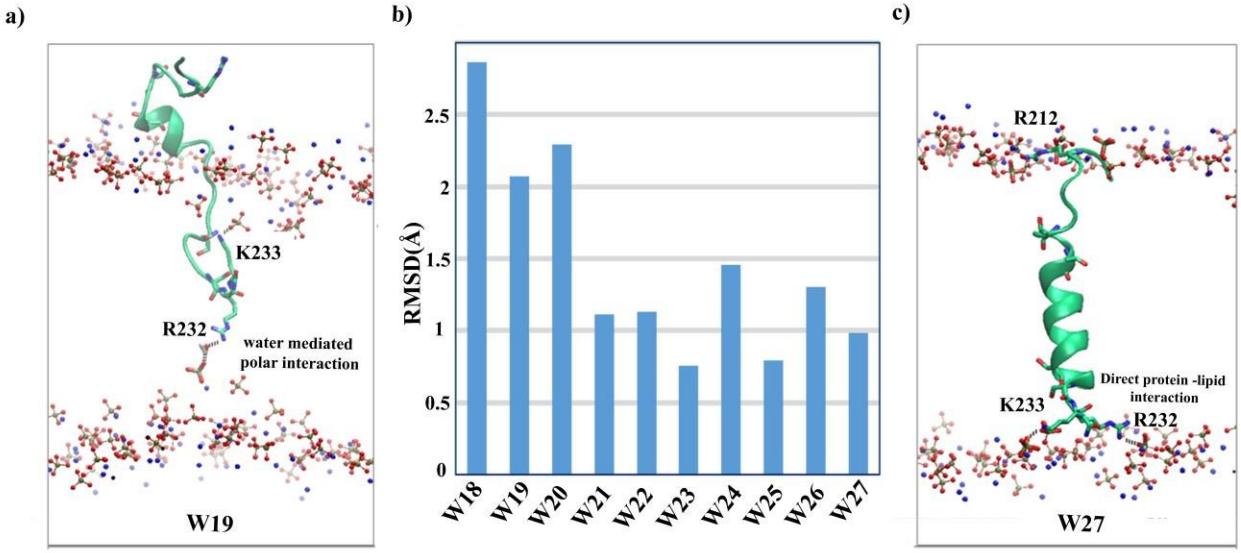


Figure 6: The decay of residence time correlation function C(t). The water molecules present within a 8 Å cutoff of the peptide inside the hydrophobic region of the membrane were monitored as a function of time to compute the correlation function; the data has been averaged over different time origins. C_t has been shown for three representative trajectories. The bulk refers to the same for the water around the same peptide solvated in water, in absence of membrane.



Physical Chemistry Chemical Physics Accepted Manuscript

Figure 7 : (a) and (c) show the polar interaction of R232 and K233 with the lipid head group in two representative windows w19 and w27 respectively. The atoms of the lipid head-groups have been shown in yellow (phosphorus), red (oxygen) and blue (nitrogen). Peptide is represented in green cartoon and few important residues are shown in sticks. (b) Average RMSD of R232 and K233 over last 5ns simulation in few representative windows (w18 – w27).

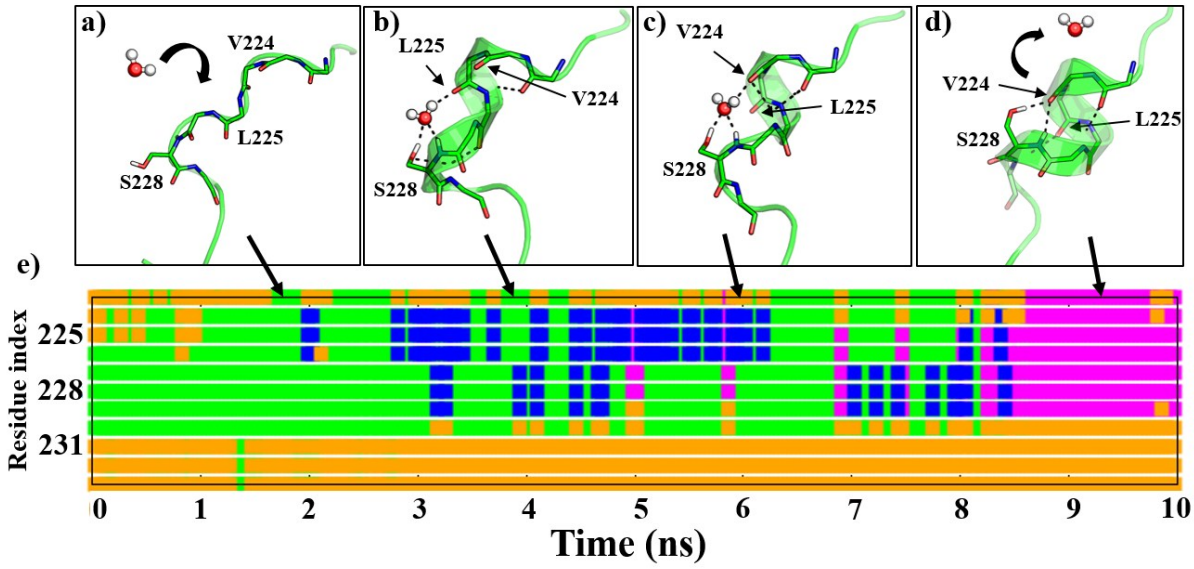


Figure 8: Representative snapshots showing that a single water moelcule is catalyzing the secondary structure formation. Figures (a) – (d) describes a gradual change from a random coil to the formation of one complete turn of α-helix. The water molecule initially bridges the backbone interactions (between L224/V225 and S228) to bring them close to each other and then the water is released once the residues can directly form i, i+4 Hbonds. The secondary structure evolution has been shown in (e) and the corresponding time points of the snapshots (a - d) have been indicated with arrows; the color scheme for secondary structure elements is same as in figure 4.

Physical Chemistry Chemical Physics Accepted Manuscript

References:

1. A. Gross, J. M. McDonnell and S. J. Korsmeyer, *Genes Dev*, 1999, 13, 1899-1911.
2. J. M. Adams and S. Cory, *Science*, 1998, 281, 1322-1326.
3. T. Kaufmann, S. Schlipf, J. Sanz, K. Neubert, R. Stein and C. Borner, *J Cell Biol*, 2003, 160, 53-64.
4. F. Edlich, S. Banerjee, M. Suzuki, M. M. Cleland, D. Arnoult, C. X. Wang, A. Neutzner, N. Tjandra and R. J. Youle, *Cell*, 2011, 145, 104-116.
5. B. Schellenberg, P. Wang, J. A. Keeble, R. Rodriguez-Enriquez, S. Walker, T. W. Owens, F. Foster, J. Tanianis-Hughes, K. Brennan, C. H. Streuli and A. P. Gilmore, *Mol Cell*, 2013, 49, 959-971.
6. Y. T. Hsu, K. G. Wolter and R. J. Youle, *Proceedings of the National Academy of Sciences of the United States of America*, 1997, 94, 3668-3672.
7. J. A. Losonczi, E. T. Olejniczak, S. F. Betz, J. E. Harlan, J. Mack and S. W. Fesik, *Biochemistry*, 2000, 39, 11024-11033.
8. C. M. Franzin, J. Choi, D. Zhai, J. C. Reed and F. M. Marassi, *Magn Reson Chem*, 2004, 42, 172-179.
9. Y. Yao, L. M. Fujimoto, N. Hirshman, A. A. Bobkov, A. Antignani, R. J. Youle and F. M. Marassi, *Journal of Molecular Biology*, 2015, 427, 2262-2270.
10. M. Sattler, H. Liang, D. Nettesheim, R. P. Meadows, J. E. Harlan, M. Eberstadt, H. S. Yoon, S. B. Shuker, B. S. Chang, A. J. Minn, C. B. Thompson and S. W. Fesik, *Science*, 1997, 275, 983-986.
11. A. Maity, S. Yadav, C. S. Verma and S. Ghosh Dastidar, *PLoS One*, 2013, 8, e76837.
12. H. Lu and P. J. Booth, *J Mol Biol*, 2000, 299, 233-243.
13. S. Y. Jeong, B. Gaume, Y. J. Lee, Y. T. Hsu, S. W. Ryu, S. H. Yoon and R. J. Youle, *EMBO J*, 2004, 23, 2146-2155.
14. M. C. Wiener and S. H. White, *Biophys J*, 1992, 61, 434-447.
15. J. U. Bowie, *Nature*, 2005, 438, 581-589.
16. F. Cymer, G. von Heijne and S. H. White, *J Mol Biol*, 2015, 427, 999-1022.
17. S. Park and K. Schulten, *J Chem Phys*, 2004, 120, 5946-5961.
18. B. Roux, B. Prod'hom and M. Karplus, *Biophys J*, 1995, 68, 876-892.
19. J. Comer, J. C. Gumbart, J. Henin, T. Lelievre, A. Pohorille and C. Chipot, *J Phys Chem B*, 2015, 119, 1129-1151.
20. M. Suzuki, R. J. Youle and N. Tjandra, *Cell*, 2000, 103, 645-654.
21. N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper and A. Sali, *Curr Protoc Bioinformatics*, 2006, Chapter 5, Unit 5 6.
22. S. Jo, J. B. Lim, J. B. Klauda and W. Im, *Biophys J*, 2009, 97, 50-58.
23. S. E. Horvath and G. Daum, *Prog Lipid Res*, 2013, 52, 590-614.
24. B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J Comput Chem*, 2009, 30, 1545-1614.
25. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J Comput Chem*, 2005, 26, 1781-1802.

Physical Chemistry Chemical Physics Accepted Manuscript

26. R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. Mackerell, Jr., *J Chem Theory Comput*, 2012, 8, 3257-3273.
27. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, Jr. and R. W. Pastor, *J Phys Chem B*, 2010, 114, 7830-7843.
28. S. E. Z. Feller, Y.; Pastor, R. W.; Brooks, B. R, *J. Chem. Phys.* , 1995, 103, 4613– 4621.
29. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys*, 1977, 23, 327.
30. T. Y. Darden, D.; Pedersen, L. G, *J. Chem. Phys.* , 1993, 98, 10089– 10092.
31. G. Fiorin, M. L. Klein and J. Henin, *Molecular Physics*, 2013, 111, 3345-3362.
32. J. Valleau and D. Card, *The Journal of Chemical Physics*, 1972, 57, 5457-5462.
33. J. Henin, G. Fiorin, C. Chipot and M. L. Klein, *J Chem Theory Comput*, 2010, 6, 35-47.
34. B. Roux, *Biophysical Journal*, 1999, 77, 139-153.
35. S. G. Dastidar, A. Madhumalar, G. Fuentes, D. P. Lane and C. S. Verma, *Theoretical Chemistry Accounts*, 2010, 125, 621-635.
36. S. J. Marrink and H. J. C. Berendsen, *Journal of Physical Chemistry*, 1994, 98, 4155-4168.
37. A. Finkelstein, *J Gen Physiol*, 1976, 68, 127-135.
38. R. O. Potts and M. L. Francoeur, *Proc Natl Acad Sci U S A*, 1990, 87, 3871-3873.
39. T. Jansson and N. P. Illsley, *J Membr Biol*, 1993, 132, 147-155.
40. S. G. Dastidar and C. Mukhopadhyay, *Physical Review E*, 2003, 68.
41. S. G. Dastidar and C. Mukhopadhyay, *Physical Review E*, 2004, 70.
42. A. C. Johansson and E. Lindahl, *Biophys J*, 2006, 91, 4450-4463.

Physical Chemistry Chemical Physics Accepted Manuscript