Wrapping up viruses at multiscale resolution: Optimizing PACKMOL and SIRAH execution for simulating the Zika virus.

Martín Soñora[1], Leandro Martínez[2], Sergio Pantano[1], Matías R. Machado[1]*

[1]Biomolecular Simulations Group, Institut Pasteur de Montevideo, Montevideo, Uruguay. Mataojo 2020, CP 11400.

[2]Institute of Chemistry and Center for Computational Engineering & Science, University of Campinas, Campinas, SP, Brazil.

*To whom correspondence should be addressed:

Matías R. Machado, Tel/Fax: +598-2522 0910, e-mail: mmachado@pasteur.edu.uy

RUNNING TITLE

Modeling enveloped viruses with PACKMOL and SIRAH Force Field.

ABSTRACT

Simulating huge biomolecular complexes of million atoms at relevant biological timescales is becoming accessible to the broad scientific community. That proves to be crucial for urgent responses against emergent diseases in real time. Yet, there are still issues to sort regarding the system setup so that Molecular Dynamics (MD) simulations can be run in a simple and standard way. Here, we introduce an optimized pipeline for building and simulating enveloped virus-like particles (VLP). First, the membrane packing problem is tackled with new features and optimized options in PACKMOL. This allows preparing accurate membrane models of thousands of lipids in the context of a VLP within a few hours using a single CPU. Then, the assembly of the VLP system is done within the multiscale framework of the coarse-grained SIRAH force field. Finally, the equilibration protocol provides a system ready for production MD simulations within a few days on broadly accessible GPU resources. The pipeline is applied to study the Zika Virus as a test case for large biomolecular systems. The VLP stabilizes at approximately 0.5 μs of MD simulation, reproducing correlations greater than 0.90 against experimental density maps from cryo-electron microscopy. Detailed structural analysis of the protein envelope also shows very good agreement in root mean square deviations and B-factors with the experimental data. The level of details attained shows for the first time a possible role of anionic phospholipids in stabilizing the envelope. Combining an efficient and reliable setup procedure with an accurate coarse-grained force field provides a valuable pipeline for simulating arbitrary viral systems or sub-cellular compartments, paving the way towards whole-cell simulations.

INTRODUCTION

Technological advances on computers and experimental techniques are allowing to simulate and analyze huge biological systems [1,2], pushing forward the field of Computational Virology to a new era [3,4]. The present capability may be further enhanced by the use of multiscale strategies, which combines atomic with coarse-grained (CG) resolutions [5–8]. Yet, building and setting up rather intricate cellular systems remain a computational challenge. The experimental resolution and the distribution of different constituents in confined compartments may hinder the proper setup, equilibration and reproducibility of the simulations. Particularly, state-of-the-art experimental techniques are achieving high resolution structures of multi-protein complexes such as viral capsids and envelopes, however membrane components are still hard to solve with enough atomic resolution for the accurate generation of molecular models [9]. In the best case scenario only raw spatial information about geometrical boundaries and lipid densities of membranes can be obtained [10]. This issue is of paramount relevance as advances in the computer software and hardware are allowing the modeling of progressively more realistic biological systems, with the introduction of multiple components in the description of membranes and solutions [11].

The membrane packing problem is an NP-hard problem, which may be solved by self-assembly or templating strategies [12]. While the former methods mimic the natural way by which molecules assemble in the cell [13], it may be cumbersome or very expensive to apply to big and complex systems. On the other hand, the second solution optimizes the arrangement of molecules according to some topological restrictions. However, it requires precise knowledge about the actual number and composition of lipids on each leaflet as well as their orientation and distribution. Some software for building realistic membrane systems by templating are CHARMM-GUI [14], HTMD [15], cellPACK [16], MemBuilder [17], MemGen [18], Insane [19], MERMAID [20], LipidWrapper [21], TS2CG [22], BUMPy [23] and CmME [24].

Some of these tools are being used to model representative virus structures in realistic membrane environments [25]. Nevertheless, building molecular systems of entire enveloped viruses and related virus-like particles (VLP) remains a rather intricate procedure [26,27]. In this regard, we identify three main problems for the multiscale modeling of these systems. The first issue consists of building accurate protein-membrane models according to experimental restraints in confined conditions. The second issue is generating a multiscale representation of the system. And finally, the third issue is having a simple and robust protocol for equilibrating and simulating the multiscale systems.

PACKMOL is a versatile tool for generating molecular systems for simulation by solving a general packing-optimization problem avoiding short-ranged intermolecular repulsive interactions [28–30]. In PACKMOL, different kinds of spatial constraints can be defined for each constituent of the system, to represent the desired overall molecular arrangement. Therefore, it stands as an appealing tool for solving the first two aforementioned issues. Indeed, it was recently used in automated pipelines for building protein-membrane systems [31]. However, there is still room to optimize PACKMOL to efficiently build

enveloped VLPs. In particular, setting the correct membrane thickness, area per lipid and hydrophobic mismatch requires solving lipid-lipid and lipid-protein clashes as well as defects in the lipid distributions. On the other hand, a multiscale scheme implies the coexistence of different granularities or resolutions at the same time in the system. Most notably, the radii of molecular elements (atoms or effective beads, for example), may vary widely, directly implying that the short-ranged packing function must be customized.

Defining general protocols for performing multiscale simulations is a must for spreading the usage of such computational techniques among the scientific community. That requires easy to follow procedures with step-by-step details on each equilibration stage, including hints to overcome common issues [32,33]. They should also be flexible enough to allow for a broad spectrum of applications.

Here, we address the setup and simulation of enveloped VLPs at multiscale resolution. New features and improved options for PACKMOL are introduced to efficiently solve the membrane packing problem and setting up multiscale systems. A systematic and optimized pipeline based on the CG SIRAH force field [34] is designed to equilibrate and simulate the system by Molecular Dynamics (MD). The pipeline is validated on the Zika virus (ZIKV), which represents a stringent test case of highly dense and confined arrangement of heterogeneous components. The strategy may be easily applied to study arbitrary VLPs and facilitate the setup of membrane systems of ever increasing molecular complexity [11].

METHODOLOGY

Virus-like particle model.

The VLP model is based on the structure of the mature virion of ZIKV. The coordinates of ZIKV proteins are taken from the refined structure of the mature particle at 3.1 Å resolution (PDB id. 6CO8 [35]). The atomistic structures of ZIKV proteins are curated according to protocols described in [5]. Briefly, missing side chain atoms and protons in the asymmetric unit are added by PDB2PQR [36] at pH 7. The glycosidic modifications are not considered. The following Cysteine pairs are assumed to form disulfide bridges in E protein monomers: 3:30, 60:121, 74:105, 92:116, 190:291, 308:339. A 5000 steps energy minimization in vacuum is performed using the AMBER 14SB force field [37]. The entire envelope is then generated by applying the symmetry transformations (BIOMT) provided in the corresponding PDB file.

Definition of the VLP membrane.

Location and amount of lipids composing the membrane are estimated from cryo-electron microscopy (cryo-EM) studies on West Nile Virus [10], which is a *Flavivirus* of similar radius of gyration to ZIKV. The experimental evidence shows that the location of the outer leaflet roughly matches the amphiphilic α-helixes at the stem of E protein (residues 406 to 423 and 437 to 460). However, the bilayer follows the hydrophobic mismatch at the different symmetry axes, resulting in a variable membrane thickness and

local curvature pattern [10]. To simplify the computational setup, the bilayer is initially built as a spherical shell. A practical geometrical rule for setting the membrane in mature *Flavivirus* is defining the position of phosphate atoms in the outer leaflet at the radius of gyration of the stem, which corresponds to 191 Å in ZIKV, and using a membrane thickness of 34 Å. The experimentally estimated number of lipid molecules in the outer and inner leaflets ranges from 2700 to 3600 and 3600 to 4800, respectively [10]. Such variability may respond to the very nature of the viral assembly process and is indicative of a biological plasticity in the lipid coverage for achieving correctly/functional assembled particles. In this study, the maximum expected amount of lipids is used to build the VLP of ZIKV. The membrane complexity is defined by the lipid composition. In case of *Flavivirus*, experimental data suggest that their membranes are enriched in phosphatidylcholine (PC), phosphatidylethanolamine (PE) and phosphatidylserine (PS) species according to a 6:3:1 proportion [38]. As specific phospholipid models, we use palmitoyl, oleoyl variants of PC, PE and PS (POPC, POPE and POPS). This lipid diversity is enough to correctly represent the main structural features of *Flavivirus'* membranes by MD simulations [27].

Setting the membrane in other VLPs may depend on the available experimental information. When the lipid density is unknown, the amount of lipids at each leaflet may be inferred from known in:out ratios of pure vesicles [39].

Computational details.

The initial packing of the system is done with PACKMOL (version 18.169 or greater, http://m3g.iqm.unicamp.br/packmol) [28]. The SIRAH force field 2.0 is used to represent the proteins [40] and the lipids [41] in combination with a multiscale solvent model [5]. All MD simulations are performed with the GPU code of GROMACS (version 2018.4, http://www.gromacs.org) [42]. A reference temperature of 300 K is set by coupling solute and each solvent separately to the V-rescale thermostat [43] with coupling times of 2 ps. The pressure is kept at 1 bar by means of a Parrinello-Rahman barostat [44,45] with a coupling time of 8 ps. A minimum cutoff for nonbonded interactions of 12 Å is set. Long-range electrostatics are evaluated using Particle Mesh Ewald [46,47] each 10 integration steps, the same time at which neighbor searching is performed. Newton's equations of motion are solved using a leap-frog integrator algorithm. A time step of 2 fs is used during the first nanosecond of each MD simulation, then it is switched to 20 fs. Snapshots are recorded every 100 ps for analysis.

Analyzed properties.

Membrane thickness and area per lipid are calculated by FATSLiM [48]. Radial distribution functions (RDF) are measured from VLP's center of geometry using the tool g_rdf available in the GROMACS package. The normalized density is calculated according to the maximum value of each component. Radius of gyration (RGYR), root mean square deviations (RMSD) and fluctuations (RMSF) are computed on Cα atoms taking as reference the experimental structure obtained from cryo-EM (PDB id.

6CO8 [35]). RMSF values are calculated from the last 0.1 μs of the trajectory. The average B-Factor of the *i*-th residue in E or M proteins is estimated as $B_i = 8 * \pi^2 * \langle RMSF_i^2 \rangle / 3$ over all corresponding protomers. This value is normalized by the mean and standard deviation along the protein: $B_i' = (B_i - \langle B \rangle) / \sigma(B)$. Average volumetric occupancy and density maps are calculated with the VolMap plugin of VMD [49]. Weighted density maps by the occupancy are generated for the entire VLP (protein and membrane) or separated components from the CG coordinates by using an atom size of 2.1 at grid resolutions of 3.1, 7.0 or 10.0 Å. Averages correspond to time windows of 0.1 μs, unless otherwise stated. Correlations against the cryo-EM density map (EMD-7543) are calculated by the Fit in Map tool of Chimera (http://www.cgl.ucsf.edu/chimera) [50], without performing local and global optimizations. All molecular representations are generated by VMD [49].

RESULTS

Solving the membrane packing problem with PACKMOL.

Assembling of the membrane at atomic level is done with PACKMOL [28]. Owing to the complexity of the membrane packing problem, PACKMOL does not converge to a perfect packing [31]. Indeed, there may not exist a solution fulfilling all conditions for packing complex rigid objects in such dense arrangement, requiring some flexibility (internal rotations) to achieve that goal. However, it is possible to generate good enough arrangements for energy minimization and simulation. To do that, we introduce an improved heuristic to surf the solution space in a way that the algorithm focuses on relevant parts of the problem while avoiding to spend time in less important aspects of the system. Optimized and newly implemented options to ensure a good and efficient protein-membrane packing are listed in Table 1 and discussed below.

We first introduce improvements for building up pure membrane systems. Initially, we describe the construction of a reference flat membrane of 140 Å per XY side composed of 306 POPC molecules in the top leaflet and a mixture of 223 POPE and 111 POPS lipids in bottom leaflet, as reported in ref. [41]. The atomic structure of each lipid is provided within the SIRAH force field package (http://www.sirahff.com). The orientation of lipids along the Z axis is defined by two planes, one restraining the position of phosphate atoms of the phosphoglycerol group and the other the terminal acyl tail, such that these two groups are at least 19 Å apart (Figure 1A). Phosphate atoms on each leaflet are set to face the solvent, while the distance between phosphate's planes aims at a membrane thickness of 38 Å. These are standard geometrical restraints used for building membranes with PACKMOL [31]. Default and optimized options for packing lipids are then tested.

Table 1. List of keywords for optimizing the PACKMOL execution.

[1]Input section	Keyword	Type	Description	Default value
general	movebadrandom		Place worst scored molecules at new random positions instead of nearby well packed molecules.	Not used
	maxit		Maximum number of iterations of the local optimizer (GENCAN) per loop.	20
	nloop		Maximum number of optimization loops.	
	movefrac		Fraction of moved molecules per optimization loop.	0.05
	tolerance		Maximum interatomic distance defining a clash (Å).	2.0
	use_short_tol	new	Use a penalty tolerance for short distances.	Not used
	short_tol_dist	new	Distance from which the penalty is applied (Å).	none
	short_tol_scale	new	Scaling value of the short distance penalty term.	none
	packall	new	Skip initial individual packing steps	Not used
general/ structure	restart_to	new	Save packing state information to file.	Not used
	restart_from	new	Retrieve packing state information from file.	Not used
structure	radius	new	Atom radius (Å).	Not used
	fscale	new	Scaling value of the full distance penalty term.	1

[1] *Structure* is a keyword wrapping other keywords

One difficulty in packing very large systems efficiently is the identification of empty spaces. PACKMOL tries to generate an homogeneous distribution of the molecules and uses increased packing radii in initial packing iterations to occupy empty spaces, but the default heuristics are insufficient to avoid uneven lipid distributions in the systems considered here (Figure 1B left). Two parameters modulating these issues are the size of the beads and the packing heuristic. Setting an atom radius of 1.5 Å to Hydrogen-less lipid molecules, instead of 1.0 Å (keyword *tolerance*, Table 1), renders a better reproduction of the area per lipid in flat membranes. In addition, randomly placing the worst scored molecules (keyword *movebadrandom*, Table 1) instead of moving them nearby well packed molecules (the default option), avoids the formation of artifactual lipid clusters of same species, while promoting their better distribution and spread along the entire membrane surface (Figure 1B right). Clashes are naturally amended by the GENCAN algorithm [51], but due to the way the problem is partitioned in space according to the defined molecule sections, a poor treatment of intra- and/or inter-leaflet clashes is performed, leading to excessive entanglement and interdigitation issues in output structures. By default, each type of molecule (as defined by the *structure* keyword, Table 1) is initially packed independently in the defined spatial regions, and overall resolution of inter-species clashes is performed afterwards. Since the membrane constituents occupy the same regions in space, such procedure is not beneficial here. This issue is magnified by the limited number of optimization loops, the diversity of constituents, the size and complexity of the system. To solve this issue, an option is introduced to allow for packing all molecular groups together at once (keyword *packall*, Table 1). Although this strategy may slow

down the convergence in systems with a clear phase separation, it speeds up the solution search in case of mixtures or soft interfaces (Figure 1B).

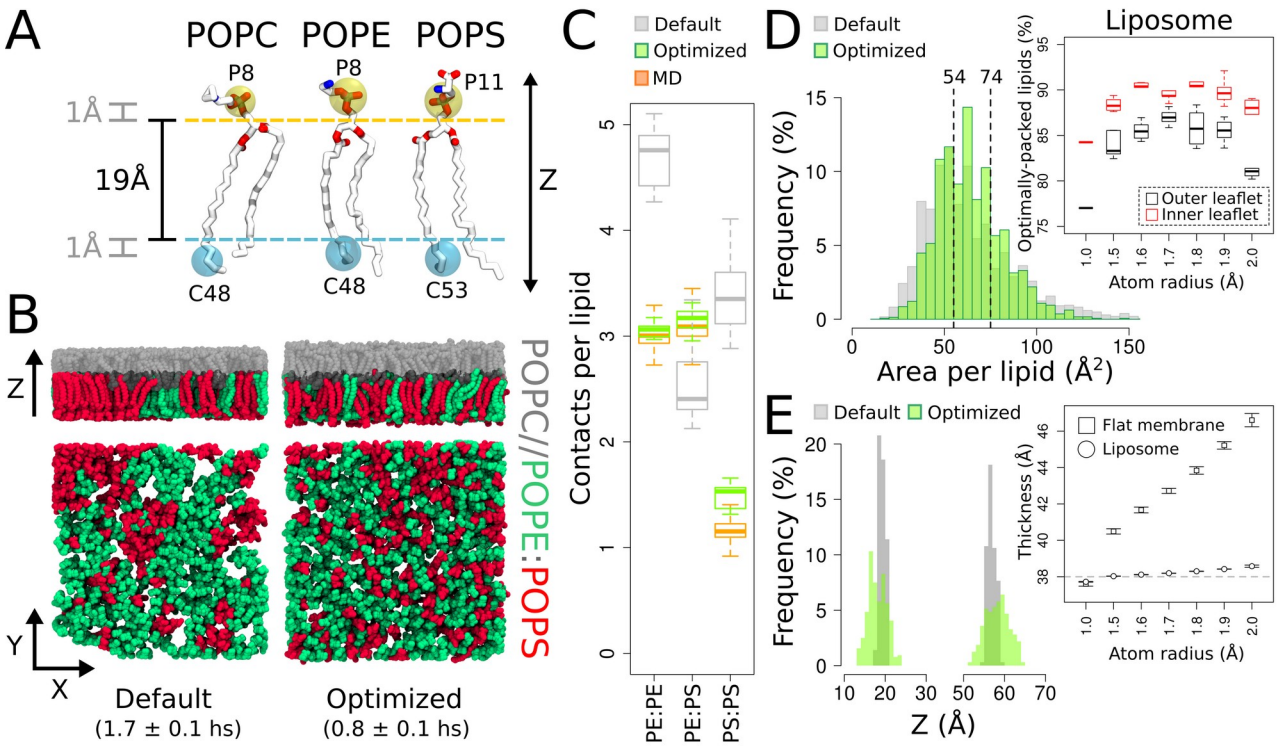


Figure 1. Optimizing the packing of pure lipid membranes with PACKMOL. A) Lipid conformations and geometrical restraints defining a leaflet. Restricted atoms are represented as transparent spheres of the corresponding color to the geometrical restraint. B) Representative configuration of resulting membranes using default and optimized packing protocols. Average execution times on a single CPU Intel Core i7-5930K, 3.5GHz are listed. C) Lipid-Lipid contacts for membranes in panel B. Lipids are identified by their head groups. Contacts are measured from the central bead of the glycerol moiety using a cutoff distance of 11 Å. MD values are taken from the last 0.5 μs of simulation in ref. [41]. Values are normalized by the number of molecules of the second lipid species in the pair. D) Distribution of area per lipid values for individual POPC molecules in flat membranes of panel B. The inset shows the effect of the used atom radius on the percentage of lipids with ideal areas in the packing of a referential liposome system. E) Phosphate distributions along the Z axis in flat membranes of panel B. The inset shows the dependency of the membrane thickness on the used atom radius for packing a flat membrane or a liposome. PACKMOL results correspond to 10 replicates.

Besides the molecular shape, other physicochemical features of the lipids and particular characteristics of membrane, such as its geometry, composition and protein context, may influence the bilayer organization. Interesting, the optimized PACKMOL options clearly improve the distribution of lipid in the studied membrane system as evidenced by the lipid-lipid contacts (Figure 1C). Particularly, contacts among POPE, POPE-POPS and POPS in the optimized membrane are comparable to those obtained after 1.0 μs of CG simulation with SIRAH force field at 310 K. This is very important considering that ~ 0.2 μs of MD are required to equilibrate the lipid mixture, implying an important gain in computational time for this system (see Figure 5 from ref. [41]). Similarly, the distribution of individual areas per lipid is well centered on a target value of 64 ± 10 Å^2 (Figures 1D). Using default options produces a broad distribution, which is shifted to values below 50 Å^2 due to the presence of over-packed

regions, while having 8.5 % of values larger than 100 Å^2 due to depleted regions. However, we observe that optimized options generate less compact membranes (thickness of 40.1 ± 0.2 Å), which are ~ 2.0 Å thicker than using default options (37.7 ± 0.1 Å) and the target value of 38 Å (Figure 1E). Both membrane thickness and area per lipid depend on the lipid conformation and the atom size. While the volume of head groups mainly contributes to the correct lipid distribution and area per lipid, the tail's volume and conformation restrict the available space for proper interdigitation between lipid layers, modulating the thickness. In particular, an atom radius of 1.5 Å is the best choice for the used lipid structures, as larger values greatly impact the final thickness of flat membranes after packing (Figure 1E inset).

Flat and curved membranes may imply different lipid conformations and distributions. Hence, to explore the effects of the system's geometry on the selected parameters, a liposome with an outer radius of 191 Å is packed assuming the same thickness and lipid composition of the previously introduced flat membrane. The liposome size is set to mimic the ZIKV membrane, containing 7729 and 6268 lipids in the outer and inner leaflet, respectively (see Methodology). The packing uses the optimized options, applying the same geometrical restraint in Figure 1A to define each leaflet. The quality of the packing is evaluated by measuring the percentage of lipids with areas of 60 ± 10 Å^2 or 47 ± 10 Å^2 for outer and inner leaflets, respectively. Using an atom radius of 1.5 Å for the lipids already produces very good packing results, while increasing the value provides some gains mainly at the inner layer due to its larger curvature (Figure 1D inset). Setting the atom radius larger than 1.9 Å is detrimental to the packing quality. On the contrary to flat membranes, the atom radius shows minor impacts on the target thickness (Figure 1E inset). Based on these results, we propose to use an atom radius between 1.5 and 1.9 Å for packing spherical vesicles.

The presented optimized options for PACKMOL represent a good choice for building high quality membrane systems. These options are kept in the following section to improve the solution of the protein-membrane packing problem.

Solving the protein-membrane packing problem with PACKMOL.

We next focus on improving the packing of lipids at the protein's interface. Packing all components together may be a hard and time consuming task, requiring tweaks to be efficient. Packmol is implemented to only allow rigid-body rotations and translations of each molecule type. Hence, the use of a particular molecular conformation may restrict or compromise the packing-solution space. As pointed out for flat membranes, this aspect is relevant for reproducing the membrane thickness and the protein-membrane hydrophobic mismatch. To simplify the input file, a single and rather straight lipid conformation is used for each species (Figure 1A), but applying different spatial constraints to better account/compensate for the reduction in degrees of freedom. In this case, the orientation of the lipid in each leaflet is set by restraining the phosphate atom of the phosphoglycerol group to be within a slab of 2.0 Å thick, while the tip of one tail is set to be at least 12 Å below the phosphate by restraining the tip

of the acyl tail (Figure 2A). The constraining boundaries are planes in case of flat membranes or nanodiscs, but spherical shells in liposomes and enveloped VLPs. The membrane thickness is defined between the middle radii of phosphate slabs at outer and inner leaflets. This setup grants a broad spectrum of tilting angles which allows for a better fit to the protein's shape, and facilitates the optimization of clashes without affecting the specified thickness (see below).

Embedded proteins containing Hydrogen atoms are treated as fixed steric constraints with atom radius of 1.5 Å by setting keyword *tolerance* to 3.0 Å (Table 1). A radius of 1.8 Å is assigned to the lipid atoms (keyword *radius*, Table 1). We observed that, given the huge number of protein atoms, a common issue is that some lipids remain entangled with the protein core, preventing adequate termination of the packing process. Increasing the radii of the protein atoms is a possible workaround that shifts the distribution of lipids away from the protein surface. However, it creates solvent pockets and over-packing lipids at inter-protein regions, thickening the membrane. Then, instead of increasing the radii of protein atoms, a new parameter is introduced to increase the weight of the overlap penalty for atoms of a given structure, named f_{scale} (Table 1, Figure 2B left). The effect of using a larger penalty on the protein is that lipids entangled with it will have associated the largest overlap penalties, and will be translated to new random positions eventually. Additionally, we implement a new short-ranged atom-pair penalty that can be summed up to the standard penalty function (Figure 2B right). While the former penalization only reflects the size of the particles, the new short-ranged penalty improves the mitigation of bad contacts, which are critical for the simulation of the system. A schematic pseudo-code of the overlap part of the objective function in the new PACKMOL implementation is, therefore,

```
f_overlap = 0
for i in (system atoms)
  for j in (system atoms)
    if (molecule of i) ≠ (molecule of j)
      d_ij = ||x_j-x_i||
      f_overlap ← f_overlap + f_scale × max{0,(d_tol²-d_ij²)}²
        if ( use_short_tol(i) or use_short_tol(j) )
          f_overlap ← f_overlap + s_scale × (d_tol/s_tol)⁴ × max{0,(s_tol²-d_ij²)}²
```

where the terms in bold correspond to new additions to the PACKMOL objective function. The overlap penalty is calculated from the interatomic distance d_{ij}. Parameters f_{scale} and d_{tol} are the weight and sum of radii of the two atoms, and s_{scale} and s_{tol} the equivalent parameters of the new short-ranged overlap penalization potential. The weights of each overlap penalty, f_{scale} and s_{scale} are obtained by the product of the corresponding user-defined parameters for each structure type (Table 1). Analytical derivatives are implemented for all terms. The complete objective function is the sum of the overlap function to the

penalties associated with the regions of the space to which each structure must be located [28]. Efficient implementation of this objective function is achieved using a linked-cell approach [52].

By tweaking the new parameters it is possible to prioritize the optimization of specific contacts (Figure 2B). In particular we empirically find that particles closer than 0.21 Å can not be minimized at CG level by MD engines. Similarly, avoiding lipid-protein contacts below 2.0 Å prevents lipid miss-placements inside the proteins' core. Both criteria are used to define the number of remaining critical contacts after packing.

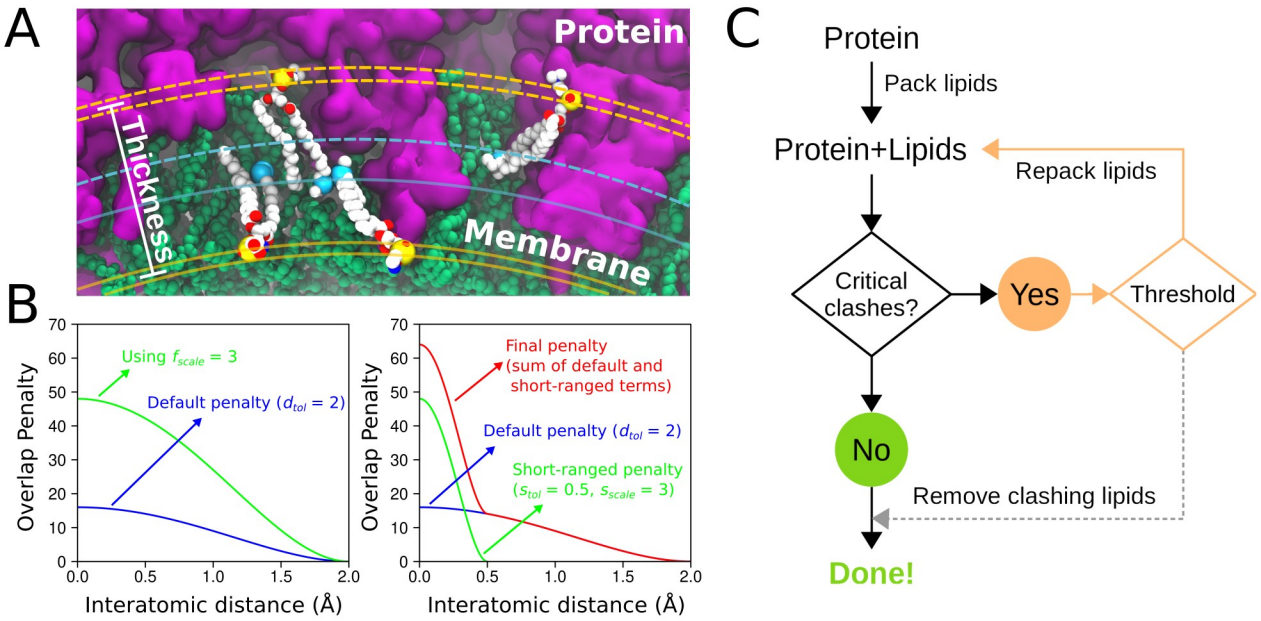


Figure 2. Solving the protein-membrane packing problem with PACKMOL. A) Geometrical restraints defining each leaflet in a VLP system are indicated by solid and dashed lines. Restricted atoms are represented as transparent spheres of the corresponding color to the geometrical restraint. The packing of four lipids is highlighted using white spheres. B) New features in the PACKMOL scoring function. Left and right panels show the effect of applying the f_{scale} parameter or including a short-distance penalty (s_{tol} and s_{scale} parameters) in the default penalty, respectively. C) Optimized workflow for packing protein-membrane systems with PACKMOL.

An enveloped VLP based on the mature virion of ZIKV is used as a representative case for optimizing the protein-membrane packing protocol. The structure of the VLP, including the protein and membrane definition, is described in the Methodology. Table 2 shows an exhaustive exploration of the parameter space in PACKMOL to find the best combination of options for packing lipids within this challenging VLP context. Using default packing options renders a large number of critical clashes in the final structure (~ 18 %), which corresponds to bad lipid-protein contacts (pack #1 from Table 2). Hence, the output from a default packing protocol can not be use to perform MD simulations of VLP systems. To circumvent this problem we applying a penalty value of 100 to any protein related contact. That greatly reduces the amount of critical clashes (<1 %, pack #2 from Table 2). Importantly, a fine balance between solving protein-lipid and lipid-lipid clashes is required to correctly pack all constituents. That is because of the difficulty to shape the penalty function for efficiently optimizing every aspect of the

packing at the same time. For example, using a short distance cutoff for pruning bad lipid-lipid contacts at this stage is counterproductive (compare pack #2 against #3-5 in Table 2). Similarly, increasing the protein-lipid penalty too much increases the number of lipid-lipid clashes (compare packs #8-10 form Table 2). However, a significant boost is obtained by reducing the fraction of moved lipids per optimization loop (keyword *movefrac*, compare packs #2 and #6 from Table 2). Similarly, equivalent results can be obtained in half time by reducing the number of optimization loops (keyword *nloop*, compare packs #2 and #7 in Table 2). Hence, going through extensive optimization loops, while moving many molecules on each round, is not an efficient packing strategy for large or dense systems. Resetting too many positions just resets the packing problem in other regions of the system. Indeed, moving 5 % of lipids each time along 200 optimization loops, is equivalent to move all membrane's constituents 10 times through the packing procedure, in the extreme case. Further reducing the optimization exhaustiveness reduces the packing time at expenses of lowering the precision of the solution (compare packs #7 and #8 in Table 2). On the other hand, increasing the maximum number of local iterations per optimization loop (keyword *maxit*, Table 1) while keeping the same exhaustiveness has no significant effect on the final solution quality nor running time (compare pack pairs #6 and #13, #7 and #12, #8 and #11 from Table 2).

Table 2. Parameter optimization of the first protein-membrane packing cycle.[1]

	PACKMOL parameter						Lipid-Lipid contacts < Å					[3] Lipid-Protein contacts	[4] Critical clashes		[5] Time (h)
Pack #	[2] fscale	movefrac	maxit	nloop	short_tol_dist (Å)	short_tol_scale	0.21	0.3	0.4	0.5	1.0				
1	-	0.05	40	200	-	-	0	2	6	14	229	1514	1514	(18.0)	10.1
2	100	0.05	40	200	-	-	14	64	146	273	1735	33	47	(0.6)	8.8
3	100	0.05	40	200	0.5	3	4	307	750	1375	4821	224	224	(2.7)	9.4
4	100	0.05	40	200	0.5	10	4	223	620	1132	4580	215	219	(2.6)	9.0
5	100	0.05	40	200	0.5	100	2	404	1246	2255	6199	353	354	(4.2)	9.4
6	100	0.01	40	200	-	-	8	22	77	153	1078	10	18	(0.2)	9.5
7	100	0.01	40	100	-	-	22	68	130	245	1730	17	39	(0.5)	4.4
8	100	0.01	40	50	-	-	42	114	251	471	2548	25	67	(0.8)	2.3
9	200	0.01	40	50	-	-	59	171	433	827	3770	37	95	(1.1)	2.5
10	500	0.01	40	50	-	-	93	291	694	1182	4547	14	105	(1.3)	2.5
11	100	0.01	80	25	-	-	46	126	317	553	2657	31	75	(0.9)	2.3
12	100	0.01	80	50	-	-	22	58	120	244	1719	8	30	(0.4)	4.6
13	100	0.01	80	100	-	-	8	18	54	109	958	13	21	(0.3)	10.3
14	100	0.01	80	200	-	-	2	16	50	109	839	13	15	(0.2)	18.3

[1] The optimization is based on ZIKV system. The protein is treated as a fixed restraint. The used *keywords*/parameters in all tested packs are: *packall*, *movebadrandom*, *tolerance* 3.0 (Å), lipid atoms with *radius* 1.8 (Å).
[2] The scaling coefficient is applied to the protein by setting the keyword within its structure section to modify the scoring of contacts against other molecules.
[3] Lipids closer than 2 Å to the protein.
[4] Critical clashes are defined as lipid molecules closer that 2.0 Å to the protein or 0.21 Å to other lipid. As the same molecule may be clashing both the protein and a lipid at the same time, the metric may not match the sum of lipid-lipid and lipid-proteins contacts. Percentages are shown between brackets relative to the total amount of lipids in the membrane (8400 molecules).
[5] Running time on a single CPU (Intel Core i7-5930K, 3.5 GHz).

In summary, we find that exhaustive packing protocols #6 and #13 render the best solutions for the explored parameters. In addition, protocols #7 (or #12) and #8 (or #11) reach very good solutions, presenting good potential as fast methods. However, in these last cases a single packing protocol may not be enough to solve all critical clashes, requiring to iterate over repacking cycles to improve the solution's quality (Figure 2C). Ideally, the optimization cycle stops when no critical clashes exist, but we observe that 2 iterations are enough to achieve a very low level of critical clashes (below 0.3 %) in a

short running time. This is easily done by using restart files (keywords *restart_to* and *restart_from,* Table 1), a feature also implemented for the current work. Remaining problematic lipid molecules are removed from the structure to get the final protein-membrane system. We compare the repacking from two possible scenarios to render a *good-compromised* (pack #7 then #23) and a *quick* (pack #8 then #32) packing strategy (see Tables 2 and 3). These protocols differ on the execution time and final membrane quality, which is measured in terms of all lipid-lipid contacts <1 Å. During repacking cycles new option values are set to account for neglected aspects in previous iterations. Briefly, in all cases, the best repacking protocol consists of further reducing the fraction of moved lipids per optimization loop and the number of loops to focus on very critical contacts while avoiding the introduction of new issues by chance. The protein-lipid penalty is increased and a short distance penalty below 5.0 Å is introduced. As an example, packing 8400 lipids within an intricate arrangement of proteins takes about 5 hs running the good-compromised protocol on a single CPU. It is important to note that repacking from more exhaustive protocols (e.g. pack #6), which already provides critical clashes below 0.3 %, renders little gains or even drawbacks (Table 3). Hence this quality threshold is about the limit of the current PACKMOL's heuristic to solve the protein-membrane packing problem. The described equilibration protocol for the VLP is based on the *good-compromised* packing solution (pack #7 then #23, Tables 2 and 3).

Table 3. Parameter optimization of the second protein-membrane packing cycle.[1]

Restart from Pack #	Repack #	PACKMOL parameter						Lipid-Lipid contacts < Å					[3]Lipid-Protein contacts	[4]Critical clashes		[5]Time (h)
		[2]fscale	movefrac	maxit	nloop	short_tol_dist (Å)	short_tol_scale	0.21	0.3	0.4	0.5	1.0				
6	15	100	0.01	40	20	-	-	20	54	133	244	1352	15	35	(0.4)	1.2 (10.7)
	16	100	0.005	40	20	-	-	6	16	56	122	985	6	12	(0.1)	1.3 (10.8)
	17	100	0.005	40	20	0.3	3	4	26	56	122	935	10	14	(0.2)	1.3 (10.8)
	18	100	0.005	40	20	0.5	3	0	30	76	149	1048	14	14	(0.2)	1.2 (10.7)
	19	200	0.005	40	20	0.5	3	0	22	94	164	1166	11	11	(0.1)	1.3 (10.8)
	20	500	0.005	40	20	0.5	3	0	28	88	185	1226	6	6	(0.1)	1.4 (10.9)
7	21	100	0.005	40	10	0.5	3	0	48	109	213	1724	16	16	(0.2)	0.5 (5.0)
	22	100	0.005	40	20	0.5	3	0	48	109	213	1724	16	16	(0.2)	1.0 (5.4)
	23	500	0.005	40	20	0.5	3	0	49	132	260	1753	1	1	(0.0)	1.2 (5.6)
	24	100	0.005	80	10	0.5	3	0	36	70	154	1424	23	23	(0.3)	1.0 (5.4)
	25	500	0.005	80	10	0.5	3	0	66	175	334	1956	13	13	(0.2)	0.8 (5.2)
8	26	100	0.005	40	10	-	-	38	106	235	415	2390	17	54	(0.6)	0.9 (3.1)
	27	100	0.005	40	10	0.3	3	28	78	185	363	2257	19	47	(0.6)	0.7 (2.9)
	28	100	0.005	40	10	0.5	3	4	96	228	401	2440	19	23	(0.3)	0.6 (2.9)
	29	100	0.005	40	20	-	-	20	96	223	450	2280	12	32	(0.4)	1.3 (3.5)
	30	100	0.005	40	20	0.3	3	24	72	141	317	2082	15	39	(0.5)	1.2 (3.5)
	31	100	0.005	40	20	0.5	3	4	96	228	401	2440	19	23	(0.3)	1.2 (3.5)
	32	500	0.005	40	20	0.5	3	6	109	258	450	2549	10	16	(0.2)	1.2 (3.5)
	33	100	0.005	80	10	0.5	3	0	95	210	440	2490	22	22	(0.3)	0.6 (2.9)
	34	200	0.005	80	10	0.5	3	2	94	220	432	2311	14	16	(0.2)	0.8 (3.1)
	35	500	0.005	80	10	0.5	3	0	89	229	484	2511	8	8	(0.1)	1.2 (3.5)

[1] The optimization is based on ZIKV system. The protein is treated as a fixed restraint. The used *keywords*/parameters in all tested packs are: *packall*, *movebadrandom*, *tolerance* 3.0 (Å), lipid atoms with *radius* 1.8 (Å).
[2] The scaling coefficient is applied to the protein by setting the keyword within its structure section to modify the scoring of contacts against other molecules.
[3] Lipids closer than 2 Å to the protein.
[4] Critical clashes are defined as lipid molecules closer that 2.0 Å to the protein or 0.21 Å to other lipid. As the same molecule may be clashing both the protein and a lipid at the same time, the metric may not match the sum of lipid-lipid and lipid-proteins contacts. Percentages are shown between brackets relative to the total amount of lipids in the membrane (8400 molecules).
[5] Running time on a single CPU (Intel Core i7-5930K, 3.5 GHz). The accumulated time of the two packing cycles is shown between brackets.

Setting up multiscale systems with PACKMOL.

To further reduce the computational cost of the simulation, the solvent is represented at two resolution levels by using a CG water model (WT4) [53] and a supra-CG solvent (WLS) [5]. The aqueous environment around the VLP is described by a shell of WT4 and CG ions of 25 Å from the solute. The remaining space of the computational box is filled by WLS to render an onion like configuration of the system (Figure 3A). The initial arrangement of the WT4 shell and the inner WLS phase is built with PACKMOL in absence of the solute. The amount of solvent molecules is calculated from the volume of each shell by considering the molecular weight of WT4 (200 a.u.) and WLS (1000 a.u.) as well as their density (1000 g L^{-1}). Building multiscale systems is allowed by using different radii for each particle's model resolution in the system during packing (keyword *radius*, Table 1). Specifically, beads of WT4 and WLS are set a radius of 2.2 and 3.4 Å, respectively. Due to the rather spherical symmetry of the solvent molecules, successful solutions for the packing problem can be found. Table 4 shows the optimization of parameters to improve the execution time of PACKMOL, being pack #5 the best found protocol.

Table 4. Parameter optimization for packing the multiscale solvent.[1]

	PACKMOL parameter			
Pack #	packall	movebadrandom	movefrac	[2] Time (h)
1	Yes	Yes	0.05	[3] 58.0
2	Yes	No	0.05	[3] 53.4
3	No	No	0.05	8.0
4	No	No	0.01	5.3
5	Yes	No	0.01	2.7
6	Yes	Yes	0.01	7.5

[1] The packed system contains 201649 WT4, 180 CG ions and 6144 WLS molecules, which definition is based on ZIKV.
[2] Running time on a single CPU (Intel Core i7-5930K, 3.5 GHz).
[3] No successful packing is found.

Setting up the solvent box.

For technical convenience, the external WLS is added using GROMACS' tools from small pre-equilibrated solvation boxes, which are provided within the SIRAH force field. This method is not only faster than packing individual molecules, but better for defining the most efficient orientation for the computational box according to the used MD engine. In addition, any issue in the outer solvent density is rapidly solved during NPT equilibrations. A slab of at least 50 Å width (considering PBC images) is required to ensure the proper behavior of WLS [5]. To match that requirement, an octahedral box of length 600 Å is used in case of ZIKV. The multiscale solvent is merged with the VLP and equilibrated according to the protocol described in the next section (Figure 3A).

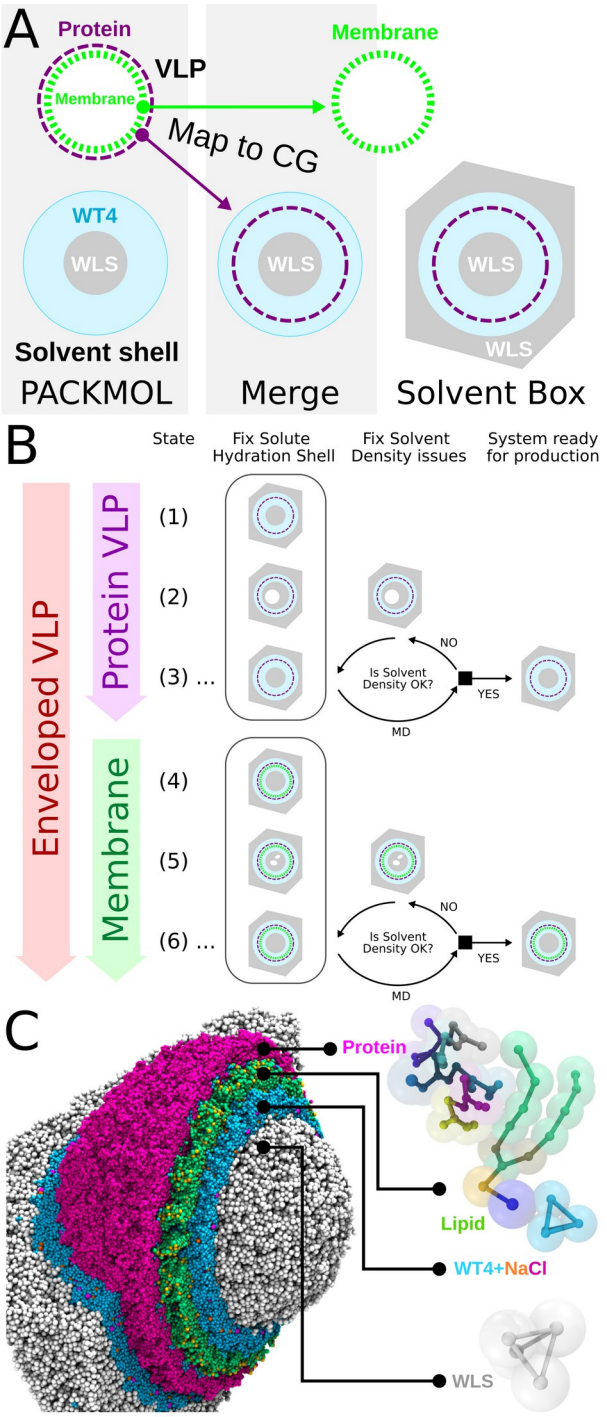


Figure 3. Pipeline for building and simulating enveloped VLP systems at multiscale resolution. A) Building phase. Proteic and lipidic components of the VLP are packed. The initial multiscale solvation shell is generated. The VLP is map to CG. The protein shell of the VLP is merged to the multiscale solvent and the simulation box is created. Membrane coordinates are used at later stages of the equilibration protocol. B) Stepwise protocol for equilibrating enveloped VLPs. The equilibration is split into *Protein VLP* and *Membrane* protocols. C) Resulting enveloped VLP of ZIKV after executing the pipeline described in panels A and B. Molecular components of the system are shown according to their relative size.

Pipeline for multiscale simulations of VLPs.

The multiscale simulation strategy was first introduced in ref. [5] for proteinaceous VLPs, and it is now extended to describe enveloped VLPs. The pipeline for setting up proteinaceous and enveloped systems consists of a building phase followed by an equilibration phase. During the building phase (Figure 3A), an atomistic model of the enveloped VLP is generated by packing the viral proteins with the lipids. At the same time, an initial multiscale solvation shell, involving WT4 and WLS water molecules, is also packed. The enveloped VLP system is then mapped to its CG representation using SIRAH Tools [54]. The protein shell of the VLP is superposed with the initial multiscale solvation shell. Solvent molecules in close contact with the solute are removed by using a distance cut-off of 3.0 Å, as recommended in ref. [40]). The computational box is generated by adding WLS molecules and counterions at WT4 phase if needed. Membrane coordinates are used at later stages of the protocol.

The equilibration phase starts with the *Protein VLP protocol*, which aims at fixing solvation issues at the protein's hydrophilic surface (Figure 3B, state #1). The protein shell from the building phase of the VLP is minimized using 5000 steps of steepest-descent/conjugate-gradient. During the minimization, positional restraint of 2.4 kcal mol^{-1} $Å^{-2}$ are applied on backbone beads of the protein to improve the residues' side chain hydration, as shown in ref. [40]. In case of enveloped VLPs, all beads in the transmembrane region are also restrained with the same force constant to preserve the original side chain conformations, which render the packing quality of membrane generated at building phase. Then a second minimization of 5000 steps is done by only restraining the transmembrane regions. The system is equilibrated by 6 ns of MD applying restraints on backbone beads and transmembrane regions as before. As a result of movement and reorganization of water molecules around the protein's surface, density issues (e.g. vacuum bubbles) in the multiscale solvent structure may arise (Figure 3B, state #2). Such a problem is magnified by the impossibility to exchange molecules between inner and outer compartments. Hence, the multiscale solvent structure is improved by keeping the first hydration shell of WT4 (10 Å) around the protein and resolvating the VLP with the same initial solvent structure generated at the building phase (Figure 3B, state #3). New solvent molecules clashing with the equilibrated hydration shell are removed. In case of naked VLPs, the equilibration proceeds by fixing eventual density issues at the inner WLS phase, as described in [5]. On the other hand, the equilibration of enveloped VLPs proceeds with the *Membrane protocol*, which aims at fixing solvation issues in lipidic vesicles. Splitting protein and membrane equilibration steps avoids introducing solvent gaps at the protein-membrane interface, which may lead to artifactual adsorptions of lipids on the protein during the simulations. The membrane is added to the system by merging the lipids' coordinates generated at the building phase to the resulting structure at state #3, leading to state #4 (Figure 3B). Solvent molecules in close contact with the membrane (less than 3.0 Å) are removed. Counterions are added to WT4 phase if needed. The system is minimized applying positional restraints on all backbone beads of the proteins. The membrane is relaxed by two equilibration steps of 6 ns, first applying positional restraints of 2.4 kcal mol^{-1} $Å^{-2}$ on phosphate beads of the outer leaflet and then without

restraining the lipids. During both equilibration steps, the protein's backbone is restrained as before. As already pointed out, solvent reorganization may generate density issues in the multiscale representation (Figure 3B, state #5). As before, the multiscale solvent structure is improved by keeping the first hydration shell around the protein and membrane and resolvating the system again with the same initial solvent structure of the building phase (Figure 3B, state #6). At this state, 0.15 M of NaCl is added to the WT4 phase according to the SPLIT method [55], by randomly replacing solvent molecules. The system is minimized and equilibrated by 6 ns of MD, applying restraints on backbone beads as before. Then, 11 ns of unrestrained MD is performed. The solvent density is checked along the simulation box. If the density is correct, then the system is ready for production MD. Otherwise, it is indicative of vacuum bubbles, in which case it is required to re-solvate the system by using small pre-equilibrated solvent boxes of WLS provided within the SIRAH force field. The system is minimized, equilibrated and simulated as before. The re-solvation step is repeated until the solvent density is correct, which in most of the cases is achieved after the first iteration of the loop, according to our experience.

This general pipeline can be implemented in a practical way on different MD engines depending on their specific features. In particular, we tune the pipeline for GROMACS in order to preserve the compatibility of topologies and coordinates along the protocol while minimizing the number of executed commands and generated intermediate files. In our experience, the whole pipeline takes about 7 to 10 days for ZIKV running on an Intel Core i7-5930K 3.5 GHz with a GPU Tesla K40c. A structural representation of the finally equilibrated ZIKV system is shown in Figure 3C. The following section describes the results from a production simulation of 1.5 μs for this system.

Multiscale simulation of ZIKV.

As a highly non-trivial example of application, we show the modeling and simulation of the ZIKV. ZIKV is an enveloped positive-stranded RNA virus from the genus Flavivirus, which has recently risen important worldwide health concerns due to its neurological effects upon infection to human hosts [56]. The mature virion is composed of the genetic material stabilized by the structural protein named Capsid (C), which are surrounded by a lipid membrane and 180 copies of both the Membrane (M) and Envelope (E) proteins [57]. As imperfect icosahedral symmetry of C proteins impairs their accurate solution and the genome in the mature viral particle [58], only the coordinates of the E and M proteins are available. In the mature particle, E and M heterodimers are arranged in an icosahedral symmetry to produce a VLP of nearly 210 Å radius [35]. Other naturally occurring or engineered Flavivirus-like particles have also been characterized, some of which have applications in vaccinology. [59–61]

Following the above-described pipeline, we construct and simulate a multiscale model of an enveloped VLP based on the mature virion of ZIKV. As shown in Figure 3C, the VLP is arranged in concentric shells of components. The calculation of the different components' radial distribution during the simulation shows the expected partition of molecular constituents (Figure 4A). The inner core of supra CG water (WLS) is surrounded by CG water and ions, followed by the lipid bilayer. The protein spans

from the membrane to the second shell of CG solvent and the whole system is surrounded by supra CG water. Detailed analysis of the protein's distribution shows a main peak corresponding to the ectodomain, which nicely matches the experimental radius of gyration. Moreover, the two shoulders in the distribution can be ascribed to transmembrane and amphipathic helices in E and M. Our protocol correctly reproduces water molecules' presence in the interstices between the membrane and the ectodomain. The presence of solvent in these interstices, indicated by a minor peak in WT4's distribution, is fundamental to solvate amphipathic helices and avoid the ectodomain's collapse onto the membrane.

The examination of global dynamic descriptors suggests that the accurate construction of the initial multi-component configuration leads to a swift stabilization on gross features of the system. Indeed, the radius of gyration of the VLP stabilizes already within nanoseconds, converging to values within 2 Å of the experimental value (Figure 4B). The same observation holds valid when individually considering E or M proteins, highlighting the quality of the model building on lipids and aqueous solvent. In contrast, the protein's RMSD calculated on the VLP shows a slower stabilization, which likely reflects quaternary structure movements associated with thermal fluctuations that drive the system out of the perfectly symmetrized icosahedral structure. Indeed, the RMSDs calculated on individual proteins evidence smaller values for E and M chains, with slightly faster convergence. Despite this, comparing the protein's fluctuations observed during the multiscale simulation against the experimentally derived B-factors provides a quasi quantitative agreement (Figure 4C). The correspondence of values from experimental and simulated structures at transmembrane regions of E and M proteins is notorious, pointing to a good quality of the protein-membrane model. Moreover, the flexibility at regions triggering the infection process and the immune escape, such as the of fusion loop (FL) and variable glycosylation loop (VGL) [62,63], is well reproduced.

To acquire a more stringent comparison against experimental data, we directly compare the density map measured during the simulation to the electron density obtained from cryo-EM experiments [35]. To increase the statistics obtained from the simulation, we first filter the trajectories of individual trimers of E-M heterodimers (in the *Flavivirus*' jargon, this is called a protein raft, Figure 5A left). Then, we concatenate individual raft's trajectories from the entire VLP. On this concatenated trajectory, we calculate the averaged volumetric occupancy. As seen from Figure 5A right and Figure 5B, the matching between both density maps is outstanding. The excellent agreement extends to the protein part and the bilayer, which shows a flattening in the regions away from the transmembrane helices.

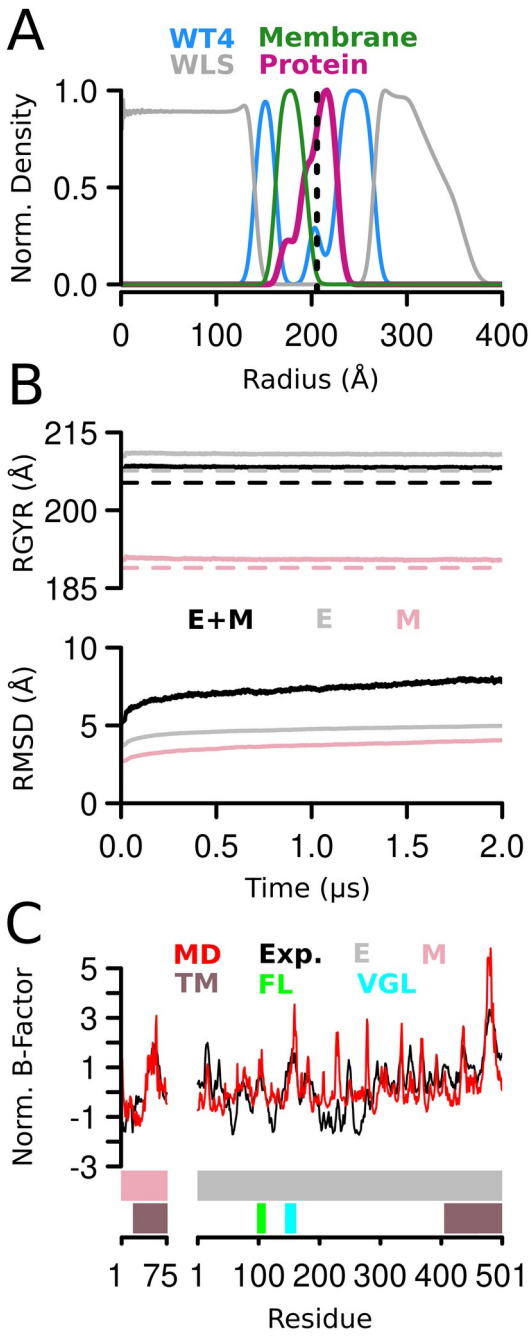


Figure 4. Global MD descriptors in the multiscale simulation of ZIKV. All experimental values are taken from PDB id. 6CO8 [35]. A) Normalized density of each component in the simulated system. Values are averaged over the last 0.1 μs of trajectory. The dashed line is the experimental RGYR of the VLP, same as indicated in panel B. B) Structural features of the VLP along the simulation. Values are reported for the entire VLP (E+M proteins) and averages over individual monomers (E or M). C) Normalized B-factors of E and M proteins. Locations of the transmembrane region (TM), the fusion loop (FL) and the variable glycosylation loop (VGL) are indicated in the sequence.

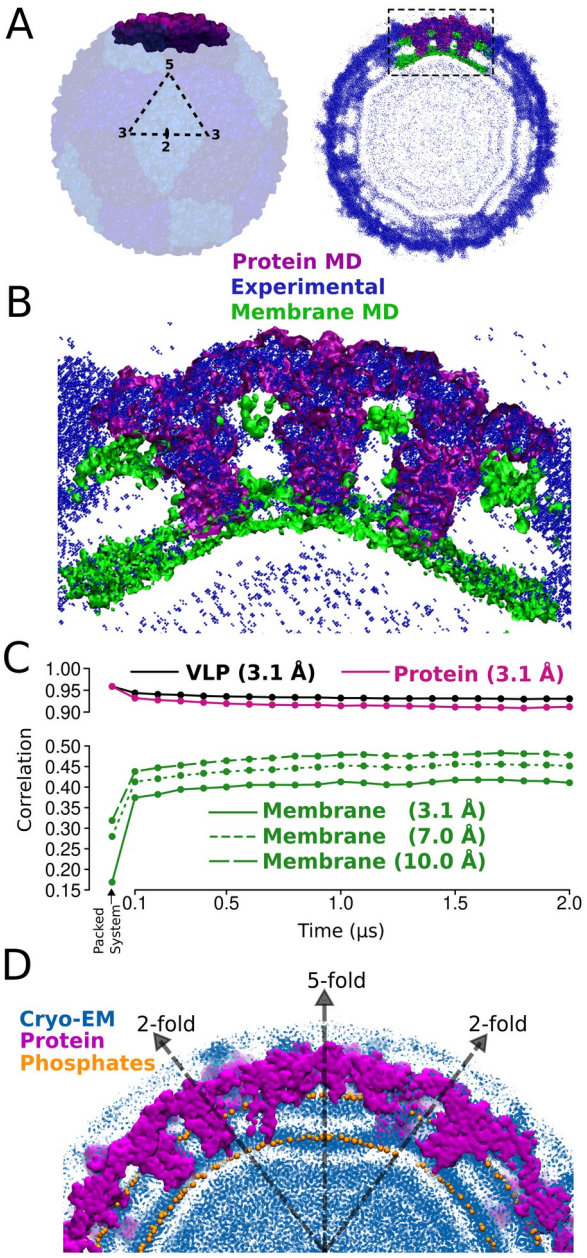


Figure 5. Experimental vs simulated ZIKV. A) Left, molecular representation of the ZIKV envelope showing the organization of E proteins into rafts along the VLP. The 2-, 3- and 5-fold symmetry axes of the structure are indicated. Right, overlap between the cryo-EM density map EMD-7543 and an average raft from the last 0.5 μs of trajectory (dashed square). B) Zoom into the inset of the right panel A. C) Correlation of density maps from the simulation and the cryo-EM structure. Values at the first point correspond to the packed VLP by PACKMOL, while the rest are obtained averaging windows of 0.1 μs of simulation after the equilibration phase. D) Structural superposition of the experimental density map in panel A with the packed membrane by PACKMOL showing the placement of phosphates. Symmetry axes are depicted.

To acquire a quantitative assessment on the quality of the model, we calculate the global correlation between the density maps of the simulated VLP and the cryo-EM (Figure 5C). The initial correlation coefficient of the entire VLP results as high as 0.96 at the experimental resolution. As mentioned previously, temperature effects drive the VLP away from the perfect symmetry, progressively reducing the correlation along the simulation. Nevertheless, even at seemingly high RMSD values, our simulation continues to reflect the experimental data to a large extent, as the correlation for the whole VLP remains over 0.90 (Figure 5C). Calculating the correlations against MD density maps using only the protein or the membrane reveals that the decrease in correlation can be mainly ascribe to the separation of protein from the perfect symmetric configuration. Indeed, a similar trend is also reported in all-atom simulations of mature *Flavivirus* envelopes [64]. Conversely, calculating the correlation only on the membrane shows a marked increase from the initial conformation. The presence of the protein envelope imposes significant distortions to the lipid bilayer. As a result, the spherical packing solution produces a better match in the neighborhood of the 2-fold axes (i.e., near the transmembrane helices), which coincides with the radius used to place the lipids. In contrast, the bilayer regions close to the 5-fold axes, which are flatter in the cryo-EM map, show a significant mismatch (Figure 5D). However, the combination of an appropriate force field with a robust equilibration protocol rapidly fixes these small issues (Figure 5C). Correlation values for the membrane are expected to be low because the comparison is done against the whole experimental density. In that regard, the MD density map at grid resolution of 3.1 Å already renders very good correlations compared to previous works [27]. Obviously, using MD density maps at lower grid resolutions further improves the comparison as the experimental density on that region is very low. Such an agreement with the experimental data provides a validation of the building protocol and the multiscale simulation to explore the dynamics of the VLP.

To further characterize the membrane's dynamics we analyze of the lipid-lipid contacts per molecular species. During the simulation, lipid species rapidly reach characteristic numbers of contacts at each leaflet (Figure 6A). The relative abundance of POPC molecules around POPC or POPS lipids takes longer times to stabilize. That is something expected being the most and the least represented species in the mixture. Worth noting, average contacts from the last 0.5 μs of trajectory differs at most in about one unit from the packed membrane (Table 5). That difference is already reduced to less than 0.5 contacts during the equilibration phase of the membrane (Figure 3B). Once again, these results highlight the convenience of the packing procedure and the equilibration strategy introduced here.

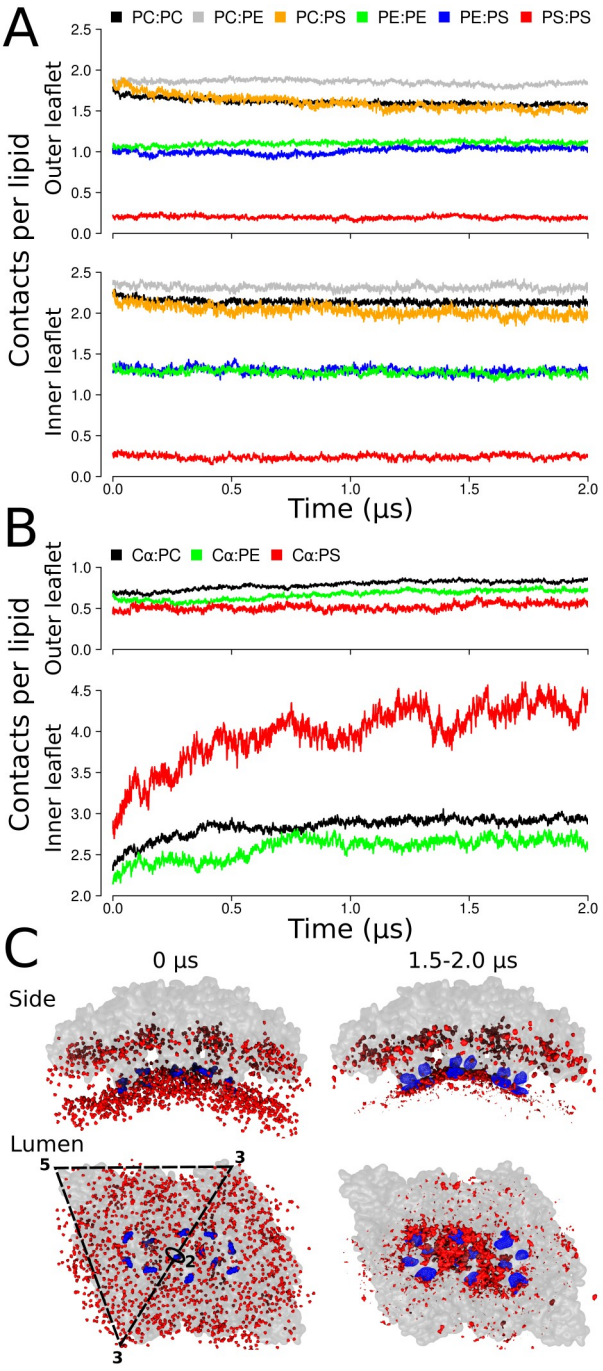


Figure 6. A) Time evolution of lipid-lipid contact per molecular species at each leaflet. Lipids are identified by their head groups. Contacts are measured from the central bead of the glycerol moiety using a cutoff distance of 11 Å. Values are normalized by the number of molecules of the second lipid species in the pair. After the equilibration phase. B) Lipid-protein contact per lipid spices. Contacts are measured from Cα atom in the transmembrane region of E and M proteins using the same criteria of panel A. C) Side and lumen views of phosphate distributions of POPS lipids from average rafts are shown in red. The occupancy of Lys480 in E and Lys60 in M proteins are in blue. Proteins are shown as gray surfaces. Densities at 0 μs are measured from the packed VLP by PACKMOL. Symmetry axes are indicated.

Table 5. Lipid-lipid and lipid-protein contacts per molecule in the packed and simulated system.[1]

Leaflet	Contact[2]	PACKMOL	MD[3]
Outer	PC:PC	2.5	1.6
	PC:PE	2.6	1.8
	PC:PS	2.7	1.5
	PE:PE	1.3	1.1
	PE:PS	1.3	1.0
	PS:PS	0.4	0.2
	Cα:PC	0.7	0.8
	Cα:PE	0.8	0.7
	Cα:PS	0.6	0.6
Inner	PC:PC	3.1	2.1
	PC:PE	3.2	2.3
	PC:PS	3.3	2.0
	PE:PE	1.7	1.3
	PE:PS	1.6	1.3
	PS:PS	0.6	0.2
	Cα:PC	0.5	2.9
	Cα:PE	0.6	2.7
	Cα:PS	0.6	4.3

[1] Contacts are measured from the central bead of the glycerol moiety or the Cα atom using a cutoff distance of 11 Å. Values are normalized by the number of molecules at the corresponding leaflet of the second lipid species in the pair.
[2] Lipids are identified by their head groups. The transmembrane region of E and M proteins is used in the analysis.
[3] Average values are calculated from the last 0.5 μs of trajectory in Figure 6.

Although an in-depth exploration of the dynamical properties of the Zika's VLP goes beyond the scope of this manuscript, we provide an example of the capability of the method to expand our structural knowledge of *Flavivirus*. As stated before, we use a 6:3:1 proportion of POPC, POPE, and POPS phospholipids for representing the membrane. Although minimalist, this population allows studying specific phospholipids' fingerprinting. To this aim, we analyze the lipid-protein contacts at the transmembrane region of E and M proteins (Figure 6B). Despite having few contacts per lipid at both leaflets in the initially packed membrane (Table 5), only contacts at the inner leaflet show a remarkably increase during the simulation. In particular POPS exhibits the largest increase, suggesting their recruitment to the neighborhood of the transmembrane segments. To gain a deeper insight on this effect, we compare the initial (random) distribution of lipids with their averaged occupational density in the last 0.5 μs of simulation. While POPC and POPE showed no evident patterns, POPS tends to cluster around the luminal part of the transmembrane helices (Figure 6C). Interestingly, this clustering is

mediated by the presence of a conserved set of Lysine residues (namely Lys480 at E protein, and Lys60 at M protein), which face to the luminal side of the VLP. Since the relative amount of POPS is similar to that of the Endoplasmic Reticulum, the simulations point to POPS's functional role in stabilizing viral proteins' assembly.

CONCLUSIONS

In this work we provide solutions to three main problems limiting the broad application of multiscale strategies to study enveloped VLP systems. These problems are: building accurate protein-membrane models, setting up multiscale representations and equilibrating and simulating the systems.

The first two issues are solved by PACKMOL. We implement new input parameters and objective function terms optimizing the heuristics to build up protein-membrane systems. That leads to advantages over existing solutions. It is simple, only one conformation per lipid species is used and its orientation, distribution and packing are optimized at the same time in an efficient way, making the extension to other lipids straightforward. That simplifies the prerequisites for including new molecules, while avoiding intermediate computational steps, which may require some user expertise. For example, the creation of system-specific grids of pseudo-atoms, which are then filled by molecular insertion, replacement and refinement from a predefined library of lipid conformations during tempering (e.g. CHARMM-GUI [14]). It is fast, enveloped VLP models are obtained within a few hours using a single CPU. It is flexible, creating multiscale representations is very simple by defining the distribution and granularity (size) of each component in the system. It is accurate, rendering lipid distributions in good agreement with experiments and simulations, saving computational time during the equilibration. Starting from conformations as close as possible to the experimental structure may be particularly relevant in confined systems, which lack enough room for fully relaxing the membrane during the equilibration. That may be the case of *Flaviviruses*, in which the protein coverage of outer and inner leaflets is about 60 % and 25 %, respectively [10]. As pointed out before, the local membrane thickness is affected by the presence of proteins in comparison to lipid-only systems [65]. In that sense, packing strategies based on picking conformations from pure membrane simulations (e.g. LipidWrapper [21]) may fail to represent highly crowded bilayers. In contrast, the present strategy provides solutions very close to the experimental data, with high correlations against cryo-EM density maps of ZIKV.

Despite starting from a good initial configuration may be beneficial to any molecular model used in MD simulations, the extent to which a given model profits from the initial condition may also depend on the equilibration procedure. Hence, the third issue is solved by optimizing the equilibration and simulation protocol within the multiscale framework of the SIRAH force field. The SIRAH force field has several particular features that make it very suitable for this approach. It has a constantly growing repertory of biomolecules that covers a wide diversity of biological applications [34]. Despite using a simplified topology, SIRAH provides an unbiased description of the protein dynamics without the need for tailor-made forces to keep an explore the native state [40,66]. It uses partial charges and long range electrostatics.

The lipid-protein interactions have been validated against different membrane compositions [41]. The explicit polar solvent model (WT4) permits describing relevant properties of the aqueous environment [53]. A compatible supra-CG solvent model (WLS) allows to generate multiscale systems to further boost MD simulations. In that sense, a modular pipeline based on a strategy already introduced for naked VLPs [5] is extended to enveloped VLPs and tested on a representative model of ZIKV. In particular, explicitly including the solvent during equilibration steps and fixing the hydration shell at different stages helps reproducing the complex thickness pattern of the *Flavivirus* membrane. The quality of the multiscale representation is validated against different structural descriptors such as radial distribution functions, RMSD, RGYR, B-factors. The combination of an unbiased CG force field with an optimized pipeline for generating and simulating enveloped VLPs allows to characterize fine details in the protein-membrane interaction, leading to point specific lipid fingerprints.

As detailed all-atom simulations of the full HIV-1 capsid clearly show, different molecular phenomena occurring in a virus have different characteristic times [67]. In that respect, we provide a solution to extend the accessible time windows for studying some new phenomena. Additionally, the general guidelines in the presented strategy can be applied to simplify and improve the system setup and equilibration by other molecular models, such as the MARTINI force field [68]. Hence, we expect this pipeline for wrapping up viruses at multiscale resolution constitute a robust and cost effective framework to keep pushing forward the field of computational virology.

ACKNOWLEDGMENTS

This work is partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11. M.R.M. and S.P. belong to the SNI program of ANII. M.S. is a postgraduate fellowship of ANII. Some of the graphic cards used in this research were donated by the NVIDIA Corporation. We thank E.E. Barrera for useful comments on the manuscript, and P. Cossio and J.S. Ortiz for helping us in the analysis of Cryo-EM data. Production simulations were run at the HPC facility of LNCC - Santos Dumont, Petropolis, Brazil, grant VICBF1. L.M. thanks the São Paulo Foundation for Research (FAPESP), grants 2010/16947-9, 2018/24293-0, 2013/08293-7, and 2018/14274-9.

REFERENCES

(1) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74. https://doi.org/10.1016/j.sbi.2015.03.007.

(2) González-Arias, F.; Reddy, T.; Stone, J. E.; Hadden-Perilla, J. A.; Perilla, J. R. Scalable Analysis of Authentic Viral Envelopes on FRONTERA. *Comput. Sci. Eng.* **2020**, *22*, 11–20. https://doi.org/10.1109/MCSE.2020.3020508.

(3) Jefferys, E. E.; Sansom, M. S. P. Computational Virology: Molecular Simulations of Virus Dynamics and Interactions. In *Physical Virology: Virus Structure and Mechanics*; Greber, U. F., Ed.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2019; pp 201–233. https://doi.org/10.1007/978-3-030-14741-9_10.

(4) Marzinek, J. K.; Huber, R. G.; Bond, P. J. Multiscale Modelling and Simulation of Viruses. *Curr. Opin. Struct. Biol.* **2020**, *61*, 146–152. https://doi.org/10.1016/j.sbi.2019.12.019.

(5) Machado, M. R.; González, H. C.; Pantano, S. MD Simulations of Viruslike Particles with Supra CG Solvation Affordable to Desktop Computers. *J. Chem. Theory Comput.* **2017**, *13*, 5106–5116. https://doi.org/10.1021/acs.jctc.7b00659.

(6) Liu, Y.; De Vries, A. H.; Barnoud, J.; Pezeshkian, W.; Melcr, J.; Marrink, S. J. Dual Resolution Membrane Simulations Using Virtual Sites. *J. Phys. Chem. B* **2020**, *124*, 3944–3953. https://doi.org/10.1021/acs.jpcb.0c01842.

(7) Ayton, G. S.; Voth, G. A. Multiscale Computer Simulation of the Immature HIV-1 Virion. *Biophys. J.* **2010**, *99*, 2757–2765. https://doi.org/10.1016/j.bpj.2010.08.018.

(8) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *bioRxiv* **2020**, 2020.10.02.323915. https://doi.org/10.1101/2020.10.02.323915.

(9) Wen, P.-C.; Mahinthichaichan, P.; Trebesch, N.; Jiang, T.; Zhao, Z.; Shinn, E.; Wang, Y.; Shekhar, M.; Kapoor, K.; Chan, C. K.; Tajkhorshid, E. Microscopic View of Lipids and Their Diverse Biological Functions. *Curr. Opin. Struct. Biol.* **2018**, *51*, 177–186. https://doi.org/10.1016/j.sbi.2018.07.003.

(10) Zhang, W.; Kaufmann, B.; Chipman, P. R.; Kuhn, R. J.; Rossmann, M. G. Membrane Curvature in Flaviviruses. *J. Struct. Biol.* **2013**, *183*, 86–94. https://doi.org/10.1016/j.jsb.2013.04.005.

(11) Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **2019**, *119*, 6184–6226. https://doi.org/10.1021/acs.chemrev.8b00460.

(12) Sommer, B. Membrane Packing Problems: A Short Review on Computational Membrane Modeling Methods and Tools. *Comput. Struct. Biotechnol. J.* **2013**, *5*. https://doi.org/10.5936/csbj.201302014.

(13) Sharma, S.; Kim, B. N.; Stansfeld, P. J.; Sansom, M. S. P.; Lindau, M. A Coarse Grained Model for a Lipid Membrane with Physiological Composition and Leaflet Asymmetry. *PLOS ONE* **2015**, *10*, e0144814. https://doi.org/10.1371/journal.pone.0144814.

(14) Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S.-J.; Patel, D. S.; Beaven, A. H.; Lee, K. I.; Rui, H.; Park, S.; Lee, H. S.; Roux, B.; MacKerell, A. D.; Klauda, J. B.; Qi, Y.; Im, W. CHARMM-GUI 10 Years for Biomolecular Modeling and Simulation. *J. Comput. Chem.* **2017**, *38*, 1114–1124. https://doi.org/10.1002/jcc.24660.

(15) Doerr, S.; Giorgino, T.; Martínez-Rosell, G.; Damas, J. M.; De Fabritiis, G. High-Throughput Automated Preparation and Simulation of Membrane Proteins with HTMD. *J. Chem. Theory Comput.* **2017**, *13*, 4003–4011. https://doi.org/10.1021/acs.jctc.7b00480.

(16) Johnson, G. T.; Autin, L.; Al-Alusi, M.; Goodsell, D. S.; Sanner, M. F.; Olson, A. J. CellPACK: A Virtual Mesoscope to Model and Visualize Structural Systems Biology. *Nat. Methods* **2015**, *12*, 85–91. https://doi.org/10.1038/nmeth.3204.

(17) Ghahremanpour, M. M.; Arab, S. S.; Aghazadeh, S. B.; Zhang, J.; van der Spoel, D. MemBuilder: A Web-Based Graphical Interface to Build Heterogeneously Mixed Membrane Bilayers for the GROMACS Biomolecular Simulation Program. *Bioinformatics* **2014**, *30*, 439–441. https://doi.org/10.1093/bioinformatics/btt680.

(18) Knight, C. J.; Hub, J. S. MemGen: A General Web Server for the Setup of Lipid Membrane Simulation Systems. *Bioinformatics* **2015**, *31*, 2897–2899. https://doi.org/10.1093/bioinformatics/btv292.

(19) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with Insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144–2155. https://doi.org/10.1021/acs.jctc.5b00209.

(20) Damre, M.; Marchetto, A.; Giorgetti, A. MERMAID: Dedicated Web Server to Prepare and Run Coarse-Grained Membrane Protein Dynamics. *Nucleic Acids Res.* **2019**, *47*, W456–W461. https://doi.org/10.1093/nar/gkz416.

(21) Durrant, J. D.; Amaro, R. E. LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry. *PLOS Comput. Biol.* **2014**, *10*, e1003720. https://doi.org/10.1371/journal.pcbi.1003720.

(22) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping Triangulated Surfaces to Coarse-Grained Membrane Models. *Nat. Commun.* **2020**, *11*, 2296. https://doi.org/10.1038/s41467-020-16094-y.

(23) Boyd, K. J.; May, E. R. BUMPy: A Model-Independent Tool for Constructing Lipid Bilayers of Varying Curvature and Composition. *J. Chem. Theory Comput.* **2018**, *14*, 6642–6652. https://doi.org/10.1021/acs.jctc.8b00765.

(24) Sommer, B.; Dingersen, T.; Gamroth, C.; Schneider, S. E.; Rubert, S.; Krüger, J.; Dietz, K.-J. CELLmicrocosmos 2.2 MembraneEditor: A Modular Interactive Shape-Based Software Approach To Solve Heterogeneous Membrane Packing Problems. *J. Chem. Inf. Model.* **2011**, *51*, 1165–1182. https://doi.org/10.1021/ci1003619.

(25) Woo, H.; Park, S.-J.; Choi, Y. K.; Park, T.; Tanveer, M.; Cao, Y.; Kern, N. R.; Lee, J.; Yeom, M. S.; Croll, T. I.; Seok, C.; Im, W. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B* **2020**, *124*, 7128–7137. https://doi.org/10.1021/acs.jpcb.0c04553.

(26) Reddy, T.; Sansom, M. S. P. The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virion Envelope. *Structure* **2016**, *24*, 375–382. https://doi.org/10.1016/j.str.2015.12.011.

(27) Marzinek, J. K.; Holdbrook, D. A.; Huber, R. G.; Verma, C.; Bond, P. J. Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations. *Structure* **2016**, *24*, 1410–1420. https://doi.org/10.1016/j.str.2016.05.014.

(28) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164. https://doi.org/10.1002/jcc.21224.

(29) Martínez, J. M.; Martínez, L. Packing Optimization for Automated Generation of Complex System's Initial Configurations for Molecular Dynamics and Docking. *J. Comput. Chem.* **2003**, *24*, 819–825. https://doi.org/10.1002/jcc.10216.

(30) Andreani, R.; Birgin, E. G.; Martínez, J. M.; Schuverdt, M. L. On Augmented Lagrangian Methods with General Lower-Level Constraints. *SIAM J. Optim.* **2007**, *18*, 1286–1309. https://doi.org/10.1137/060654797.

(31) Schott-Verdugo, S.; Gohlke, H. PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein–Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, *59*, 2522–2528. https://doi.org/10.1021/acs.jcim.9b00269.

(32) Wong-ekkabut, J.; Karttunen, M. The Good, the Bad and the User in Soft Matter Simulations. *Biochim. Biophys. Acta BBA - Biomembr.* **2016**, *1858*, 2529–2538. https://doi.org/10.1016/j.bbamem.2016.02.004.

(33) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2018**, *1*, 5957. https://doi.org/10.33011/livecoms.1.1.5957.

(34) Machado, M. R.; Zeida, A.; Darré, L.; Pantano, S. From Quantum to Subcellular Scales: Multi-Scale Simulation Approaches and the SIRAH Force Field. *Interface Focus* **2019**, *9*, 20180085. https://doi.org/10.1098/rsfs.2018.0085.

(35) Sevvana, M.; Long, F.; Miller, A. S.; Klose, T.; Buda, G.; Sun, L.; Kuhn, R. J.; Rossmann, M. G. Refinement and Analysis of the Mature Zika Virus Cryo-EM Structure at 3.1 Å Resolution. *Structure* **2018**, *26*, 1169-1177.e3. https://doi.org/10.1016/j.str.2018.05.006.

(36) Unni, S.; Huang, Y.; Hanson, R. M.; Tobias, M.; Krishnan, S.; Li, W. W.; Nielsen, J. E.; Baker, N. A. Web Servers and Services for Electrostatics Calculations with APBS and PDB2PQR. *J. Comput. Chem.* **2011**, *32*, 1488–1491. https://doi.org/10.1002/jcc.21720.

(37) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(38) Zhang, Q.; Hunke, C.; Yau, Y. H.; Seow, V.; Lee, S.; Tanner, L. B.; Guan, X. L.; Wenk, M. R.; Fibriansah, G.; Chew, P. L.; Kukkaro, P.; Biuković, G.; Shi, P.-Y.; Shochat, S. G.; Grüber, G.; Lok, S.-M. The Stem Region of Premembrane Protein Plays an Important Role in the Virus Surface Protein Rearrangement during Dengue Maturation. *J. Biol. Chem.* **2012**, *287*, 40525–40534. https://doi.org/10.1074/jbc.M112.384446.

(39) Marquardt, D.; Geier, B.; Pabst, G. Asymmetric Lipid Membranes: Towards More Realistic Model Systems. *Membranes* **2015**, *5*, 180–196. https://doi.org/10.3390/membranes5020180.

(40) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719–2733. https://doi.org/10.1021/acs.jctc.9b00006.

(41) Barrera, E. E.; Machado, M. R.; Pantano, S. Fat SIRAH: Coarse-Grained Phospholipids To Explore Membrane–Protein Dynamics. *J. Chem. Theory Comput.* **2019**, *15*, 5674–5688. https://doi.org/10.1021/acs.jctc.9b00435.

(42) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2015; pp 3–27. https://doi.org/10.1007/978-3-319-15976-8_1.

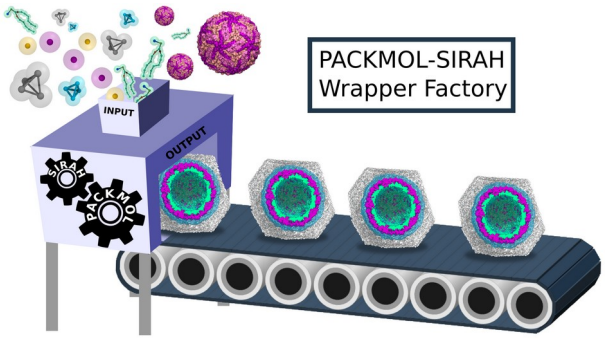
(43) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. https://doi.org/10.1063/1.2408420.

(44) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055–1076. https://doi.org/10.1080/00268978300102851.

(45) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. https://doi.org/10.1063/1.328693.

(46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. https://doi.org/10.1063/1.464397.
(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. https://doi.org/10.1063/1.470117.
(48) Buchoux, S. FATSLiM: A Fast and Robust Software to Analyze MD Simulations of Membranes. *Bioinformatics* **2017**, *33*, 133–134. https://doi.org/10.1093/bioinformatics/btw563.
(49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5.
(50) Goddard, T. D.; Huang, C. C.; Ferrin, T. E. Visualizing Density Maps with UCSF Chimera. *J. Struct. Biol.* **2007**, *157*, 281–287. https://doi.org/10.1016/j.jsb.2006.06.010.
(51) Birgin, E. G.; Mario Martínez, J. Large-Scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients. *Comput. Optim. Appl.* **2002**, *23*, 101–125. https://doi.org/10.1023/A:1019928808826.
(52) The Linked Cell Method for Short-Range Potentials. In *Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications*; Griebel, M., Zumbusch, G., Knapek, S., Eds.; Texts in Computational Science and Engineering; Springer: Berlin, Heidelberg, 2007; pp 37–111. https://doi.org/10.1007/978-3-540-68095-6_3.
(53) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *J. Chem. Theory Comput.* **2010**, *6*, 3793–3807. https://doi.org/10.1021/ct100379f.
(54) Machado, M. R.; Pantano, S. SIRAH Tools: Mapping, Backmapping and Visualization of Coarse-Grained Models. *Bioinformatics* **2016**, *32*, 1568–1570. https://doi.org/10.1093/bioinformatics/btw020.
(55) Machado, M. R.; Pantano, S. Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 1367–1372. https://doi.org/10.1021/acs.jctc.9b00953.
(56) Pierson, T. C.; Diamond, M. S. The Emergence of Zika Virus and Its New Clinical Syndromes. *Nature* **2018**, *560*, 573–581. https://doi.org/10.1038/s41586-018-0446-y.
(57) Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T. C.; Rossmann, M. G.; Kuhn, R. J. The 3.8 Å Resolution Cryo-EM Structure of Zika Virus. *Science* **2016**, *352*, 467–470. https://doi.org/10.1126/science.aaf5316.
(58) Therkelsen, M. D.; Klose, T.; Vago, F.; Jiang, W.; Rossmann, M. G.; Kuhn, R. J. Flaviviruses Have Imperfect Icosahedral Symmetry. *Proc. Natl. Acad. Sci.* **2018**, *115*, 11608–11612. https://doi.org/10.1073/pnas.1809304115.
(59) Krol, E.; Brzuska, G.; Szewczyk, B. Production and Biomedical Application of Flavivirus-like Particles. *Trends Biotechnol.* **2019**, *37*, 1202–1216. https://doi.org/10.1016/j.tibtech.2019.03.013.
(60) Shen, W.-F.; Galula, J. U.; Liu, J.-H.; Liao, M.-Y.; Huang, C.-H.; Wang, Y.-C.; Wu, H.-C.; Liang, J.-J.; Lin, Y.-L.; Whitney, M. T.; Chang, G.-J. J.; Chen, S.-R.; Wu, S.-R.; Chao, D.-Y. Epitope Resurfacing on Dengue Virus-like Particle Vaccine Preparation to Induce Broad Neutralizing Antibody. *eLife* **2018**, *7*, e38970. https://doi.org/10.7554/eLife.38970.
(61) Morrone, S. R.; Chew, V. S. Y.; Lim, X.-N.; Ng, T.-S.; Kostyuchenko, V. A.; Zhang, S.; Wirawan, M.; Chew, P.-L.; Lee, J.; Tan, J. L.; Wang, J.; Tan, T. Y.; Shi, J.; Screaton, G.; Morais, M. C.; Lok, S.-M. High Flavivirus Structural Plasticity Demonstrated by a Non-Spherical Morphological Variant. *Nat. Commun.* **2020**, *11*, 3112. https://doi.org/10.1038/s41467-020-16925-y.
(62) Rey, F. A.; Stiasny, K.; Vaney, M.-C.; Dellarole, M.; Heinz, F. X. The Bright and the Dark Side of Human Antibody Responses to Flaviviruses: Lessons for Vaccine Design. *EMBO Rep.* **2018**, *19*, 206–224. https://doi.org/10.15252/embr.201745302.
(63) Carbaugh, D. L.; Lazear, H. M. Flavivirus Envelope Protein Glycosylation: Impacts on Viral Infection and Pathogenesis. *J. Virol.* **2020**, *94*. https://doi.org/10.1128/JVI.00104-20.
(64) Pindi, C.; Chirasani, V. R.; Rahman, M. H.; Ahsan, M.; Revanasiddappa, P. D.; Senapati, S. Molecular Basis of Differential Stability and Temperature Sensitivity of ZIKA versus Dengue Virus Protein Shells. *Sci. Rep.* **2020**, *10*, 8411. https://doi.org/10.1038/s41598-020-65288-3.
(65) Corradi, V.; Sejdiu, B. I.; Mesa-Galloso, H.; Abdizadeh, H.; Noskov, S. Yu.; Marrink, S. J.; Tieleman, D. P. Emerging Diversity in Lipid–Protein Interactions. *Chem. Rev.* **2019**, *119*, 5775–5848. https://doi.org/10.1021/acs.chemrev.8b00451.
(66) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J. Chem. Theory Comput.* **2015**, *11*, 723–739. https://doi.org/10.1021/ct5007746.

(67) Perilla, J. R.; Schulten, K. Physical Properties of the HIV-1 Capsid from All-Atom Molecular Dynamics Simulations. *Nat. Commun.* **2017**, *8*, 15959. https://doi.org/10.1038/ncomms15959.
(68) J. Marrink, S.; Peter Tieleman, D. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822. https://doi.org/10.1039/C3CS60093A.



TOC Graphic