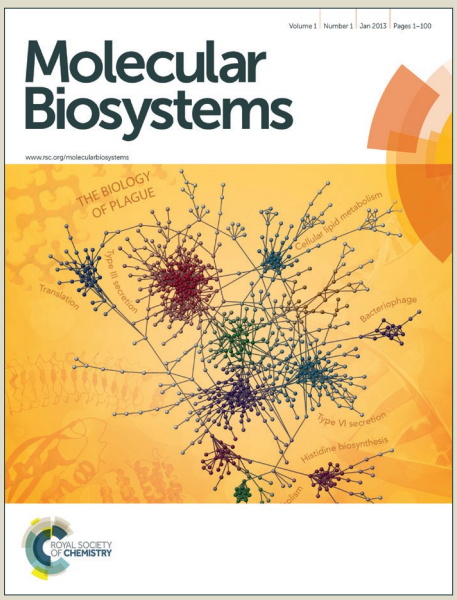
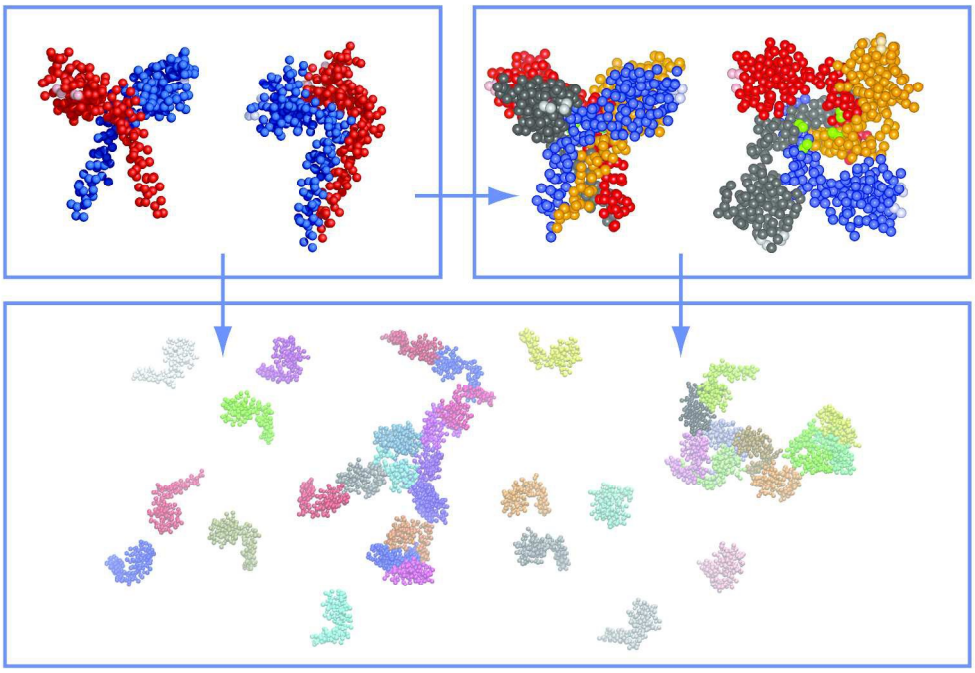
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



288x202mm (300 x 300 DPI)

Patch formation of a viral channel forming protein within a lipid membrane – Vpu of HIV-1.

Meng-Han Lin, Chin-Pei Chen[1], Wolfgang B. Fischer[*]

Institute of Biophotonics, School of Biomedical Science and Engineering and Biophotonics & Molecular Imaging Research Center (BMIRC), National Yang-Ming University, Taipei 112, Taiwan

[1]current address: Genomics Research Center, Academia Sinica, Taipei 115, Taiwan

*Correspondence to: W. B. Fischer, Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang-Ming University, 155, Li-Non St., Sec. 2, Taipei, 112, Taiwan

E-mail address: wfischer@ym.edu.tw

Tel: +886-2-2826-7394, Fax: +886-2-28235460

Keywords: Vpu of HIV-1, viral membrane protein, protein dynamics, protein – protein interactions, lipid membranes, coarse grained molecular dynamics simulations

Abstract

Ion channels and their viral companions are defined by their quaternary structure. The individual sub-units have to assemble into homo- or hetero-oligomers. Using Vpu of HIV-1 a putative viral channel forming protein (VCP) as a test case, the formation of a quaternary structure is monitored using coarse grained molecular dynamics simulations. Full length Vpu is generated by combining the helical transmembrane domain (TMD) with the cytoplasmic domain derived from NMR spectroscopy. Patches of 2 to 6 as well as patches with 16 and 32 Vpu proteins, Vpu-WT, not containing unphosphorylated serines 52 and 56 are used in coarse grained molecular dynamics (CGMD) simulations to study assembly dynamics. The same patches are simulated for Vpu double mutant, Vpu-DD, in which the two serines 52 and 56 are replaced by aspartic acid. Serines 52 and 56 in Vpu-WT allow short lived contacts between the cytoplasmic domains. Dimer formation is the first step for long lasting assemblies and is induced by the EYR motif. Roll-over movements allow rearrangement within the dimer. Independent of the number of Vpu proteins, Vpu-DD prefers smaller aggregates than Vpu-WT. In case of simulation of 4 Vpu-WT proteins a pore like assembly is directly identified with the TMD Ser-23 pointing towards a putative central pore axis.

Introduction

Membrane proteins are inserted into the lipid membrane *via* the translocon machinery and released into the membrane of the endoplasmic reticulum (ER) through a side passage [1-3]. A consecutive step is to assemble the proteins into fully functional units adopting a quaternary structure. Focussing on ion channels, several assembly pathways are proposed to form heteroligomeric assemblies of the individual subunits [4]. Following their host companions, viral channel forming proteins also have to assemble in order to develop a quaternary structure which renders the lipid membrane permeable for ions and small molecules [5-7]. A common feature for all the membrane proteins is, that the fold and its changes when escaping the translocon as a monomer and the final oligomerization state, is unknown.

In case of membrane proteins forming ion channels [4] or toxins generating pores [8, 9] the task is to finally generate a passage which enables ions and compounds to pass the hydrophobic barrier imposed by the lipid membrane. Once assembled, the question arises whether the assembly is unique and can be gated in a sophisticated manner or whether it simply resembles a loose and flexible arrangement.

Viral channel proteins found in many viruses are an interesting tool to investigate the pressing questions about the formation of ion channels due to their small size [10, 11]. However, the size imposes a challenge for structural experimental investigations and thus only a few of these VCPs have unravelled their membrane embedded parts/structure [12-17]. For one of them, Vpu of HIV-1, in addition to its transmembrane domain (TMD) [18], its extramembrane part is resolved [19-22]. Most recently, the structure of unphosphorylated full length Vpu is reported [22]. Only for one protein, p7 of HCV, a NMR spectroscopic structure of the full length protein in

an oligomeric state is available [16]. Computational structural modelling serves as a valuable tool in structural modelling of these proteins and to support the elucidation of the mechanism of function of these proteins [23-26]. Furthermore it allows to investigate protein internal dynamics as well as diffusion dynamics of the proteins.

Vpu is an 81 amino acid type-I integral membrane protein encoded by HIV-1 [27, 28]. It is bitopic with a single TMD at its N terminus and a large cytoplasmic domain which includes two phosphorylation sites, Ser-52 and Ser-56. Vpu is expressed in the late stage of viral infection and considered to be an auxiliary protein which amplifies viral release (reviewed in [29]). Functionally the two domains of Vpu act independently fulfilling two different roles [30]. The TMD is found to oligomerise showing channel activity in artificial membrane systems as well when expressed in cell systems [31, 32]. The cytoplasmic domain renders CD4, one of receptor proteins necessary for cell entry of HIV-1, susceptible for proteasomic degradation [33]. In recent years evidence has raised that Vpu acts through interaction with a series of host factors independent of its role to form channels. It counteracts e.g. the roles of BST-2 [34, 35] and NTB-A [36] to support the viral infectivity cycle. Thus, its role of forming channels is put to test [37]. Structural data are available for Vpu for (i) solely its TMD (e.g. [18]), (ii) the cytoplasmic domain [12, 19-22], and most recently (iii) the full length protein [22].

In this study a structural model of Vpu is constructed and used in multi micro-second coarse grained molecular dynamics (CGMD) simulations. The TMD is taken as an α-helix and linked to an experimentally derived structure of the cytoplasmic domain. Embedded in a lipid membrane comprised of POPC the dynamics of randomly inserted Vpu proteins are investigated. The number of embedded Vpu proteins ranges from 2 to 6 of them and includes patches of 16 as well as 36 Vpu proteins. The smaller patch size leans on current assumptions

that small sized homo-oligomers with defined geometry form the pore. The larger patch size is attributed to experimental findings that the protein is also found in larger oligomers (Chen and Ma, personal communication). The oligomerization ratio is monitored and compared to the same simulations with a mutant Vpu in which the two serines 52 and 56 are changed to two aspartic acids (Vpu-DD). In this respect, Vpu-DD is seen as a more realistic representation of the *in vivo* wild-type Vpu since the Vpu WT structure (referred to as Vpu-WT) used in this study does not contain the phosphates. The investigations propose mechanical features involved in dimer formation and identify patterns for higher oligomers.

Materials and Methods

Construction of the Vpu models

An ideal helical structure of the first 52 amino acids of Vpu (Vpu_{1-52}, HV1S1, UniProt entry: P19554) was generated using the MOE software suit (www.chemcomp.com). The helix was bent around residues Glu-28 to Ile-32 so that the helical stretch from residues Leu-33 to Ser-52 aligned with the membrane surface (as reported in [38]). Asp-39 and Arg-48 were pointing towards the bilayer surface according to experimental findings [39]. Two of these structures (each 530 atoms including united atoms), were embedded in a hydrated lipid bilayer so that one of them is inverted in respect to the other one. The last frame of a 100 ns MD simulation of one of the Vpu_{1-52} was chosen to generate full-length Vpu_{1-81}.

The first structure out of the 20 structures of the models deposited in the PDB data bank (PDB ID: 2K7Y, HV1H2, P05919; residues 36 to 81) [21] was merged with Vpu_{1-52} from the MD simulations so that a part of the helix ranging from residues Ile-39 to Arg-45, were merged with the helical motif of residues Ile-38 to Arg-44 of Vpu_{1-52} on the level of the Cα atoms. The merged structure full-length Vpu, Vpu_{1-81}, is henceforth referred to as Vpu-WT.

QPIPIVAIVA[10] LVVAIIIAIV[20] VWSIVIIEYR[30] KILRQRKIDR[40] LIDRLIERAE[50] DSGNESEGDQ[60] EELSALVERG[70] HLAPWDVDDL[80].

Two of these structures (790 atoms including united atoms) were also embedded into a fully hydrated lipid bilayer as mentioned above. The last frame of the 100 ns molecular dynamic (MD) simulation of one of the Vpu_{1-81} was chosen to generate a coarse grained (CG) model Vpu-WT. Mutation of the serines at positions 52 and 56 were done on the level of the equilibrated (100 ns) Vpu-WT structure prior for use in CGMD. Serines 53 and 56 were replaced by aspartic acids (Ds) and the respective structure named Vpu-DD.

Classical MD simulations

MD simulations were carried out with GROMACS 4.5.5 using Gromos96 (ffG45a3) force field (ff) with an integration step size of 2 fs. Protein, lipid, and the water molecules were separately coupled to a Berendsen thermostat at 310K with a coupling time of 0.1 ps. A semi isotropic pressure coupling was applied with a coupling time of 1.0 ps and a compressibility of 4.5e-5 bar^{-1}. Long-range electrostatics were treated using the particle-mesh Ewald (PME) algorithm with grid dimensions of 0.12 nm and interpolation order 4. Lennard-Jones and short-range Coulomb interactions were cut off at 1.4 and 1nm, respectively.

Vpu_{1-52} and Vpu_{1-82} proteins were embedded into lipid bilayers consisting of 228/228 lipids (11856/11856 atoms) hydrated with 11671/11473 water molecules (35013/34419 atoms). Overlapping lipids with the peptide were removed. After minimization (5000 steps of steepest descent and 5000 steps of conjugate gradient) the systems were equilibrated by gradually increasing the temperature from 100 K to 200 K and then to 310 K, whilst keeping the peptide fully restrained (k = 1000 kJ mol^{-1} nm^{-2}). The first two simulations (at 100 K and 200 K) were run for 200 ps, the last simulation (at 310 K) was run for 8.5 ns. The total equilibration time summed up to 8.65 ns. Space between helix 2 and lipid membrane did not contain any water molecules, since the hydrophobic residues were pointing toward to the lipids. The restraints, imposed by a force constant k on the peptide, were released in two

steps (k = 500 kJ mol^{-1} nm^{-2}, k = 250 kJ mol^{-1} nm^{-2}), running each of the steps for 500 ps and holding the systems at 310 K. The unconstrained systems were submitted to production runs of 100 ns.

Coarse grained MD (CGMD) simulations

The long-time-scale and large-system simulations were performed using CGMD simulations which simplified the system by replacing the groups of approximately four heavy atoms into a single bead and its extension to protein. The simulations were performed using the MARTINI force fields v2.0 for water and v2.1 for protein [40, 41]. The Martini script was used to convert Vpu into coarse-graining model. Maintainance of the dynamics and preservation of the internal flexibility was achieved by application of an elastic network. The simulations were performed with the GROMACS software. The integration time step dt = 30 fs. Periodic boundary conditions were applied and the non-bonded interactions were cut off at a distance of 1.2 nm. Proteins, lipids, and the water molecules were separately coupled to a Berendsen thermostat at 310 K with a coupling time of 1.0 ps. A semiisotropic pressure coupling was applied with a coupling time of 12.0 ps and a compressibility of $3e^{-5}$ bar^{-1}.

A pre-equilibrated lipid POPC membrane hydrated by CG water molecules (W) was used as a starting point. Three sizes of hydrated lipid patches were prepared consisting of 512/8478, 1152/19098, 2048/33912 POPCs/Ws in the proper ratio for simulation. Different numbers, e.g. two, three, four, five and six of Vpu-WTs and Vpu-DDs were embedded in a POPC membrane with the same orientation and Na-ions added to neutralize the systems (Suppl. Figure 1). The systems were energy minimized (500 steps of steepest decent) and equilibrated with proteins being restrained (k=500 kJ mol^{-1} nm^{-2}) for a total of 2.7 ns. The systems with unrestrained proteins were finally submitted to production runs of 10 and 15 μs.

Reverse mapping was used to connect the different length scales. Initially, united-atom particles were positioned close to their reference CG beads. A restrained simulated annealing was used to generate low-energy structures. During the annealing, the temperature was raised to 1300 K and a restraining potential kept at the centre of mass of groups of united-atom particles fixed to the corresponded CG

beads. The complete potential is the composition of united-atom force field and restrained harmonic potential. The simulation was carried out in the NVT ensemble. At the end of the simulated annealing, the restraints were released.

Data analysis

The oligomerization ratio from the computational data was calculated from the quotient of numbers of monomers found in an oligomeric state and the total number of monomers. Structures are considered as oligomers when the distance between 10 pairs of CG-atoms of different Vpu structures was below 5 Å and observed continuously for more than 10 times steps between the proteins. The respective data were fitted with a logarithmic growth function. Non-linear regression was performed by using non-linear curve fitting of OriginLab 9.0. The initial values were set to a = 1, b = 1, and c = 0.1. Iteration was conducted until the difference between reduced $\chi 2$ values of two successive iterations was less than a specified tolerance value, here 10^{-9} by default.

The tilt of the TMDs is calculated as the angle between the helix axis and the membrane normal (z-axis). The in respect to normal is obtained from averaged Cα-positions (z-axis) of four of the amino acids on either end of the TMD.

Results

EYR-motif triggers initial contact followed by roll overs into stable dimers.

Dimerisation of Vpu-WT and Vpu-DD is simulated in a hydrated lipid bilayer. After 0.5 μs Vpu-WT attaches to each other by the cytoplasmic domain (Figure 1). The two proteins dissociate and re-attach after around 3 μs with their transmembrane domains (TMDs). The contact of the two TMDs does not lead to any dissociation but leaves the two proteins

dimerized for the rest of the simulation time. Similarly, Vpu-DD dimers are formed after 1 μs and remaining together hither forth.

During the contact of the two cytoplasmic domains of Vpu-WT, which is denoted as mode I (M-I) type of binding, the two domains adopt a large angle between them (~ 134.8 ± 1.9 °) whilst the tilt of the two TMDs is 48.2 ± 1.5 ° (Figure 1A). When attaching *via* the TMDs, first there is a second binding mode (M-II) in which the two cytoplasmic domains are apart from each other spanning a large angle and the TMDs adopt a tilt of 47.2 ± 0.5 °. The two proteins 'roll around each other' along the TMD (as marked by the arrow in Figure 1) adopting a third mode of binding (M-III) in which units, cytoplasmic domain are lining up almost parallel (22.2 ± 0.5 °). This alignment leads to a straitening of the TMDs adopting a tilt of 11.3 ± 0.4 °. During M-III occasional changes of the tilt angle to values of 8.9 ± 0.7 ° (interim of the former two) are observed.

Vpu-DD dimerises *via* M-II and M-III modes of bindings (Figure 1B). Interactions between the cytoplasmic domains (M-I) are not observed. M-II is the first kind of contact mode at around 1 μs, characterised by an almost 180 ° orientation of the cytoplasmic domains and a tilt of 32.8 ± 1.2 °, which is lower than the tilt observed for Vpu-WT in this binding mode. Without any dissociation the two proteins 'roll over' into M-III which is characterized by an angle of 76.5 ± 0.3 ° between the cytoplasmic domains. The tilt of the TMDs switches frequently between two main stages of 10.4 ± 0.6 ° and 33.4 ± 0.3 °. Eventually a third stage is screened with a tilt of 29.5 ± 0.3 °. Beyond 5 μs of the simulation the dimer adopts M-II and remains in this mode with occasional changes of the contact angle. The dimeric structures of the two Vpu protein CG models at the end of the simulations differ with a RMSD of 0.64 (Cα-atoms).

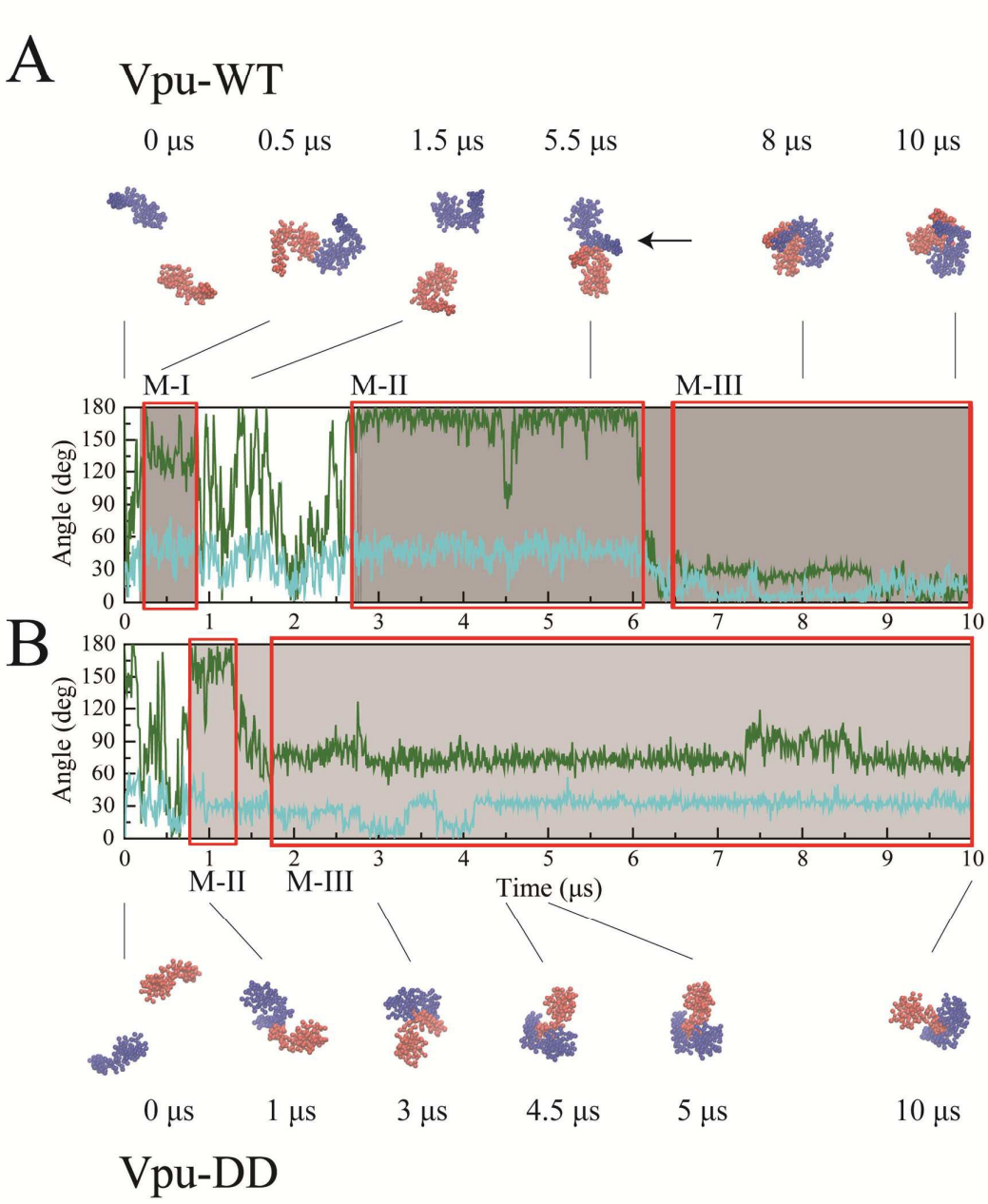


Fig. 1: Structural features of Vpu-WT (A) and Vpu-DD (B) during a 10 µs trajectory. Angles between two cytoplasmic domains (green) and the tilt of the TMDs (cyan) are calculated for each time frame. The grey areas indicate the time of contact between the two Vpu proteins in each of the simulations. The tilt angle represents an average value of the two TMDs. Two Vpu proteins (red and blue) are shown in a top view along the membrane normal towards the cytoplasmic domain as CG models. The arrow marks the region of the TMD for a putative 'rolling-over' step as indicated in the text. Lipid and water molecules are omitted for clarity.

As a result, in M-III Vpu-DD cytoplasmic domains adopt bigger angles than in Vpu-WT, indicating that they are not so tightly bound to each other. There is also a stronger step-like fluctuation of the tilt angle. In M-II, the angle spanned between the cytoplasmic domains is similar for Vpu-WT and Vpu-DD. The tilt in Vpu-DD is less than the tilt in Vpu-WT. M-I identified for Vpu-WT does not occur in Vpu-DD. M-II and M-III are putatively interchangeable by a kind of 'rolling-over' step around the surface of the TMD, no dissociation is needed. For both mutants the duration of the dimers being in either mode is M-III > M-II > M-I (Table 1).

Tab. 1: Approximate time of existence of the binding modes M-I, M-II and M-III and identified closest contact areas and residues. Closest contact residues are calculated using MOE.

	M-I	M-II	M-III
Time			
WT-Vpu Vpu-DD	~ 0.5 μs -	~ 3 μs ~ 0.5 μs	> 4 μs > 3.5 μs
Contact region (WT-Vpu)	cytoplasm	hinge	Cyto < TM
Contact residues	Arg-48 – Glu-57 Ser-52 – Ser-56 Glu-55 – Glu-55 Ser-56 – Ser-52 Glu-57 – Arg-48	Val-25 – Lys-31 Glu-28 – Arg-30 Tyr-29 – Tyr-29 Arg-30 – Glu-28 Lys-31 – Val-25	Val-6 – Leu-11 Val-9 – Leu-11 Val-12 – Ile-15 Val-13 – Ile-15 Ile-16 – Trp-22 Val-20 – Ala-18 Ile-42 – Val-21

Inspection of the closest contacts between pairs of amino acids of the three binding modes shows that M-I is characterised by electrostatic interactions between residues Arg-48 and Glu-57 (Figure 2A and Table 1). The respective spheres of Arg-48 and Glu-57 in the CG model are about 0.52 nm apart of each other (0.52 ± 0.03 nm for 0.5 ns to 0.8 ns during M-I formation), which marks the closest distance they can achieve. The distance between the nitrogen atom of Arg-48 and the oxygen atom of Glu-57 in the respective united atom model have been calculated to be 0.35 nm. Since Ser-52 and -56 are also found, it can be speculated that hydrogen binding could be also a source for binding. The EYR motif is responsible for M-II interaction which seems to initiate the binding of the two proteins *via* the TMDs (Figure 2B). M-III has the largest number of interactions between the two proteins (Figure 2C). The contact is mostly due to pairs of hydrophobic residues including those from helix-2 of the cytoplasmic domain (Ile-42).

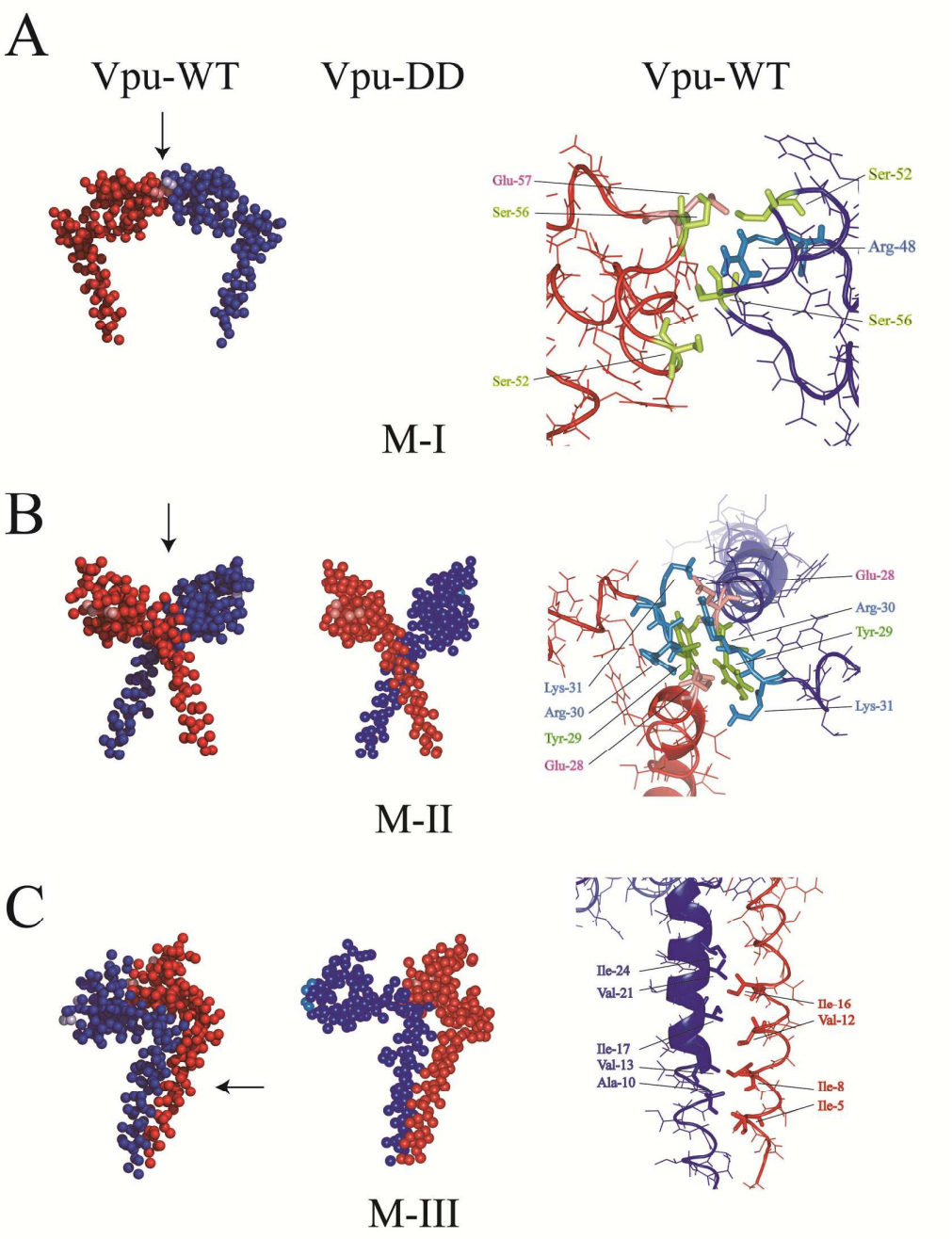


Fig. 2: Representative structures of the individual binding modes, M-I (A), M-II (B) and M-III (C) of Vpu-WT and Vpu-DD in CG representation (two left columns) and close-up view (right). The arrow indicates the viewing direction for the structures on the right hand side. Closest contact residues are identified using by MOE. For M-I: Arg-48 and Glu-57 are shown in blue and red, respectively. Ser-52 and -56 are shown in light green. For M-II: positively charged residues Arg-30/Lys-31 and negatively charged residue Glu-28 are shown in blue and pink, respectively. Tyr-29 is shown in green. For M-III: residues are shown either in blue or red for the individual structures.

The energy distribution amongst the three modes of binding for Vpu-WT indicates a narrow valley for M-III (~ -11 kJ/mol) and shallow valleys for M-I (~ -7 kJ/mol) (Figure 3A). M-II (~ -8 kJ/mol) comprises a small minimum with a small energy barrier to M-III. Similarly, M-III of Vpu-DD is the narrowest one (Figure 3B). A flank of the valley of M-III (-11.1 kJ/mol) is attributed to changes of the angle for the cytoplasmic domains (Figure 1B green line in the range 7 – 9 μs), marked as 'a' in Figure 3B, and a small valley next to the valley of M-III, marked as 'b' in Figure 3B, is attributed to change in the tilt angles of M-III (Figure 1B cyan line for > 2.5 μs). M-II (~ -5 kJ/mol) is represented by a collection of small minima readily separated from M-III (Figure 3B). It is less favourable than M-II of Vpu-WT.

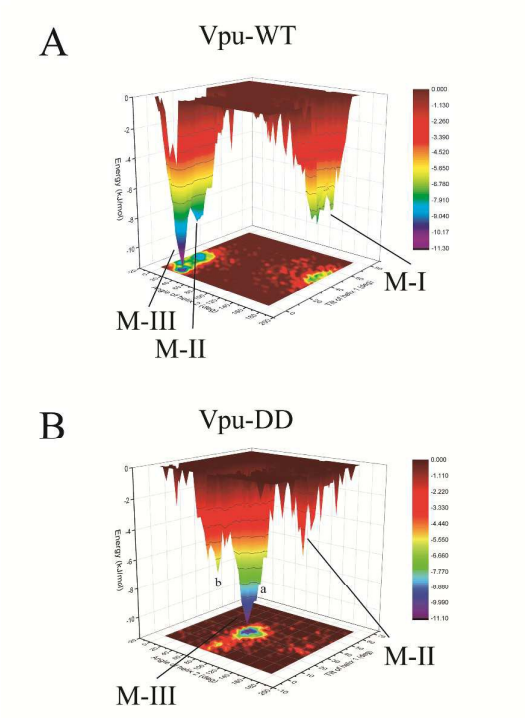


Fig. 3: Energy landscape of the assembly trajectory of Vpu-WT (A) and Vpu-DD (B). The angle of helix 2 of each dimer is correlated with the tilt of helix 1. Areas in (B) are marked where the angle between the cytoplasmic domains (a) and the tilt angle (b) of the two TMDs are changing.

Stable tetramer via direct dimer-dimer assembly

Investigation of a larger number of Vpu proteins in the simulation box reveals that the type of binding is similar to the modes described above. Calculation of the formation of the individual oligomers including the total oligomerization rate over time reveals a dynamic process of the oligomerization over a certain time period ending in mostly stable aggregates at the end of the simulation period (Suppl. Figure 2). The characteristics for each simulation can be summarized as following.

Simulation of trimers, Vpu-WT forms a trimer after 2 μs into the simulation (Figure 4A, upper panel and Suppl. Figure 2A), whilst it takes Vpu-DD about 3 μs to form a trimer (Figure 4A, lower panel and Suppl. Figure 2A). Before those time periods, dimers are formed. Whilst Vpu-WT forms stable tetramers after about 2 μs (Figure 4B, upper panel), Vpu-DD does not form any tetramer at all (Figure 4B, lower panel). Instead, there is a frequent association-dissociation of tetramers into dimers. The binding energy can be estimated to be around -0.6 kcal/mol. During the simulation of four Vpu proteins,

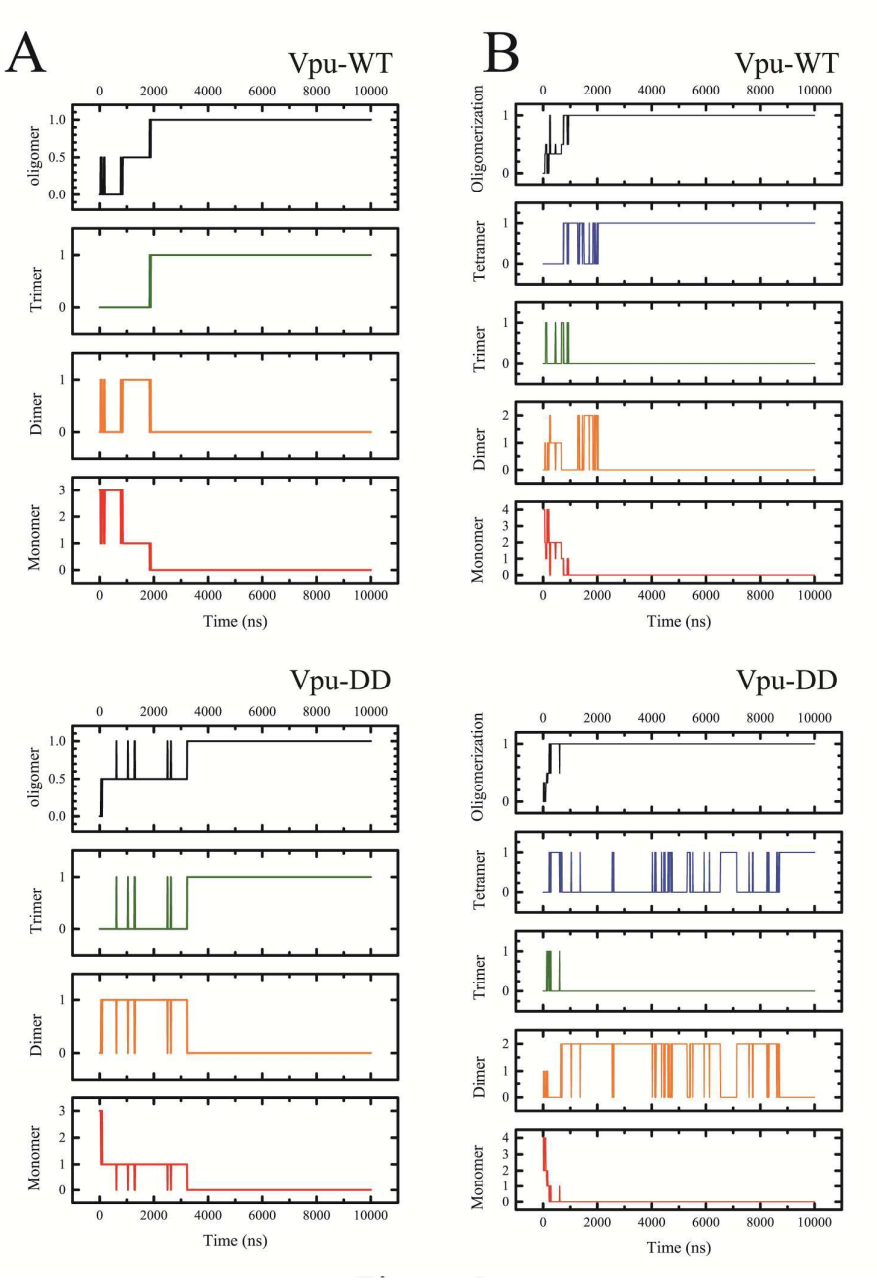


Fig. 4: Time-resolved oligomerization rate of the individual protein assemblies as well as the total rate from a simulation with three (A) and four Vpu proteins (B). Data for Vpu-WT (top) and Vpu-DD are shown (bottom). Data of the rates are shown in red for the monomers, orange for the trimers and blue for the tetramers. The total oligomerization rate is represented in a black curve.

Vpu-WT oligomerises into a tetramer *via* dimer-dimer formation (Figure 5). The hydrophilic residue Ser-22 of all four proteins is pointing towards the centre of the assembly. The cytoplasmic domains with residues Ser-52 and -56 are facing to the outside. The tilt of the TMDs is calculated to be 22.4 ± 0.5 °.

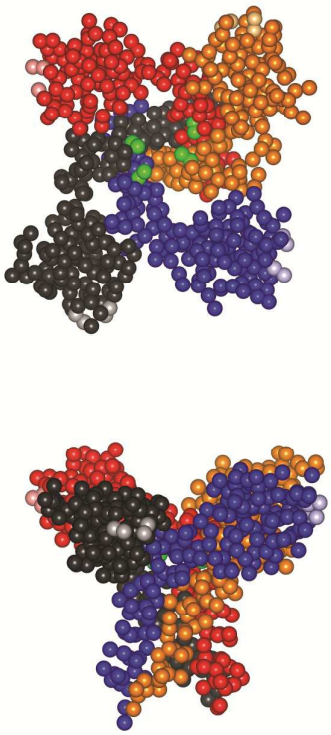


Fig. 5: Tetrameric Vpu-WT from a 10 μs simulation of a lipid patch containing four Vpu-WT proteins in a top view along the membrane normal onto the cytoplasmic domains (top) and in a side view (bottom). Each monomer in its CG representation is shown in a different colour. Lipid and water molecules are omitted for clarity. Green spheres show Ser-22 pointing towards the centre of the tetramer. Light spheres represent Ser-52 and -56 which are phosphorylated *in vivo*.

Simulation of five Vpu proteins leads to the formation of a Vpu-WT pentamer after about 8 μs of simulation (Figure 6A, upper panel). There are two dimers formed first followed by a trimer formation after 1 μs at the cost of one dimer. The pentamer does not resemble a pore like structure as observed for the tetramer. Vpu-DD also forms dimers first, followed by an occasional trimer formation within the first 5 μs (Figure 6A, lower panel). After a short transition into a tetramer/monomer formation, dimer and trimer remain, no pentamer is formed.

Six Vpu proteins in the simulation box generate slightly more complex dynamics during the assembly for Vpu-WT (Figure 6B, upper panel). As dimer formation occurs first, there is flickering between all oligomers possible (< 1.8 μs up to tetramer, 1.8 μs < also pentamer) up to 7 μs of simulation time. Beyond that time a stable hexamer is formed. Similar to the pentamer, the hexamer is not in a stage which would allow to expecting a pore to be formed. Vpu-DD follows the pattern of dimerization first (< 4 μs), but does not form any oligomer larger than a tetramer and a dimer after about 6 μs (Figure 6B, lower panel).

All simulated systems reach a stage where all Vpu-WT monomers are assembled inot the highest possible stage. None of these assemblies form relevant pore-like structures except the tetramer.

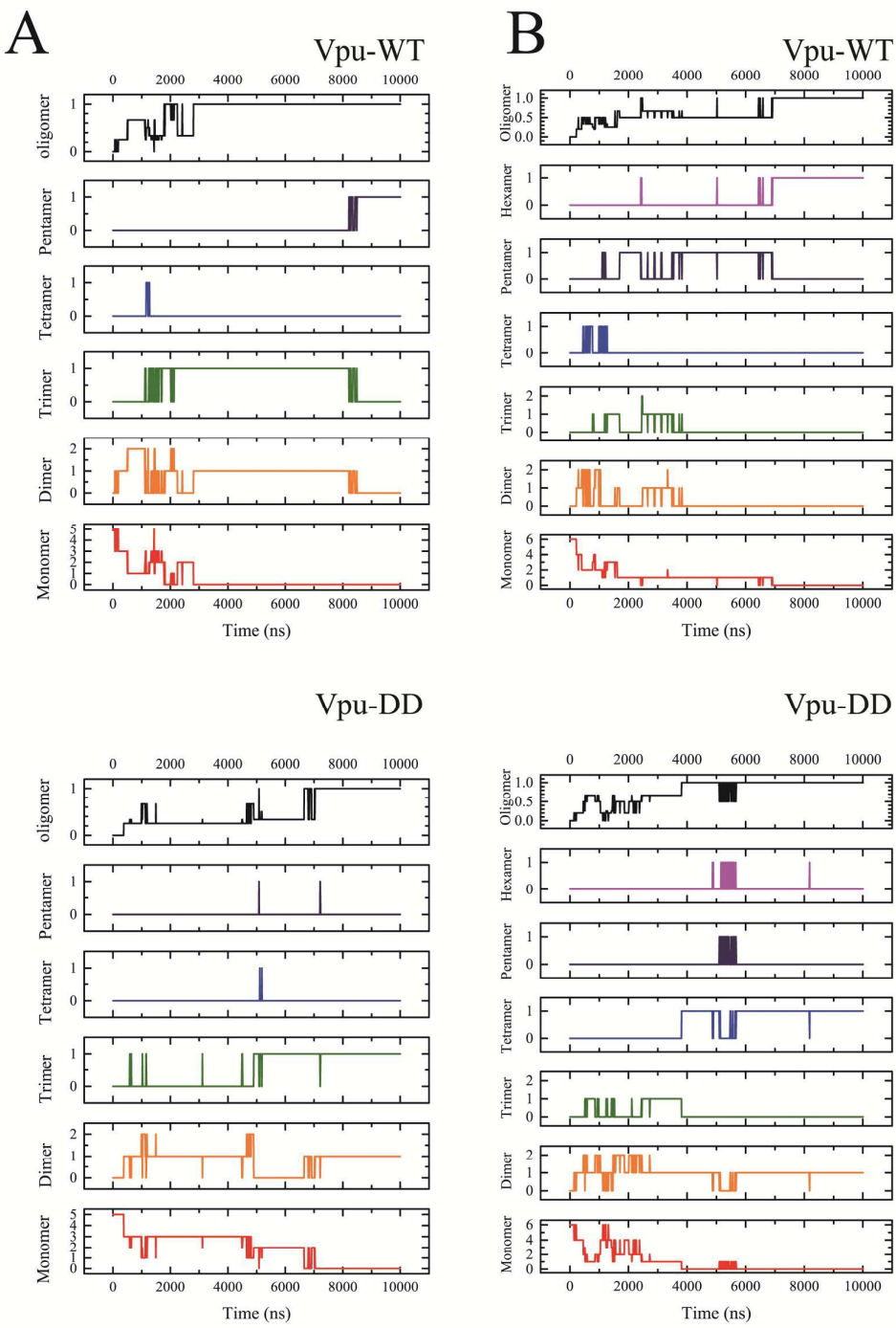


Fig. 6: Time-resolved oligomerization ratio of the individual protein assemblies as well as the total rate from a simulation with five (A) and six Vpu proteins (B). Data for Vpu-WT (top) and Vpu-DD are shown (bottom). Data of the rates are shown in red for the monomers, orange for the trimers, blue for the tetramers, dark blue for the pentamer and pink for the hexamer. The total oligomerization ratio is represented in a black curve.

Evidence for threshold number for patch-formation

Aggregation of 16 Vpu-WT proteins also starts with the formation of dimers first followed by trimers and tetramers (< 1μs) (Suppl. Figure 3). The temporary formation of medium sized aggregates (8-mer to 11-mer) follow leading to even larger assemblies of 14-mer to 16-mer (1 – 6 μs). Towards the end of the simulation two patches, 12-mer and 4-mer, remain. Vpu-DD shows a similar pattern at shorter time steps forming up to 9-mer. Higher aggregates only exist sporadically (12-mer and 13-mer). At the end of the simulation two patches of 9-mer and 6-mer remain. The overall oligomerization ratio of Vpu-DD does not reach 100 % as does the oligomerization ratio of Vpu-WT.

In a system with 36 Vpu-WT proteins at a lipid protein ratio 9 : 1 during the first 3 μs also dimers and trimers are both formed almost instantaneously (Suppl. Figures 2B,I and 4A). Higher oligomers (tetramer to 12-mer) are temporarily formed as well. Up to 6 μs into the simulation, temporarily tetramers as well as one 27-mer, 28-mer and a 32-mer are formed. An aggregation of all 36 Vpu-WT exists from 6 μs onwards towards the end of the simulation. Simulations with Vpu-DD result in the same sequence of oligomers being formed as afore mentioned with the difference that the oligomers associate and dissociate more frequently (Suppl. Figures 2B,II and 4B). The last frame (10 μs) contains a 23-mer, 10-mer and three monomers.

Systems of 36 Vpu proteins simulated at a lipid : protein ration 20 : 1 (diluted) both proteins, Vpu-WT (Figure 7A and Suppl. Figure 5A) and Vpu-DD (Figure 7B and Suppl. Figure 5B) barely show oligomers with more than 16 proteins. The last frame of the simulation with Vpu-WT releases one 14-mer and one 10-mer as well as 12 monomers, whilst the simulation

with Vpu-DDs releases single 16-mer, 6-mer, trimer as well as 11 monomers. Similar to the simulations at higher protein concentration, simulations with Vpu-DD show larger association/dissociation frequencies especially for the smaller oligomers (≤ 6).

The simulations indicate that larger numbers of Vpu proteins do not form pore-like structures but rather irregular patches under the current computational conditions (Figure 7). It emerges that the number of monomers forming a patch does not exceed a specific value, e.g. ~16 monomers.

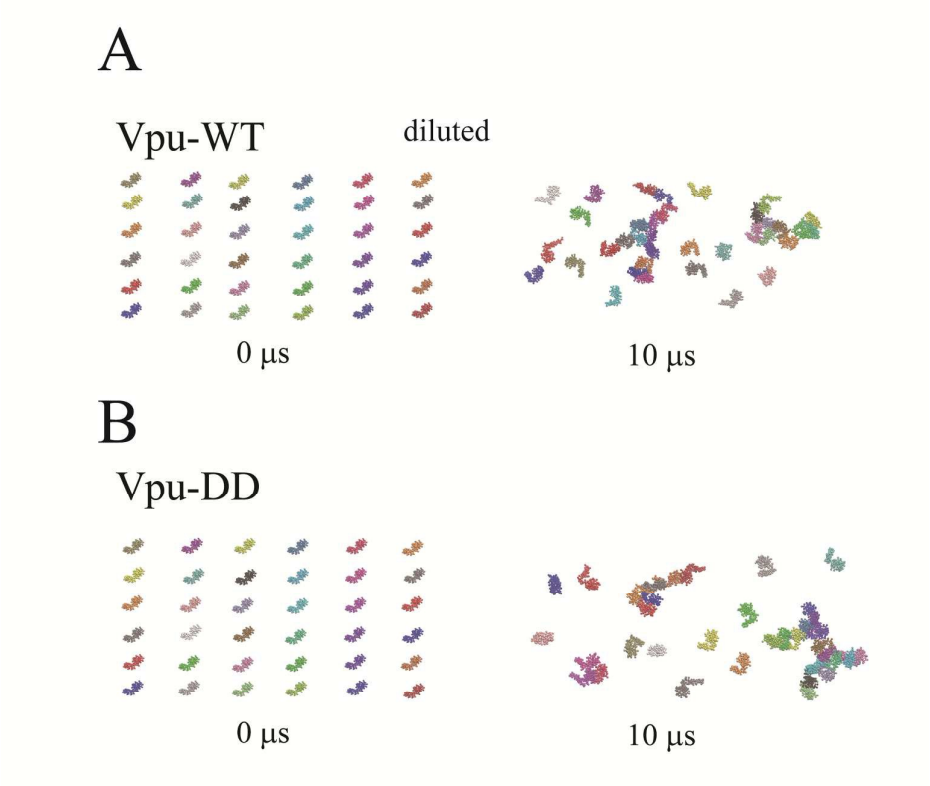


Fig. 7: Top view along the membrane normal onto the cytoplasmic domain of 36 Vpu-WT (A) and Vpu-DD (B) in a in their starting configuration (0 μs) and at the final conformation at the end of each of the simulations. The lipid : protein ratio is 20 : 1. Lipid and water molecules are omitted for

Dilution of protein lowers maximum oligomerization ratio

Oligomerization of 36 Vpu-WT proteins is dissected into contributions of the TMD or the cytoplasmic domain and the evolution of the oligomerization ration calculated over time (Suppl. Figure 6A). The oligomerization ratio of the TMDs constantly rises up to ~0.5 and is higher than the rate of the cytoplasmic domain which remains at around 0.2 after a short rise (Suppl. Figure 6A,I). Upon 'dilution' of the proteins into a patch of higher lipid : protein ratio, the two values level of at an approximate ration of 0.2 (Suppl. Figure 6A,II). A similar feature emerges for the simulations with Vpu-DD proteins. In the concentrated simulation the ration for the TMD rises to ~0.7 above the values for the cytoplasmic domains (Suppl. Figure 6B,I). Upon dilution, the TMD ratio of Vpu-DD is lower and levels off at a later stage during the simulation than for Vpu-WT at ~0.2 (Suppl. Figure 6B,II). The ratio is still higher than the values of the cytoplasmic domain indicating that the cytoplasmic domains repel each other and the TMD is responsible for the assembly. Overall, in simulations of low lipid : protein ratio Vpu-WT associates much more than Vpu-DD (Suppl. Figure 6C,I) In simulations of a higher lipid : protein ratio association for both proteins remains low. The difference between Vpu-WT and Vpu-DD vanishes (Suppl. Figure 6C,II).

Simulations under 'diluted' conditions reveal lower oligomerization ratio within the 10 μs simulations. Independent of the simulation condition, the TMD is the key domain for driving oligomerization.

Discussion

The model: The computational model leans on experimental data from NMR spectroscopy in as much as the TMD is helical and slightly kinked or bent [42]. Structural data of the cytoplasmic domain is taken from solution NMR spectroscopic experiments with the domain dissolved in DPC (dodecylphosphatidylcholine) to mimic the cytoplasm/membrane attachment in a most likely way [21]. Structural analysis of isolated domains is already used in structural biology. Alignment of the second helix of Vpu with its long axis perpendicular to the membrane normal is shown in solid state NMR spectroscopic experiments [12]. During the generation of the relevant features between the TMD and the second helix, care has been taken that the orientation of helix 2 is based on experimental findings [39]. Thus, the Vpu model resembles a most likely model of the monomeric protein. The chosen structural computational full length unphosphorylated Vpu model is in contrast to the experimentally reported model which has been measured in DHPC (1,2-dihexanoyl-sn-glycero-3-phosphatidylcholine) [22]. In the recent reported model helix 2 is also positioned at the membrane surface but helix 3 is proposed to be within the hydrophobic slab of the bilayer. In the model chosen in this study both cytoplasmic helices, termed 2 and 3, are placed above the lipid bilayer.

Using the novel Martini ff, Martini 2.2 for the protein, care is taken to include protein flexibility as much as possible. Diffusion dynamics are accelerated by approximately a factor of 4 when using CGMD [43]. An elastic network (EN) model is mixed with the MARTINI ffs to enable structural fluctuations which screen local and global deformations to the same extent as in classical MD simulations [44]. Due to the inherent nature of the EN model the structure maintains the reference structure and large scale structural changes considered to

fold the protein cannot be observed. Nevertheless, CGMD is applied to investigate membrane protein association [45-48]. In this respect the simulations comprise an interrogation into the assembly dynamics of a specifically formed rigid body.

Interpretation of the data is on the basis that Vpu-WT resembles Vpu-protein in which the phosphate groups have been removed. Vpu-DD, with its negative parameters for the aspartic acids, is more likely to resemble in vivo Vpu-WT in its phosphorylated form using this naming for 'in vitro wild type Vpu'. Aspartic acid is chosen as mutation for the serines site 52 and 56 in as much experimental data are available for this mutant and because the aspartic acid removes the voluminous phosphates but leaves the charges in the protein.

Based on the data, it is proposed that Vpu forms large unstructured aggregates with and without phosphorylated serines. In the presence of the negative charges the assembly is channelled to mostly deliver dimers, tetramers and eventually higher oligomers in more defined numbers than uncharged Vpu would do. The maximum size of the patches could be limited to a certain threshold number of monomers involved. This threshold number could depend on thermodynamic conditions as well as on biochemical conditions within specific sub-cellular organelles or compartments. It is anticipated that Vpu can form relevant pore-like structures within these patches due to conformational rearrangements of its monomers. On the other hand, a route of forming pore-like structures directly *via* specific step-wise assembly is still possible. Thus, two pathways are suggested, pore-formation by (i) direct assembly or (ii) conformational rearrangement within larger patches. During the 'direct assembly' pathway Vpu follows dimer-first as reported experimentally (Chen, Lin, Ma and Fischer submitted) and also for the formation of heterologous large ion channels [4, 49].

However, in the light of eminent interaction of Vpu with host factors larger agglomerates of Vpu could present potent binding sites to the host factors. At this stage it could be speculated that the electrostatic interactions may eventually be contra-productive. Since there is experimental evidence that Vpu is not only expressed into doubly phosphorylated Vpu, it is hypothesised that those mono-phosphorylated on fully unphosphorylated Vpu steer the aggregation and eventually the association with the host.

Triggering conformation changes into different modes is possible by a movement of the TMDs independent of the cytoplasmic domain. This may explain experimental findings of two functional modes which can be readily attributed to either of the two domains [30]. Changing tilt could mean that lateral pressure profiles of the lipid membrane or its curvature play a crucial role in triggering function. It is possible that the assembled structures attract ions in solution which then induce conformational changes in a cooperative manner finally leading to ion conducting pores.

Conclusions

The two phosphorylation sites of Vpu do not only serve as markers to follow the proteosomal degradation pathway upon contact with host proteins but also as modulator for the dynamics of aggregation of Vpu into larger assemblies. Modulation involves events of initial contacts of monomers as well as to limit protein patches in their size. Pore formation, seen here as a formation of monomers around a pseudo symmetry axis, can follow two pathways *via* direct assembly or structural rearrangement within assembled patches.

Supporting Information Available:

SI contains pictures of the starting and end configurations of the proteins used in this manuscript: three to six Vpu-WT and Vpu-DD as well as representations of the respective simulations with 36 proteins (Suppl. Fig. 1). Time-resolved numbers of oligomers and total oligomerization rate of the individual protein assemblies are given as well as the oligomerization ratio of simulation of 16 and 36 Vpu-WT as well as Vpu-DD (Suppl. Figures 3 – 5). For the simulations with 36 Vpu proteins, both Vpu-WT and Vpu-DD, the ratio is dissected into contributions of the transmembrane domain and the cytoplasmic domain (Suppl. Figure 6). This material is available at free of charge *via* the Internet at http://pubs.acs.org

Acknowledgement

WBF thanks the National Science Council (NSC-101-2112-M-010-002-MY3), Taiwan, for financial support. MHL acknowledges a student-fellowship of National Yang-Ming University. We thank Y.-T. Wang, and R. D. Mahato (Taipei, TW) as well as R. Schilling (Heidelberg, D) for helpful discussions.

References

1. D. J. Schnell and D. N. Hebert, *Cell*, 2003, **112**, 491-505.
2. J.-F. Ménétret, R. S. Hedge, S. U. Heinrich, P. Chandramouli, S. J. Ludtke, T. A. Rapoport and C. W. Akey, *J. Mol. Biol.*, 2005, **348**, 445-457.
3. S. Habdenteufel, M.-C. Klein, A. Melnyk and R. Zimmermann, *Biochem. Cell Biol.*, 2014, **92**, 499-509.
4. W. N. Green, *J. Gen. Physiol.*, 1999, **113**, 163-169.
5. L. Carrasco, *Adv. Virus Res.*, 1995, **45**, 61-112.
6. W. B. Fischer, Y.-T. Wang, C. Schindler and C.-P. Chen, *Int. Rev. Cell Mol. Biol.*, 2012, **294**, 259-321.
7. J. L. Nieva, V. Madan and L. Carrasco, *Nat. Rev. Microbiol.*, 2012, **10**, 563-574.
8. M. R. Gonzales, M. Bischofberger, L. Pernot, F. G. Van der Goot and B. Frêche, *Cell. Mol. Life Sci.*, 2008, **65**, 493-507.
9. H. Bayley, *Nature*, 2009, **459**, 651-652.
10. K. Wang, S. Xie and B. Sun, *Biochim. Biophys. Acta*, 2010, **1808**, 510-515.
11. B. OuYang and J. J. Chou, *Biochim. Biophys. Acta*, 2014, **1838**, 1058-1067.
12. C. Ma, F. M. Marassi, D. H. Jones, S. K. Straus, S. Bour, K. Strebel, U. Schubert, M. Oblatt-Montal, M. Montal and S. J. Opella, *Prot. Sci.*, 2002, **11**, 546-557.
13. C. Li, H. Qin, F. P. Gao and T. A. Cross, *Biochim. Biophys. Acta*, 2007, **1768**, 3162-3170.
14. G. A. Cook and S. J. Opella, *Eur. Biophys. J.*, 2010, **39**, 1097-1104.

15. R. Montserret, N. Saint, C. Vanbelle, A. G. Salvay, J. P. Simorre, C. Ebel, N. Sapay, J.-G. Renisio, A. Böckmann, E. Steinmann, T. Pietschmann, J. Dubuisson, C. Chipot and F. Penin, *J. Biol. Chem.*, 2010, **285**, 31446-31461.

16. B. OuYang, S. Xie, M. J. Berardi, X. Zhao, J. Dev, W. Yu, B. Sun and J. J. Chou, *Nature*, 2013, **498**, 521-525.

17. T. L. Foster, G. S. Thompson, A. P. Kalverda, J. Kankanala, M. Bentham, L. F. Wetherill, J. Thompson, A. M. Barker, D. Clarke, M. Noerenberg, A. R. Pearson, D. J. Rowlands, S. W. Homans, M. Harris, R. Foster and S. Griffin, *Hepatology*, 2014, **59**, 408-422.

18. F. M. Marassi, C. Ma, H. Gratkowski, S. K. Straus, K. Strebel, M. Oblatt-Montal, M. Montal and S. J. Opella, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14336-14341.

19. T. Federau, U. Schubert, J. Floßdorf, P. Henklein, D. Schomburg and V. Wray, *Int. J. Peptide Protein Res.*, 1996, **47**, 297-310.

20. D. Willbold, S. Hoffmann and P. Rösch, *Eur. J. Biochem.*, 1997, **245**, 581-588.

21. M. Wittlich, B. W. Koenig, M. Stoldt, H. Schmidt and D. Willbold, *FEBS J.*, 2009, **276**, 6560-6575.

22. H. Zhang, E. C. Lin, B. B. Das, Y. Tian and S. J. Opella, *Biochim. Biophys. Acta*, 2015, **1848**, 3007-3018.

23. V. Lemaitre, D. Willbold, A. Watts and W. B. Fischer, *J. Biomol. Struct. Dyn.*, 2006, **23**, 485-496.

24. G. Patargias, N. Zitzmann, R. Dwek and W. B. Fischer, *J. Med. Chem.*, 2006, **49**, 648-655.

25. D. E. Chandler, F. Penin, K. Schulten and C. Chipot, *PLOS Comput. Biol.*, 2012, **8**, e1002702.

26. M. M. Kalita, S. Griffin, J. J. Chou and W. B. Fischer, *Biochim. Biophys. Acta*, 2015, **1848**, 1383-1392.

27. K. Strebel, T. Klimkait and M. A. Martin, *Science*, 1988, **241**, 1221-1223.

28. E. A. Cohen, E. F. Terwilliger, J. G. Sodroski and W. A. Haseltine, *Nature*, 1988, **334**, 532-534.

29. S. Bour and K. Strebel, *Microbes and Infections*, 2003, **5**, 1029-1039.

30. U. Schubert, S. Bour, A. V. Ferrer-Montiel, M. Montal, F. Maldarelli and K. Strebel, *J. Virol.*, 1996, **70**, 809-819.

31. U. Schubert, A. V. Ferrer-Montiel, M. Oblatt-Montal, P. Henklein, K. Strebel and M. Montal, *FEBS Lett.*, 1996, **398**, 12-18.

32. G. D. Ewart, T. Sutherland, P. W. Gage and G. B. Cox, *J. Virol.*, 1996, **70**, 7108-7115.

33. M.-Y. Chen, F. Maldarelli, M. A. Martin and K. Strebel, *J. Virol.*, 1993, **67**, 3877-3884.

34. S. J. D. Neil, T. Zang and P. D. Bieniasz, *Nature*, 2008, **451**, 425-431.

35. N. van Damme, D. Goff, C. Katsura, R. L. Jorgensen, R. Mitchell, M. C. Johnson, E. B. Stephens and J. Guatelli, *Cell Host Microbe*, 2008, **3**, 1-8.

36. A. H. Shah, B. Sowrirajan, Z. B. Davis, J. P. Ward, E. M. Campbell, V. Planelles and E. Barker, *Cell Host Microbe*, 2010, **8**, 397-409.

37. K. Strebel, *Biochim. Biophys. Acta*, 2014, **1838**, 1074-1081.

38. I. Sramala, V. Lemaitre, J. D. Faraldo-Gomez, S. Vincent, A. Watts and W. B. Fischer, *Biophys. J.*, 2003, **84**, 3276-3284.

39. P. Henklein, R. Kinder, U. Schubert and B. Bechinger, *FEBS Lett.*, 2000, **482**, 220-224.

40. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812-7824.

41. L. Monticelli, S. K. Kandasamy, X. Perioloe, R. G. Larson, D. P. Tieleman and S. J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819-834.

42. S. H. Park, A. A. Mrse, A. A. Nevzorov, M. F. Mesleh, M. Oblatt-Montal, M. Montal and S. J. Opella, *J. Mol. Biol.*, 2003, **333**, 409-424.

43. S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750-760.

44. X. Periole, M. Cavalli, S.-J. Marrink and M. A. Ceruso, *J. Chem. Theory Comput.*, 2009, **5**, 2531-2543.

45. A. Y. Shih, A. Arkhipov, P. L. Freddolino and K. Schulten, *J. Phys. Chem. B*, 2006, **110**, 3674-3684.

46. X. Periole, T. Huber, S. J. Marrink and T. P. Sakmar, *J. Am. Chem. Soc.*, 2007, **129**, 10126-10132.

47. T. Carpenter, P. J. Bond, S. Khalid and M. S. P. Sansom, *Biophys. J.*, 2008, **95**, 3790-3801.

48. D. Sengupta and S. J. Marrink, *Phys. Chem. Chem. Phys.*, 2010, **12**, 12987-12996.

49. W. N. Green and N. S. Millar, *Trends Neurosci.*, 1995, **18**, 280-287.