



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Computational modelling of flavivirus dynamics: The ins and outs

Roland G. Huber^{a,1}, Jan K. Marzinek^{a,1}, Priscilla L.S. Boon^{a,b,d}, Wan Yue^c, Peter J. Bond^{a,d,*}^a Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix #07-01, 138671, Singapore^b NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore (NUS), University Hall, Tan Chin Tuan Wing #04-02, 119077, Singapore^c Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), 60 Biopolis Street, Genome #02-01, 138672, Singapore^d Department of Biological Sciences (DBS), National University of Singapore (NUS), 16 Science Drive 4, Building S3, Singapore

ARTICLE INFO

Keywords:

Enveloped virus

Dengue

Zika

Coronavirus

Multiscale modelling

Molecular simulations

Structural genomics

ABSTRACT

Enveloped viruses such as the flaviviruses represent a significant burden to human health around the world, with hundreds of millions of people each year affected by dengue alone. In an effort to improve our understanding of the molecular basis for the infective mechanisms of these viruses, extensive computational modelling approaches have been applied to elucidate their conformational dynamics. Multiscale protocols have been developed to simulate flavivirus envelopes in close accordance with biophysical data, in particular derived from cryo-electron microscopy, enabling high-resolution refinement of their structures and elucidation of the conformational changes associated with adaptation both to host environments and to immunological factors such as antibodies. Likewise, integrative modelling efforts combining data from biophysical experiments and from genome sequencing with chemical modification are providing unparalleled insights into the architecture of the previously unresolved nucleocapsid complex. Collectively, this work provides the basis for the future rational design of new antiviral therapeutics and vaccine development strategies targeting enveloped viruses.

1. Introduction

Viral diseases impose a severe burden on human health and economy. Despite recent advances in the medical and pharmacological sciences, viral infections continue to cause millions of human deaths every year, and also have a major impact on livestock and agricultural productivity. Furthermore, viral pandemics such as HIV [1] pose a continuing global risk to public health. There is a constant threat of emerging viral pathogens which can cause epidemics, as exemplified by the 2015 outbreak of Zika virus (ZIKV) and its association with birth defects and neurological issues. Most recently, in late 2019, the first reports of an unknown respiratory infection emerged from Wuhan, China. The source of that infection was a novel coronavirus, related to those that had previously caused outbreaks of Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV). Since the emergence of the illness resulting from the new SARS-CoV-2 virus, COVID-19, millions of infections and hundreds of thousands of deaths have been reported worldwide [2]. Both viruses led the World Health Organization (WHO) to declare a Public Health Emergency of International Concern.

Flaviviruses are typically transmitted by mosquitoes or ticks, and encompass around 50 known pathogens that cause human disease. Dengue (DENV) is the archetypal member, infecting up to ~400 million people per year worldwide [3]. All flaviviruses adopt a similar structure, composed of only three structural proteins, namely the envelope (E) and membrane (M) proteins – embedded within a lipid bilayer vesicle and arranged in an icosahedral fashion – and the capsid (C) proteins, which form a complex with the single-stranded, positive-sense RNA genome inside the virion core. 180 copies of the E and M proteins are arranged with icosahedral symmetry on the surface of the virion particle [4]. In the mature particle, these are found as three sets of parallel dimers or “rafts” which are arranged in a herringbone-like pattern. The E protein is exposed on the virion surface, and is responsible both for binding to cell receptors during infection, as well as fusion with endosomal membranes inside the cell, enabling genome release [5].

The discovery of new antiviral drugs is hampered by numerous challenges including the integration of the viral life cycle with that of the host, potential development of resistance, and the ongoing threat of newly arising pandemics. Vaccination is a potentially effective strategy

* Corresponding author at: Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix #07-01, 138671, Singapore.

E-mail address: peterjb@bii.a-star.edu.sg (P.J. Bond).

¹ Contributed Equally.

<https://doi.org/10.1016/j.ymeth.2020.06.004>

Received 18 December 2019; Received in revised form 24 April 2020; Accepted 4 June 2020

1046-2023/ © 2020 Elsevier Inc. All rights reserved.

to combat viral disease, but can be complicated by pathogen-specific factors such as antibody-dependent enhancement (ADE) [6] in which non-neutralizing antibodies actually facilitate viral infectivity. This is a particularly relevant issue for DENV, which exists as four distinct serotypes, and a recently available vaccine has been reported to worsen the outcome of patients who have not previously been infected [7]. To counter the challenges in viral disease therapy, research efforts continue to focus on better understanding the molecular mechanisms associated with infection. These have been facilitated by advances in structural biology, including the ever increasing resolutions attainable by cryo-electron microscopy (cryo-EM) [8] for whole virion particles, particularly when resultant density maps are combined judiciously with available X-ray crystallographic or solution NMR structures of individual viral proteins. Furthermore, the molecular dynamics (MD) simulation technique has now come of age. With the ever increasing computational power available and specialized hardware, along with appropriate multiscale approaches that enable accurate but coarse-grained (CG) representations of systems or components thereof, it is now possible to simulate entire viruses, and to use such simulations to refine integrative models incorporating experimental data [9].

As described in the following sections, such computational models have enabled the description of the conformational dynamics of flaviviruses during progression along their life cycle in unparalleled detail, as well as furthering our understanding of their interactions with host components such as antibodies. In addition, the architecture of the flavivirus nucleocapsid complex containing the RNA genome is not well understood, because it appears not to be organized in a clear geometric pattern amenable to cryo-EM, and because a significant component of the C protein is intrinsically disordered and hence challenging to study via traditional approaches, depending on conditions and maturation state [10]. Thus, we also describe our recent efforts to integrate biophysical experiments with simulation to characterize the C protein conformational ensemble, and to use high-throughput sequencing (HTS) with chemical modification techniques in combination with modelling to predict genomic RNA structures.

2. Life cycle of flaviviruses and gaps in our knowledge

Flavivirus infection is triggered through receptor mediated endocytosis, when the virus is internalized inside an endosome. The acidic endosomal pH triggers the E proteins to undergo a major structural rearrangement and a dimer-to-trimer transition. Through this process, hydrophobic fusion peptides (FPs), which are positioned at the tip of each trimeric E protein “spike”, become exposed. This allows them to interact with the endosomal membrane, leading to its fusion with the viral envelope [11]. We and others have reported efforts to model the flavivirus lipid bilayer, and in particular, we describe in subsequent sections protocols to model the entire envelope particle in accordance with cryo-EM data, and to leverage such models to understand different stages of the viral life cycle.

Once the capsid and RNA genome are released into the cell, this leads to polyprotein translation and viral replication [4,12,13]. Subsequently, the assembly of the nascent immature dengue virus (immDENV) proceeds via budding of its components into the endoplasmic reticulum (ER). The immDENV is rendered non-infectious due to the complex formed between the prM (precursor of membrane) protein with the E protein in the viral membrane. The immDENV envelope is composed of 60 trimeric spikes, in which each monomer consists of an E:prM complex [14,15]. Under the low pH conditions in the *trans*-Golgi network (TGN), the maturation process is triggered, involving cleavage by furin protease of pr fragments to form the mature M proteins, and a large-scale icosahedral DENV surface rearrangement with a trimer-to-dimer transition of the E proteins. Nevertheless, the low pH prevents the dissociation of pr molecules from dissociating from the E protein [15,16]. This serves to prevent the virus particle from fusing within the cell prior to release, as the E protein FP regions are “capped” by the pr

fragments.

Upon release of the virus from the cell, the neutral pH of the extracellular environment causes dissociation of pr molecules, thus completing the maturation process. Maturation is inefficient and a substantial number of DENV particles released from the cell are either partially mature or immature [17]. However, when they are complexed with specific antibodies, this can lead to ADE, rendering them infectious by facilitating the endosomal internalization of the immDENV:antibody complex. In this context, although furin can cleave the pr:M complex from the antibody-bound virion inside the endosome [18], it has remained unclear how pr dissociates in the acidic pH environment. Thus, we also describe below our recent use of simulations of the immature viral envelope in combination with cryo-EM, co-floitation assays, and hydrogen deuterium exchange mass spectrometry (HDX-MS) to rationalize an example of ADE at the molecular level.

It is also emerging that flaviviruses undergo major structural and dynamic changes outside of the cell, in response to different environmental triggers. In particular, the envelope is modulated when exposed to higher temperatures that may be associated with fever. Such viral conformational dynamics have been referred to as “breathing” [19–22]. The most prominent temperature-mediated transitions have been observed in specific strains of DENV2. The virus envelope has been observed to expand and become “bumpy” at 37 °C, in contrast to 28 °C where it exhibits a “smooth” envelope surface [19,23,24]. Although the existence of such variable morphologies may not be correlated with infectivity [23], they likely have important consequences for antibody binding and neutralization, and represent a challenge for therapeutics development. Interestingly, flavivirus “breathing” was initially thought to be irreversible, but as described below, a combination of multiscale modelling, cryo-EM, single-molecule fluorescence, and HDX-MS has enabled us to observe the process of viral expansion in atomic detail, and to elucidate the mechanism by which divalent cations enable the virion to reverse this process [25].

In the following, we summarize computational and integrative modelling methods, guided by biophysical data, that have yielded molecular insights into the “ins” and “outs” of enveloped flaviviruses. In Section 3, we begin by highlighting areas of ongoing interest in the context of viral envelope dynamics, which have been tackled recently using a combination of multiscale simulation and experiment. We then go on to detail efforts to characterize the virion’s nucleocapsid cargo via simulation (Section 4) and RNA sequencing-guided modelling (Section 5), before reflecting upon where future efforts may lie in better understanding enveloped viruses (Section 6).

3. Methods to study the structure and dynamics of flavivirus envelopes

3.1. Simulations of flavivirus systems

Numerous all-atom and CG simulations of the flavivirus envelope have been reported. For example, several all-atom simulation studies investigated viral envelope conformational changes occurring upon exposure to low pH, both during the fusion and maturation processes [26–34]. The capacity for individual E/M proteins or their complexes to initiate membrane fusion [35–40] and host membrane bending [41] has also been extensively studied via simulation. Nevertheless, it is important to consider the dynamics of the ~50 nm diameter flavivirus envelope in its entirety [42,43]. The sheer size of the virus particle, comprising 180 monomeric E/M proteins, represents a significant computational cost in attempting to simulate relevant timescales [42], leading to the frequent use of the well-established Martini CG force field, in which four heavy atoms map to approximately one CG particle [44–47]. Furthermore, the flavivirus shell poses additional challenges as a result of the need to model a lipid membrane in which the viral proteins are embedded. In the following sections, we summarize key steps required in our protocol [47] for generating a complete flavivirus

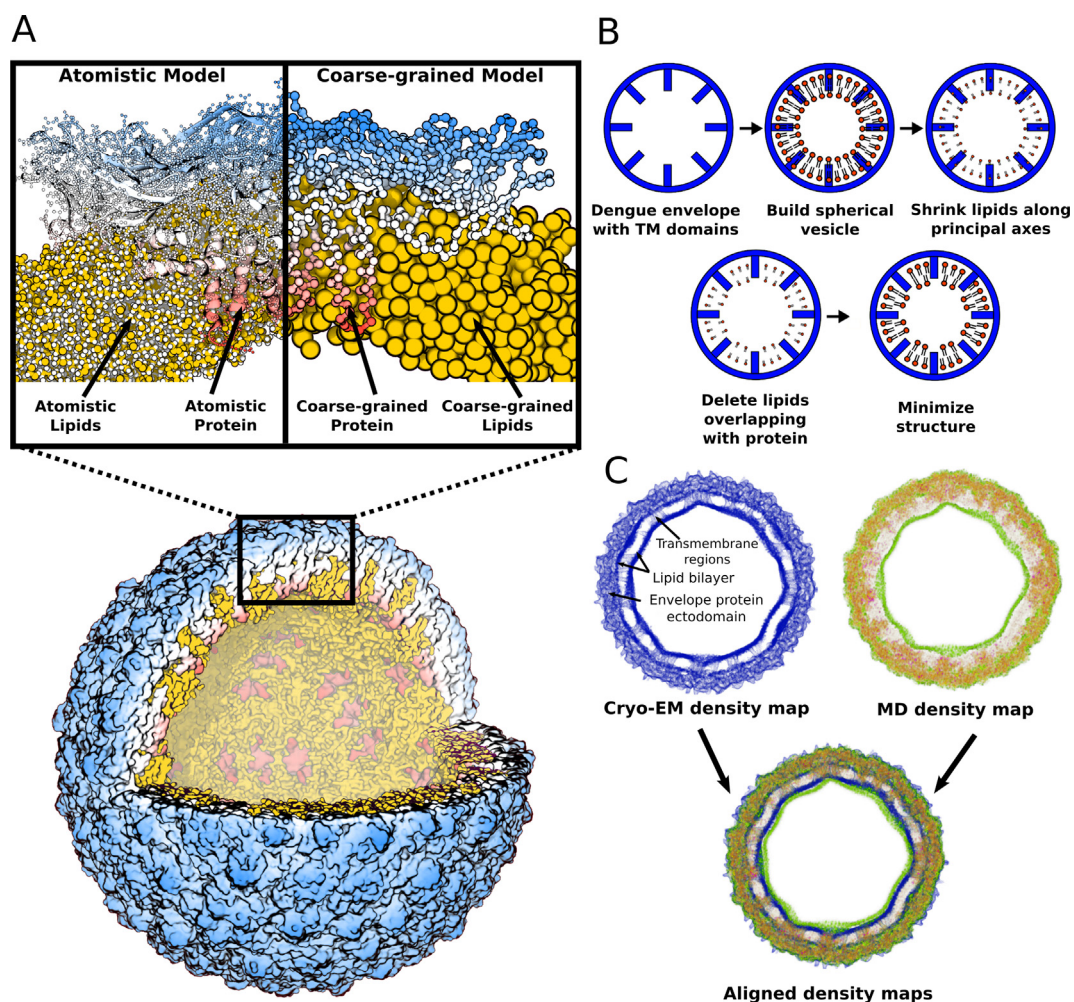


Fig. 1. Multiscale protocol for modelling flavivirus envelope particle and simulation-based refinement against cryo-EM data. (A) Multiscale representation of dengue virus envelope complex. Protein is shown in surface format (bottom), while the zoomed snapshot (top) shows an E protein dimer in all-atom and CG resolution. Proteins are colored according to their radial distance from the virus center of mass. Lipids are colored in orange. (B) Schematic of the step-by-step protocol for embedding viral proteins in a lipid vesicle. (C) Alignment of MD simulation-averaged (blue) and experimental (orange/green) electron density maps.

envelope model, for multiscale refinement of its structure in direct accordance with cryo-EM data, which may subsequently be leveraged to explore conformational dynamics associated with the viral life cycle.

3.2. CG representation of flavivirus proteins

Particle types are approximately transferable across systems in the Martini force field as a result of their being parameterized according to partitioning free energies [48–49]. On the other hand, several approaches are possible for the treatment of protein conformation, including application of an elastic network model [50] (ENM) between backbone beads, on top of secondary structure dependent dihedrals, to preserve higher order structure. Our experience suggests that this tends to be system specific, and calibration against all-atom reference simulations is desirable. Thus, we first simulated the solvated DENV dimeric E protein ectodomain in atomic resolution for several hundred nanoseconds, and generated equivalent CG trajectories [47] (Fig. 1A). Default Martini parameters along with a range of ENMs were tested, validated by comparing protein conformational properties from all-atom and CG simulations, including backbone root mean square deviations (RMSDs) with respect to the X-ray structure, protein radius of gyration, and pairwise residue contact maps. An ENM with non-standard cut-off distances and force constants were ultimately required to accurately describe the E protein conformational dynamics [47].

3.3. Constructing the flavivirus vesicle

The viral lipid bilayer vesicle was subsequently constructed using the CHARMM-GUI [51] Martini-builder, with a diameter chosen in accordance with estimates from cryo-EM density maps. The DENV2 membrane composition included palmitoyl-oleoyl phosphatidylcholine (PC), palmitoyl-oleoyl phosphatidylethanolamine (PE), and palmitoyl-oleoyl phosphatidylserine (PS) in a 6:3:1 ratio, as guided by lipidomics data measured for C6/36 mosquito cells [52]. The complete protein coordinates of 180 E/M proteins representing the entire virion envelope, obtained from the cryo-EM structure, were next converted to CG representation. These were then directly overlaid with the lipid vesicle, and overlapping lipids deleted (Fig. 1B). To facilitate insertion whilst reducing steric clashes, we employed a “shrinking” strategy for the lipids. This was performed for the coordinates of lipid tails by scaling their atomic positions onto the lipid tail centres of mass in a plane perpendicular to the tail principle axes [47]. This step did not involve any simulation, but was followed by multiple sets of energy minimization whilst “freezing” the coordinates of the protein backbone; this may be performed several times in an iterative manner to obtain a stable starting model of the lipid/protein envelope complex. This may subsequently be solvated (in a sensibly chosen simulation unit cell, e.g. dodecahedron or truncated octahedron), with energy minimization and protein-restrained equilibrations in the NVT ensemble to relax the system. We would advise that a short equilibration simulation time step

(e.g. ~ 2 fs) be chosen initially, which can be gradually increased in progressive $\sim 100,000$ -step equilibration runs. Finally, this may then be followed by NPT ensemble equilibration and unrestrained production runs.

3.4. CG simulation refinement of flavivirus envelope against cryo-EM data

The MDAanalysis [53] or GROMaps [54] tools may be used to extract a grid-based theoretical density map from a single simulation frame, or from numerous averaged frames extracted from progressive equilibration or production runs. Both protein and lipid should be included for density map calculations, and it is recommended to use a voxel grid-spacing close to the resolution of the cryo-EM map of interest. The Chimera [55] visualization and analysis tool (<https://www.cgl.ucsf.edu/chimera/>) enables map-in-map fitting of the theoretical MD generated and experimental density maps. This is performed via local optimization using steepest ascent to maximize the overlap or correlation between two sets of map grid points. The correlation between two maps corresponds to the mean cosine similarity across voxels, and varies from -1 to 1 . In our case the correlation reached ~ 0.75 – 0.8 during the simulations of the DENV particle, dependent upon precise system conditions (Fig. 1C) [47].

3.5. Mapping conformational changes associated with virion “breathing”

Due to the low resolution of the cryo-EM map obtained for the “bumpy” expanded state of DENV, the membrane-embedded components of the proteins (namely, E protein transmembrane helices and entire M proteins) could not be resolved. To elucidate the complete expanded virus structure and to explore the “breathing” pathway, a targeted MD (TMD) [56,57] approach was taken to trigger the transition, using our refined structure of the smooth DENV virion as a starting point (Fig. 2). TMD is currently available in Gromacs [58] when used in combination with PLUMED [59], as well as in NAMD [60], and employs a time dependent harmonic potential in which the difference in RMSD between the initial first-step and final-step target structures evolves linearly over a specified simulation timescale. Over the TMD simulation, the force on each atom is given by the gradient of the potential:

$$U_{TMD} = \frac{k}{2}(\text{RMSD}(t) - \text{RMSD}^*(t))^2$$

where $\text{RMSD}(t)$ is the instantaneous RMSD value measured between the current and target coordinates, $\text{RMSD}^*(t)$ evolves linearly from the first to the final step, and k is the associated force constant.

The force constant used in the initial TMD simulations for triggering DENV “breathing” were $1,000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ applied to the E protein backbone beads; this was sufficiently strong that the structure follows the centre of the harmonic spring closely, but not weak enough to avoid

introducing numerical instabilities. Following an initial 100 ns of TMD simulation, the E protein coordinates of the entire virion typically reached ~ 0.4 – 0.6 nm RMSD with respect to the target, expanded virus structure. To further refine the structure, it is advised to run additional, shorter follow-up TMD simulations with ~ 3 – 5 times increased force constant. Thus, an additional 10 ns simulation with a force constant of $4,000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ yielded a virion structure within ~ 0.1 nm RMSD with respect to the expanded experimental structure. The TMD simulations revealed the gradual opening of protrusions at the 3- and 5-fold vertices on the icosahedral surface of the virus, consistent with cryo-EM images [23].

3.6. Identification of ion-binding sites that govern DENV “breathing” reversibility

Fluorescence and HDX-MS data suggested an influence of divalent magnesium or calcium cations upon the reversibility of DENV expansion, upon lowering the temperature from 37°C back to 28°C , [25] and localized the effect as occurring at peptide segments around the virion 5-fold vertices. In order to investigate the molecular basis for this, we “back-mapped” our expanded CG DENV envelope model to all-atom representation. To do this, geometric projection/reconstruction was performed using a library of mapping definitions (available at <http://cgmartini.nl/index.php/tools2/resolution-transformation>), followed by several all-atom simulation relaxations [61]. From the back-mapped, relaxed virion structure, we extracted an E protein pentamer and its associated M proteins arranged around a 5-fold vertex. This was performed to reduce the system size in all-atom representation. The virus membrane was approximated as a flat lipid bilayer. This was built using CHARMM-GUI [62,63] and corresponded to a biologically relevant membrane composition, similarly to the preceding whole virion model, of a PC/PE/PS mixture in a $\sim 6:3:1$ ratio [52]. Subsequently, we embedded the pentamer with its transmembrane (TM) regions inserted into the membrane, and removed clashing lipids within ~ 0.3 nm of protein atoms. The system was then minimized in vacuum and solvated. In order to retain the higher order pentameric “expanded” structure, with its ectodomain positioned above the membrane according to cryo-EM data, we applied position restraints with a low force constant to the protein $\text{C}\alpha$ atoms [25] during simulations totalling 200 ns. Atomistic simulations revealed that the emergent spaces at the 5-fold vertices are accessible to water and salt, and that divalent cations can “soak” the protein 5-fold interface and competitively interact with acidic residues to break inter-chain salt bridges. This is expected to reduce the stability of the expanded state, enabling reversible contraction of the virion to its smooth form at low temperatures [19,23,24].

More recently, a combination of atomic-resolution simulations and cryo-EM revealed how point mutations at the dimeric E protein interface modulate the threshold for virion “breathing”, which is proposed

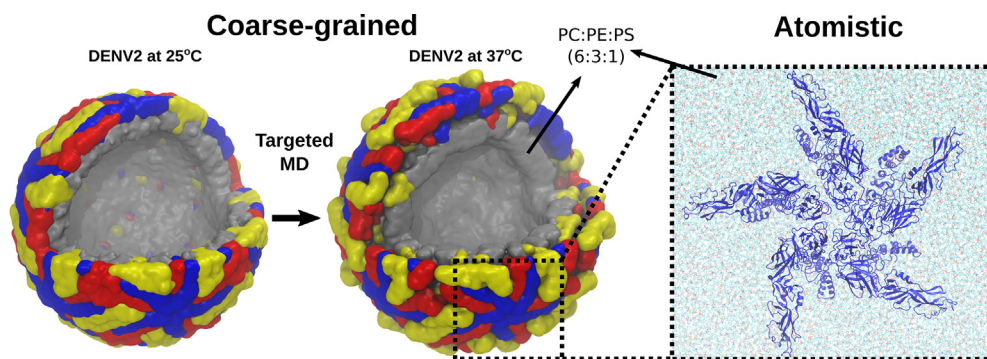


Fig. 2. Multiscale simulation approach to study DENV “breathing”. On the left, the application of TMD in CG resolution is illustrated for the transition between the smooth DENV2 (typical of lower temperatures as found in mosquitoes) and expanded DENV2 (typical of higher temperatures as found in the human host). The envelope is shown in surface representation, with lipids in grey, and proteins in yellow, red, or blue for the 2-, 3- and 5-fold vertices, respectively. On the right, an atomistic representation for an isolated E protein pentamer is shown, following-back mapping from the CG expanded virion structure. Protein is shown as blue cartoon, with lipids (PC:PE:PS in a 6:3:1 ratio) in CPK lines format.

to be a mechanism by which DENV adapts to its host environment to evade antibody recognition [64].

3.7. Integrative modelling of DENV maturation and ADE

ADE can result from the recognition of prM proteins by non-neutralizing antibodies on the surface of immDENV particles, facilitating virion endocytosis. However, following cleavage by furin of the pr:M complexes inside the endosome, it is unclear how pr dissociates in the acidic pH environment, such that the E protein FP is exposed and the virus can infect the cell. To investigate this further, low resolution (~12–25 Å) cryo-EM structural intermediates along the maturation pathway were solved of such an immDENV:antibody complex [65]. To improve the resolution and to resolve TM regions, a higher-resolution structure (of a fully immature virion, PDB: 4B03) [66] was overlaid onto the fully immature virion solved at pH = 8, via superimposition onto each of the individual 180 E protein ectodomains, using PyMOL [67]. Subsequently, a similar procedure to that described in Section 3.3 was used to generate equilibrated CG models for the proteins of the immature virion particle embedded within an explicit lipid bilayer envelope; the resulting theoretical density maps overlapped well with the equivalent experimental ones, reproducing the spherical membrane vesicle (Fig. 3A).

The equilibrated model of the immDENV:pr (pH = 8) complex was subsequently used as a starting point for TMD simulations, to explore the molecular pathway towards the intermediate-mature structure solved at pH = 5 (Fig. 3B). The initial RMSD between the ectodomain backbone in the initial versus target structures was ~5 nm. Initially, all E protein chains were simultaneously biased towards the target structure, but this led to significant clashes between E protein trimers (blue) and pentamers (red) (Fig. 3B). Thus, we instead applied a sequential series of biases to the E proteins; this involved first biasing all chains between RMSD values from 5 to 3.5 nm along the pathway, followed by a bias of the blue and green chains only between RMSD values from ~3.5 to 2.5 nm, and finally a bias of all chains from ~2.5 nm RMSD until the end of the simulation (Fig. 3C). These simulations revealed that the neighbouring blue E protein molecule was approached by the pr:red-E protein complexes, dislodging the pr molecule from the 5-fold symmetry axes; based on additional modelling studies, this was predicted to be accelerated in the presence of bound antibodies due to clashes between Fab domains. This extent of pr binding site occlusion on the red E protein was not observed for the blue or green molecules (Fig. 3C). Similar conclusions were drawn when running TMD simulations in which biases were instead applied independent either to red (5-fold), blue (3-fold), or green (2-fold) E protein molecules alone. These results matched nicely with cryo-EM data, and combined with co-floatation assays and HDX-MS measurements, this study indicates that pr:Fab complexes dissociate as a complex from DENV E proteins due to steric clashes, thereby facilitating maturation and infectivity [65].

It should be noted that whilst the predictions regarding conformational pathways in breathing or maturation are likely to be accurate, not least because of the incorporation of multiple experimental intermediate “snapshots” which help to validate the structural transitions, the kinetics of these processes are more difficult to accurately characterize. In part, this is because of the TMD approach, in which the specified timescale during the linear evolution of the structure as well as associated biasing force constant will largely determine the rate of change along the pathway of interest. On top of this, the nature of the CG methodology means that timescales should be treated with great care. Like other CG forcefields, Martini simplifies the energy landscape, speeding up sampling rates, but this speed-up is not easily predictable and is unlikely to be consistent across different degrees of freedom, or different types of molecules [68]. From the experimental perspective, the rates of whole-virion conformational changes are similarly not well established, though single-molecule approaches are beginning to help measure such phenomena [43].

4. Simulations and integrative modelling of the flavivirus C protein

The flavivirus C protein is a small but multifunctional protein. A large part of its multifunctionality is derived from the intrinsically disordered regions (IDRs) at the N-terminus of the protein. Due to the flexible nature of the N-terminal IDRs, the structure and dynamics of these regions have remained undefined: X-ray crystallographic and NMR structures of flavivirus C proteins either solve the structure without the N-terminal tails or are unable to resolve the atoms in the region [69–71]. However it remains an important task to characterise and understand the dynamics of these IDRs, and indeed, intrinsically disordered proteins (IDPs) are abundantly found in nature [72,73].

Using MD simulations in conjunction with biophysical experiments, the dynamics of these IDRs can be better understood at atomic resolution. In recent years, the need for accurate treatment of IDRs in simulations have seen a rise in IDP-specific parameterised force fields [74–77]. We recently showed that certain force fields are more suited to the DENV C protein than others [78]. Small angle X-ray scattering (SAXS) provides low resolution information on the shape of proteins in solution and is suited to IDPs due to their fuzzy and dynamic nature [79]. Selecting conformations from a simulation has the benefit of selecting structures that have been relaxed and equilibrated and are closer to the experimental conditions. This is as opposed to selecting conformations from a random pool of conformations for the IDRs derived from X-ray structures or NMR structures that may have unresolved clashes or are in non-aqueous conformations. Using an ensemble of conformations derived from MD simulations, we were able to fit the SAXS data better than using a randomly generated pool [78]. The SAXS data was used to validate the most representative ensemble of conformations from the different force fields by fitting a pool of structures derived from the simulations, and selected via a genetic algorithm that minimises the chi-squared value. The force field with the smallest chi-squared value was the IDP-parameterized force field, amber03ws. The resultant conformations that best fit the SAXS data showed that the alpha 1 helices towards the N-terminus of the protein may come together in a clamp-like motion to occlude a hydrophobic patch at the core of the DENV C protein. This is a refinement of the previously solved NMR structure, and may be a better representation of the protein in an aqueous environment.

HDX-MS may also be useful to elucidate the dynamics of IDRs. The protein is incubated in the presence of deuterated water and then the reaction is quenched. During the time of incubation in the presence of deuterium, the backbone amide hydrogens exchange with the deuterium, and this change in mass can be detected through mass spectrometry. IDPs and IDRs show greater exchange of their backbone amide hydrogens with deuterium than do proteins regions that are structured. The correlation of backbone hydrogen bond formation or solvent accessibility with amide backbone deuteration may also be used to validate the ensemble of structures derived from different forcefields [78]. Again, trajectories based on the amber03ws force field exhibited the highest correlations for both backbone hydrogen bond formation and solvent accessibility with experimentally measured amide backbone deuteration, thus validating our calculated ensemble of C protein conformations.

5. Methods for RNA genome structure determination

Inherent structural complexity and variability of viral genomic material make it an elusive target for common structural biology techniques. While it is possible to study genomic structures of simple viruses with cryo-EM [80–83], more complex viral systems such as flaviviruses remain a challenge [84–88]. Determining the structure and organization of viral genomes is crucial for understanding the function of this core component of viruses at the molecular level [89–96]. In the face of challenges to classic structural biology techniques, a new set of

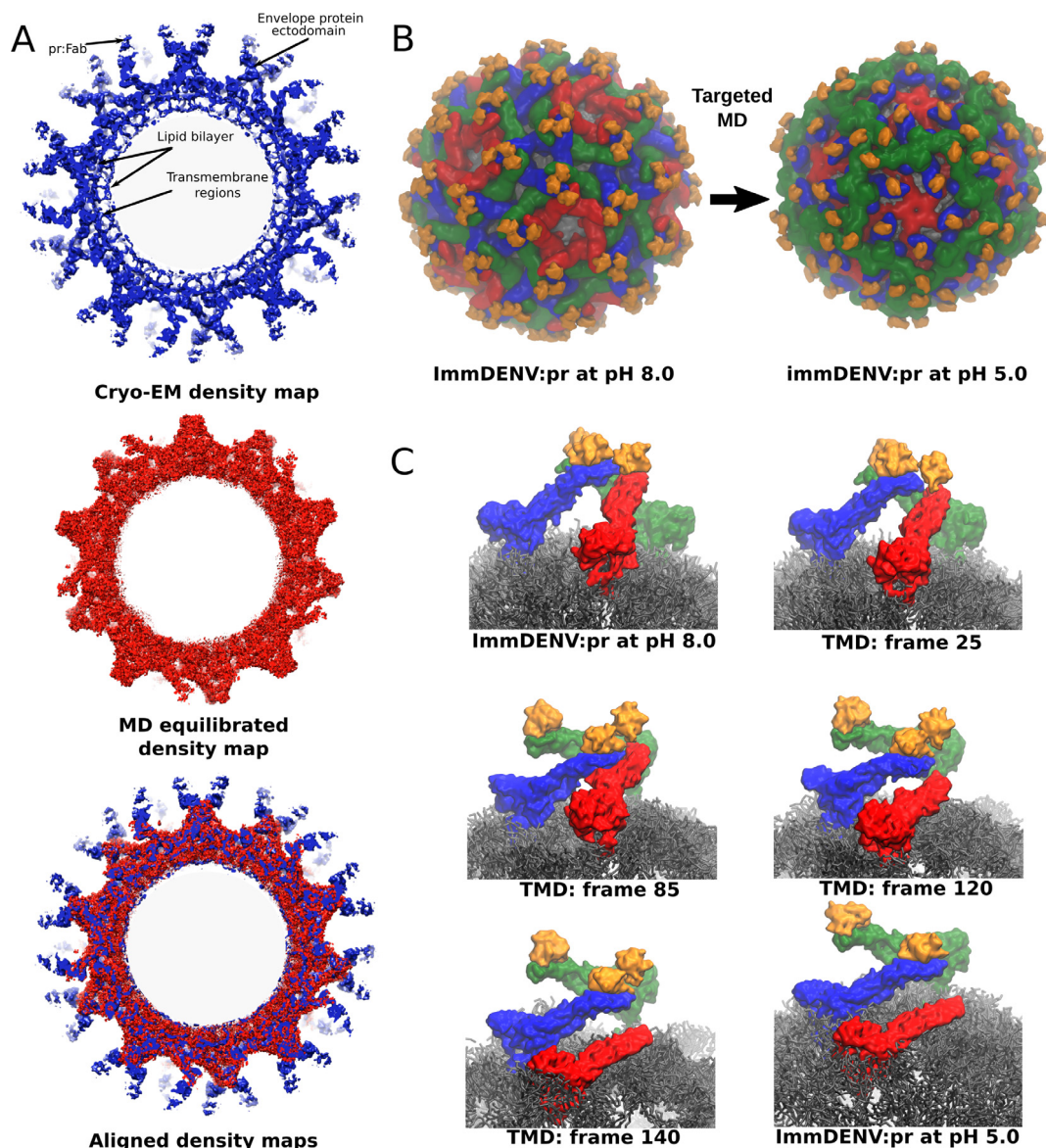


Fig. 3. Integrative modelling of maturation and ADE. (A) Cross-sections of the cryo-EM map of the immDENV:pr:Fab complex at pH 5.0 (shown in blue, top), the MD simulation generated density map of the immDENV:pr complex (shown in red, middle), and an alignment of both (bottom). Fab molecules were not present during simulations, hence the different appearance of the theoretical versus experimental maps in the outer part of the virion (bottom). (B) The initial (left) and final (right) frames of TMD simulations to sample the transition pathway during maturation. (C) Representative snapshots during a TMD simulation transitioning between experimental states (immDENV:pr at pH 8.0 and 5.0), during which a pr fragment is spontaneously dislodged. In the target experimental structure (bottom right) the pr molecule on the pentamer (red) was not resolved. In (B) and (C), proteins are shown in surface representation (the elongated structures are E proteins, whilst the smaller, compact structures are the pr fragments, shown in orange), coloured according to the symmetry axis for red (pentamer), blue (trimer), and green (dimer) molecules, and lipids are shown in grey.

methods is called for to investigate this key area of viral biology. Recent advances in genome sequencing have allowed us to leverage structural probing techniques that were previously confined to small RNA systems on the whole genome level, particularly for viral genomes [89,94,97–103]. Combining information obtained through structural probing and cross-linking experiments with computational modelling techniques allows us to gain insights into genome-level interactions between different loci that are of functional importance for the virus [95]. Structural probing can be performed at distinct phases within the viral life cycle, e.g. at a virion stage or within infected cells, and hence a time-resolved picture of genomic organization can be obtained. This allows us to identify key structural features at the stages of translation, replication, or packaging, as recently demonstrated for the ~11 kb RNA genome of DENV and ZIKV, potentially opening up novel therapeutic strategies targeting flaviviruses [95] (Fig. 4).

5.1. Structure prediction and modelling

RNA is capable of adopting a wide range of different structures that serve a regulatory or catalytic [104] purpose including translation, splicing decay, and RNA transport [105]. As such, studying how RNAs fold and their energetic landscapes is important for understanding gene regulation [106]. The problem of predicting RNA secondary structure from sequence is challenging as a large search space of possible base pairs needs to be traversed and the energetics assessed with high accuracy. A variety of tools exist that allow the prediction of likely RNA secondary structure based on folding free energy estimates and capable of including chemical reactivity information [107–110]. Generally, these approaches use energetic terms for base pairing and stacking interactions. The challenge for predicting the structure of long RNAs increases exponentially as the search space increases and hence accuracy

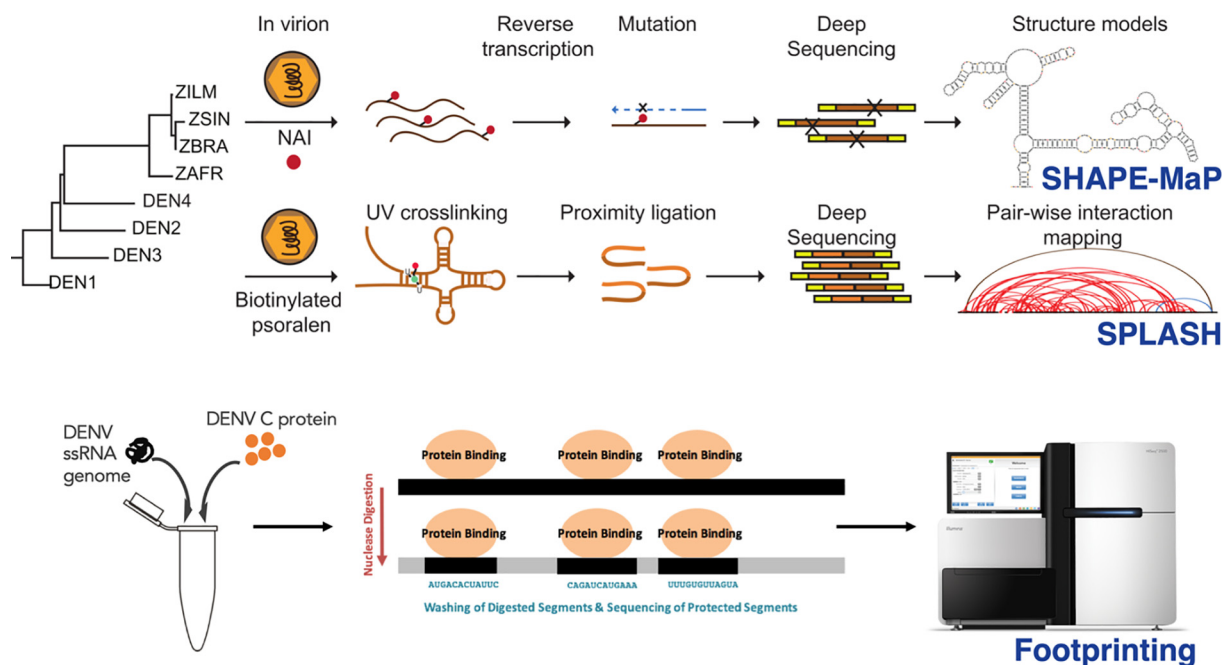


Fig. 4. Experimental sources of data to augment RNA structure prediction. The search space for the prediction of long RNA molecules is too large to obtain accurate predictions of structure using computational techniques alone. Incorporation of experimental data can increase the prediction accuracy dramatically. SHAPE data provides information on local structure by assessing the differential reactivity of the 2'-hydroxyl group in flexible versus constrained bases. SPLASH and other cross-linking techniques can provide information on the macro-organization of RNA by forming chimeric reads. Footprinting techniques can identify regions of RNA interacting with proteins by observing an increased resistance to digestion that correlate with increasing protein concentrations.

of prediction for long RNAs without the inclusion of additional experimental information is limited [100]. A common strategy to mitigate this effect is to limit the maximum distance between paired bases to a few hundred nucleotides, thus reducing the search space considerably. The trade-off involved in this approach results in more accurate predictions of local secondary structure at the cost of neglecting longer-range interactions that are responsible for the macro-scale organization of RNA molecules. Unfortunately, it is evident from studies into viral systems that such long-range interactions are often key functional elements, as is highlighted e.g. by the Flavivirus 5'-3' circularization motifs that span the length of the entire genome [111]. At present, only a combination of different methods is able to capture long-range and short-range interactions with high accuracy [95,112,113].

While it has been shown that de-novo prediction of RNA structures from sequences alone is unreliable for larger RNAs [114,115], this problem can be overcome to some degree by including experimental information on secondary structure in the modelling process [100]. Selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) is an established protocol yielding such secondary structure information for all of the four bases (A, U, C, G) [116]. The concept underlying SHAPE probing is that local RNA structure, especially the difference between flexible and physically constrained bases, affects the reactivity of the 2'-hydroxyl group of constituent nucleotides. Flexible bases are able to orientate themselves to react with SHAPE chemicals to result in acylation at the 2'-hydroxyl of the RNA. Hence, local secondary structure can be read by observing the degree of 2'-hydroxyl acylation at a specific position. The readout can be performed on a full genome level, as 2'-hydroxyl modified nucleotides tend to induce skips with a significantly increased likelihood of reverse transcription over unmodified nucleotides [99,100]. By sequencing a reverse-transcribed library of chemically modified RNA, an increase in mutations at modified positions can be observed. From this data, local secondary structure can be inferred by comparing the local mutation rate to the mutation rate of a mock (usually DMSO) modified sample. The modification ratio thus obtained can be used as a probabilistic penalty term in secondary structure prediction [100,101,118]. It should be noted that whereas

structuredness precludes high reactivity signals at a specific position, low reactivity signals are still more likely to occur even at an open or flexible region [118]. In addition, low reactivity signals could also be protein-interacting sites, whereby the protein binding protects the region from interacting with SHAPE compounds. However, in general, reactivity can be translated into an odds ratio [118] that contributes significantly to the prediction accuracy of RNA secondary structure. Whereas naïve prediction accuracy for the 16S subunit of the *E. coli* ribosome reaches around 45% of base pairs and helices, including SHAPE data allows an increase in accuracy to well over 90% [100]. This technique has been successfully applied to probe the genomic structure of a range of viral systems, e.g. HIV [98,101,117] or flaviviruses [95,112,119].

Whereas SHAPE-MaP data is able to improve the prediction accuracy of local secondary structure significantly, there remains a level of structural uncertainty related to the macro-organization of genomic RNAs. Ensembles generated with common structure prediction tools yield a diversity of macro-patterns. Chemical cross-linking information is invaluable to narrow down the space of these patterns to fewer possibilities by providing pair-wise RNA interaction information [95]. Several methods for cross-linking based structure probing exist, mainly based on psoralen intercalation into helical segments and subsequent UV-based covalent cross-linking [89,112–113,119,120]. One way of enriching for the cross-linked RNA interactions is to utilize a biotinylated version of psoralen, which allows the selective extraction of cross-linked segments using streptavidin beads. Subsequent proximity ligation steps followed by PCR amplification and deep sequencing produces chimeric reads that reveal intra-genomic structures at various distances. Moreover, if conducted in the presence of host RNA, virus-host RNA-RNA interactions can be captured as well. Interestingly, a significant number of positions tend to form multiple interactions that are mutually exclusive, indicating the presence of either structural heterogeneity in the virus population, or constant fluctuations in virus structures [95]. Observing mutually exclusive interactions in the full ensemble is to be expected if structural heterogeneity in the underlying population exists, i.e. if there is more than one way that the genome can

be arranged within the mature virions (or in infected cells for that matter). Unfortunately, to date, we have been technically unable to conduct an experiment on a single virion, which would answer this question. In case the observed ensemble is the result of in-virion dynamics, time averaging during the experiment may still reveal competing positions. If structural heterogeneity exists only across virions with a specific frozen state in each individual virion, the single-virion experiment would not show any competing positions, but a subset of the full ensemble. Combining ensembles of structures that have been predicted with high local accuracy using SHAPE-constrained modelling with information about the macro-organization obtained through cross-linking experiments allows for a high-confidence selection of plausible RNA structures from the local to global scale of organization.

5.2. Integrative modelling of RNA and RNA-capsid interactions

RNA molecules do not exist in isolation inside cells or even inside virion particles, and can closely interact with proteins, e.g. in viral capsids or the ribosome. Such RNA-protein interactions have the potential to materially affect the structure of RNA, but are currently impossible to account for in free-energy-based structure prediction algorithms. The integrative modelling of RNA and protein-RNA interactions in the context of the whole viral nucleocapsid is still a new and challenging field [45]. Due to the flexible nature of RNA and the range of secondary and tertiary structures that RNAs can adopt, modelling the RNA genome is a non-trivial problem at the atomistic and even CG level. There have been recent efforts to develop tools to model the 3D structure of RNA and these methods range from atomistic representations, CG representations, and graph representations of RNA (i.e. helices as edges and loops as vertices); these have been reviewed extensively elsewhere [121–122].

There are also currently a few methods that can model the interactions of large RNAs and multiple proteins [123,124]. Two such methods are the integrative modelling platform (IMP) [125,126], which allows the definition of proteins and RNA as large beads whose spatial arrangement can be subjected to minimization to satisfy user-defined restraints, and PyRy3D (<http://genesilico.pl/pyry3d/>), which also employs an integrative modelling approach. These restraints can come from knowledge of protein-binding regions along the viral genome, length, size, base-pairing, long-range interactions, electrostatic interactions between the capsid and viral genome, and volume restraints of the viral particle.

5.3. Verification of RNA structure

Thorough verification of identified structural motifs is a crucial part of structural genomics research. Beyond the identification of structural peculiarities, identifying functional structures is a key goal. Functional structures, either existing within viral genomes or formed via virus-host RNA interactions in infected cells may offer targets for therapeutic intervention. A key method to validate structural motifs arises from investigating the evolutionary history of a particular virus [127,128]. By identifying covarying bases in the genetic history of a virus, evolutionarily conserved structural motifs can be identified [127–129]. A prerequisite for such an analysis is the availability of a large number of viral sequences, ideally spanning as diverse a number of time points and geographies as possible. Such a wealth of sequence information is usually only available for well-studied systems, e.g. HIV, Influenza, or DENV. Emerging threats such as the 2016 Zika pandemic virus prove more difficult to investigate by reference to the historical genetic record. While there is a large number of full-genome sequences available for Zika viruses, most of these samples focus on the 2016 South American outbreak and exhibit only a limited amount of genetic diversity, making the identification of covariation for structure validation challenging. In contrast, analysing dengue viruses that have been subjected to much more intense surveillance over time and geographic

regions allowed us to validate predicted structural motifs with high confidence [95].

Ultimately, verifying the functional significance of structural motifs relies on the ability to create viruses with mutations that disrupt the structures of interest and subsequently measuring the functional impact of these modifications via fitness assays in cells or in animals. This poses a particular challenge when the motifs are present in coding regions, as opposed to untranslated regions, as possible modifications are constrained by a number of factors such as the need to preserve protein sequences through synonymous mutations and the need to avoid rare codons that would impact translation efficiency independently of any structural effects at the genome level. Moreover, the ability to design compensatory mutations is heavily constrained in coding regions due to the same factors and the need for both modifications to align.

6. Conclusions

Viruses have complex life cycles which are dependent upon interactions with a variety of host factors and are constantly adapting, making therapeutics development challenging. Continuing increases in computational power, widespread use of multiscale approaches, and growing efforts to integrate diverse structural, biophysical, and HTS data into theoretical models have made possible the accurate description of entire virus particles, including component protein, lipid, and nucleic acid material. As outlined here for flaviviruses, this is enabling us to refine envelope structure, predict its dynamic interplay with environmental and immunological triggers, elucidate capsid conformational ensembles, and map genomic architecture. Similar approaches are now being applied to the emerging virus SARS-CoV-2 that is responsible for the COVID-19 pandemic, and indeed, numerous coronavirus structures of the trimeric pre-fusion glycoprotein spike responsible for infection have been published [130]. In the years to come, this will provide promising routes to novel antiviral and vaccine design strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by A*STAR. JKM and PJB thank NRF (NRF2017NRF-CRP001-027) for funding.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymeth.2020.06.004>.

References

- [1] K.F. Ortblad, R. Lozano, C.J.L. Murray, The Burden of HIV: Insights from the Global Burden of Disease Study 2010, *AIDS* 27 (13) (2013) 2003–2017.
- [2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, et al., A Novel Coronavirus from Patients with Pneumonia in China, 2019, *N. Engl. J. Med.* 382 (2020) 727–733.
- [3] S. Bhatt, P.W. Gething, O.J. Brady, J.P. Messina, A.W. Farlow, C.L. Moyes, J.M. Drake, J.S. Brownstein, A.G. Hoen, O. Sankoh, et al., The Global Distribution and Burden of Dengue, *Nature* 496 (2013) 504–507.
- [4] S. Bressanelli, K. Stiasny, S.L. Allison, E.A. Stura, S. Duquerroy, J. Lescar, F.X. Heinz, F.A. Rey, Structure of a Flavivirus Envelope Glycoprotein in Its Low-PH-Induced Membrane Fusion Conformation, *EMBO J.* 23 (2004) 728–738.
- [5] R.J. Kuhn, W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, et al., Structure of dengue virus: implications for flavivirus organization, maturation, and fusion, *Cell* 108 (5) (2002) 717–725.
- [6] D.M. Morens, Antibody-dependent enhancement of infection and the pathogenesis

- of viral disease, *Clin. Infect. Dis.* 19 (3) (1994) 500–512.
- [7] M. Aguiar, N. Stollenwerk, S.B. Halstead, The risks behind dengvaxia recommendation, *Lancet Infect. Dis.* 16 (8) (2016) 882.
- [8] E. Callaway, Revolutionary Cryo-EM Is Taking over Structural Biology, *Nature* 578 (7794) (2020) 201.
- [9] J.K. Marzinek, R.G. Huber, P.J. Bond, Multiscale Modelling and Simulation of Viruses, *Curr. Opin. Struct. Biol.* 61 (2020) 146–152.
- [10] T.Y. Tan, G. Fibriansah, V.A. Kostyuchenko, T.S. Ng, X.X. Lim, S. Zhang, X.N. Lim, J. Wang, J. Shi, M.C. Morais, et al., Capsid protein structure in zika virus reveals the flavivirus assembly process, *Nat. Commun.* 11 (1) (2020) 1–13.
- [11] J.M. Smit, B. Moesker, I. Rodenhuis-Zybert, J. Wilschut, Flavivirus Cell Entry and Membrane Fusion, *Viruses* 3 (2) (2011) 160–171.
- [12] Y. Modis, S. Ogata, D. Clements, S.C. Harrison, Structure of the Dengue Virus Envelope Protein after Membrane Fusion, *Nature* 427 (6972) (2004) 313–319.
- [13] Y. Modis, S. Ogata, D. Clements, S.C. Harrison, A ligand-binding pocket in the dengue virus envelope glycoprotein, *Proc. Natl. Acad. Sci.* 100 (12) (2003) 6986–6991.
- [14] L. Li, S.-M. Lok, L.-M. Yu, Y. Zhang, R.J. Kuhn, J. Chen, M.G. Rossmann, The flavivirus precursor membrane-envelope protein complex: structure and maturation, *Science* 319 (2008) 1830–1834.
- [15] Y. Zhang, J. Corver, P.R. Chipman, W. Zhang, S.V. Pletnev, D. Sedlak, T.S. Baker, J.H. Strauss, R.J. Kuhn, M.G. Rossmann, Structures of immature flavivirus particles, *EMBO J.* 22 (11) (2003) 2604–2613.
- [16] L.M. Yu, W. Zhang, H.A. Holdaway, L. Li, V.A. Kostyuchenko, P.R. Chipman, R.J. Kuhn, M.G. Rossmann, J. Chen, Structure of the immature dengue virus at low pH primes proteolytic maturation, *Science* 319 (5871) (2008) 1834–1837.
- [17] J. Junjhon, T.J. Edwards, U. Utaipat, V.D. Bowman, H.A. Holdaway, W. Zhang, P. Keelapang, C. Puttikhant, R. Perera, P.R. Chipman, et al., Influence of Pr-m cleavage on the heterogeneity of extracellular dengue virus particles, *J. Virol.* 84 (16) (2010) 8353–8358.
- [18] V. Claus, A. Jahraus, T. Tjelle, T. Berg, H. Kirschke, H. Faulstich, G. Griffiths, Lysosomal Enzyme Trafficking between Phagosomes, Endosomes, and Lysosomes in J774 Macrophages: Enrichment of Cathepsin H in Early Endosomes, *J. Biol. Chem.* 273 (16) (1998) 9842–9851.
- [19] S.M. Lok, V. Kostyuchenko, G.E. Nybakken, H.A. Holdaway, A.J. Battisti, S. Sukupolvi-Petty, D. Sedlak, D.H. Fremont, P.R. Chipman, J.T. Roehrig, et al., Binding of a Neutralizing Antibody to Dengue Virus Alters the Arrangement of Surface Glycoproteins, *Nat. Struct. Mol. Biol.* 15 (3) (2008) 312–317.
- [20] K.A. Dowd, S. Mukherjee, R.J. Kuhn, T.C. Pierson, Combined Effects of the Structural Heterogeneity and Dynamics of Flaviviruses on Antibody Recognition, *J. Virol.* 88 (20) (2014) 11726–11737.
- [21] K.A. Dowd, T.C. Pierson, Antibody-Mediated Neutralization of Flaviviruses: A Reductionist View, *Virology* 411 (2) (2011) 306–315.
- [22] F.A. Rey, Dengue virus: two hosts, two structures, *Nature* 497 (7450) (2013) 443–444.
- [23] G. Fibriansah, T.-S. Ng, V. Kostyuchenko, J. Lee, S. Lee, J. Wang, S.-M. Lok, Structural Changes in Dengue Virus When Exposed to a Temperature of 37°C, *J. Virol.* 87 (13) (2013) 7585–7592.
- [24] X. Zhang, J. Sheng, P. Plevka, R.J. Kuhn, M.S. Diamond, M.G. Rossmann, Dengue structure differs at the temperatures of its human and mosquito hosts, *Proc. Natl. Acad. Sci.* 110 (17) (2013) 6795–6799.
- [25] K.K. Sharma, X.X. Lim, S.N. Tantirimudalige, A. Gupta, J.K. Marzinek, D. Holdbrook, X.Y.E. Lim, P.J. Bond, G.S. Anand, T. Wohland, Infectivity of Dengue Virus Serotypes 1 and 2 Is Correlated with E-Protein Intrinsic Dynamics but Not to Envelope Conformations, *Structure* 27 (4) (2019) 618–630.
- [26] D.S. Mueller, T. Kampmann, R. Yennamalli, P.R. Young, B. Kobe, A.E. Mark, Histidine protonation and the activation of viral fusion proteins, *Biochem. Soc. Trans.* 36 (1) (2008) 43–45.
- [27] C.A. Fuzo, L. Degrève, The pH dependence of flavivirus envelope protein structure: insights from molecular dynamics simulations, *J. Biomol. Struct. Dyn.* 32 (10) (2014) 1563–1574.
- [28] K.D. Dube, A.K. Chaubey, R.P. Ojha, Stability of Trimeric DENV Envelope Protein at Low and Neutral pH: An Insight from MD Study, *Biochim. Biophys. Acta - Proteins Proteomics* 1834 (1) (2013) 53–64.
- [29] K.D. Dube, A.K. Chaubey, R.P. Ojha, Role of pH on Dimeric Interactions for DENV Envelope Protein: An Insight from Molecular Dynamics Study, *Biochim. Biophys. Acta - Proteins Proteomics* 1814 (12) (2011) 1796–1801.
- [30] L. Degrève, C.A. Fuzo, Structure and Dynamics of the Monomer of Protein E of Dengue Virus Type 2 with Unprotonated Histidine Residues, *Genet. Mol. Res.* 12 (1) (2013) 348–359.
- [31] M.K. Prakash, A. Barducci, M. Parrinello, Probing the Mechanism of pH-Induced Large-Scale Conformational Changes in Dengue Virus Envelope Protein Using Atomistic Simulations, *Biophys. J.* 99 (2) (2010) 588–594.
- [32] R. Fritz, K. Stiasny, F.X. Heinz, Identification of Specific Histidines as pH Sensors in Flavivirus Membrane Fusion, *J. Cell Biol.* 183 (2) (2008) 353–361.
- [33] J. Sun, Y. Li, P. Liu, J. Lin, Study of the Mechanism of Protonated Histidine-Induced Conformational Changes in the Zika Virus Dimeric Envelope Protein Using Accelerated Molecular Dynamic Simulations, *J. Mol. Graph. Model.* 74 (2017) 203–214.
- [34] E.R.A. Oliveira, R.B. de Alencastro, B.A.C. Horta, New Insights into Flavivirus Biology: The Influence of pH over Interactions between PrM and E Proteins, *J. Comput. Aided. Mol. Des.* 31 (11) (2017) 1009–1019.
- [35] J.M. Vanegas, F. Heinrich, D.M. Rogers, B.D. Carson, S. La Bauve, B.C. Vernon, B. Akgun, S. Satija, A. Zheng, M. Kielian, et al., Insertion of Dengue E into Lipid Bilayers Studied by Neutron Reflectivity and Molecular Dynamics Simulations, *Biochim. Biophys. Acta - Biomembr.* 1860 (5) (2018) 1216–1230.
- [36] D.M. Rogers, M.S. Kent, S.B. Rempe, Molecular Basis of Endosomal-Membrane Association for the Dengue Virus Envelope Protein, *Biochim. Biophys. Acta* 1848 (4) (2015) 1041–1052.
- [37] J.K. Marzinek, R. Lakshminarayanan, E. Goh, R.G. Huber, S. Panzade, C. Verma, P.J. Bond, J.K. Marzinek, R. Lakshminarayanan, E. Goh, R.G. Huber, S. Panzade, C. Verma, P. B. Characterizing the Conformational Landscape of Flavivirus Fusion Peptides via, Simulation and Experiment. *Sci. Rep.* 6 (2016) 19160.
- [38] J.K. Marzinek, N. Bag, R.G. Huber, D.A. Holdbrook, T. Wohland, C.S. Verma, P.J. Bond, A Funneled Conformational Landscape Governs Flavivirus Fusion Peptide Interaction with Lipid Membranes, *J. Chem. Theory Comput.* 14 (7) (2018) 3920–3932.
- [39] J. Pan, C.B. Lai, W.R.P. Scott, S.K. Straus, Synthetic Fusion Peptides of Tick-Borne Encephalitis Virus as Models for Membrane Fusion, *Biochemistry* 49 (2) (2010) 287–296.
- [40] Y.S. Mendes, N.S. Alves, T.L.F. Souza, I.P. Sousa, M.L. Bianconi, R.C. Bernardi, P.G. Pascutti, J.L. Silva, A.M.O. Gomes, A.C. Oliveira, The Structural Dynamics of the Flavivirus Fusion Peptide-Membrane Interaction, *PLoS One* 7 (10) (2012) e47596.
- [41] *J. Phys.: Condens. Matter* 29 (50) (2017) 504002, <https://doi.org/10.1088/1361-648X/aa99c6>.
- [42] R.G. Huber, J.K. Marzinek, D.A. Holdbrook, P.J. Bond, Multiscale Molecular Dynamics Simulation Approaches to the Structure and Dynamics of Viruses, *Prog. Biophys. Mol. Biol.* 128 (2017) 121–132.
- [43] K.K. Sharma, J.K. Marzinek, S.N. Tantirimudalige, P.J. Bond, T. Wohland, Single-Molecule Studies of Flavivirus Envelope Dynamics: Experiment and Computation, *Prog. Biophys. Mol. Biol.* 143 (2019) 38–51.
- [44] T. Reddy, M.S.P. Sansom, The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virus Envelope, *Structure* 24 (3) (2016) 375–382.
- [45] T. Reddy, M.S.P. Sansom, Computational Virology: From the inside Out, *Biochim. Biophys. Acta - Biomembr.* 1858 (7) (2016) 1610–1618.
- [46] C. Pindi, V.R. Chirasani, M.H. Rahman, M. Ahsan, P.D. Revanasiddappa, S. Senapati, Breaking the “Unbreakable” ZIKA Virus Envelope, *bioRxiv* (2018) 345884.
- [47] J.K. Marzinek, D.A. Holdbrook, R.G. Huber, C. Verma, P.J. Bond, Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations, *Structure* 24 (8) (2016) 1410–1420.
- [48] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, S.J. Marrink, The MARTINI Coarse-Grained Force Field: Extension to Proteins, *J. Chem. Theory Comput.* 4 (5) (2008) 819–834.
- [49] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. De Vries, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, *J. Phys. Chem. B* 111 (27) (2007) 7812–7824.
- [50] X. Periole, M. Cavalli, S.J. Marrink, M.A. Ceruso, Combining an Elastic Network with a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition, *J. Chem. Theory Comput.* 5 (9) (2009) 2531–2543.
- [51] Y. Qi, H.I. Ingólfsson, X. Cheng, J. Lee, S.J. Marrink, W. Im, CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field, *J. Chem. Theory Comput.* 11 (9) (2015) 4486–4494.
- [52] Q. Zhang, C. Hunke, Y.H. Yau, V. Seow, S. Lee, L.B. Tanner, X.L. Guan, M.R. Wenk, G. Fibriansah, P.L. Chew, et al., The Stem Region of Premembrane Protein Plays an Important Role in the Virus Surface Protein Rearrangement during Dengue Maturation, *J. Biol. Chem.* 287 (48) (2012) 40525–40534.
- [53] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations, *J. Comput. Chem.* 32 (10) (2011) 2319–2327.
- [54] R. Briones, C. Blau, C. Kutzner, B.L. de Groot, C. Aponte-Santamaría, GROMACS: A GROMACS-Based Toolset to Analyze Density Maps Derived from Molecular Dynamics Simulations, *Biophys. J.* 116 (1) (2019) 4–11.
- [55] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A Visualization System for Exploratory Research and Analysis, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [56] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, A. Wollmer, Targeted Molecular Dynamics Simulation of Conformational Change - Application to the T ↔ R Transition in Insulin, *Mol. Simul.* 10 (2–6) (1993) 291–308.
- [57] J. Schlitter, M. Engels, P. Krüger, Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions, *J. Mol. Graph.* 12 (2) (1994) 84–89.
- [58] Spoel, D. Van Der; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* 2005, 26 (16), 1701–1718.
- [59] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New Feathers for an Old Bird, *Comput. Phys. Commun.* 185 (2) (2014) 604–613.
- [60] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.* 26 (16) (2005) 1781–1802.
- [61] T.A. Wassenaar, K. Pluhackova, R.A. Bockmann, S.J. Marrink, D.P. Tieleman, Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models, *J. Chem. Theory Comput.* 10 (2) (2014) 676–690.
- [62] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM, *J. Comput. Chem.* 29 (11) (2008) 1859–1865.
- [63] E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, et al., CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations, *J. Comput. Chem.* 35 (27) (2014) 1997–2004.

- [64] X.-N.N. Lim, C. Shan, J.K. Marzinek, H. Dong, T.S. Ng, J.S.G.G. Ooi, G. Fibriansah, J. Wang, C.S. Verma, P.J. Bond, et al., Molecular Basis of Dengue Virus Serotype 2 Morphological Switch from 29°C to 37°C, *PLoS Pathog.* 15 (9) (2019) e1007996.
- [65] M. Wirawan, G. Fibriansah, J.K. Marzinek, X.X. Lim, T.-S.S. Ng, A.Y.L.L. Sim, Q. Zhang, V.A. Kostyuchenko, J. Shi, S.A. Smith, et al., Mechanism of Enhanced Immature Dengue Virus Attachment to Endosomal Membrane Induced by PrM Antibody, *Structure* 27 (2) (2019) 253–267.
- [66] V.A. Kostyuchenko, Q. Zhang, J.L. Tan, T.-S. Ng, S.-M. Lok, Immature and Mature Dengue Serotype 1 Virus Structures Provide Insight into the Maturation Process, *J. Virol.* 87 (13) (2013) 7700–7707.
- [67] L. Schrödinger, The PyMOL Molecular Graphics System, Version 1 (2015) 8.
- [68] S.J. Marrink, D.P. Tieleman, Perspective on the Martini Model, *Chem. Soc. Rev.* 42 (16) (2013) 6801–6822.
- [69] L. Ma, C.T. Jones, T.D. Groesch, R.J. Kuhn, C.B. Post, Solution Structure of Dengue Virus Capsid Protein Reveals Another Fold, *Proc. Natl. Acad. Sci. United States Am.* 101 (10) (2004) 3414–3419.
- [70] Z. Shang, H. Song, Y. Shi, J. Qi, G.F. Gao, Crystal Structure of the Capsid Protein from Zika Virus, *J. Mol. Biol.* 430 (7) (2018) 948–962.
- [71] T. Dokland, M. Walsh, J.M. Mackenzie, A.A. Khromykh, K.-H. Ee, S. Wang, West Nile Virus Core Protein: Tetramer Structure and Ribbon Formation, *Structure* 12 (7) (2004) 1157–1163.
- [72] B. Xue, D. Blocquel, J. Habchi, A.V. Uversky, L. Kurgan, V.N. Uversky, S. Longhi, Structural Disorder in Viral Proteins, *Chem. Rev.* 114 (13) (2014) 6880–6911.
- [73] V.N. Uversky, The Most Important Thing Is the Tail: Multitudinous Functionalities of Intrinsically Disordered Protein Termini, *FEBS Lett.* 587 (13) (2013) 1891–1901.
- [74] R.B. Best, W. Zheng, J. Mittal, Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association, *J. Chem. Theory Comput.* 10 (11) (2014) 5113–5124.
- [75] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmüller, A.D. MacKerell, CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins, *Nat. Methods* 14 (1) (2016) 71–73.
- [76] J. Huang, A.D. MacKerell, Force Field Development and Simulations of Intrinsically Disordered Proteins, *Curr. Opin. Struct. Biol.* 48 (2018) 40–48.
- [77] D. Song, R. Luo, H.F. Chen, The IDP-Specific Force Field Ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins, *J. Chem. Inf. Model.* 57 (5) (2017) 1166–1178.
- [78] P.L.S. Boon, W.G. Saw, X.X. Lim, P.V. Raghuvamsi, R.G. Huber, J.K. Marzinek, D.A. Holdbrook, G.S. Anand, G. Grüber, P.J. Bond, Partial Intrinsic Disorder Governs the Dengue Capsid Protein Conformational Ensemble, *ACS Chem. Biol.* 13 (6) (2018) 1621–1630.
- [79] A.G. Kikhney, D.I. Svergun, A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins, *FEBS Lett.* 589 (19) (2015) 2570–2577.
- [80] X. Zhang, L. Jin, Q. Fang, W.H. Hui, Z.H. Zhou, 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry, *Cell* 141 (3) (2010) 472–482.
- [81] D.K. Clare, E. Orlova, V. 4.6 Angstrom Cryo-EM Reconstruction of Tobacco Mosaic Virus from Images Recorded at 300 KeV on a 4k x 4k CCD Camera, *J. Struct. Biol.* 171 (3) (2010) 303–308.
- [82] I. Gutsche, A. Desfosses, G. Effantin, W.L. Ling, M. Haupt, R.W.H. Ruigrok, C. Sachse, G. Schoehn, Near-Atomic Cryo-EM Structure of the Helical Measles Virus Nucleocapsid, *Science* 348 (6235) (2015) 704–707.
- [83] E.L. Hesketh, Y. Meshcheriakova, K.C. Dent, P. Saxena, R.F. Thompson, J.J. Cockburn, G.P. Lomonosoff, N.A. Ranson, Mechanisms of Assembly and Genome Packaging in an RNA Virus Revealed by High-Resolution Cryo-EM, *Nat. Commun.* 6 (2015) 10113.
- [84] D.L. Hurdiss, E.L. Morgan, R.F. Thompson, E.L. Prescott, M.M. Panou, A. Macdonald, N.A. Ranson, New Structural Insights into the Genome and Minor Capsid Proteins of BK Polyomavirus Using Cryo-Electron Microscopy, *Structure* 24 (4) (2016) 528–536.
- [85] D. Sirohi, Z. Chen, L. Sun, T. Klose, T.C. Pierson, M.G. Rossmann, R.J. Kuhn, The 3.8 Å Resolution Cryo-EM Structure of Zika Virus, *Science* 3516 (March) (2016) 1–7.
- [86] V.A. Kostyuchenko, E.X.Y. Lim, S. Zhang, G. Fibriansah, T.-S. Ng, J.S.G. Ooi, J. Shi, S.-M. Lok, Structure of the Thermally Stable Zika Virus, *Nature* 533 (7603) (2016) 425–428.
- [87] X. Zhang, P. Ge, X. Yu, J.M. Brannan, G. Bi, Q. Zhang, S. Schein, Z.H. Zhou, Cryo-EM Structure of the Mature Dengue Virus at 3.5-Å Resolution, *Nat. Struct. Mol. Biol.* 20 (1) (2013) 105–110.
- [88] G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, et al., Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics, *Nature* 497 (7451) (2013) 643–646.
- [89] J.G.A. Aw, Y. Shen, A. Wilm, M. Sun, X.N. Lim, K.-L. Boon, S. Tapsin, Y.-S. Chan, C.-P. Tan, A.Y.L. Sim, et al., In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation, *Mol. Cell* 62 (4) (2016) 603–617.
- [90] S. Welsch, S. Miller, I. Romero-Brey, A. Merz, C.K.E. Bleck, P. Walther, S.D. Fuller, C. Antony, J. Krijnse-Locker, R. Bartschlag, Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites, *Cell Host Microbe* 5 (4) (2009) 365–375.
- [91] J.D. Perlmutter, M.R. Perkett, M.F. Hagan, Pathways for Virus Assembly around Nucleic Acids, *J. Mol. Biol.* 426 (18) (2014) 3148–3165.
- [92] D.E. Alvarez, A.L. De Lella Ezcurrea, S. Fucito, A.V. Gamarnik, Role of RNA Structures Present at the 3'UTR of Dengue Virus on Translation, RNA Synthesis, and Viral Replication, *Virology* 339 (2) (2005) 200–212.
- [93] W.D. Wilson, L. Ratmeyer, M. Zhao, L. Strekowski, D. Boykin, The Search for Structure-Specific Nucleic Acid-Interactive Drugs: Effects of Compound Structure on RNA versus DNA Interaction Strength, *Biochemistry* 32 (15) (1993) 4098–4104.
- [94] D. Ferhadian, M. Contrant, A. Printz-Schweigert, R.P. Smyth, J.C. Paillart, R. Marquet, Structural and Functional Motifs in Influenza Virus RNAs, *Front. Microbiol.* 9 (2018) 559.
- [95] R.G. Huber, X.N. Lim, W.C. Ng, A.Y.L. Sim, H.X. Poh, Y. Shen, S.Y. Lim, K.B. Sundstrom, X. Sun, J.G. Aw, et al., Structure Mapping of Dengue and Zika Viruses Reveals Functional Long-Range Interactions, *Nat. Commun.* 10 (1) (2019) 1408.
- [96] A.P. Russ, S. Lampel, The Druggable Genome: An Update, *Drug Discov. Today* 10 (23–24) (2005) 1607–1610.
- [97] A. Tambe, J. Doudna, L. Pachter, Identifying RNA Contacts from SHAPE-MaP by Partial Correlation Analysis, *arXiv Prepr. arXiv1412.3800* (2014, 20–22.).
- [98] Siegfried, N. a; Busan, S.; Rice, G. M.; Nelson, J. a E.; Weeks, K. M. RNA Motif Discovery by SHAPE and Mutational Profiling (SHAPE-MaP). *Nat. Methods* 2014, 11 (9), 959–965.
- [99] M.J. Smola, G.M. Rice, S. Busan, N.A. Siegfried, K.M. Weeks, Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling (SHAPE-MaP) for Direct, Versatile and Accurate RNA Structure Analysis, *Nat. Protoc.* 10 (11) (2015) 1643–1669.
- [100] K.E. Deigan, T.W. Li, D.H. Mathews, K.M. Weeks, Accurate SHAPE-Directed RNA Structure Determination, *Proc. Natl. Acad. Sci.* 106 (1) (2009) 97–102.
- [101] J.T. Low, K.M. Weeks, SHAPE-Directed RNA Secondary Structure Prediction, *Methods* 52 (2) (2010) 150–158.
- [102] D.M. Mauger, K.M. Weeks, Toward Global RNA Structure Analysis, *Nat. Biotechnol.* 28 (11) (2010) 1178–1179.
- [103] M. Kertesz, Y. Wan, E. Mazor, J.L. Rinn, R.C. Nutter, H.Y. Chang, E. Segal, Genome-Wide Measurement of RNA Secondary Structure in Yeast, *Nature* 467 (7311) (2010) 103–107.
- [104] E.A. Doherty, J.A. Doudna, Ribozyme Structures and Mechanisms, *Annu. Rev. Biochem.* 69 (1) (2000) 597–615.
- [105] L.R. Ganser, M.L. Kelly, D. Herschlag, H.M. Al-Hashimi, The Roles of Structural Dynamics in the Cellular Functions of RNAs, *Nat. Rev. Mol. Cell Biol.* 20 (8) (2019) 474–489.
- [106] J.A. Doudna, Structural Genomics of RNA, *Nat. Struct. Biol.* 7 (11) (2000) 954–956.
- [107] D.H. Mathews, D.H. Turner, Prediction of RNA Secondary Structure by Free Energy Minimization, *Curr. Opin. Struct. Biol.* 16 (3) (2006) 270–278.
- [108] J.S. Reuter, D.H. Mathews, RNAstructure: Software for RNA Secondary Structure Prediction and Analysis, *BMC Bioinformatics* 11 (1) (2010) 129.
- [109] Tafer, H.; Höner zu Siederdissen, C.; Stadler, P. F.; Bernhart, S. H.; Hofacker, I. L.; Lorenz, R.; Flamm, C. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 2011, 6, 26.
- [110] Z. Ouyang, M.P. Snyder, H.Y. Chang, SeqFold: Genome-Scale Reconstruction of RNA Secondary Structure Integrating High-Throughput Sequencing Data, *Genome Res.* 23 (2) (2013) 377–387.
- [111] C.V. Filomatori, M.F. Lodeiro, D.E. Alvarez, M.M. Samsa, L. Pietrasanta, A.V. Gamarnik, A 5' RNA Element Promotes Dengue Virus RNA Synthesis on a Circular Genome, *Genes Dev.* 20 (16) (2006) 2238–2249.
- [112] P. Li, Y. Wei, M. Mei, L. Tang, L. Sun, W. Huang, J. Zhou, C. Zou, S. Zhang, C.-F. Qin, et al., Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity, *Cell Host Microbe* 24 (6) (2018) 875–886.
- [113] Z. Lu, Q.C. Zhang, B. Lee, R.A. Flynn, M.A. Smith, J.T. Robinson, C. Davidovich, A.R. Gooding, K.J. Goodrich, J.S. Mattick, RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure, *Cell* 165 (5) (2016) 1267–1279.
- [114] K.J. Doshi, J.J. Cannone, C.W. Cobaugh, R.R. Gutell, Evaluation of the Suitability of Free-Energy Minimization Using Nearest-Neighbor Energy Parameters for RNA Secondary Structure Prediction, *BMC Bioinformatics* 5 (2004) 105.
- [115] F. Ding, S. Sharma, P. Chalasani, V.V. Demidov, N.E. Broude, N.V. Dokholyan, Ab Initio RNA Folding by Discrete Molecular Dynamics: From Structure Prediction to Folding Mechanisms, *RNA* 14 (6) (2008) 1164–1173.
- [116] K.A. Wilkinson, E.J. Merino, K.M. Weeks, Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution, *Nat. Protoc.* 1 (3) (2006) 1610.
- [117] J.M. Watts, K.K. Dang, R.J. Gorelick, C.W. Leonard, J.W.B. Jr, R. Swanstrom, C.L. Burch, K.M. Weeks, Architecture and Secondary Structure of an Entire HIV-1 RNA Genome, *Nature* 460 (7256) (2009) 711–716.
- [118] S.R. Eddy, Computational Analysis of Conserved RNA Secondary Structure in Transcriptomes and Genomes, *Annu. Rev. Biophys.* 43 (2014) 433–456.
- [119] O. Ziv, M.M. Gabrylska, A.T.L. Lun, L.F.R. Gebert, J. Sheu-Gruttadauria, L.W. Meredith, Z.-Y. Liu, C.K. Kwok, C.-F. Qin, I.J. MacRae, et al., COMRADES Determines In Vivo RNA Structures and Interactions, *Nat. Methods* 15 (10) (2018) 785–788.
- [120] Z. Lu, H.Y. Chang, The RNA Base-Pairing Problem and Base-Pairing Solutions, *Cold Spring Harb. Perspect. Biol.* 10 (12) (2018) a034926.
- [121] M. Baaden, S.J. Marrink, Coarse-Grain Modelling of Protein-Protein Interactions, *Curr. Opin. Struct. Biol.* 23 (6) (2013) 878–886.
- [122] Z. Miao, E. Westhof, RNA Structure: Advances and Assessment of 3D Structure Prediction, *Annu. Rev. Biophys.* 46 (1) (2017) 483–503.
- [123] Ponce-Salvatierra, A.; Astha; Merdas, K.; Nithin, C.; Ghosh, P.; Mukherjee, S.; Bujnicki, J. M. Computational Modeling of RNA 3D Structure Based on Experimental Data. *Biosci. Rep.* 2019, 39 (2), 1–25.
- [124] T. Schlick, A.M. Pyle, Opportunities and Challenges in RNA Structural Modeling

- and Design, *Biophys. J.* 113 (2) (2017) 225–234.
- [125] B. Webb, S. Viswanath, M. Bonomi, R. Pellarin, C.H. Greenberg, D. Saltzberg, A. Sali, Integrative Structure Modeling with the Integrative Modeling Platform, *Protein Sci.* 27 (1) (2018) 245–258.
- [126] D. Russel, K. Lasker, B. Webb, J. Velázquez-Muriel, E. Tjioe, D. Schneidman-Duhovny, B. Peterson, A. Sali, Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies, *PLoS Biol.* 10 (1) (2012) 1–5.
- [127] J. Parsch, J.M. Braverman, W. Stephan, Comparative Sequence Analysis and Patterns of Covariation in RNA Secondary Structures, *Genetics* 154 (2) (2000) 909–921.
- [128] E. Rivas, J. Clements, S.R. Eddy, A Statistical Test for Conserved RNA Structure Shows Lack of Evidence for Structure in LncRNAs, *Nat. Methods* 14 (1) (2016) 45–48.
- [129] S. Lindgreen, P.P. Gardner, A. Krogh, Measuring Covariation in RNA Alignments: Physical Realism Improves Information Measures, *Bioinformatics* 22 (24) (2006) 2988–2995.
- [130] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation, *Science* 367 (6483) (2020) 1260–1263.