



**HAL**  
open science

# Comparative Normal Mode Analysis of the Dynamics of DENV and ZIKV Capsids

Yin-Chen Hsieh, Frédéric Poitevin, Marc Delarue, Patrice Koehl

► **To cite this version:**

Yin-Chen Hsieh, Frédéric Poitevin, Marc Delarue, Patrice Koehl. Comparative Normal Mode Analysis of the Dynamics of DENV and ZIKV Capsids. *Frontiers in Molecular Biosciences*, 2016, 3, pp.85. 10.3389/fmolb.2016.00085 . pasteur-02174598

**HAL Id: pasteur-02174598**

**<https://pasteur.hal.science/pasteur-02174598>**

Submitted on 5 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Comparative Normal Mode Analysis of the Dynamics of DENV and ZIKV Capsids

Yin-Chen Hsieh<sup>1</sup>, Frédéric Poitevin<sup>2,3</sup>, Marc Delarue<sup>4</sup> and Patrice Koehl<sup>1\*</sup>

<sup>1</sup> Department of Computer Science and Genome Center, University of California, Davis, Davis, CA, USA, <sup>2</sup> Department of Structural Biology, Stanford University, Stanford, CA, USA, <sup>3</sup> SLAC National Accelerator Laboratory, Stanford PULSE Institute, Menlo Park, CA, USA, <sup>4</sup> Unit of Structural Dynamics of Macromolecules, UMR 3528 du Centre National de la Recherche Scientifique, Institut Pasteur, Paris, France

Key steps in the life cycle of a virus, such as the fusion event as the virus infects a host cell and its maturation process, relate to an intricate interplay between the structure and the dynamics of its constituent proteins, especially those that define its capsid, much akin to an envelope that protects its genomic material. We present a comprehensive, comparative analysis of such interplay for the capsids of two viruses from the *flaviviridae* family, Dengue (DENV) and Zika (ZIKV). We use for that purpose our own software suite, DD-NMA, which is based on normal mode analysis. We describe the elements of DD-NMA that are relevant to the analysis of large systems, such as virus capsids. In particular, we introduce our implementation of simplified elastic networks and justify their parametrization. Using DD-NMA, we illustrate the importance of packing interactions within the virus capsids on the dynamics of the E proteins of DENV and ZIKV. We identify differences between the computed atomic fluctuations of the E proteins in DENV and ZIKV and relate those differences to changes observed in their high resolution structures. We conclude with a discussion on additional analyses that are needed to fully characterize the dynamics of the two viruses.

**Keywords:** proteins, normal modes, elastic network models, viruses, Dengue, Zika

## 1. INTRODUCTION

A major goal of molecular biology is to understand at the atomic level the functions of macromolecules and/or biological nano-machines, which are believed to be intimately related to the dynamics of their three-dimensional structures and especially their collective degrees of freedom (Koehl, 2014; Bahar et al., 2015). Our current understanding of the dynamics of macromolecules is, however, largely incomplete. This arises because only a few experimental techniques are capable of collecting time-resolved structural data, and those that can collect those data are usually limited to a narrow time window (Fromme, 2015). Similarly, state-of-the-art computational methods are limited in scope (usually in the microsecond time-scale), because of limitations in computing power (Fengand et al., 2015).

An alternate and promising approach to molecular dynamics is to infer dynamics from static structures corresponding to locally stable states (Mahajan and Sanejouand, 2015), together with reliable coarse-graining approaches to bridge the time-scale gap (Saunders and Voth, 2013; López-Blanco and Chacón, 2016). Cartesian Normal Modes, for example, represent a class of movements around a local energy minimum that are both straightforward to calculate and biologically relevant

## OPEN ACCESS

### Edited by:

Slavica Jonic,  
Université Pierre et Marie Curie,  
France

### Reviewed by:

Yves-henri Sanejouand,  
Unité Fonctionnalité et Ingénierie des  
Protéines (CNRS), France  
Xavier Siebert,  
University of Mons, Belgium

### \*Correspondence:

Patrice Koehl  
koehl@cs.ucdavis.edu

### Specialty section:

This article was submitted to  
Biophysics,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 25 October 2016

**Accepted:** 12 December 2016

**Published:** 27 December 2016

### Citation:

Hsieh Y-C, Poitevin F, Delarue M and  
Koehl P (2016) Comparative Normal  
Mode Analysis of the Dynamics of  
DENV and ZIKV Capsids.  
*Front. Mol. Biosci.* 3:85.  
doi: 10.3389/fmolb.2016.00085

(Noguti and Go, 1982; Brooks et al., 1983; Levitt et al., 1985). The low-frequency part of the spectrum of normal modes is often associated with functional transitions, for instance, between two known states of the same macromolecule such as its apo (ligand-free) or holo (bound) form. The Elastic Network Model (ENM), introduced by Tirion in 1996, offers a particularly simple and efficient way to calculate these modes, allowing fast access to the collective dynamics of large complexes with no minimization issues as it enforces that the crystal structure is already at the energy minimum (Tirion, 1996). This model was later expanded as the Gaussian Network Model (Bahar et al., 1997) and the Anisotropic Network Model (Hinsen, 1998; Tama et al., 2000; Atilgan et al., 2001), which were shown to describe conformational changes remarkably well (Tama and Sanejouand, 2001; Delarue and Sanejouand, 2002; Zheng and Doniach, 2003; Mahajan and Sanejouand, 2015).

During the past few years, several web-servers performing on-line Normal Mode Analysis (NMA) have been set up and described: ElNemo (Suhre and Sanejouand, 2004), ENCoM (Frappier et al., 2015), Webnm@ (Tiwari et al., 2014), ANM 2.0 (Eyal et al., 2015), AD-ENM (Zheng and Doniach, 2003), NMSim (Kruger et al., 2012). We have extended and updated our own server, NOMAD-REF (Lindahl et al., 2006), with a new and user-friendlier interface, including a better visual representation of the results while at the same time enlarging the performances of the core calculation of Normal Modes in the framework of the ENM representation. New features include (i) a wider array of coarse-graining levels prior to the actual building of the ENM, and (ii) variants of the ENM that are based on a cutoff-free Delaunay tessellation of the set of atoms of the molecule of interest. With these features we depart from the original Elastic Network Model (Tirion, 1996), but keep most of its salient features, as the construction of the original Tirion Elastic Model remains available. We found for example that the Elastic Network coming from a Delaunay tessellation correctly handles PDB models with isolated domains and/or dangling ends (Xia et al., 2014). In addition, the performance of the calculation of Normal Modes has been improved to a point where it can deal with 100,000 atoms routinely, making it possible, for instance, to deal with entire virus capsids without having to resort to a symmetry-specific implementation (Simonson and Perahia, 1992; van Vlijmen and Karplus, 2005; Peeters and Taormina, 2009).

In the present paper, we show an application of some of the tools implemented in DD-NMA, the updated version of NOMAD-REF, to study the dynamics of viruses of the *flaviviridae* family, namely of Dengue virus and Zika virus.

Dengue virus (DENV) is a positive-sense RNA virus responsible for dengue fever, a tropical infectious disease whose incidence has increased drastically over the last decades, for which no prophylactic treatments are known (with the exception of eliminating the vector, i.e., mosquitoes). Today, about 3.9 billion people, or 50% of the world's population, live in areas where there is a risk of dengue transmission (Brady et al., 2012). Dengue is endemic in at least 128 countries in Asia, the Pacific, the Americas, Africa, and the Caribbean (Brady et al., 2012). The World Health Organization (WHO) estimates that close to

390 million infections occur yearly, of which 96 million manifest clinically (Bhatt et al., 2013). DENV is recognized as a potential threat to public health in the USA (Morens and Fauci, 2008). Of similar concerns are the recent outbreaks of ZIKA virus (ZIKV), another *flaviviridae* virus similar to DENV. The current ZIKV epidemic in the Americas is linked to a sudden increase in the reported cases of congenital microcephaly and Guillain Barré syndrome. This led the World Health Organization (WHO) in February 2016 to declare a "public health emergency of international concern" (WHO, 2016). As no treatments currently exist for the consequences of infections with those two viruses, and as their incidence is only expected to increase, basic research on their infection mechanisms becomes highly significant.

*Flaviviridae* genomes encode for ten different proteins, three structural proteins that form the virus particle, and seven non-structural (NS) proteins that are involved in its replication (for recent review see Meng et al., 2015). Structures of all four serotypes of DENV (Perera and Kuhn, 2008 and references therein; Zhang et al., 2012; Kostyuchenko et al., 2013, 2014; Fibriansah et al., 2015) and recently two structures of the same ZIKV strain have been published (Kostyuchenko et al., 2016; Sirohi et al., 2016). Those structures show the same global architecture, with their capsids having icosahedral symmetry consisting of 60 units, with each unit containing three copies of the E protein and three copies of protein M. The E protein is known to play a central role in many parts of the virus life cycle (Perera and Kuhn, 2008). A perhaps surprising idea that has emerged from years of studies of viruses is that their biology is deeply encoded in the dynamics of these proteins. Significant structural dynamics has been shown to occur during infection cycles, both at the level of individual proteins and at the quaternary structure level of the viral particle. These dynamics can be blocked by antibody binding (Lok et al., 2008; Teoh et al., 2012; Fibriansah et al., 2015). In addition, while the overall geometry of the viral capsid is identical in all those viruses and only small differences are observed at a finer structural scale, significant differences in stability are observed between those viruses. For example, while infection with DENV is significantly affected by temperature, infection with ZIKV remains constant at even relatively high temperatures (Kostyuchenko et al., 2016). To better understand differences between those two viruses, we investigate the dynamics of their capsid E proteins. We study those proteins independently, as well as the impact of packing imposed by the icosahedral arrangement of the virus capsid. We explore whether the differences observed, if any, are consistent with the differences observed between the structures of the capsids of DENV and ZIKV and their biological activities.

The paper is organized as follows. In the next section, we describe normal mode analysis (NMA) in the context of the Elastic Network Model. We provide an overview of the theory and discuss the different options for choosing its parameters, namely the choice of coarse-graining level, the choice of the elastic force constants, and the cutoff for selecting the pairs of atoms that belong to the ENM. In the following section, we provide a description of the algorithms used to implement NMA within our new server DD-NMA, with a special focus on scalability to large molecular systems. In the Results section, we

discuss the applications of DD-NMA to study the dynamics of DENV and ZIKV, focusing on the differences and similarities of the dynamics of their capsid E protein. We conclude the paper with a brief discussion on future developments of normal mode analysis applied to viral structures.

## 2. NORMAL MODE ANALYSIS

### 2.1. Normal Mode Analysis Based on the Tirion Elastic Network Model

The Elastic Network Model (ENM) was originally introduced by Tirion (1996). It is a model that captures the geometry of the molecule of interest in the form of a network of inter-atomic connections, linked together with elastic springs. Two categories of normal mode analyses based on ENMs are widely used today, namely, the Gaussian Network Model (GNM) (Bahar et al., 1997; Haliloglu et al., 1997) and the anisotropic network model (ANM) (Tirion, 1996; Atilgan et al., 2001). Here we follow the latter model, in which the energy of the molecule is equated to the harmonic energy associated with these springs. This defines a quadratic energy on the inter-atomic distances. Let  $M$  be a biomolecule containing  $N$  atoms, with atom  $i$  characterized by its position  $X_i = (x_i, y_i, z_i)$ . The whole molecule is then described by a  $3N$  position vector  $X$ . For two atoms  $i$  and  $j$  of  $M$ , we set  $r_{ij} = |X_i - X_j|$  and  $r_{ij}^0 = |X_i^0 - X_j^0|$  to be their Euclidean distances in any conformation  $X$  and in the ground-state conformation  $X^0$  (usually the X-ray structure), respectively. The total potential  $V_{ENM}$  of the biomolecule is then set to:

$$V_{ENM}(X) = \frac{1}{2} \sum_{i=1}^N \sum_{j>i} k_{ij} (r_{ij} - r_{ij}^0)^2 \Theta(R_c - r_{ij}^0) \quad (1)$$

where  $R_c$  is a cutoff distance,  $k_{ij}$  is the force constant of the "spring" formed by the pair of atoms  $i$  and  $j$ , and  $\Theta(x)$  is the Heaviside unit step function, i.e.,  $\Theta(x) = 0$  if  $x < 0$  and  $\Theta(x) = 1$  otherwise. Both  $R_c$  and  $k_{ij}$  are discussed in detail below.

In the normal mode framework, the potential  $V_{ENM}$  is then approximated with a second-order Taylor expansion in the neighborhood of the ground state  $X^0$ :

$$V_{ENM}(X) \approx V_{ENM}(X^0) + \nabla V_{ENM}(X^0)^T (X - X^0) + \frac{1}{2} (X - X^0)^T H (X - X^0) \quad (2)$$

where  $\nabla V_{ENM}$  and  $H$  are the gradient and Hessian of  $V_{ENM}$ , respectively. Note that based on Equation 1,  $V_{ENM}(X^0) = 0$  and  $\nabla V_{ENM}(X^0)$  is the null vector (i.e.,  $X^0$  is a global minimum of  $V_{ENM}$  by definition). The ENM energy is then simply

$$V_{ENM}(X) \approx \frac{1}{2} (X - X^0)^T H (X - X^0) \quad (3)$$

The  $3 \times 3$  submatrix  $H_{ij}$  of the Hessian  $H$  corresponding to two atoms  $i$  and  $j$  that are in contact is given by:

$$H_{ij} = -\frac{k_{ij}}{(r_{ij}^0)^2} (X_i - X_j)(X_i - X_j)^T \\ = -\frac{k_{ij}}{(r_{ij}^0)^2} \begin{bmatrix} (x_i - x_j)^2 & (x_i - x_j)(y_i - y_j) & (x_i - x_j)(z_i - z_j) \\ (y_i - y_j)(x_i - x_j) & (y_i - y_j)^2 & (y_i - y_j)(z_i - z_j) \\ (z_i - z_j)(x_i - x_j) & (z_i - z_j)(y_i - y_j) & (z_i - z_j)^2 \end{bmatrix} \quad (4)$$

and the  $3 \times 3$  submatrix  $H_{ii}$  on the diagonal of  $H$  is then given by:

$$H_{ii} = -\sum_{j=1, N} H_{ij} \quad (5)$$

In Cartesian coordinates, the equations of motion defined by the potential  $V_{ENM}$  are derived from Newton's equation:

$$\frac{d^2 X}{dt^2} = -H(X - X^0) \quad (6)$$

Writing the solution to this equation as a linear sum of intrinsic motions (the "normal modes" of the system),

$$X_j = \sum_{k=k_0}^{3N} A_{jk} \alpha_k \cos(\omega_k t + \delta_k) \quad (7)$$

we get a standard eigenvalue problem,

$$HA = MA\Omega \quad (8)$$

The eigenfrequencies  $\omega$  are given by the elements of the diagonal matrix  $\Omega$ , namely  $\omega_i^2 = \Omega(i, i)$ . The eigenvectors are the columns of the matrix  $A$ , and the amplitudes and phases,  $\alpha_k$  and  $\delta_k$ , are determined by initial conditions. The matrix  $M$  is a diagonal matrix containing the masses of the atoms. We note that the first six eigenvalues in  $\Omega$  are equal to 0, as they correspond to global translations and rotations of the biomolecule. To characterize the internal motions of the biomolecule, the sum in Equation 8 runs then from  $k_0 = 7$  up to  $3N$ , the number of degrees of freedom of the system.

### 2.2. Parametrization: Choosing the Representation of the Molecule

The first requirement when building an ENM is to define the set of atoms on which it is based. Although all atoms could be used, it appears natural to lower the dimensionality of the system, namely "coarse-graining," when large biomolecules are considered, or in the context of a harmonic approximation to its energy as is the case in ENM (Tozzini, 2005). Coarse-grained models have long been used for studying protein folding and aggregation. They enable the exploration of large length scales and time scales that are usually inaccessible to all-atom models in explicit solvent (Saunders and Voth, 2013; Kmiecik et al., 2016). Combined with enhanced configuration search methods, these simplified models with various levels of granularity offer the possibility

to determine equilibrium structures and to compare folding kinetics and thermodynamics quantities with the corresponding values obtained by experimental techniques. In their pioneer work from 1976, Levitt and Warshel (1976) developed the foundation of coarse-graining for protein folding. They were able to fold the 58-residue BPTI protein within 6.5 Å from its experimental structure using a two-bead representation for each residue in the protein. This representation included the  $C\alpha$  and the centroid of the side chain to define a residue. They used an effective implicit solvent force field such that the atoms of the solvent need not be considered explicitly, and successive minimizations and normal mode thermalization to fold BPTI. Since then, various levels of granularity have been developed, from lattice representations to multi-bead representations, and from single atom to multiple-atom residue-level representations (Kmieciak et al., 2016). The positions of those beads are either defined by known atoms (usually the  $C\alpha$ ), or by fitting to capture the dynamics of the full molecular systems (Zhang et al., 2008, 2011; Li et al., 2016). For all the analyses of virus structures considered in this paper, we used the  $C\alpha$ -only representation.

### 2.3. Parametrization: Choosing the Spring Force Constants

In the original ENM introduced by Tirion, the elastic constants  $k_{ij}$  are set to be the same for all pairs of atoms. In other models,  $k_{ij}$  vary for different pairs of atoms. For example, Ming and Wall (2005) employed an enhanced ENM in which the interactions of neighboring  $C\alpha$  atoms on the backbone were strengthened to reproduce the correct bimodal distribution of density-of-states from an all-atom model. Kondrashov et al. (2006) used a strategy in which they classified residue interactions into several categories corresponding to different physical properties. The elastic constants can also be adjusted to have the fluctuations of the atoms of the molecule computed from the equations of motions given by Equation (7) to match the atomic fluctuations captured experimentally and usually reported as B-factors. Many methods have been developed for that purpose (see for example Xia et al., 2013, 2014 and references therein). Among those methods, the one proposed by Erman (2006) is worth discussing. Erman developed an iterative algorithm to update the Kirchhoff matrix of a Gaussian Network Model, in which the connections of neighboring  $C\alpha$  atoms on the backbone of the protein of interest are fixed, and the strengths of the interactions between pairs of residues are varied until a good fit between experimental B-factors and computed fluctuations is obtained. While this approach generates a really good fit between those two representations of fluctuations, a significant number of the optimized spring force constants are found to be negative. While such negative values are not forbidden, they do hint at the possibility of overfitting. This is in accordance with (Fuglebakk et al., 2013), who recently suggested that such a refinement procedure leads to overfitting, and not to a better dynamic model for the molecule. As such, in this study we assign the same value for all  $k_{ij}$ , following the initial ENM of Tirion (1996).

### 2.4. Parametrization: The Cutoff Parameter $R_c$

In standard implementations, the cutoff distance  $R_c$  and the force constant  $k$  are set constant for all pairs of residues. Their values, however, differ between the two models. For example, the cutoff distance  $R_c$  for GNM is usually set in the range of 7 to 8 Å (Kundu et al., 2002) while in ANM larger values are usually considered in the range from 13 to 15 Å (Eyal et al., 2006). There are, however, no guidelines as to which values are best and sometimes different implementations lead to contradicting optimal values. To circumvent these discrepancies, several authors have proposed to include all pairs of residues in a protein and to assign different force constants to their corresponding springs that relate to their lengths at rest (see for example Hinsen, 1998; Kovacs et al., 2004; Yang et al., 2009). In these methods, the use of a plain cutoff distance is avoided. The number of pairs of atoms considered, however, is large and scales as  $N^2$ , where  $N$  is either the total number of atoms in the biomolecule considered, or its number of residues. Such a quadratic behavior makes these methods unfit for studying large systems. To study the capsids of DENV and ZIKA, we have considered a traditional cutoff ENM, with the cutoff set to 14 Å, unless specifically noted.

## 3. MATERIALS AND METHODS

We have used our own software package, DD-NMA, to perform all the analyses discussed in the Results section. In the following, we highlight some of the elements of DD-NMA that are relevant to the analysis of large systems. We note that DD-NMA is available as a web-based service at <http://lorentz.dynstr.pasteur.fr/suny/index.php?id0=deLaunaynma#welcome>.

### 3.1. An Efficient Algorithm to Diagonalize the Hessian of the Elastic Potential $V_{ENM}$

The Hessian matrix of  $V_{ENM}$  is a  $3N \times 3N$  symmetric, real-valued matrix whose elements are described by Equation (4). The theory described above calls for diagonalizing this matrix, as its eigenvalues and eigenvectors provide the frequencies and directions, respectively, of the normal modes of the molecular systems under study. While many methods exist for solving such an eigenvalue problem, see Golub and van der Vorst (2000), many of those methods break down when  $N$  becomes large, both in terms of computing time and memory requirements. The Hessian matrix is highly sparse as only a subset of all atom pairs are usually considered (see previous section for a discussion of this point), but this is not enough to offset the computing requirements as the matrix  $A$  of eigenvectors is usually not sparse. However, in her original paper, Tirion (1996) had recognized that the lowest frequency normal modes can capture most of the dynamics of the protein of interest. This observation has since been supported by further evidence that the lowest-frequency normal modes generated from ENM conform with conformational changes observed by X-ray and NMR experiments (Kim et al., 2002; Maragakis and Karplus, 2005; Kurkcuglu et al., 2009) as well as with the results of MD

simulations (Rueda et al., 2007; Orellana et al., 2010; Leioatts et al., 2012). While it is unclear as to how many of those low frequency normal modes need to be considered (Petroni and Pande, 2006), it remains that only a small fraction of the eigenvalues and eigenvectors of the Hessian matrix need to be computed, which leads to the opportunity to use powerful iterative algorithms for computing those quantities. The most successful family of such algorithms is based on the efficient Krylov subspace method, as it allows for targeting only a subset of the eigenvalue spectrum of a matrix. We provide below the rationale behind this method to compute the eigenvalues with lowest magnitude of the Hessian matrix.

An intuitive method for finding the largest eigenvalue of a given  $N \times N$  matrix  $A$  is the power iteration. Starting with an initial random vector  $x$ , this method calculates  $Ax$ ,  $A^2x$ ,  $A^3x$ , ... iteratively, storing and normalizing the result into  $x$  at every iteration. The corresponding sequence of Rayleigh quotient  $R_i$

$$R_i = \frac{x^T A^i x}{x^T x} \quad (9)$$

converges to the largest eigenvalue of  $A$ , while  $x$  itself converges to the corresponding eigenvector. However, much potentially useful computation is wasted by using only the final result. This suggests that, instead, the so-called Krylov matrix is to be formed:

$$\mathcal{K}_n = [x \ Ax \ A^2x \ \dots \ A^{n-1}x] \quad (10)$$

The columns of this matrix are not orthogonal, but an orthogonal basis can be constructed via a stabilized Gram–Schmidt orthogonalization. The resulting vectors are a basis of the Krylov subspace,  $\mathcal{K}_n$ . The vectors of this basis give good approximations of the eigenvectors corresponding to the  $n$  largest eigenvalues of  $A$ . In a similar manner, the smallest eigenvalues of  $A$  can be computed by applying this strategy to either  $A^{-1}$ , or by applying a spectral shift, i.e., by computing the largest eigenvalues of  $A - \lambda_{max}I$ , where  $\lambda_{max}$  is the largest eigenvalue of  $A$ .

We use the ARPACK implementation of a variant of this approach, the implicitly restarted Arnoldi iteration method (Lehoucq et al., 1998).

### 3.2. Atomic Fluctuations Computed From Normal Modes

From the normal modes of the ENM, it is possible to compute the mean square fluctuations of the positions of the atoms according to:

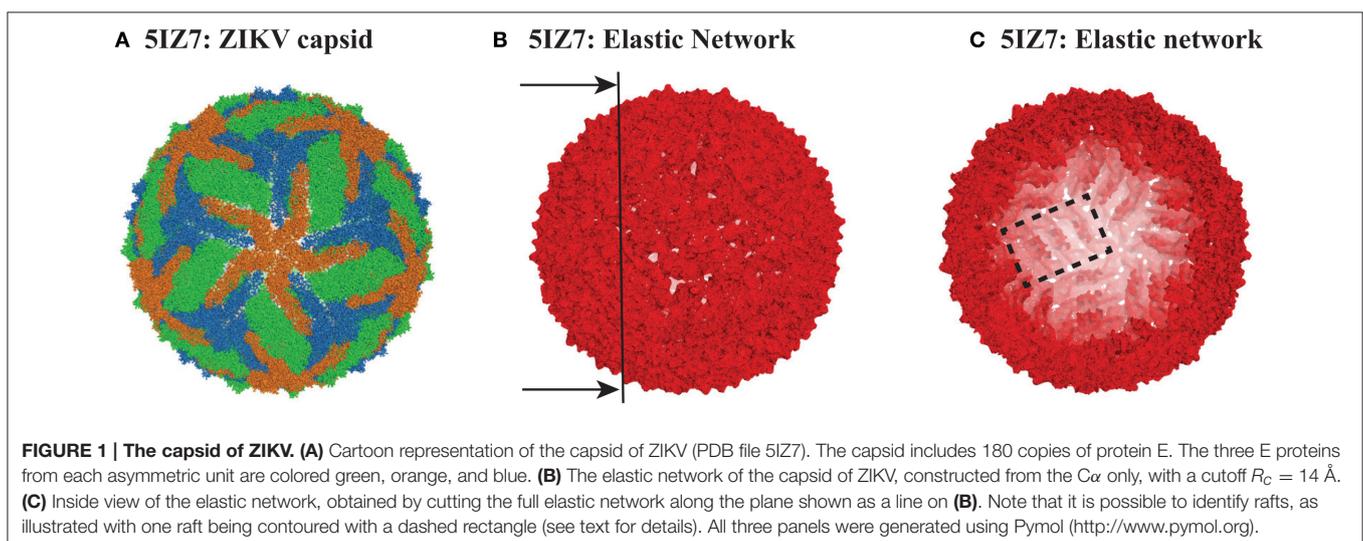
$$\langle \Delta \mathbf{X}_i^2 \rangle = \frac{k_B T}{m_i} \sum_{k=7}^m \frac{A_{ik}^2}{\omega_k^2} \quad (11)$$

where  $\Delta \mathbf{X}_i$  and  $m_i$  are the displacement vector and mass of vertex  $i$ , respectively,  $k_B$  is the Boltzmann's constant,  $T$  is the temperature considered,  $A_{ij}$  is the  $i$ -th component of the  $j$  eigenvector  $A_j$  of the Hessian, and  $\omega_i$  is its associated eigenvalue. The summation should run over all the modes of the system (excluding the six modes for rigid body transformations); it is truncated here to the first  $m = 100$  modes, as those low frequency modes are usually responsible of most of the atomic fluctuations (see above).

### 3.3. Correlated Motions Within a Biomolecule

The Boltzmann distribution for the approximate potential of the ENM (see Equation 3) is described by a multivariate Gaussian distribution with a covariance matrix proportional to the inverse of the Hessian  $H$ . Because of the six rigid motions captured by the six normal modes with 0 frequencies, the inverse of  $H$  is in fact not properly defined. We can, however, compute a pseudo-inverse by ignoring those zero energy modes; this pseudo-inverse can be regarded as a covariance matrix of internal deformation:

$$C = \sum_{k=7}^M \frac{1}{\omega_k^2} A_k A_k^T \quad (12)$$



where  $\omega_k$  and  $A_k$  are the  $k$ -th eigenvalues and eigenvectors, respectively. Note that  $C$  is a  $3N \times 3N$  matrix. The summation extends from  $k = 7$ , the first non-zero mode, to  $M$ , the highest mode considered (up to  $3N$ ). To obtain a scalar quantification of the correlation of the motions of two atoms  $i$  and  $j$ , a correlation matrix  $P$  is computed, following Ichiye and Karplus (1991):

$$P_{ij} = \frac{\text{tr}(C_{ij})}{\sqrt{\text{tr}(C_{ii})\text{tr}(C_{jj})}} \quad (13)$$

The values  $P_{ij}$  range from  $-1$  to  $+1$ , with a negative correlation value indicating an anticorrelated motion, and a positive correlation value identifying a correlated pattern of dynamics between the two atoms considered. These values are stored into a cross-correlation matrices CCM that is used to visualize correlations of motion within the molecule under study.

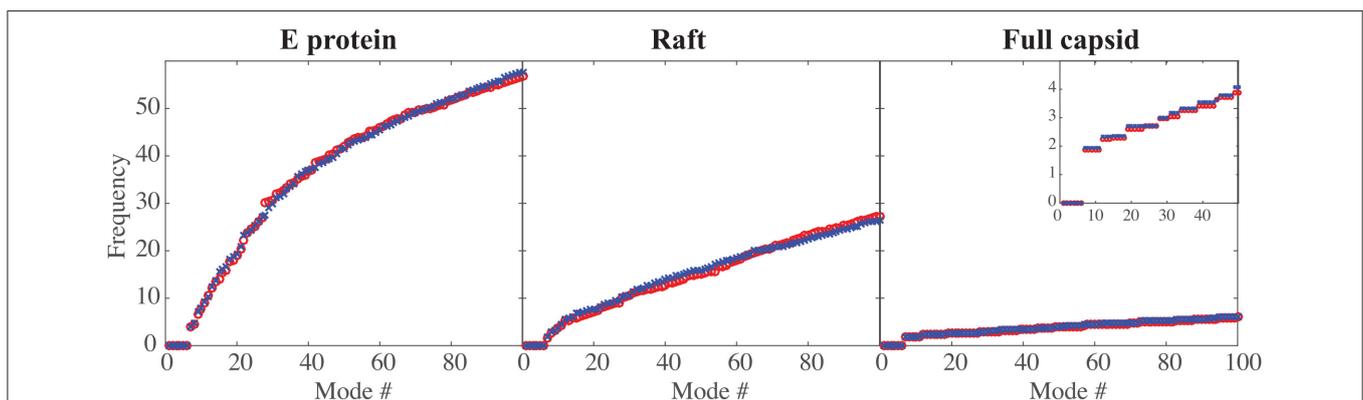
## 4. RESULTS AND DISCUSSION

DENV and ZIKV are both members of the *flaviviridae* family. DENV serotype 1 and ZIKV (which are the focus of this study) share 53% sequence identity (Kostyuchenko et al., 2016). Their particles share a common fold, with their capsids having icosahedral symmetry. Those capsids are formed of 60 asymmetrical units, with each unit containing three copies of E protein (495 and 504 residues in DENV and ZIKV, respectively) and three copies of the membrane protein M (74 and 75 residues in DENV and ZIKV, respectively). The high resolution cryo-EM structures of all four serotypes of DENV, as well as the structure of one strain of ZIKV, are available in the Protein Data Bank (Zhang et al., 2012; Kostyuchenko et al., 2013, 2014; Fibriansah et al., 2015; Kostyuchenko et al., 2016; Sirohi et al., 2016). Here we focus on the structure of the mature form of serotype 1 of DENV, with PDB code 4CCT (Kostyuchenko et al., 2013), and of the equivalent mature form of ZIKV, as given by one of the recently published structures, with PDB code 5IZ7 (Kostyuchenko et al., 2016). Those two structures were shown to be very similar,

with only small differences that will be discussed in light of their dynamics. A cartoon representation of ZIKV is given in **Figure 1A**. The DENV capsid shows the same architecture.

The PDB file for 4CCT only contains  $C\alpha$  atoms. For consistency, we used  $C\alpha$  only representations of 4CCT and 5IZ7. We isolated from those two files all the  $C\alpha$  atoms of the viral capsid. For both viruses, we considered E protein in four different environments: isolated, MONO, (corresponding to chain A in the asymmetric unit of 4CCT and chain B of the asymmetric unit of 5IZ7), within the asymmetric unit, UNIT, within a raft, RAFT, and within the whole capsid structure, FULL. The corresponding complexes MONO, UNIT, RAFT, and FULL contain 495, 1707, 3414, and 102420 residues for 4CCT, respectively, and 504, 1737, 3474, and 104220 residues for 5IZ7, respectively. We generated elastic networks for all those eight complexes using a cutoff procedure, with the cutoff set to 14 Å. **Figures 1B,C** illustrate the elastic network for the FULL complex for ZIKV (5IZ7). We note that this elastic network follows the surface of the capsid virus and does not include any edges that cross the interior of the capsid; this is a direct consequence of the cutoff that is used. The inside of the geometric structure formed by the elastic network reveals the presence of rafts (one such raft is shown inside a rectangle in **Figure 1C**), namely three dimers of E protein lying parallel to each other. Once the elastic networks were established, we computed the hundred lowest normal modes for each of them, using the procedure detailed in the Methods section.

We emphasize that the elastic networks for the full capsids were computed using the empty protein shells, following previous studies of viral particles using ENM and their normal modes (Tama and Brooks III, 2002, 2005; Kim et al., 2003; Chennubotla et al., 2005; Rader et al., 2005; Polles et al., 2013). This setting is expected to be satisfactory as the stability of the empty capsid is guaranteed by the geometric construction of the ENM, which makes up for the missing stabilizing interactions of the coat proteins and RNA. We note that the latter were not resolved in the cryo-EM structures we considered.



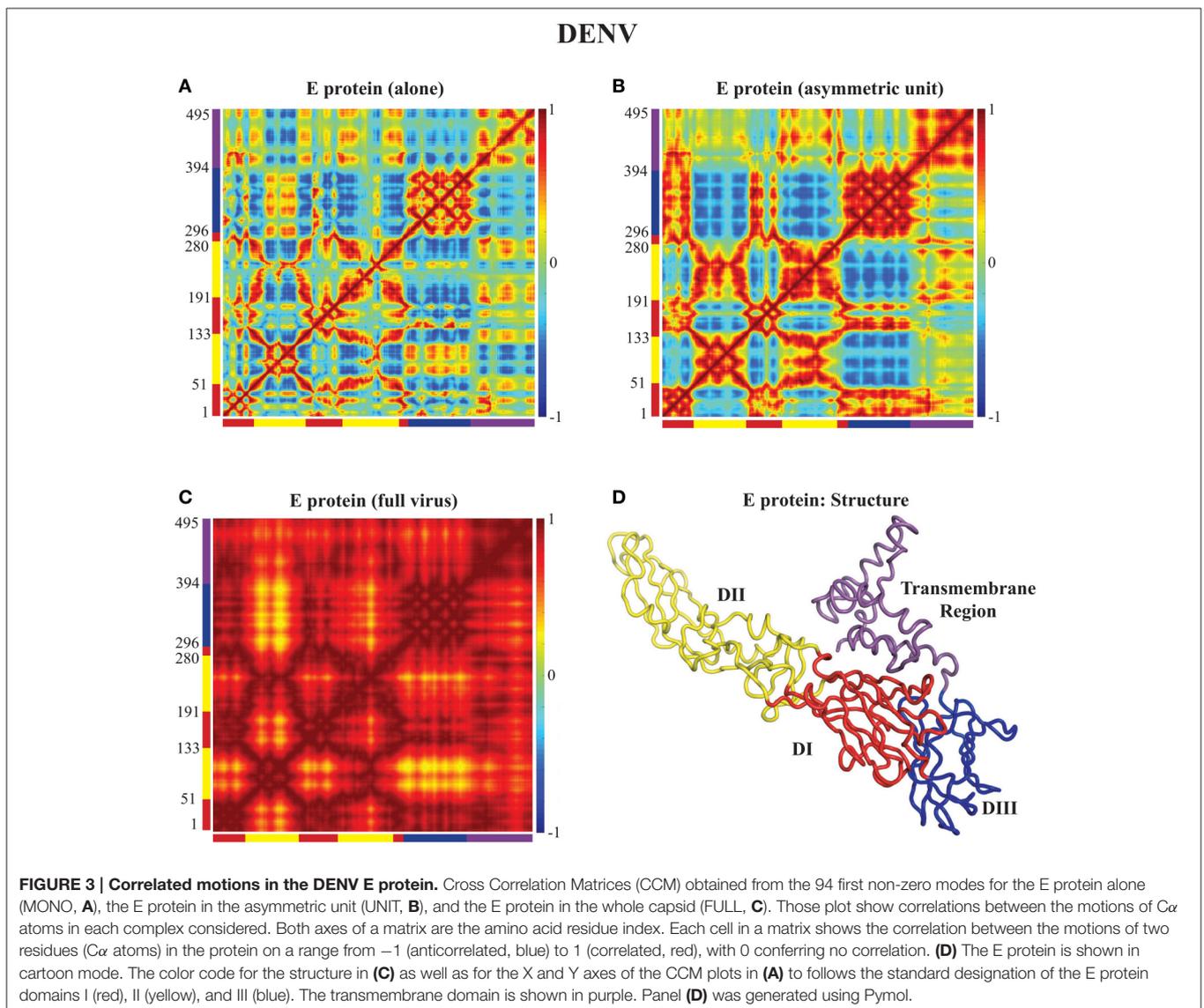
**FIGURE 2 | Comparing the low frequencies of the normal modes of DENV and ZIKA.** The frequencies of the first hundred normal modes of DENV (red circle, o) and ZIKV (blue cross, x) are plotted against the normal mode index (#), for the E protein by itself (left), for a raft (middle), and for the full capsid (right). The frequencies are in arbitrary units, as the force constants are also in arbitrary units. Note the decrease in the amplitude of those frequencies as the size of the complex increases. The insert in the right panel shows an enlargement for the first 50 normal modes; it highlights the degeneracy of the normal modes for a full capsid.

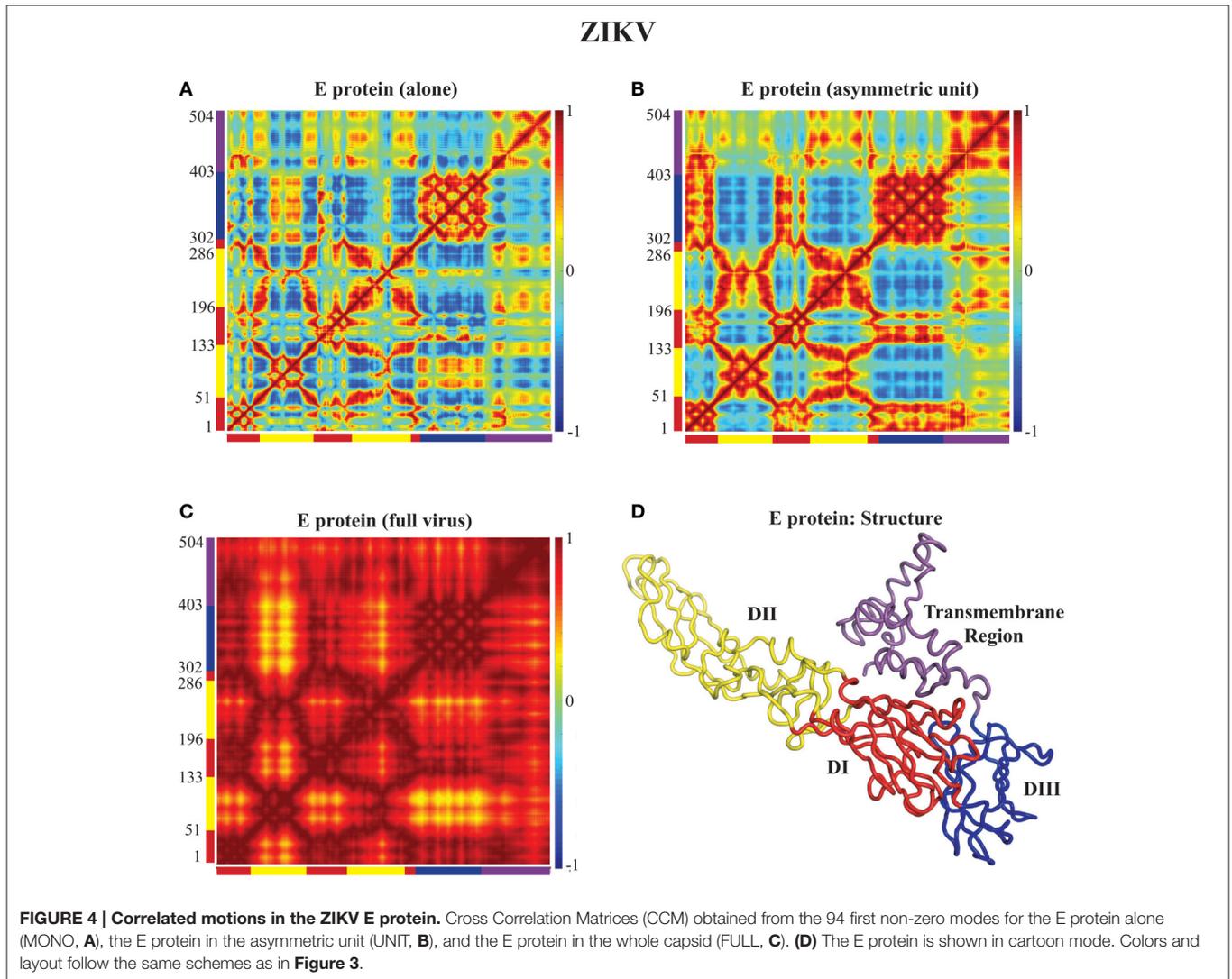
#### 4.1. Characterizing the Low Frequency Normal modes of DENV and ZIKV

In **Figure 2** we compare the frequencies of the first hundred normal modes of the MONO, RAFT, and FULL complexes of DENV and ZIKV. As expected, the first six frequencies are found equal to zero, for all complexes considered, as those frequencies correspond to the rigid motions (three translations and three rotations). The larger the protein complex, the more the spectra of frequencies of the normal modes are shifted toward lower frequencies, indicating the presence of more collective motions in protein oligomers. The spectra of frequencies for the full capsids reveal the presence of degeneracy, namely repeating frequencies, that correspond to symmetries in the capsid. All the differences observed in the three complexes are conserved between DENV and ZIKV. We note also the nearly perfect match between the normal mode frequency spectra of the two viruses.

#### 4.2. Correlated Dynamics of E Proteins in the Capsids of DENV and ZIKV

In **Figures 3, 4** we assess the extent to which packing influences the dynamics of the E protein of DENV and ZIKV, respectively. For both viruses, the cross correlation matrices (CCM) for E protein vary significantly between the MONO, UNIT, and FULL complexes. The CCM for the E protein alone reveals significant positive correlations within each of the three domains I, II, and III. Domains II and III exhibit both positive and negative correlations in their atomic fluctuations, while the motions of domain I are only weakly correlated to the motions of domain II and III. When the dynamics of the E protein are studied in the context of the asymmetric unit, the same positive correlations are observed within each of the three domains. The interactions between the domains change significantly, however. In the UNIT complex, the dynamics of domain II are strongly anticorrelated to the dynamics of domain III, while domain





I is correlated positively with domain III. In the full viral capsid, the internal dynamics of the E protein remain mostly as observed in the asymmetric unit. The only difference is the addition of a global positive correlation over the full protein that comes from concerted dynamics within the capsid. In all three oligomeric states, the transmembrane domain shows weak positive correlation with domain II.

All the differences in dynamics observed between isolated E proteins and E proteins in the whole capsid are conserved between DENV and ZIKV.

### 4.3. Correlated Dynamics of Rafts of E Proteins in the Capsids of DENV and ZIKV

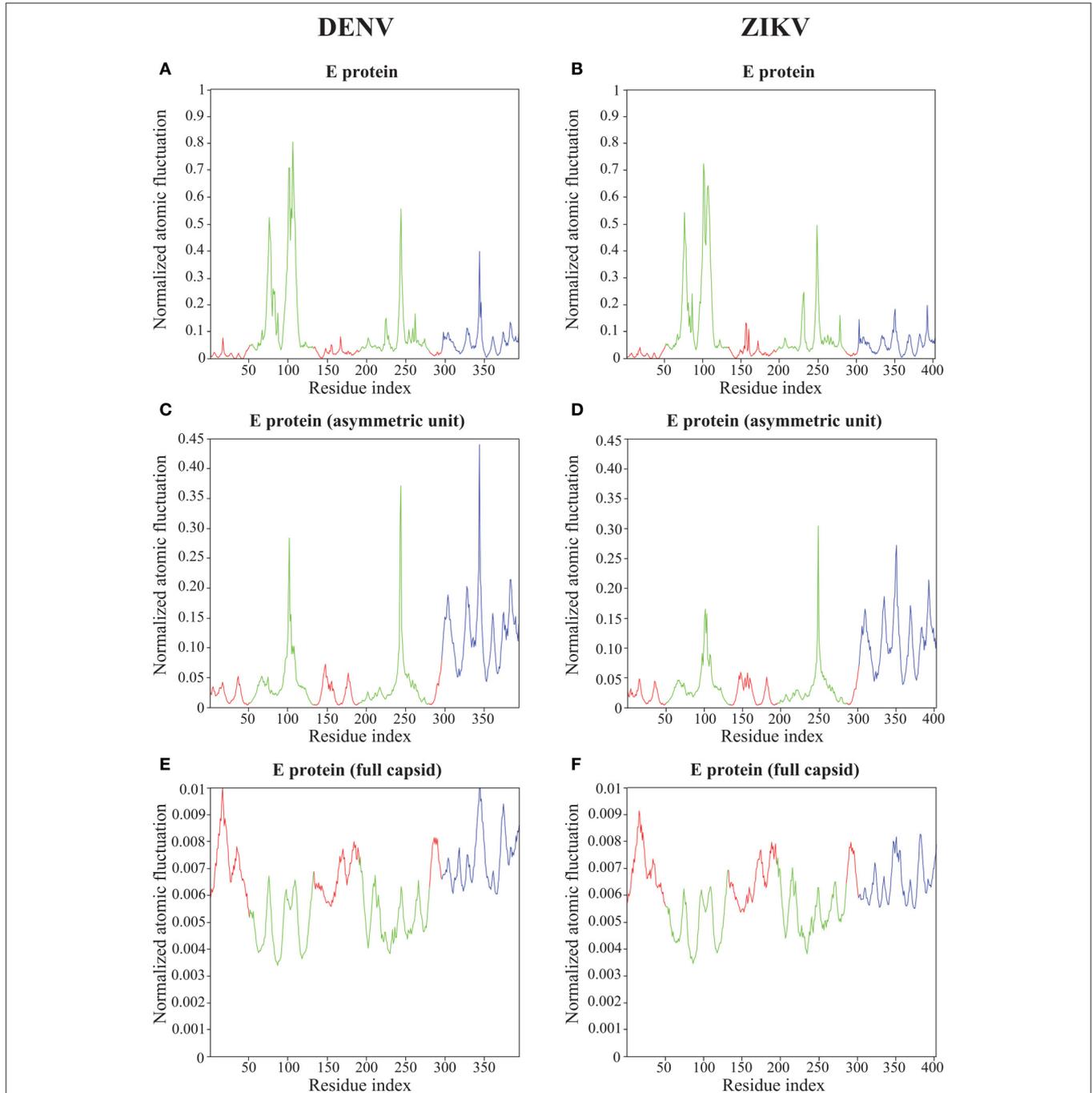
**Figures 3, 4** reveal the effects of packing in the viral capsid on the dynamics of one E protein. We performed a similar analysis on a larger structure of the capsid, namely a raft. A raft is formed from six E proteins forming 3 dimers arranged in a parallel manner, resulting from the combination of two asymmetrical units (see **Figure 5E**). The whole capsid contains 30 such rafts. In **Figures 5A–C**, we assess the extent to which

packing influences the dynamics of such rafts for both DENV and ZIKV. In the CCM for the raft alone (**Figures 5A,B** for DENV and ZIKV, respectively) we clearly identify the six E proteins along the diagonal. Each of those E proteins exhibits dynamics correlation patterns equivalent to those observed in the E protein when it is in the asymmetrical unit. The interactions between the E proteins are consistent with the structure of the raft. The first E proteins of the two asymmetrical units, proteins E1A and E1B, show strong positively correlated dynamics. Those two proteins form a dimer in the raft. In contrast, proteins E2A and E3A in Unit A, and proteins E2B and E3B in Unit B have a pattern of interactions that include both positively correlated and negatively correlated motions, depending on their domains: for example, domains III have negative correlations between the two proteins, while domains II are positively correlated between the two proteins. The pair of proteins (E2A, E3A) shows weak correlated dynamics with the pair of proteins (E2B, E3B), with a chessboard pattern (i.e., positive correlations between E2A and E2B, and negative correlations between E3A and E3B).



The CCMs for a raft included in the whole capsid (**Figures 5C,D** for DENV and ZIKV, respectively) reveal different patterns than those described for the raft alone, highlighting again the impact of packing in the virus environment. There is a high level of positive correlation of motions within each of the

units A and B. The proteins E1A and E1B that form a dimer at the center of the raft are mostly interacting with themselves in the raft alone, while they show strong levels of positive correlations with all three E proteins of the opposing unit in the raft when considered within the whole capsid. In contrast, the pairs of



**FIGURE 6 | Atomic fluctuations in the DENV and ZIKV E proteins.** The atomic displacement fluctuations obtained from the 94 first non-zero modes for the E protein alone (MONO, **A,B**), the E protein in the asymmetric unit (UNIT, **C,D**), and the E protein in the whole capsid (FULL, **E,F**) are plotted as a function of the residue number for both DENV (PDB file 4CCT) and ZIKV (PDB file 5IZ7). The Y axis represents normalized displacements (see text for details). The color code follows the standard designation of the E protein domains for domains I (red) and III (blue), while domain II has been colored green to enhance visibility.

proteins (E2A, E3A) and (E2B, E3B) present significantly lower correlation when considered in the whole capsid compared to the raft alone. Such a behavior would favor concentration of concerted internal motions in a few E protein dimers at the center of the rafts in the whole viral capsid instead of a more uniform spread of concerted motions.

Similar to the findings for the dynamics of the E proteins, the differences in dynamics observed between isolated rafts and rafts in the whole capsid are conserved between DENV and ZIKV.

#### 4.4. Atomic Fluctuations within the E Proteins of the Capsids of DENV and ZIKV

The normalized squared atomic fluctuations for each  $C\alpha$  atom in the E protein of DENV and ZIKV were calculated as the sum of their displacements along the first 94 non-zero modes, weighted by the reciprocal of the eigenvalues, as given by Equation (11). For both viruses, the calculation was performed in three states for the E protein, namely the MONO, UNIT, and FULL complexes described above. The absolute values of the amplitudes of the fluctuations computed using Equation (11) are somewhat arbitrary, as they depend on the parametrization of the elastic network, namely on the cutoff values  $R_c$  and the strength of the force constants  $k_{ij}$ . While it is possible to select those parameters such that a good fit is obtained between the computed fluctuations and experimental B-factors, we prefer not to, following the advice of Fuglebakk et al. (2013) that warned on possible overfitting problems. Instead, we just normalize the computed fluctuations for an atom  $i$  using:

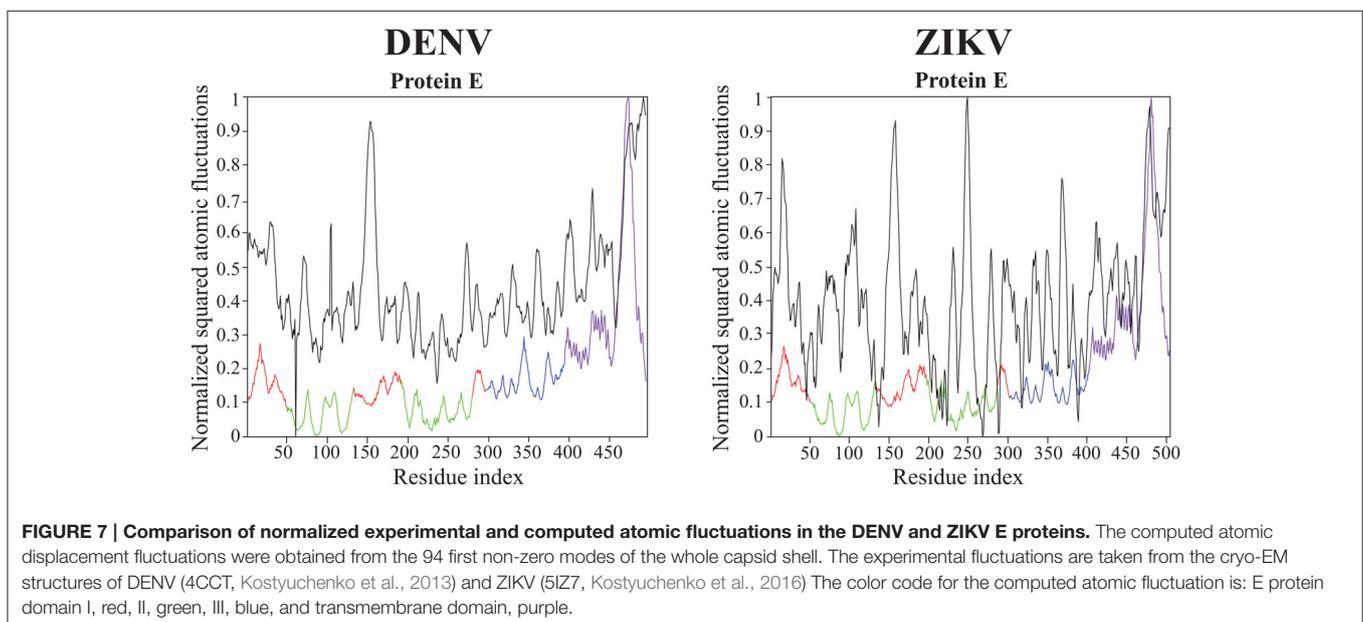
$$\langle \Delta_N X_i^2 \rangle = \frac{\langle \Delta X_i^2 \rangle - \min(\langle \Delta X^2 \rangle)}{\max(\langle \Delta X^2 \rangle) - \min(\langle \Delta X^2 \rangle)} \quad (14)$$

where the min and max values are computed over all  $C\alpha$  atoms of the molecule considered. To enable comparison, we computed

the min and max values from the fluctuations observed in the E protein alone, and applied those to normalize the fluctuations of all three states considered, i.e., MONO, UNIT, and FULL. Results for DENV and ZIKV are shown in **Figure 6**.

Not unexpectedly, the amplitude of the atomic fluctuations within the E protein decreases as the protein is more constrained, from a (normalized) range between 0 and 1 in the E protein alone (**Figures 6A,B**), to a range between 0 and 0.01 in the full capsid (**Figures 6E,F**). Of significance is the change in dynamics observed in the kl-loop between domains I and II (the DI-DII hinge, residues 280–290) between the stand alone E protein and the capsid. In the former, this loop region is predicted to be rigid, while in the latter it is found to be significantly more dynamic. This hinge is thought to be important to flip the domain DII to expose the fusion loop during the fusion event (Modis et al., 2003; Zhang et al., 2004; Kostyuchenko et al., 2016). In contrast, the HI-loop in the putative receptor binding domain DIII (residues 230–240) is found to be more dynamic in the E protein alone than in the whole capsid. DENV and ZIKV show the same dynamical behavior in both loops (the kl- and HI-loops).

The two plots showing the atomic fluctuations computed from normal modes in the E proteins are globally similar between DENV and ZIKV in all oligomeric states (**Figure 6**). There are, however, some localized differences that are worth discussing. There is a putative increase in dynamics in the region 150–160 in ZIKV compared to DENV that is most marked in the E protein monomer, but still present in its oligomeric states. This region corresponds to the Glycan loop, which contains a glycosylation site (Asn154 in ZIKV and Asn153 in DENV). It was found to be the region with the biggest structural differences (up to 6 Å) in the cryo-EM structures of ZIKV and DENV (Sirohi et al., 2016). Our calculations were performed in the absence of the sugar moieties on the Asparagine. We believe however that our results highlight an intrinsic difference in the dynamics of the



**FIGURE 7 | Comparison of normalized experimental and computed atomic fluctuations in the DENV and ZIKV E proteins.** The computed atomic displacement fluctuations were obtained from the 94 first non-zero modes of the whole capsid shell. The experimental fluctuations are taken from the cryo-EM structures of DENV (4CCT, Kostyuchenko et al., 2013) and ZIKV (5IZ7, Kostyuchenko et al., 2016). The color code for the computed atomic fluctuation is: E protein domain I, red, II, green, III, blue, and transmembrane domain, purple.

Glycan loops of DENV and ZIKV that is worth exploring further. In contrast to the Glycan loop, the region 340–350 is found to be less dynamic in ZIKV than in DENV in all oligomeric states of their E proteins. This region corresponds to the C strand and CD loop in domain III. Based on the differences in the structures of the DENV and ZIKV capsids, Kostyuchenko et al. (2016) hypothesized that the presence of an additional amino acid in the C strand in ZIKV was responsible for a rearrangement of the structure locally that is possibly responsible for the increased thermal stability of ZIKV. Our results indeed suggest a more rigid C strand in ZIKV compared to DENV. The exact relationship between this decrease in atomic fluctuations and thermal stability is unknown.

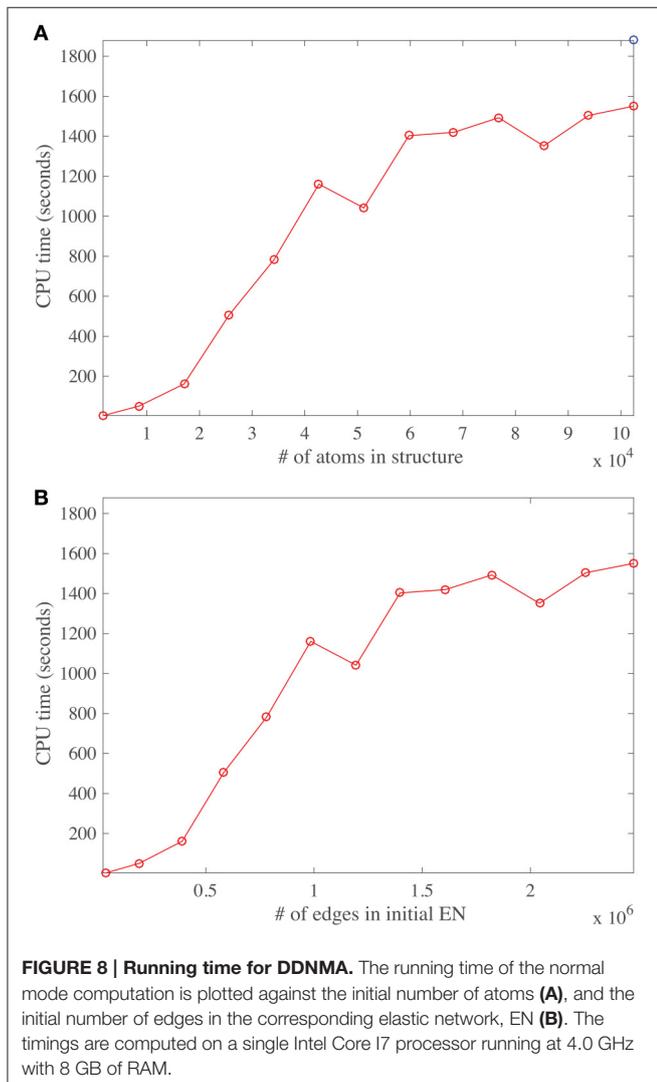
All results on dynamics presented above are based on atomic fluctuations and dynamic correlations computed from normal mode analyses. In **Figure 7** we compare those normalized computed atomic fluctuations for the  $C\alpha$  atoms of the E protein in the full capsid structure with the corresponding normalized experimental B-factors extracted from the PDB files 4CCT and

5IZ7 for DENV and ZIKV, respectively. Overall, the profiles show qualitative similarities over the full range of residues in E protein. The correlation coefficients between the experimental B-factors for DENV and ZIKV and the computed atomic fluctuations are 0.58 and 0.45, respectively. Those values are modest. We note that it would be possible to obtain significantly better correlations if the elastic constants  $k_{ij}$  assigned to the links of the networks were fitted to improve the match between B-factors and computed fluctuations. We also notice differences in relative amplitudes of the experimental and computed atomic fluctuations; these differences exist, however, between the experimental B factors for the two viruses and they could not be interpreted when analyzing the differences between the corresponding structures (Sirohi et al., 2016; Kostyuchenko et al., 2016).

#### 4.5. Computing Time

The main task performed by DD-NMA when computing the normal modes of an elastic network is the diagonalization of the Hessian. For large systems, it is not feasible to perform the full diagonalization, both because of its time and memory complexities (both of order  $O(N^3)$ , where  $N$  is the number of atoms). Instead, only partial diagonalization is performed, with only the eigenvalues with the lowest amplitudes (usually 100) being computed. The method implemented is based on an iterative procedure. As discussed in the Material and Methods section, this procedure is efficient, of order  $O(Mk + Nk^2 + k^3)$ , where  $M$  is the number of non-zero elements in the sparse representation of the Hessian matrix, and  $k$  the number of eigenvalues that are computed. The first term corresponds to the matrix vector multiplications needed at each iteration, the second term relates to the Gram-Schmidt orthogonalization required to build the Krylov basis, and the last term is the cost of diagonalizing the matrix representing this basis. To test if we observe this expected behavior on real systems, we have experimented with systems of varying size. We have applied DD-NMA on parts of the capsids of DENV, with increasing number of asymmetrical units included, from one to sixty. For all systems, we extracted the  $C\alpha$  atoms, computed an all-atom elastic network with a cutoff of 14Å, and computed the 100 lowest frequency normal modes with DD-NMA. All those experiments were performed on a iMAC Apple computer with a 4.0 GHz Intel Core I7 processor, with 8 GB of memory. The computing times for DD-NMA are plotted against the initial numbers of atoms and edges in the all-atom elastic networks in **Figure 8**.

The number of non-zero elements in the Hessian matrix is directly proportional to the number of edges in the elastic network and implicitly proportional to the number of atoms in the protein, assuming constant density of atoms. Interestingly, the curves cpu time vs. number of atoms and vs. number of edges show three different regimes. For a relatively small number of atoms (below 20,000), and for a medium number of atoms (between 20,000 and 40,000), the cpu time is found to vary linearly, as expected, but with different slopes. The different slopes come from the relative weights of the two terms  $Mk$  and  $Nk^2$  in the time complexity. For larger number of atoms, the behavior of the cpu time is found to be more erratic, with a slower rate of increase. We suspect that this behavior is due to



cache issue. The time complexity of computing the product of the Hessian with a vector using the sparse representation of the Hessian is more complex than just being proportional to  $M$ , the number of non-zero elements of the Hessian  $H$ . Indeed, for very large matrices, it depends on their storage patterns. We have not tried to optimize this storage, which is most likely the reason for the erratic behavior. It does hint to possible improvement in the computation of the normal modes, by first re-ordering the Hessian using for example METIS (Karypis and Kumar, 1999).

We note that it takes approximately 30 min to compute the first hundred normal modes for a molecular system with hundred thousand atoms, on a single core, on a desktop computer. While this is not fast *per se*, it is still manageable. We do note that part of the codes for computing the eigenvalues of the Hessian can be parallelized; we are currently working on such an improvement.

## 5. SUMMARY AND CONCLUSIONS

Understanding the dynamics of viral capsids is of fundamental interest for modeling the key steps of viral life cycles. In this paper, we have described an implementation of normal mode analysis based on elastic network models that enables such analyses. This implementation is based on the known foundations in the domain (Tirion, 1996) and does not deviate significantly from other available implementations (Zheng and Doniach, 2003; Suhre and Sanejouand, 2004; Kruger et al., 2012; Tiwari et al., 2014; Eyal et al., 2015; Frappier et al., 2015). We discuss in details its parametrization, namely the choice of the coarse graining of the molecular system, the choice of the method for computing the elastic network, and the assignment of force constants to the resulting springs, and justify the choices we have implemented. We emphasize the need for efficient and robust algorithms for computing the normal modes of elastic networks, in particular when those networks include a very large number of nodes -in the hundred of thousands-, such as those derived for virus capsids. We have illustrated the application of our method to study the dynamics of the viral capsids of DENV serotype 1 and ZIKV. We have characterized the impact of the packing imposed by the capsids on their E proteins that play essential roles in receptor binding and fusion to the membrane of the host cells. We have identified differences in the atomic fluctuations of these proteins between DENV serotype 1 and ZIKV that are consistent with the structural differences observed using high resolution cryo-EM experimental structures (Kostyuchenko et al., 2016; Sirohi et al., 2016). In the future, we will consider two types of extensions of this first study that relate to the method itself as well as to its specific application to studying DENV and ZIKV.

First, we recognize that the need for a reasonable computational cost, when applying a method such as normal mode analysis to a large molecular system such as a virus capsid, implies that some sort of coarse graining is applied to such a system. Many options exist to reduce the dimensionality of the problem by selecting subsets of atoms, “beads,” to represent the system (Kmieciak et al., 2016). The positions of those beads are either defined by known atoms (usually the  $C\alpha$ ), or by fitting to capture the dynamics of the full molecular system (Zhang

et al., 2008, 2011; Li et al., 2016). The main difficulty in coarse-graining, however, is to design potential energy functions or force fields that retain the physics of the all-atom explicit solvent system in terms of structure, thermodynamics and dynamics (Riniker et al., 2012). While significant efforts have been made to guarantee that a coarse-grained model and its potential capture the complexity of the all-atom molecular system (Riniker et al., 2012; Saunders and Voth, 2013; Na et al., 2015; Zhang, 2015), we note that much less has been done to generate a true multi-scale representation of this system, i.e., to define a hierarchy of coarse-grained models with a coupling between those models. Our intention is to generate such a hierarchy; for this purpose, we will rely on the concept of renormalization group (RG) that is well known in physics (Wilson, 1975). We have implemented in DD-NMA a beta-version of such a method that performs iterative decimation of an elastic network. We will test this method on viral capsids once we have adapted the code to deal with hundreds of thousands of atoms.

Once the representation of the molecular system is chosen, the elastic network is defined as a set of links, with a link between two residues only if the distance between their  $C\alpha$  atoms is smaller than a given cutoff. As an alternative to this cutoff model, Xia et al. (2014) proposed to use all edges of the Delaunay triangulation of the selected atoms as an alternate elastic network. We believe that the use of Delaunay triangulation to define the ENM extends the range of applicability of NMA to the realm of less globular proteins. We will proceed in this direction and test this alternate definition of ENM to study the dynamics of viruses.

Our analyses of the dynamics of DENV and ZIKV capsids were based on naked, empty shells. There are many opportunities to extend this work. We are interested in generating plausible paths between different conformations of the virus capsids, such as the “breathing” induced by increase of temperature (Fibriansah et al., 2013), and the changes observed during the maturation of the virus. We will develop new methods to find such plausible paths in very large systems such as viral capsids, where “plausible” refers to a path with minimal frustration, also defined as the Minimum Action Path (MAP) (Olender and Elber, 1996; Eastman et al., 2001; Franklin et al., 2007; Vanden-Eijnden and Heymann, 2008; Zhou et al., 2008; Chandrasekaran et al., 2016). Finally, we plan to study the impact of glycosylation of the E protein and/or antibody binding on the virus capsids onto their dynamical properties.

## AUTHOR CONTRIBUTIONS

YH and FP performed research and analyzed data. MD and PK designed research and analyzed data. PK wrote the manuscript.

## ACKNOWLEDGMENTS

PK acknowledges support from the Ministry of Education of Singapore through Grant Number: MOE2012-T3-1-008. FP acknowledges support from a Bioinfo:BipBip grant from the Agence Nationale de la Recherche (France) while he was at the Pasteur Institute.

## REFERENCES

- Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80, 505–515. doi: 10.1016/S0006-3495(01)76033-X
- Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Fold. Design* 2, 173–181. doi: 10.1016/S1359-0278(97)00024-2
- Bahar, I., Cheng, M. H., Lee, J. Y., Kaya, C., and Zhang, S. (2015). Structure-encoded global motions and their role in mediating protein-substrate interactions. *Biophys. J.* 109, 1101–1109. doi: 10.1016/j.bpj.2015.06.004
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., et al. (2013). The global distribution and burden of Dengue. *Nature* 496, 504–507. doi: 10.1038/nature12060
- Brady, O. J., Gething, P., Bhatt, S., Messina, J., Brownstein, J., Hoen, A., et al. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.* 6:e1760. doi: 10.1371/journal.pntd.0001760
- Brooks, B., Bruccoleri, R., and Olafson, B. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4, 187–217. doi: 10.1002/jcc.540040211
- Chandrasekaran, S. N., Dhas, J., Dokholyan, N. V., and Carter, C. W. Jr. (2016). A modified path algorithm rapidly generates transition states comparable to those found by other well established algorithms. *Struct. Dyn.* 3:012101. doi: 10.1063/1.4941599
- Chennubotla, C., Rader, A. J., Yang, L. W., and Bahar, I. (2005). Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* 2, S173–S180. doi: 10.1088/1478-3975/2/4/S12
- Delarue, M., and Sanejouand, Y. H. (2002). Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.* 320, 1011–1024. doi: 10.1016/S0022-2836(02)00562-4
- Eastman, P., Gronbech-Jensen, N., and Doniach, S. (2001). Simulation of protein folding by reaction path annealing. *J. Chem. Phys.* 114:3823. doi: 10.1063/1.1342162
- Erman, B. (2006). The gaussian network model: precise prediction of residue fluctuations and application to binding problems. *Biophys. J.* 91, 3589–3599. doi: 10.1529/biophysj.106.090803
- Eyal, E., Lum, G., and Bahar, I. (2015). The anisotropic network model web server at 2015 (ann 2.0). *Bioinformatics* 31, 1487–1489. doi: 10.1093/bioinformatics/btu847
- Eyal, E., Yang, L. W., and Bahar, I. (2006). Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 22, 2619–2627. doi: 10.1093/bioinformatics/btl448
- Fengand, H., Costauoc, R., Darve, E., and Izaguirre, J. A. (2015). A comparison of weighted ensemble and markov state model methodologies. *J. Chem. Phys.* 142:214113. doi: 10.1063/1.4921890
- Fibriansah, G., Ng, T. S., Kostyuchenko, V. A., Lee, J., Lee, S., Wang, J., et al. (2013). Structural changes in dengue virus when exposed to a temperature of 37°. *J. Virol.* 87, 7585–7592. doi: 10.1128/JVI.00757-13
- Fibriansah, G., Tan, J. L., Smith, S. A., de Alwis, R., Ng, T. S., Kostyuchenko, V. A., et al. (2015). A highly potent human antibody neutralizes dengue virus serotype 3 by binding across three surface proteins. *Nat. Comm.* 6:6341. doi: 10.1038/ncomms7341
- Franklin, J., Koehl, P., Doniach, S., and Delarue, M. (2007). Minactionpath: maximum likelihood trajectory for large-scale structural transitions in a coarse grained locally harmonic energy landscape. *Nucl. Acids. Res.* 35, W477–W482. doi: 10.1093/nar/gkm342
- Frappier, V., Chartier, M., and Najmanovich, R. (2015). ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. *Nucl. Acids Res.* 43, W395–W400. doi: 10.1093/nar/gkv343
- Fromme, P. (2015). XFELs open a new era in structural chemical biology. *Nat. Chem. Biol.* 11, 895–899. doi: 10.1038/nchembio.1968
- Fuglebak, E., Reuter, N., and Hinsen, K. (2013). Evaluation of protein elastic network models based on an analysis of collective motions. *J. Chem. Theory Comput.* 9, 5618–5628. doi: 10.1021/ct400399x
- Golub, G., and van der Vorst, H. (2000). Eigenvalue computation in the 20th century. *J. Comput. Applied Math.* 123, 35–65. doi: 10.1016/S0377-0427(00)00413-1
- Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79, 3090–3093. doi: 10.1103/PhysRevLett.79.3090
- Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins Struct. Func. Genet.* 33, 417–429.
- Ichiye, T., and Karplus, M. (1991). Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct. Func. Genet.* 11, 205–217. doi: 10.1002/prot.340110305
- Karypis, G., and Kumar, V. (1999). A fast and highly quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.* 20, 359–392. doi: 10.1137/S1064827595287997
- Kim, M. K., Jernigan, R. L., and Chirikjian, G. S. (2002). Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* 83, 1620–1630. doi: 10.1016/S0006-3495(02)73931-3
- Kim, M. K., Jernigan, R., and Chirikjian, G. (2003). An elastic network model of hk97 capsid maturation. *J. Struct. Biol.* 143, 107–117. doi: 10.1016/S1047-8477(03)00126-6
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chem. Rev.* 116, 7898–7936. doi: 10.1021/acs.chemrev.6b00163
- Koehl, P. (2014). Mathematics role in the grand challenge of deciphering the molecular basis of life. *Front. Mol. Biosci.* 1:2. doi: 10.3389/fmolb.2014.00002
- Kondrashov, D. A., Cui, Q., and Phillips, G. N. Jr. (2006). Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. *Biophys. J.* 91, 2760–2767. doi: 10.1529/biophysj.106.085894
- Kostyuchenko, V. A., Chew, P. L., Ng, T. S., and Lok, S. M. (2014). Near-atomic resolution cryo-electron microscopic structure of dengue serotype 4 virus. *J. Virol.* 88, 477–482. doi: 10.1128/JVI.02641-13
- Kostyuchenko, V. A., Lim, E. X., Zhang, S., Fibriansah, G., Ng, T. S., Ooi, J. S., et al. (2016). Structure of the thermally stable zika virus. *Nature* 533, 425–428. doi: 10.1038/nature17994
- Kostyuchenko, V. A., Zhang, Q., Tan, J. L., Ng, T. S., and Lok, S. M. (2013). Immature and mature dengue serotype 1 virus structures provide insight into the maturation process. *J. Virol.* 83:7700–7707. doi: 10.1128/JVI.00197-13
- Kovacs, J., Chacon, P., and Abagyan, R. (2004). Predictions of protein flexibility: first-order measures. *Proteins* 54, 661–668. doi: 10.1002/prot.20151
- Krüger, D. M., Ahmed, A., and Gohlke, H. (2012). NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucl. Acids Res.* 40, W310–W316. doi: 10.1093/nar/gks478
- Kundu, S., Melton, J. S., Sorenson, D. C., and Phillips, G. N. Jr. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J.* 83, 723–732. doi: 10.1016/S0006-3495(02)75203-X
- Kurkcuoglu, O., Turgut, O. T., Cansu, S., Jernigan, R. L., and Doruker, P. (2009). Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.* 97, 1178–1187. doi: 10.1016/j.bpj.2009.06.009
- Lehoucq, R., Sorensen, D., and Yang, C. (1998). *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. Philadelphia, PA: SIAM. doi: 10.1137/1.9780898719628
- Leioatts, N., Romo, T. D., and Grossfield, A. (2012). Elastic network models are robust to variations in formalism. *J. Chem. Theory Comput.* 8, 2424–2434. doi: 10.1021/ct3000316
- Levitt, M., Sander, C., and Stern, P. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181, 423–447. doi: 10.1016/0022-2836(85)90230-X
- Levitt, M., and Warshel, A. (1976). Computer simulation of protein folding. *Nature* 253, 694–698.
- Li, M., Zhang, J. Z., and Xia, F. (2016). A new algorithm for construction of coarse-grained sites of large biomolecules. *J. Comp. Chem.* 37, 795–804. doi: 10.1038/253694a0
- Lindahl, E., Azuara, C., Koehl, P., and Delarue, M. (2006). NORMAnDRef: visualization, deformation, and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucl. Acids. Res.* 34, W52–W56. doi: 10.1093/nar/gkl082
- Lok, S. M., Kostyuchenko, V., Nybakken, G. E., Holdaway, H. A., Sukupolvi-Petty, A. B. S., Sedlak, D., et al. (2008). Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nat. Struct. Mol. Biol.* 15, 312–317. doi: 10.1038/nsmb.1382
- López-Blanco, J. R., and Chacón, P. (2016). New generation of elastic network models. *Curr. Opin. Struct. Biol.* 37, 46–53. doi: 10.1016/j.sbi.2015.11.013

- Mahajan, S., and Sanejouand, Y.-H. (2015). On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch. Biochem. Biophys.* 567, 59–65. doi: 10.1016/j.abb.2014.12.020
- Maragakis, P., and Karplus, M. (2005). Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J. Mol. Biol.* 352, 807–822. doi: 10.1016/j.jmb.2005.07.031
- Meng, F., Badierah, R. A., Almehdar, H. A., Redwan, E. M., Kurgan, L., and Uversky, V. N. (2015). Unstructural biology of the dengue virus proteins. *FEBS J.* 282, 3368–3394. doi: 10.1111/febs.13349
- Ming, D., and Wall, M. (2005). Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95:198103. doi: 10.1103/PhysRevLett.95.198103
- Modis, Y., Ogata, S., Clements, D., and Harrison, S. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6986–6991. doi: 10.1073/pnas.0832193100
- Morens, D. M., and Fauci, A. S. (2008). Dengue and hemorrhagic fever. A potential threat to public health in the United States. *JAMA* 299, 214–216. doi: 10.1001/jama.2007.31-a
- Na, H., Jernigan, R. L., and Song, G. (2015). Bridging between nma and elastic network models: preserving all-atom accuracy in coarse-grained models. *PLoS Comput. Biol.* 11:e1004542. doi: 10.1371/journal.pcbi.1004542
- Noguti, T., and Go, N. (1982). Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature* 296, 776–778. doi: 10.1038/296776a0
- Olender, R., and Elber, R. (1996). Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* 105, 9299–9315. doi: 10.1063/1.472727
- Orellana, L., Rueda, M., Ferrer-Costa, C., Lopez-Blanco, J. R., Chacon, P., and Orozco, M. (2010). Approaching elastic network models to molecular dynamics flexibility. *J. Chem. Theory Comput.* 6, 2910–2923. doi: 10.1021/ct100208e
- Peeters, K., and Taormina, A. (2009). Group theory of icosahedral virus capsid vibrations: a top-down approach. *J. Theor. Biol.* 256, 607–624. doi: 10.1016/j.jtbi.2008.10.019
- Perera, R., and Kuhn, R. (2008). Structural proteomics of dengue virus. *Curr. Opin. Microbiol.* 11, 369–377. doi: 10.1016/j.mib.2008.06.004
- Petrone, P., and Pande, V. (2006). Can conformational change be described by only a few normal modes? *Biophys. J.* 90, 1583–1593. doi: 10.1529/biophysj.105.070045
- Polles, G., Indelicato, G., Potesio, R., Cermelli, P., Twarock, R., and Micheletti, C. (2013). Mechanical and assembly units of viral capsids identified via quasi-rigid domain decomposition. *PLoS Comput. Biol.* 9:e1003331. doi: 10.1371/journal.pcbi.1003331
- Rader, A. J., Vlad, D. H., and Bahar, I. (2005). Maturation dynamics of bacteriophage HK97 capsid. *Structure* 13, 413–421. doi: 10.1016/j.str.2004.12.015
- Riniker, S., Allison, J. R., and van Gunsteren, W. F. (2012). On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* 14, 12423–12430. doi: 10.1039/c2cp40934h
- Rueda, M., Chacon, P., and Orozco, M. (2007). Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* 15, 565–575. doi: 10.1016/j.str.2007.03.013
- Saunders, M. G., and Voth, G. A. (2013). Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* 42, 73–93. doi: 10.1146/annurev-biophys-083012-130348
- Simonson, T., and Perahia, D. (1992). Normal modes of symmetric protein assemblies. application to the tobacco mosaic virus protein disk. *Biophys. J.* 61, 410–427. doi: 10.1016/S0006-3495(92)81847-7
- Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T., Rossmann, M., et al. (2016). The 3.8 Å resolution cryo-em structure of zika virus. *Science* 352, 467–470. doi: 10.1126/science.aaf5316
- Suhre, K., and Sanejouand, Y.-H. (2004). Elnémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucl. Acids Res.* 32, W610–W614. doi: 10.1093/nar/gkh368
- Tama, F., and Brooks III, C. L. (2002). The mechanism and pathway of ph induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.* 318, 733–747. doi: 10.1016/S0022-2836(02)00135-3
- Tama, F., and Brooks III, C. L. (2005). Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.* 345, 299–314. doi: 10.1016/j.jmb.2004.10.054
- Tama, F., Gadea, F. X., Marques, O., and Sanejouand, Y. H. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 41, 1–7. doi: 10.1002/1097-0134(20001001)41:1<1::AID-PROT10>3.0.CO;2-P
- Tama, F., and Sanejouand, Y. H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14, 1–6. doi: 10.1093/protein/14.1.1
- Teoh, E. P., Kukkaro, P., Teo, E. W., Lim, A. P., Tan, T. T., Yip, A., et al. (2012). The structural basis for serotype-specific neutralization of dengue virus by a human antibody. *Sci. Transl. Med.* 4:139ra83. doi: 10.1126/scitranslmed.3003888
- Tirion, M. (1996). Large amplitude elastic motions in proteins from a single parameter, atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908. doi: 10.1103/PhysRevLett.77.1905
- Tiwari, S. P., Fuglebakk, E., Hollup, S. M., Skjaerven, L., Cragolini, T., Grindhaug, S., et al. (2014). WEBnm@ v2.0: web server and services for comparing protein flexibility. *BMC Bioinformatics* 15:427. doi: 10.1186/s12859-014-0427-6
- Tozzini, V. (2005). Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15, 144–150. doi: 10.1016/j.sbi.2005.02.005
- van Vlijmen, H., and Karplus, M. (2005). Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry. *J. Mol. Biol.* 350, 528–542. doi: 10.1016/j.jmb.2005.03.028
- Vanden-Eijnden, E., and Heymann, M. (2008). The geometric minimum action method for computing minimum energy paths. *J. Chem. Phys.* 128:061103. doi: 10.1063/1.2833040
- WHO (2016). *Zika Strategic Response Framework and Joint Operations Plan (January-June)*. Geneva: WHO.
- Wilson, K. (1975). The renormalization group: critical phenomena and the kondo problem. *Rev. Mod. Phys.* 47, 773–840. doi: 10.1103/RevModPhys.47.773
- Xia, F., Tong, D., and Lu, L. (2013). Robust heterogeneous anisotropic elastic network model precisely reproduces the experimental b-factors of biomolecules. *J. Chem. Theory Comput.* 13, 3704–3714. doi: 10.1021/ct4002575
- Xia, F., Tong, D., Yang, L., Wang, D., Hoi, S. C., Koehl, P., et al. (2014). Identifying essential pairwise interactions in elastic network model using the alpha shape theory. *J. Comp. Chem.* 35, 1111–1121. doi: 10.1002/jcc.23587
- Yang, L., Song, G., and Jernigan, R. (2009). Protein elastic nmodels and the ranges of cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12347–12352. doi: 10.1073/pnas.0902159106
- Zhang, X., Ge, P., Yu, X., Brannan, J., Bi, G., Zhang, Q., et al. (2012). Cryo-EM structure of the mature dengue virus at 3.5 Å resolution. *Nat. Struct. Mol. Biol.* 20, 105–110. doi: 10.1038/nsmb.2463
- Zhang, Y., Zhang, W., Ogata, S., Clements, D., Strauss, J. H., Baker, T. S., et al. (2004). Conformational changes of the flavivirus e glycoprotein. *Structure* 12, 1607–1618. doi: 10.1016/j.str.2004.06.019
- Zhang, Z. (2015). Systematic methods for defining coarse-grained maps in large biomolecules. *Adv. Exp. Med. Biol.* 827, 33–48. doi: 10.1007/978-94-017-9245-5\_4
- Zhang, Z., Lu, L., Noid, W. G., Krishna, V., Pfandner, J., and Voth, G. A. (2008). A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.* 95, 5073–5083. doi: 10.1529/biophysj.108.139626
- Zhang, Z., Sanbonmatsu, K. Y., and Voth, G. A. (2011). Key intermolecular interactions in the e. coli 70s ribosome revealed by coarse-grained analysis. *J. Am. Chem. Soc.* 133, 16828–16838. doi: 10.1021/ja2028487
- Zheng, W., and Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13253–13258. doi: 10.1073/pnas.2235686100
- Zhou, X., Ren W, E. W. (2008). Adaptive minimum action method for the study of rare events. *J. Chem. Phys.* 128:104111. doi: 10.1063/1.2830717

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hsieh, Poitevin, Delarue and Koehl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.