Progress in Biophysics and Molecular Biology 143 (2019) 20–37

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Coarse-grained dynamics of supramolecules: Conformational changes in outer shells of Dengue viruses

Patrice Koehl [a, *], Marc Delarue [b]

[a] Department of Computer Science, University of California, Davis, USA
[b] Unité de Dynamique Structurale des Macromolécules, UMR 3528 Du CNRS, Institut Pasteur, 75015, Paris, France

ARTICLE INFO

Article history:
Received 31 May 2018
Received in revised form
22 September 2018
Accepted 24 September 2018
Available online 28 September 2018

Keywords:
Normal modes
Coarse-grained potentials
Fast eigenvalue computation
Dengue

ABSTRACT

While structural data on viruses are more and more common, information on their dynamics is much harder to obtain as those viruses form very large molecular complexes. In this paper, we propose a new method for computing the coarse-grained normal modes of such supra-molecules, NormalGo. A new formalism is developed to represent the Hessian of a quadratic potential using tensor products. This formalism is applied to the Tirion elastic potential, as well as to a Gō like potential. When combined with a fast method for computing a select set of eigenpairs of the Hessian, this new formalism enables the computation of thousands of normal modes of a full viral shell with more than one hundred thousand atoms in less than 2 h on a standard desktop computer. We then compare the two coarse-grained potentials. We show that, despite significant differences in their formulations, the Tirion and the Gō like potentials capture very similar dynamics characteristics of the molecule under study. However, we find that the Gō like potential should be preferred as it leads to less local deformations in the structure of the molecule during normal mode dynamics. Finally, we use NormalGo to characterize the structural transitions that occur when FAB fragments bind to the icosahedral outer shell of serotype 3 of the Dengue virus. We have identified residues at the surface of the outer shell that are important for the transition between the FAB-free and FAB-bound conformations, and therefore potentially useful for the design of antibodies to Dengue viruses.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Coarse-grained models have long been used for studying protein folding, aggregation, as well as large molecular systems such as viruses (Smit et al. (1993); Shelley et al. (2001); Nielsen et al. (2004); Bond et al. (2004); Shillcock and Lipowsky (2005); Tozzini (2005); Marrink et al. (2007); Shih et al. (2007); Scott et al. (2008); Zhang et al. (2008); Riniker et al. (2012); Sinitskiy and Voth (2013); Saunders and Voth (2013); Kar and Feig (2014); Zhang (2015); Kmiecik et al. (2016); Marzinek et al. (2016); Reddy and Sansom (2016); Voth (2017); Zheng and Wen (2017)). They enable the exploration of large length scales and time scales that are usually inaccessible for all-atom models (Saunders and Voth (2013); Kmiecik et al. (2016)). Combined with enhanced configuration search (Monte Carlo) and energy minimization methods, these simplified models offer the possibility to determine equilibrium structures and to compare folding kinetics and thermodynamics quantities with the corresponding values obtained by experimental techniques. The first step in coarse-graining is to use a simplified representation of the molecular system. In their pioneering work from 1976, Levitt and Warshel (1976) developed the foundation of coarse-graining for protein folding and proposed a two-bead representation for each amino acid, namely the Cα and the centroid of the side chain. Since then, various levels of granularity have been developed, from lattice to multi-bead representations and from single atom to multiple-atom residue-level representations (Kmiecik et al. (2016)). The positions of those beads are either defined by known atoms (usually the Cα), or by fitting their positions to capture the dynamics of the full molecular systems (Zhang et al. (2008, 2011); Li et al. (2016)). In this paper we are interested in proteins and consider methods that implement coarse-graining with a Cα-only representation.

While defining the geometry of the simplified model is important, the main difficulty in coarse-graining is to design potential energy functions or force fields that retain the physics of the all-

* Corresponding author.
E-mail addresses: pakoehl@ucdavis.edu (P. Koehl), delarue@pasteur.fr (M. Delarue).

https://doi.org/10.1016/j.pbiomolbio.2018.09.004
0079-6107/© 2018 Elsevier Ltd. All rights reserved.

atom explicit solvent system in terms of structure, thermodynamics, and dynamics (Riniker et al. (2012)). Two very popular models for representing the energy of a coarse-grained system are the Gō-like models (Gō and Abe (1981); Abe and Gō (1981)) and the Elastic Network Model (ENM) (Tirion (1996)) and its variants (Lopez-Blanco and Chacon (2016)). The Gō potential mimics the semi-empirical potential used in traditional molecular mechanics (MM), with bonded and non-bonded energy terms, and a simplified representation of the molecule with either one or two atoms per residue. The significant difference between the Gō potential and a standard MM potential is that the native structure of the protein, as usually determined by X-ray crystallography, defines the ground state of the former, while standard stereochemistry is imposed by the latter. Gō-like models have been used extensively for studying protein folding (for review, see for example Dill et al. (2008)). In the ENM, a residue in a protein is usually represented by a bead located at the position of its Cα atom. Each pair of Cα atoms within a given cutoff distance R_C is connected by a harmonic spring with a uniform force constant throughout the molecule, whose equilibrium conformation is set to the experimentally-determined structure. The total potential energy of the Elastic Network is then expressed as the sum of the simple harmonic potentials of these springs.

Coarse-graining has proved useful to study dynamics of biomolecules at multiple scales (Karplus (2014); Levitt (2014); Warshel (2014)). This is particularly true within the framework of normal mode analyses, which are designed to infer dynamics from static structures corresponding to locally stable states (Mahajan and Sanejouand (2015)). Cartesian Normal Modes, for example, represent a class of movements around a local energy minimum that are both straightforward to calculate and biologically relevant (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 (Tama and Sanejouand (2001)). Normal modes have also been adapted to coarse-grained potentials. The simple quadratic form of the latter allows for a simple decomposition of the motions of the protein of interest into vibrational modes with different frequencies. This approach has been widely used in computational studies of macromolecules since its introduction nearly two decades ago (for recent reviews, see Sanejouand (2013), Sinitskiy and Voth (2013) and Lopez-Blanco and Chacon (2016)).

Computing coarse-grained normal modes for a molecular system is deceptively easy. Starting from a conformation C_0 for the molecular system, a coarse-grained potential V that is minimum at C_0, and taking a second order approximation of that potential, the Newton's equations of motion expressed in Cartesian coordinates lead to a generalized eigenvalue problem

$$Hv = \lambda Mv, \tag{1}$$

where $H \in \mathbb{R}^{3N \times 3N}$ is the Hessian matrix of the potential V computed at C_0, M is the (diagonal) mass matrix for the system, λ is one eigenvalue and v the eigenvector associated to λ. The row size of the Hessian matrix, $3N$, is equal to three times the number N of atoms in the system. When N is small, this Hessian can be stored as a dense matrix and the computation of its eigen pairs is straightforward, leading to the popularity of the method described above. However, while the contribution of normal modes techniques to study the dynamics of such small systems is well documented, the challenge is to use them to analyze much bigger molecular systems, in which case N will be large. As standard algorithms for eigen analyses have a time complexity of $O(N^3)$, they become prohibitive for such systems and have to be adapted. For example, normal mode techniques have been used extensively to study the dynamics of virus outer shells (for reviews, see for example Tama and Brooks III (2006); Dykeman and Sankey (2010); Lezon et al. (2010)). However, most of those applications resort to a symmetry-specific implementation that lead to smaller system sizes (Simonson and Perahia (1992); van Vlijmen and Karplus (2005); Peeters and Taormina (2009)), or to higher level of coarsening, as in the Rotations and Translations of Blocks (RTB) method (Tama et al. (2000)) and in the block normal mode method (Li and Cui (2002)), and only a few have used full atom representations (Dykeman and Sankey (2008, 2010); Hsieh et al. (2016)), at the cost of large computational needs.

This paper includes two components, a methodological development related to normal modes, and an application of the corresponding method to studying structural transitions in virus outer shells. On the methodological side, we propose a new implementation of the computation of coarse-grained normal modes. In particular, a new representation of the Hessian of a pairwise potential is proposed using tensors. Tensor products are tools from linear algebra that are commonly used to simplify the computation and representation of derivatives. For example, Blondel and Karplus (1996) used tensors to simplify the computation of derivatives of dihedral and improper angles in molecular mechanics force fields, showing that it even results in removal of singularities. While we use tensor products for similar simplifications when computing the Hessian of the elastic potential introduced by Tirion (1996) and of a Gō-like potential (Clementi et al. (2000)), we note that we also use the properties of such a representation and show that it leads to faster and parallelizable implementations of Hessian vector products. We believe that this is a novel contribution to the field of computational structural biology. In addition, we implement a new method for computing some of the eigen pairs of the huge Hessian associated with a supra molecule, based on a block Chebyshev Davidson algorithm with inner-outer restart (Zhou (2010)). The combination of the tensor representations for the Hessians and the fast method for computing a fraction of the eigen pairs of those Hessians leads to a new, fast implementation of normal modes that is amenable to analyzing supra molecular systems. We then compare the normal modes computed with those two potentials. We show that while the corresponding two models for coarse-grained normal modes capture similar motions both in frequencies and direction, the Gō-like potential leads to less deformations of the molecule and better preserves the stereochemistry.

We use icosahedral outer shells of Dengue viruses as very large systems to analyze the performances of our new implementations of coarse-grained normal modes as well as the differences between the two coarse-grained potentials we have considered. The outer shells of Dengue viruses are multi-protein assemblies forming an envelope that serve as containers for the genetic material of the viruses (Hagan (2014)). One of the most important roles of the outer shell is to protect its contents. The outer shell also needs to change conformation to release its content into a host cell during infection. Thus, the stability of the outer shell and its dynamics are key issues in the life cycle of a virus (Hagan (2014)). Many viral outer shells are known structurally. However, much less is known about their dynamics. We focus on the dynamics of serotype 3 of Dengue viruses, more specifically on the changes of conformation induced by the binding of an antibody to their outer shells, which we will refer to as their envelopes (Fibriansah et al. (2015)).

The paper is organized as follows. In the next section, we describe normal mode analysis for coarse grained systems. We provide an overview of the theory, discuss the two coarse-grained potentials considered, namely the elastic potential and the Gō

potential, and discuss the computations of the Hessian matrices associated with those potentials, describing a new formulation in which these matrices are expressed as a sum of tensor products. We also provide a (brief) description of the method that we have implemented for finding a small subset of the eigen pairs of a large symmetric matrix. The following section describes the implementations of those methods into a new program, NormalGo, as well as of the tools that we use to analyze coarse-grained normal modes. In the Results section, we present and discuss the applications of these methods to computing the normal modes of different conformations of the envelope of serotype 3 of Dengue virus, and analyze the possible transitions between those conformations. We conclude the paper with a discussion on future developments.

2. Normal mode analysis

Notations. We will use the following notations in all the following subsections. We write the inner (dot) and outer products of two vectors $\mathbf{u}$ and $\mathbf{v}$ as $<\mathbf{u},\mathbf{v}>$ and $\mathbf{u}\otimes\mathbf{v}$, respectively. Let B be a biomolecule containing N atoms, with atom i characterized by its position $\mathbf{r}_i = (x_i, y_i, z_i)$. The whole molecule is then described by a $3N$ position vector $\mathbf{X} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. For two atoms i and j of B, we define $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ to be the vector pointing from i towards j. We set $r_{ij} = ||\mathbf{r}_{ij}||$ and $r^0_{ij} = \left|\left|\mathbf{r}^0_{ij}\right|\right|$ to be the length of $\mathbf{r}_{ij}$ in any conformation X and in the ground-state conformation X^0 (usually the experimental structure), respectively.

2.1. Coarse-grained normal modes

The crux of normal mode analysis is to infer the dynamics of a biomolecule from a static structure corresponding to a locally stable state (Noguti and Go (1982); Brooks and Karplus (1983); Levitt et al. (1985)). Let us consider such a biomolecule B, with a stable state $\mathbf{X}_0$ for a given potential V. In the normal mode framework, the potential V is approximated with a second-order Taylor expansion in the neighborhood of X^0 to yield a quadratic potential:

$$V_{quad}(\mathbf{X}) = V\left(\mathbf{X}^0\right) + \nabla V\left(\mathbf{X}^0\right)^T\left(\mathbf{X}-\mathbf{X}^0\right) + \frac{1}{2}\left(\mathbf{X}-\mathbf{X}^0\right)^T H\left(\mathbf{X}-\mathbf{X}^0\right) \quad (2)$$

where ∇V and H are the gradient and Hessian of V, respectively. In this expression, $V(\mathbf{X}^0)$ is a constant that can be safely ignored and $\nabla V(\mathbf{X}^0) = 0$ under the assumption that $\mathbf{X}_0$ is a stable state. The quadratic energy is then simply

$$V_{quad}(\mathbf{X}) = \frac{1}{2}\left(\mathbf{X}-\mathbf{X}^0\right)^T H\left(\mathbf{X}-\mathbf{X}^0\right) \quad (3)$$

For simplicity, we will assume that each atom is assigned a mass of 1. The procedure described below can easily be expanded to the more general case of different values for the masses. In Cartesian coordinates, the equations of motion defined by the potential V_{quad} are derived from Newton's equations:

$$\frac{d^2\mathbf{X}}{dt^2} = -H\left(\mathbf{X}-\mathbf{X}^0\right) \quad (4)$$

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

$$\mathbf{X} = \sum_{k=1}^{3N}\alpha_k\mathbf{X}_k = \sum_{k=1}^{3N}\mathbf{V}_k\alpha_k cos(\omega_k t + \delta_k) \quad (5)$$

we get a standard eigenvalue problem,

$$HV = V\Omega \quad (6)$$

The eigenfrequencies ω_k are given by the elements of the diagonal matrix Ω, namely $\omega_k^2 = \Omega(k,k)$. The eigenvectors are the columns of the matrix V, and the amplitudes and phases, α_k and δ_k, are determined by initial conditions. We note that (at least) the first six eigenvalues in Ω are equal to 0, as they correspond to global translations and rotations of the biomolecule.

The input to a normal mode analysis is usually the experimental structure $\mathbf{X}_{exp}$ of the molecule of interest, most often the X-ray structure, and recently the cryo-EM structure when the molecule is large. For generic potentials V, such as those implemented in semi-empirical forcefields (Nerenberg and Head-Gordon (2018)), there are no reasons to expect that this structure is a stable conformation for V. Modifying $\mathbf{X}_{exp}$ to reach a true minimum $\mathbf{X}_0$ for V is unfortunately a difficult task, especially for large biomolecules. Starting from a conformation $\mathbf{X}_0$ that does not satisfy $V(\mathbf{X}^0) = 0$ is, however, problematic, as it may lead to the Hessian computed at $\mathbf{X}_0$ to be negative, in which case it has negative eigenvalues that are not meaningful with respect to dynamics. To circumvent this problem, Tirion (1996) proposed a simplified elastic potential that has a single minimum set to be the input structure $\mathbf{X}_{exp}$. This simple idea has led to the popularity of coarse-grained normal mode analyses. In the following, we review this simplified potential, and discuss another coarse-grained Gō-like potential that exhibits similar properties.

2.2. The Tirion elastic potential

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); Tama and Sanejouand (2001). Here we follow the latter model, in which the energy of the molecule is equated with the harmonic energy associated with these springs. This defines a quadratic energy on the inter-atomic distances, defined as

$$V(\mathbf{X}) = \frac{1}{2}\sum_{(i,j)}k_{ij}\left(r_{ij}-r^0_{ij}\right)^2 \quad (7)$$

when the biomolecule is in conformation $\mathbf{X}$. In this equation, k_{ij} is the force constant of the "spring" formed by the pair of atoms i and j. The sum includes all pairs of atoms (i,j) that satisfies $r^0_{ij} < R_c$, where R_c is a cutoff distance. In the original ENM (Tirion (1996)), the force constants k_{ij} are set constant for all pairs of residues. 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 Cα atoms on the backbone of a protein were strengthened to reproduce the correct bimodal distribution of density-of-states from an all-atom model, while Kondrashov et al. (2006) used a strategy in which they classified residue interactions into several categories corresponding to different

physical properties. In parallel, the values R_c differ between models. Those values are usually considered in the range from 12 to 15 Å if only one atom per residue (usually Cα) is considered, and in the range from 6 to 8 Å if all atoms are considered (Eyal et al. (2006)). We will use constant $k_{ij} = 1$ and $R_c = 15$ for Cα-only models.

Computing the Hessian of the elastic potential $V(X)$. Let us rewrite the quadratic potential for the elastic network as:

$$V_{ENM}(X) = \frac{1}{2}\sum_{(i,j)} V_{ij}(X) \tag{8}$$

where the summation extends to all pairs of atoms (i,j) that satisfy the cutoff criterium (see above). We compute the derivatives and Hessian of this potential in vector form.

We define the vector $\mathbf{U}_1(i,j)$ such that

$$\mathbf{U}_1(i,j) = \left(0,\ldots,0,\frac{\mathbf{r}_{ij}}{r_{ij}},0,\ldots,0,-\frac{\mathbf{r}_{ij}}{r_{ij}},0,\ldots,0\right), \tag{9}$$

namely, $\mathbf{U}_1(i,j)$ is a vector in $\mathbb{R}^{3N}$ that is zero everywhere, except at positions i and j where it is equal to the unit vector associated with the difference vector $\mathbf{r}_{ij}$ and with the difference vector $\mathbf{r}_{ji} = -\mathbf{r}_{ij}$, respectively.

Let us first analyze the pairwise potential $V_{ij}(X)$. Its gradient in $\mathbb{R}^{3N}$ at a position $\mathbf{X}$ is given by:

$$\nabla V_{ij}(\mathbf{X}) = k_{ij}\left(r_{ij} - r_{ij}^0\right)\mathbf{U}_1(i,j) \tag{10}$$

and its Hessian at the same position $\mathbf{X}$ is given by:

$$H_{ij}(\mathbf{X}) = k_{ij}\left(r_{ij} - r_{ij}^0\right)\frac{\delta \mathbf{U}_1(i,j)}{\delta \mathbf{X}} + k_{ij}\mathbf{U}_1(i,j)\otimes\mathbf{U}_1(i,j) \tag{11}$$

Note that both terms in the expression of the Hessian are matrices of size $3N \times 3N$. For normal mode analyzes, the gradient and Hessian are evaluated at $\mathbf{X}^0$:

$$\nabla V_{ij}\left(\mathbf{X}^0\right) = 0 \tag{12}$$

and

$$H_{ij}\left(\mathbf{X}^0\right) = k_{ij}\mathbf{U}_1(ij)\otimes\mathbf{U}_1(i,j) \tag{13}$$

The total Hessian of the elastic potential is then given by:

$$H = H\left(\mathbf{X}^0\right) = \sum_{(i,j)} k_{ij}\mathbf{U}_1(i,j)\otimes\mathbf{U}_1(i,j) \tag{14}$$

2.3. Normal mode analysis based on Gō-like potential

The elastic potential given by Equation (7) is a simple pairwise geometric potential. As such, it does not account for the stereochemistry of the molecule explicitly. In particular, it may not preserve bond lengths, bond angles, and preferences in dihedral angles, as a full atom, semi-empirical forcefield would. While it is possible to account for those specific interactions to some extent by modulating the values of the spring constants k_{ij} (Ming and Wall (2005); Kondrashov et al. (2006)), an alternate approach is to redesign the potential to include those bonded interactions explicitly. The elastic potential is based on the native structure of the molecular system, as it only includes pairs of atoms that are in close contact in that native structure. The Gō-like potential has the same foundation as it takes into account only native interactions (Ueda et al. (1975); Clementi et al. (2000)). It has been adapted to the framework of coarse-grained normal modes (Lin and Song (2010); Na et al. (2014)). In those papers, the presentation of the computation of the Hessian of their Gō-like potential was somewhat complicated. We provide a simplified derivation below, using the framework of tensor products defined above.

The Gō-like potential only considers the Cα of all residues in the molecule of interest. Let b_i be the length of the pseudo-bond between the Cαs of the consecutive residues i and $i+1$, θ_i be the virtual bond angle formed by the Cαs of the consecutive residues i, $i+1$, and $i+2$, and let ϕ_i be the virtual dihedral angle formed by the Cαs of the consecutive residues i, $i+1$, $i+2$, and $i+3$.

The Gō-like potential at a conformation $\mathbf{X}$ is then defined as

$$\begin{aligned} V_G(\mathbf{X}) &= V_{bond}(\mathbf{X}) + V_{angle}(\mathbf{X}) + V_{dih}(\mathbf{X}) + V_{nb}(\mathbf{X}) \\ &= \frac{1}{2}\sum_{i=1}^{N-1} K_r\left(b_i - b_i^0\right)^2 + \frac{1}{2}\sum_{i=1}^{N-2} K_\theta\left(\theta_i - \theta_i^0\right)^2 + \sum_{i=1}^{N-3}\left[K_{\phi 1}\left(1 - \cos\left(\phi_i - \phi_i^0\right)\right) + K_{\phi 3}\left(1 - \cos 3\left(\phi_i - \phi_i^0\right)\right)\right] \\ &\quad + \sum_{(i<j-3)} \varepsilon\left[5\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^0}{r_{ij}}\right)^{10}\right] \end{aligned} \tag{15}$$

where the superscript 0 refers to the values of the variables for the equilibrium conformation $\mathbf{X}^0$. The first three terms refer to (pseudo) bonded interactions, while the last term corresponds to non bonded interactions. When the molecular system considered includes multiple chains, special care is needed to only include bonds, angles, and dihedral angles that exist within a chain. In all computations included in this paper, we will use $K_r = 100\varepsilon$, $K_\theta = 20\varepsilon$, $K_{\phi 1} = \varepsilon$, $K_{\phi 3} = 0.5\varepsilon$, and $\varepsilon = 0.36$. In addition, the non-bonded term V_{nb} is computed over all pairs of non-bonded atoms that satisfy $r_{ij}^0 < R_g$, where R_g is a cutoff distance set to 25 Å.

In the normal mode framework, the Gō-like potential V_G is approximated with a second-order Taylor expansion in the neighborhood of the ground state X^0:

$$\begin{aligned} V_{GM}(\mathbf{X}) &= V_G\left(\mathbf{X}^0\right) + \nabla V_G\left(\mathbf{X}^0\right)^T\left(\mathbf{X} - \mathbf{X}^0\right) \\ &\quad + \frac{1}{2}\left(\mathbf{X} - \mathbf{X}^0\right)^T H_G\left(\mathbf{X} - \mathbf{X}^0\right) \end{aligned} \tag{16}$$

where ∇V_G and H_G are the gradient and Hessian of V_G, respectively. Note that based on Equation (15), $\nabla V(\mathbf{X}^0) = 0$, and $V_G(\mathbf{X}^0) = -\varepsilon N_{NB}$, where N_{NB} is the number of non-bonded pairs of atoms considered. As the latter is independent of the conformation $\mathbf{X}$ (as it only depends on $\mathbf{X}^0$), we can safely ignore its contribution to normal modes. The approximate Gō-like potential V_{GM} can then be written as

$$V_{GM}(\mathbf{X}) = \frac{1}{2}\left(\mathbf{X} - \mathbf{X}^0\right)^T H_G\left(\mathbf{X} - \mathbf{X}^0\right) \tag{17}$$

To compute the Hessian H_G, we note first that V_G is the sum of four terms. As the Hessian is a linear operator, we need to compute four Hessians, H_{bond}, H_{angle}, H_{dih}, and H_{nb}, for the bond, angle, dihedral angle, and non-bonded terms included in V_G, respectively.

Computing the Hessian of the bond potential $V_{bond}(X)$. The bond potential is similar to the elastic potential discussed above, limited

to pairs of consecutive atoms. The computation of its Hessian is then straightforward. For an atom i, we set $j = i+1$. Using the same definition of $U_1(i,j)$ introduced for the elastic potential, we find,

$$H_{bond} = \sum_i H_{ij}\left(\mathbf{X}^0\right) = \sum_i K_r \mathbf{U}_1(i,j) \otimes \mathbf{U}_1(i,j) \tag{18}$$

where the sum extends to all atoms i in $[1, M-1]$ and where $j = i+1$.

Computing the Hessian of the angle potential $V_{angle}(X)$. Let us consider first the pseudo angle θ_j formed by three consecutive Cα, i, $j = i+1$, and $k = i+2$, and centered at j. Those three atoms form a triangle (see Fig. 1A). Using the law of cosines,

$$r_{ik}^2 = r_{ij}^2 + r_{jk}^2 - 2 r_{ij} r_{jk} \cos(\theta_j) \tag{19}$$

we get:

$$\cos(\theta_j) = \frac{r_{ij}^2 + r_{jk}^2 - r_{ik}^2}{2 r_{ij} r_{jk}},$$

$$\sin(\theta_j) = \sqrt{1 - \cos^2(\theta_j)}. \tag{20}$$

Using Equation (20), we obtain:

$$\mathbf{v}_i = \frac{\delta\theta_j}{\delta \mathbf{r}_i} = \frac{-r_{ij}\mathbf{r}_{jk} - r_{jk}\cos(\theta_j)\mathbf{r}_{ij}}{\sin(\theta_j) r_{ij}^2 r_{jk}}$$

$$\mathbf{v}_k = \frac{\delta\theta_j}{\delta \mathbf{r}_k} = \frac{r_{jk}\mathbf{r}_{ij} + r_{ij}\cos(\theta_j)\mathbf{r}_{jk}}{\sin(\theta_j) r_{ij} r_{jk}^2}$$

$$\mathbf{v}_j = \frac{\delta\theta_j}{\delta \mathbf{r}_j} = -\mathbf{v}_i - \mathbf{v}_k \tag{21}$$

We define the vector $\mathbf{U}_2(i,j,k)$ such that

$$\mathbf{U}_2(i,j,k) = (0, \ldots, 0, \mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k, 0, \ldots, 0), \tag{22}$$

namely $\mathbf{U}_2(i,j,k)$ is a vector in $\mathbb{R}^{3N}$ that is zero everywhere, except at positions $i, j = i+1$, and $k = i+2$.

Let $V_{angle,j}$ be the potential energy associated with θ_j:

$$V_{angle,j} = \frac{1}{2} K_\theta \left(\theta_j - \theta_j^0\right)^2 \tag{23}$$

The gradient of $V_{angle,j}$ at a position $\mathbf{X}$ is then given by:

$$\nabla V_{angle,j}(\mathbf{X}) = K_\theta \left(\theta_j - \theta_j^0\right) \mathbf{U}_2(i,j,k) \tag{24}$$

and its Hessian at the same position $\mathbf{X}$ is given by:

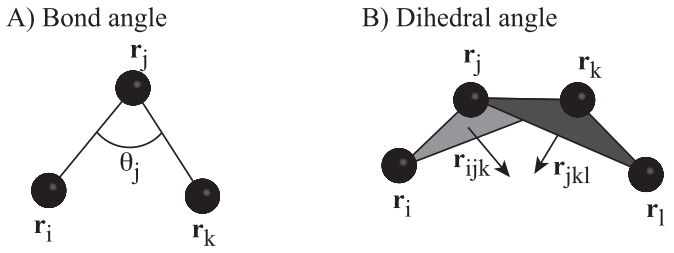

Fig. 1. Bond angle (A) and dihedral angle (B) for consecutive Cαs.

$$H_j(\mathbf{X}) = K_\theta \left(\theta_j - \theta_j^0\right) \frac{\delta \mathbf{U}_2(i,j,k)}{\delta \mathbf{X}} + K_\theta \mathbf{U}_2(i,j,k) \otimes \mathbf{U}_2(i,j,k) \tag{25}$$

For normal mode analyzes, the gradient and Hessian are evaluated at $\mathbf{X}^0$:

$$\nabla V_{angle,j}\left(\mathbf{X}^0\right) = 0$$

$$H_j\left(\mathbf{X}^0\right) = K_\theta \mathbf{U}_2(i,j,k) \otimes \mathbf{U}_2(i,j,k) \tag{26}$$

The total Hessian of the angle potential is then given by:

$$H_{angle} = H_{angle}\left(\mathbf{X}^0\right) = \sum_i K_\theta \mathbf{U}_2(i,j,k) \otimes \mathbf{U}_2(i,j,k) \tag{27}$$

where the sum extends to all atoms i in $[1, M-2]$ and where $j = i+1, k = i+2$.

Computing the Hessian of the dihedral angle potential $V_{dih}(X)$. Let us consider first the pseudo dihedral ϕ_j formed by four consecutive Cα, i, $j = i+1$, $k = i+2$, and $k = i+3$, and centered around the pseudo bond $j-(j+1)$. ϕ_j is the angle between the planes $i-j-k$ and $j-k-l$ (see Fig. 1B). Let us define the normals $\mathbf{r}_{ijk}$ and $\mathbf{r}_{jkl}$ to those two planes.

$$\mathbf{r}_{ijk} = \mathbf{r}_{ij} \times \mathbf{r}_{kj}$$

$$\mathbf{r}_{jkl} = \mathbf{r}_{kj} \times \mathbf{r}_{kl}, \tag{28}$$

and their 2-norms, r_{ijk}, and r_{jkl}. The angle ϕ_j is the angle between those two vectors and can be obtained using

$$\cos(\phi) = \frac{<\mathbf{r}_{ijk}, \mathbf{r}_{ijl}>}{r_{ijk} r_{jkl}} \tag{29}$$

From this definition of the dihedral angle, one obtains (van Schaik et al. (1993)).

$$\mathbf{w}_i = \frac{\delta\phi_j}{\delta \mathbf{r}_i} = \frac{r_{jk}}{r_{ijk}^2}\mathbf{r}_{ijk}$$

$$\mathbf{w}_l = \frac{\delta\phi_j}{\delta \mathbf{r}_l} = -\frac{r_{jk}}{r_{jkl}^2}\mathbf{r}_{jkl}$$

$$\mathbf{w}_j = \frac{\delta\phi_j}{\delta \mathbf{r}_j} = \left(-\frac{<\mathbf{r}_{ij}, \mathbf{r}_{jk}>}{r_{jk}^2} - 1\right)\mathbf{v}_i + \frac{<\mathbf{r}_{kl}, \mathbf{r}_{jk}>}{r_{jk}^2}\mathbf{v}_l$$

$$\mathbf{v}_k = \frac{\delta\phi_j}{\delta \mathbf{r}_k} = -\mathbf{v}_i - \mathbf{v}_j - \mathbf{v}_l. \tag{30}$$

We next define the vector $\mathbf{U}_3(i,j,k,l)$ such that

$$\mathbf{U}_3(i,j,k,l) = (0, \ldots, 0, \mathbf{w}_i, \mathbf{w}_j, \mathbf{w}_k, \mathbf{w}_l, 0, \ldots, 0), \tag{31}$$

namely $\mathbf{U}_3(i,j,k,l)$ is a vector in $\mathbb{R}^{3M}$ that is zero everywhere, except at positions $i, j = i+1$, $k = i+2$, and $l = i+3$.

Let $V_{dih,j}$ be the potential energy associated with the dihedral angle ϕ_j:

$$V_{dih,j} = K_{\phi 1}\left(1 - \cos\left(\phi_i - \phi_i^0\right)\right) + K_{\phi 3}\left(1 - \cos 3\left(\phi_i - \phi_i^0\right)\right) \tag{32}$$

Using the chain rule, we find that the gradient of $V_{dih,j}$ at a position $\mathbf{X}$ is given by:

$$\nabla V_{dih,j}(\mathbf{X}) = K_{\phi 1}\sin\left(\phi_i - \phi_i^0\right)\mathbf{U}_3(i,j,k,l) + 3K_{\phi 3}\sin 3\left(\phi_i - \phi_i^0\right)\mathbf{U}_3(i,j,k,l) \quad (33)$$

and its Hessian at the same position $\mathbf{X}$ is given by:

$$H_j(\mathbf{X}) = K_{\phi 1}\sin\left(\phi_i - \phi_i^0\right)\frac{\delta \mathbf{U}_3(i,j,k,l)}{\delta \mathbf{X}} + K_{\phi 1}\cos\left(\phi_i - \phi_i^0\right)\mathbf{U}_3(i,j,k,l)\otimes\mathbf{U}_3(i,j,k,l) + 3K_{\phi 3}\sin 3\left(\phi_i - \phi_i^0\right)\frac{\delta \mathbf{U}_3(i,j,k,l)}{\delta \mathbf{X}} + 9K_{\phi 3}\cos\left(\phi_i - \phi_i^0\right)\mathbf{U}_3(i,j,k,l)\otimes\mathbf{U}_3(i,j,k,l) \quad (34)$$

For normal mode analyzes, the gradient and Hessian are evaluated at $\mathbf{X}^0$:

$$\nabla V_{dih,j}\left(\mathbf{X}^0\right) = 0$$

$$H_j\left(\mathbf{X}^0\right) = (K_{\phi 1} + 9K_{\phi 3})\mathbf{U}_3(i,j,k,l)\otimes\mathbf{U}_3(i,j,k,l) \quad (35)$$

The total Hessian of the dihedral potential is then given by:

$$H_{dih} = H_{dih}\left(\mathbf{X}^0\right) = \sum_i (K_{\phi 1} + 9K_{\phi 3})\mathbf{U}_3(i,j,k,l)\otimes\mathbf{U}_3(i,j,k,l) \quad (36)$$

where the sum extends to all atoms i in $[0, M-3]$ and where $j = i+1, k = i+2, l = i+3$.

Computing the Hessian of the non bonded potential $V_{nb}(X)$. Let us consider two Cα atoms from residues i and j, such that $i < j-3$ and $r_{ij} < R_g$. The non-bonded potential between these two atoms is

$$V_{nb,i,j} = \varepsilon\left[5\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^0}{r_{ij}}\right)^{10}\right] \quad (37)$$

Using the chain rule, we find that the gradient of $V_{nb,i,j}$ at a position $\mathbf{X}$ is given by:

$$\nabla V_{nb,i,j}(\mathbf{X}) = -\frac{60\varepsilon}{r_{ij}}\left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}}\right)^{10}\right]\mathbf{U}_1(i,j) \quad (38)$$

where $U_1(i,j)$ is the same vector in $\mathbb{R}^{3N}$ introduced for the elastic potential and for the bond term of the Gō-like potential. Using again the chain rule, its Hessian at the same position $\mathbf{X}$ is given by:

$$H_{nb,i,j}(\mathbf{X}) = \frac{60\varepsilon}{r_{ij}^2}\left[13\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 11\left(\frac{r_{ij}^0}{r_{ij}}\right)^{10}\right]\mathbf{U}_1(i,j)\otimes\mathbf{U}_1(i,j) - \frac{60\varepsilon}{r_{ij}}\left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}}\right)^{10}\right]\frac{\delta \mathbf{U}_1(i,j)}{\delta \mathbf{X}} \quad (39)$$

As for the other terms, the gradient and Hessian are evaluated at $\mathbf{X}^0$:

$$\nabla V_{nb,i,j}\left(\mathbf{X}^0\right) = 0$$

$$H_{nb,i,j}\left(\mathbf{X}^0\right) = \frac{120\varepsilon}{r_{ij}^2}\mathbf{U}_1(i,j)\otimes\mathbf{U}_1(i,j) \quad (40)$$

The total Hessian of the non-bonded potential is then given by:

$$H_{nb} = H_{nb}\left(\mathbf{X}^0\right) = 120\varepsilon\sum_{i,j}\frac{1}{r_{ij}^2}\mathbf{U}_1(i,j)\otimes\mathbf{U}_1(i,j) \quad (41)$$

where the sum extends to all atoms i and j such that $i < j-3$ and $r_{ij} < R_g$.

2.4. Advantages of a tensor representation of the Hessian

Expressing the Hessians of the elastic potential and of the Gō-like potential as a (weighted) sum of tensor products provides for simpler computations of Hessian-vector multiplications. We illustrate this property below.

Each of the Hessian matrices considered for the elastic potential and of the Gō-like potential can be written as $H = \sum_l \alpha_l \mathbf{T}_l \otimes \mathbf{T}_l$, where α_l is a coefficient that depends on the specific Hessian considered, and the tensor T_l is either based on pairs of non bonded atoms, (i,j), or on 2, 3, or 4 bonded atoms, starting at an atom i. Let $\mathbf{W}$ be a vector in $\mathbb{R}^{3N}$. The product $H\mathbf{W}$ is given by:

$$H\mathbf{W} = \sum_l \alpha_l(\mathbf{T}_l\otimes\mathbf{T}_l)\mathbf{W} \quad (42)$$

Using the property $(\mathbf{T}_l \otimes \mathbf{T}_l)\mathbf{W} = \langle \mathbf{T}_l, \mathbf{W}\rangle \mathbf{T}_l$, we get:

$$H\mathbf{W} = \sum_l \alpha_l \left\langle \mathbf{T}_l, \mathbf{W} \right\rangle \mathbf{T}_l \quad (43)$$

For the Hessians related to non-bonded terms in the potentials, $T_l = U_1(i,j)$, and the $3N$-dimensional dot product $\langle \mathbf{U}_1(i,j), \mathbf{W}\rangle$ is simply the 3-dimensional dot-product:

$$\left\langle \mathbf{U}_1(i,j), \mathbf{W} \right\rangle = \frac{k_{ij}}{r_{ij}^2}\left\langle X_i - X_j, W_i - W_j \right\rangle \quad (44)$$

where W_i is the 3D vector containing the components of W at indices $3i+1, 3i+2$, and $3i+3$, with a similar definition for W_j. This dot product only requires 3 multiplications. Multiplying this dot product with $U_1(i,j)$ requires an additional 3 multiplications. Similar simple expressions exist for the bonded terms of the Gō-like potential. In addition, the computation of the product $H\mathbf{X}$ is embarrassingly parallelizable, as it can be easily divided over all the pairs of interacting atoms (i,j) in the molecular system for the non-bonded terms, and over all atoms for the bonded terms.

We note that the tensor representation of the Hessians provides an easy way to verify that those Hessians are (semi-definite) positive. Indeed, for any non zero vectors W in $\mathbb{R}^{3N}$, based on equation (43), we have:

$$\left\langle W, H\mathbf{W} \right\rangle = \sum_l \alpha_l \left\langle \mathbf{T}_l, \mathbf{W}\right\rangle^2 \quad (45)$$

All terms in this sum are positive, provided that the coefficients are positive, which is true for both the elastic potential and the Gō-like potential. Note that for both potentials H is semi-definite positive, as it has zero as an eigenvalue, with a null space of size

(at least) 6.

3. Implementation

We have implemented the different methods described above in a stand-alone application, NormalGo, written mainly in C++. The source code is available upon request to the authors. In the following, we highlight some of the elements of NormalGo that are relevant to the analysis of large systems.

3.1. Building the network of interacting atoms

The program reads in first the conformation of a biomolecule from a file in the PDB format (Berman et al. (2000)). Each line starting with the "ATOM" tag within this PDB file is processed and interpreted to inform on the positions of the atoms of the molecule. One of the parameters provided to NormalGo enables the user to decide between an "all-atom" representation or a coarse representation in which only one atom is picked for each residue, its central Cα. The latter is enforced for the Gō potential. This structural information is then used to defined either the Elastic Network Model for the elastic potential, or the list of interacting pairs for the non-bonded part of the Gō potential. In its current implementation NormalGo only allows for a simple cutoff model. Namely, the network is defined as a set of links, with a link between two atoms only if the corresponding interatomic distance is smaller than a given cutoff value R_c or R_g(see section 2 above). If the number of atoms in the molecule is N, a brute force approach to defining this network has a complexity of $O(N^2)$. This complexity can easily be reduced to $O(N)$ by using the linked-list cell MD algorithm (Quentrec and Brot (1973)), and even reduced further by distributing the cells to be checked uniformly over the p processors available for the computation, leading to a complexity of $O(N/p)$. We have implemented this solution without further refinement as the average clock time observed for a system of one million atoms is of the order of a few seconds on a desktop computer, a time that is negligible compared to the computing time required to find the eigen pairs of the Hessian of the potential.

3.2. Computing selected eigenvalues of large symmetric real matrices

We note first that the goal is to compute a small subset of the eigen pairs of a large, sparse Hessian matrix H described above. Tirion (1996) had reported that the lowest frequency normal modes can capture most of the dynamics of the molecular system; therefore, we are interested in the smallest eigenvalues of H. Second, we note that the matrix H is positive, semi-definite, both for the elastic potential, and the Gō potential. Finally, we note that there are not reasons for H to be diagonally dominant. In practice, we have observed that it is not.

The most popular methods for computing a subset of eigen pairs of a large matrix are based on the concept of Krylov sub-space, and more specifically on the Lanczos method (Arnoldi (1951)). The Jacobi-Davidson method, originally introduced by Sleijpen and van der Vorst (1996) solves the same eigenvalue problem using a different point of view. Given a target eigenvalue τ and an approximate eigenpair (λ, u) close to this target, where u belongs to a search subspace U, the idea is to expand U with a correction s such that a better approximation for the eigenpair can be found in the expanded space. This correction vector s is usually found by solving a linear system of equations, using an iterative method that can take time to converge (see for example Hochstenbach and Notay (2006)). There is however an alternative approach that has proved more efficient (Zhou and Saad (2006)) and is of interest for normal modes calculations. The exact solution of the correction equation for s gives an expansion vector proportional to $(H-\lambda I)^{-1}u$. Zhou and Saad (2006) noticed that this is equivalent to applying a polynomial filtering on the current vector u by the rational polynomial $\phi(x) = 1/(x-\lambda)$. Applying this filtering has the advantage of magnifying the direction of the eigenvector corresponding to the Ritz value λ, which is the current best approximation to a wanted eigenvalue of H. Chebyshev filtering offers an alternative approach which was shown to improve global convergence as well as robustness. The corresponding Chebyshev-Davidson-Jacobi algorithm is fully described in Zhou and Saad (2006). Zhou (2010) further improved this algorithm and developed a block version with inner-outer restart to improve convergence and reduce its memory footprint. We have implemented this algorithm, BlockChebDav, for computing the smallest amplitude eigenvalues of the Hessian matrices related to either the elastic potential, or the Gō like potential.

3.3. Analyzing the coarse-grained normal modes

3.3.1. Atomic motions and correlated motions within a biomolecule

The Boltzmann distribution associated with a coarse-grained potential is described by a multivariate Gaussian distribution with a covariance matrix proportional to the inverse of the Hessian H of that potential. Because of the six rigid motions captured by the six normal modes associated with the eigenvalue 0, the inverse of H is not properly defined. We can, however, compute a pseudo-inverse by ignoring those zero energy modes; this pseudo-inverse defines the covariance matrix for internal deformations:

$$C = \sum_{k=7}^{K} \frac{1}{\lambda_k} \mathbf{V}_k \mathbf{V}_k^T \tag{46}$$

where λ_k and V_k are the $k-th$ eigenvalue and eigenvector of the Hessian H, respectively. Note that C is a $3N \times 3N$ matrix. The summation extends from $k = 7$, the first non-zero mode, to K, the highest mode considered (up to $3N$).

The submatrix C_{ii} of size 3×3 of C defines the internal motion of atom i. In particular, the mean squared thermal fluctuation of i is given by:

$$f_i = \left\langle \Delta \mathbf{r}_i^2 \right\rangle = \frac{k_B T}{m_i} tr(C_{ii}), \tag{47}$$

where k_B is the Boltzmann's constant, T is the temperature considered, and m_i the mass of atom i. Note that f_i is related to the thermal B-factor by $B_i = \frac{8\pi^2}{3} f_i$.

To compute the correlation of the motions of two atoms i and j, a cross correlation matrix, CCM, is computed, following Ichiye and Karplus (1991):

$$CCM_{ij} = \frac{tr(C_{ij})}{\sqrt{tr(C_{ii})tr(C_{jj})}} \tag{48}$$

The values CCM_{ij} range from -1 to $+1$, with a negative value indicating an anticorrelated motion, and a positive value identifying a positively correlated pattern of dynamics between the two atoms considered.

3.3.2. Identifying important residues in structural changes as captured by normal modes

To identify the residues that are critical to the dynamics

captured by a normal mode (also called mechanical "hot-spot residues"), we use a structural perturbation method (SPM) (Zheng et al. (2005, 2007)). The basic premise of SPM is that, for a given mode k, the dynamic importance of the i-th residue can be assessed by the response to a local perturbation of the potential at i. The perturbation, which in the context of coarse-grained models is realized by small changes in the force constants of those springs (bonded, or non-bonded) that connect i to its neighbors, is similar to a point mutation in probing experiments. This perturbation, or response, is measured in terms of a normalized score $S_{i,k}$ given by

$$S_{i,k} = \frac{N}{2\omega_k^2} \sum_j < V_k, H_{ij} V_k > \tag{49}$$

where N is the total number of atoms in the molecule, ω_k^2 and V_k are the $k-th$ eigenvalue and eigenvector of the Hessian H of the coarse-grained model, respectively, and the summation extends over atoms j that are in contact with i, according to the model considered. Note that each term in this sum can be computed easily using Equation (45).

3.3.3. Comparing two sets of eigenvectors

Eigenvectors provide the direction of motions described by normal modes. Different conformations of a molecule lead to different internal motions and therefore different eigenvectors. Comparing those allows then to identify modes that are robust to conformational changes. Let $\mathbf{V}^A$ and $\mathbf{V}^B$ be two sets of eigenvectors computed from the same potential V over two different conformations $\mathbf{X}^A$ and $\mathbf{X}^B$ of a biomolecule. The similarity I_{kl} between two eigenvectors $\mathbf{V}_k^A$ and $\mathbf{V}_l^B$ is defined as the cosine between their direction:

$$I_{kl} = < \mathbf{V}_k^A, \mathbf{V}_l^B > \tag{50}$$

Note that eigenvectors have unit norm.

In addition, we compute the effective number of normal modes computed from $\mathbf{V}^B$ that are involved in the description of one mode $\mathbf{V}_k^A$ as follows (Brüschweiler (1995); Nicolay and Sanejouand (2006))

$$Neff_k^A = \exp\left[-\sum_{l=1}^{K} \alpha I_{kl}^2 \log\left(\alpha I_{kl}^2\right) \right] \tag{51}$$

where the normalization factor α satisfies $\alpha \sum_{l=1}^{K} I_{kl}^2 = 1$, and K is the number of modes considered. In effect, $Neff_k^A$ is the number of modes from $\mathbf{V}^B$ that have a non-zero overlap with $\mathbf{V}_k^A$; it ranges from 1 to K.

The same analysis can be performed when the two sets of eigenvectors $\mathbf{V}^A$ and $\mathbf{V}^B$ are computed from the same conformation $\mathbf{X}$, with two different coarse-grained potentials. The modes that are identified as conserved in two different models for the energy of the same molecule form the robust modes of the system (Nicolay and Sanejouand (2006)).

3.3.4. Mapping conformational changes onto normal modes

Let us consider a molecular system S with N atoms for which we have two conformations, $\mathbf{X}^A$ and $\mathbf{X}^B$. The change in structure between those two conformations is captured by a displacement vector, $\mathbf{D}$, such that $\mathbf{D} = \mathbf{X}^B - \mathbf{X}^A$.

Let us now consider a set of K normal modes for S computed from its conformation $\mathbf{X}^A$. These normal modes have been computed based on the eigenvalues Ω and eigenvectors V of the Hessian of a coarse-grained potential V. Under the normal mode model, the dynamics of $\mathbf{X}^A$ can be described as a linear superposition of the fundamental motions described by those eigenvectors. The corresponding dynamics that will bring $\mathbf{X}^A$ closer to $\mathbf{X}^B$ is obtained by assigning the weights $\mathbf{W}$ of the modes in this superposition through projections of the displacement vector onto the eigenvectors:

$$\mathbf{W} = V^t \mathbf{D} \tag{52}$$

The contribution of mode k to this optimal collective change of conformation can then be measured as the absolute value of the cosine of the angle between the displacement, and the direction of the mode, given by column $\mathbf{V}_k$ in the matrix V:

$$O_k = \frac{|\langle \mathbf{V_k}, \mathbf{D} \rangle|}{||\mathbf{V_k}|| ||\mathbf{D}||} \tag{53}$$

O_k takes values between 0 and 1, with small values indicating that the mode i contributes little to the conformational change, while large values indicate a significant contribution. We note that $\sum_{k=1}^{3N} O_k^2 = 1$, as the V_k are normalized to 1 and are orthogonal to each other. Then, $SO_K = \sum_{k=1}^{K} O_k^2$ is a measure of the contribution of the first K normal modes to the total overlap between the normal modes of $\mathbf{X}^A$ and the displacement between $\mathbf{X}^A$ and $\mathbf{X}^B$.

Using the optimal weights defined in equation (52), the new conformation $\mathbf{X}_K^A$ of S obtained by applying K of the $3N$ modes of $\mathbf{X}^A$ is given by:

$$\mathbf{X}_K^A = \mathbf{X}^A + \sum_{k=1}^{K} w_k \mathbf{V}_k, \tag{54}$$

where $w_k = \mathbf{W}(k)$. With some basic linear algebra, taking into account the orthogonality of the eigenvectors, it can be shown that the coordinate root mean square (cRMS) distance between $\mathbf{X}_K^A$ and the target conformation $\mathbf{X}^B$, denoted $cRMS_K$, is given by:

$$cRMS_K^2 = cRMS\left(\mathbf{X}^A, \mathbf{X}^B\right)^2 \left(1 - \sum_{k=1}^{K} O_k^2\right) \tag{55}$$

where $cRMS(\mathbf{X}^A, \mathbf{X}^B)^2$ is the square of the coordinate RMS between $\mathbf{X}^A$ and $\mathbf{X}^B$, computed as $\frac{||\mathbf{X}^B - \mathbf{X}^A||^2}{N} = \frac{||\mathbf{D}||^2}{N}$, after application of an optimal rigid body transformation of $\mathbf{X}^B$ onto $\mathbf{X}^A$ to minimize this RMS value. Note that the more we include normal modes, the smaller $cRMS_K$ becomes.

4. Results and discussions

We present results associated with the two topics of this paper, namely an analysis of the new program NormalGo that was developed for computing coarse-grained normal modes of supramolecular systems, including a comparison of the two coarse-grained potentials we have implemented, the elastic potential and a Gō-like potential, and applications of NormalGo to study the structural transitions induced by the binding of antibody fragments onto the envelope of serotype 3 of Dengue virus. All normal modes calculations are performed with standard values for the force constants of the interatomic networks, namely $k_{ij} = 1$ for all edges

(i,j) in the elastic networks, and the force constants of the Gō potential set according to Na and colleagues (Lin and Song (2010); Na et al. (2014); see Methods for details). We note that those values could be optimized to provide better fit with dynamics observed experimentally. Such optimizations are beyond the scope of this paper.

All experiments described below were conducted on an Apple computer with an Intel i7 4 GHz processor with 4 cores (and two hyperthreads per core) and 64 Gb of RAM. The program NormalGo was compiled using the Intel icpc and ifort compilers for its C++ and Fortran components, respectively, and linked against the Intel MKL library for all BLAS and LAPACK procedures that are used within the code.

4.1. The molecular system

DENV is a member of the *flaviviridae* family. There are four serotypes known for DENV. The particles for those different strains share a common structural fold, with their outer shells, or envelopes, having icosahedral symmetry. Those envelopes are formed of 60 asymmetrical units, with each unit containing three copies of the E (i.e. envelope) protein (493 residues in serotype 3 of DENV) and three copies of the M (i.e. membrane) protein (74 residues in DENV3). The high resolution cryo-EM structures of all four serotypes 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 3 of DENV, available at two different temperatures in the PDB database, with codes 3J6S and 3J6T for its conformations at 28 °C and 37 °C, respectively (Fibriansah et al. (2015)). A cartoon representation of the two structures is given in Fig. 2. In this structure, each asymmetrical unit contains 1695 residues (3×493 E protein residues and 3×72 M protein residues) that are formed of 13226 atoms. The full envelope includes 60 copies of this unit, for a total of 101,700 residues. The structure of the envelope of DENV 3 is also known when it is complexed with FAB fragments reflecting the effects of antibody binding to a viral outer shell, PDB code 3J6U (Fibriansah et al. (2015)).

We emphasize that the envelopes were studied using the empty protein shells only, following previous studies of viral particles using normal modes (Tama and Brooks III (2002, 2005); Rader et al. (2005); Chennubotla et al. (2005); Kim et al. (2003); Polles et al. (2013)). This setting is expected to be satisfactory as the stability of the empty envelope is guaranteed by the geometric construction of the coarse-grained model, 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.

We extracted from the three PDB files mentioned above all copies of the E and M proteins. As only the Cα atoms were present in the three PDB files, we got three conformations of the viral envelope of DENV3, referred to as DENV3-28C, DENV3-37C, and DENV3-FAB, each with 101,700 atoms.

4.2. Computing a subset of the normal modes of very large virus outer shells

Computational methods applied to molecular systems that rely on second order analyses of a potential, usually in the form of a Hessian, are often limited to small systems, because of the time and space complexity necessary to manage such a Hessian. This is unfortunately the case for normal mode calculations. The simplified coarse-grained models considered here enable applications of methods that scale linearly with the number of atoms, as they rely on short range interactions only through the introduction of a cutoff. This does not solve however the time complexity associated with computing even a small number of normal modes of such elastic networks when the size of the system becomes large. To assess the importance of this problem, we computed the first 1000 modes associated with the Tirion elastic potential and the Gō-like potential for the DENV3-37C envelope using two different eigenvalue algorithms, the popular Lanczos method, and the Block Chebyshev Jacobi Davidson method described in the Implementation section above. Comparisons of the computing times required by those two methods are provided in Fig. 3.

As a reminder, the envelope of DENV3-37C includes 101,700 Cα atoms. Its elastic network for the Tirion potential was computed using a cutoff $R_c = 15$ Å. It includes 2948042 edges; its corresponding Hessian contains 53,980,056 non-zero values (only 8,844,126 values need to be stored when the tensor representation of this Hessian is used). The non-bonded network for the Gō-like potential was computed using a cutoff $R_g = 25$ Å. We note that the initial implementation of the Gō-like potential did not consider a cutoff and included all pairs of Cα atoms in the molecule of interest (Lin and Song (2010); Na et al. (2014)). We cannot use the same strategy for a large molecule and therefore implemented a cutoff, that was set arbitrarily large. The corresponding non-bonded network includes 10856528 edges, approximately 3.5 times more than in the elastic network for the Tirion potential.

The Lanczos method is the most common method used to compute a small subset of eigenvalues and eigenvectors of a Hessian for normal mode calculations. We have used here what is considered its most robust implementation, ARPACK, with default values for its parameters, and setting the size of the Krylov search

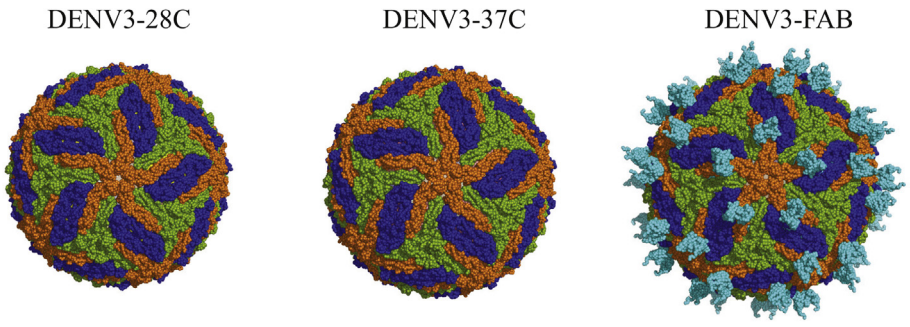

Fig. 2. Outer shells of Dengue serotype 3. Space filling, Cα model of the mature form of the envelope of Dengue Serotype 3 at 28 °C (PDB file 3J6S), at 37 °C (PDB file 3J6T), and in complex with human antibody 5J7 Fab (PDB code 3J6U). The envelope includes 180 copies of both the E protein and the M protein. We omit the latter in the display as they point inside the envelope. The three E proteins from each asymmetric unit are colored green, orange, and blue. The FAB fragments are shown in cyan. All three panels were generated using Pymol (http://www.pymol.org).

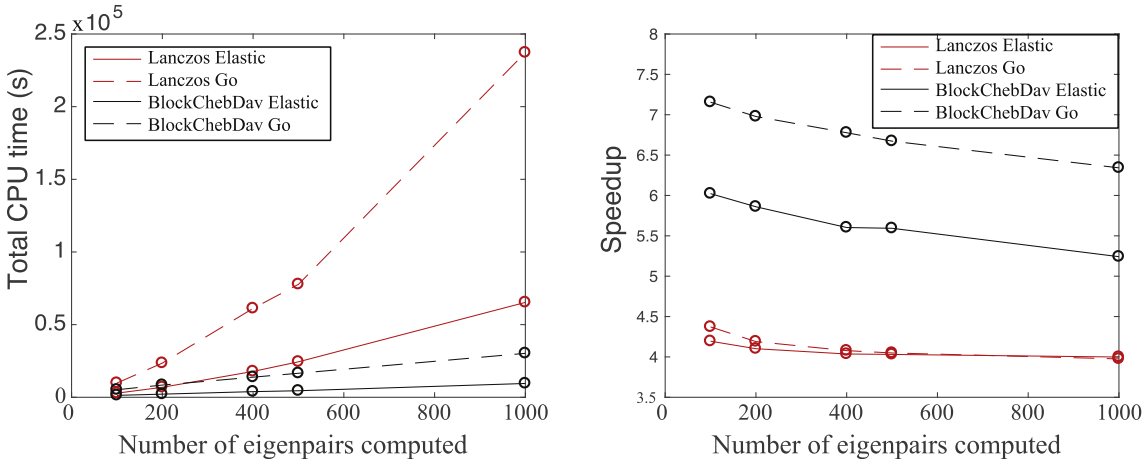

Fig. 3. Computing the first 1000 normal modes for coarse-grained potentials applied on a virus outer shell. A. The total CPU times for two potentials, Elastic and Gō, and two methods for computing eigenvalues of a large symmetric matrix, namely the Lanczos method and the Block Chebyshev-Davidson method, are plotted against the number of eigenvalues computed. **B.** The speedups (computed as the ratio of total CPU time over clock time) are plotted against the same number of eigenvalues computed. All calculations were performed on an Intel i7 4 GHz processor with 4 cores, and two (hyper) threads per core.

space to be twice the number of requested eigenvalues, namely 2× 1000. We find that the Lanczos method becomes inefficient for large viral systems, especially when a "large" number of eigen pairs is needed. The approximate total CPU time to compute the first 1000 modes for the whole DENV3 viral envelope is 65000 s, with a clock time of 16500 s (approximately 4.5 h) for the Tirion potential, and 240,000 s, with a clock time of 60,000 s (approximately 16.5 h) for the Gō-like potential. Of concern is the limited speedup induced by parallelization of the Hessian matrix products and by the use of optimized BLAS and LAPACK routines. This speedup is of the order of 4 for both potentials, compared to an ideal value of 8 on the desktop computer that was used.

The Chebyshev-filtered, block version of the Jacobi Davidson method, BlockChebDav, was found to be significantly more performant. It was run with a block size k_{block} set to 16, a Chebyshev polynomial order of 100, and the total size for the search space set to 2000. Using this method, the approximate CPU time to compute the first 1000 modes for the whole DENV3 viral envelope is 9500 s, with a clock time of 1800 s (approximately 0.5 h, i.e. a speedup of a factor 9 compared to the Lanczos method) for the Tirion potential, and 30,200 s, with a clock time of 4800 s (approximately 1 h 20 min, i.e. a speedup of a factor of 12 compared to the Lanczos method) for the Gō-like potential. The use of a block procedure in BlockChebDav enables efficient parallelization as all the Hessian-vector multiplications can be distributed efficiently over the different processors and as all matrix-matrix products benefit from optimized BLAS3 routines. The corresponding speedup is of the order of 6.5 for the Gō-like potential, and 5.2 for the Tirion potential, both compared to an ideal value of 8. The difference between the two potentials comes from the difference in the size of their corresponding Hessian matrices. As mentioned above, the Hessian for the Gō like potential contains close to 3.5 times more non-zero values than the Hessian for the Tirion potential. As matrix-vector products are highly parallelized using the tensor representations of the Hessians, and their contributions to the total computing costs are higher for the Gō-like potential, we observe a better speedup ratio.

4.3. Comparing the normal modes computed with the Tirion potential and the Gō like potential

The Tirion elastic potential and the Gō like potential are both coarse-grained, geometric based potentials that can be used to compute collective motions within biomolecules, as captured by normal modes. They are both constructed such that any conformation of the molecular system under study is compared to a reference structure for that system, taken to be the input experimental conformation. For both potentials this input conformation defines their unique minimum. However, their similarities stop there. The Tirion potential only included pairwise interactions, with no distinction of local, and non-local contacts. The Gō like potential is formed of a combination of local interactions and global interactions, with the former including pair-wise and higher order interactions, namely three-atom terms for bond angles, and four-atom terms for dihedral angles. Even the non-bonded terms in the Gō like potential differ from the pair-wise terms in the Tirion potential as they are represented with a 12-10 Lennard-Jones functional, rather than with a Hooke potential. While both potentials are first approximated with a second-order Taylor expansion prior to being used for normal mode analysis, it is unclear whether they capture the same dynamics for the system of interest. In this section, we compare the normal modes generated by those potentials on viral outer shells, focusing on the spectra of their eigenvalues, the differences (and similarities) of their eigenvectors, the dynamics they generate, and the geometric deformations that can result from these coarse-grained models.

4.3.1. Eigenvalue spectra for the Tirion potential and the Gō like potential

In Fig. 4 we compare the frequencies of the first fifty normal modes of the Tirion potential and the Gō-like potential of the E-protein, and of the full envelope of DENV3-37C. The E protein includes 393 Cα atoms, while the whole envelope is comprised of 101,700 atoms. The potentials were computed as described in the previous section. As expected, the first six frequencies are found equal to zero, for all systems considered, as those frequencies correspond to the rigid motions (three translations and three rotations). The spectra of frequencies of the normal modes computed for the full virus envelope are shifted towards lower frequencies compared to the E-protein alone, indicating the presence of more collective motions in protein oligomers. Of significance is the presence of degeneracies in the spectra of the full envelope, namely repeating frequencies with order 1, 2, 3, 4, or 5, as expected from the icosahedral symmetries of the envelopes. The same

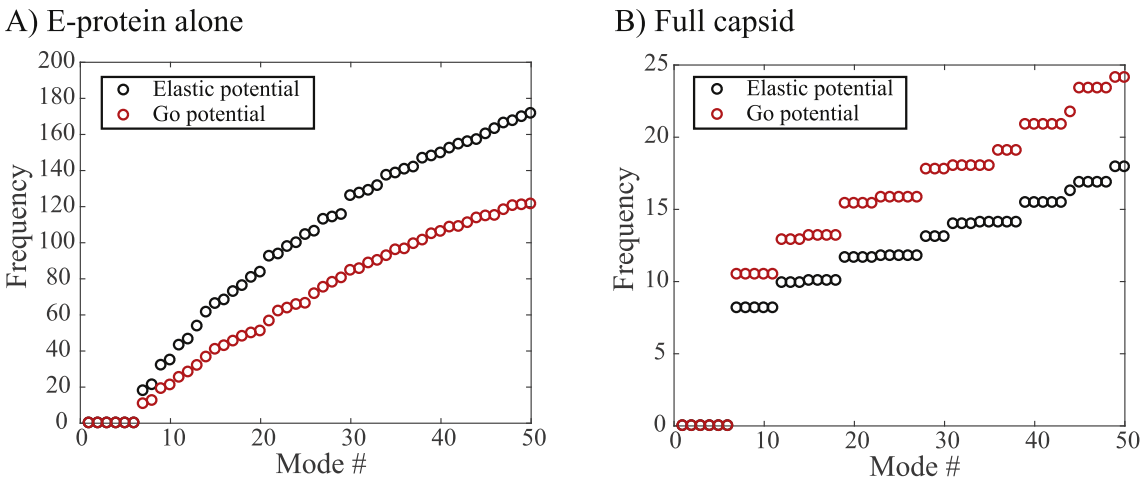

Fig. 4. Normal modes of DENV3. The frequencies of the first fifty normal modes of one E-protein (panel A) and of the full viral envelope (panel B) of DENV3 are plotted against the mode number, for both the elastic potential (black), and the Gō-like potential (red). Note in (B) the similarity between the two spectra, both in terms of degeneracies of the modes, and in terms of relative differences between the frequencies. For both panels the frequencies are in arbitrary units, as the force constants that define the potentials are also in arbitrary units.

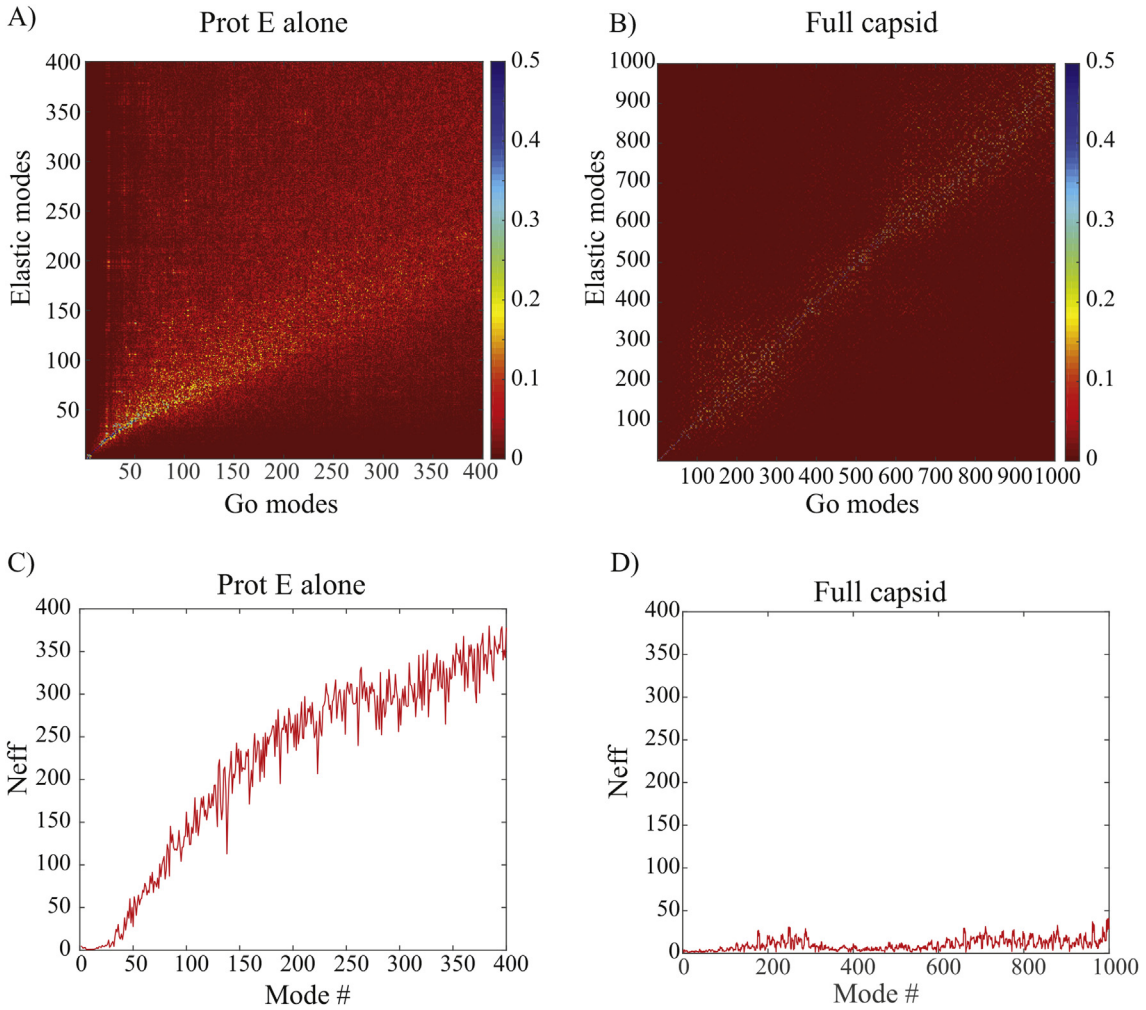

Fig. 5. Comparing the eigenvectors associated with the Tirion potential and the Gō potential. (A) and (B) Overlaps (computed as dot products) between the eigenvectors of the Hessian of the elastic Tirion potential (y axis), and the Gō-like potential (x axis) for one E protein in DENV3-37C, and for the corresponding full viral envelope.(C) and (D) The effective number of elastic Tirion modes needed to describe one Gō mode, as a function of the Gō mode number, for one E protein, and for the full viral envelope.

degeneracies are observed with both potentials. Overall, these two potentials lead to normal modes with eigen frequencies of similar magnitude, and similar spectra, at least in the low frequency range, despite being computed with significantly different models and parameters.

4.3.2. Comparing the eigenvectors of the Hessians of the Tirion potential and Gō like potential

The spectra shown in Fig. 4 provide a comparison of the eigenvalues of the Hessians of the two coarse grained models, namely of the frequencies of the normal modes of the molecule of interest. Of similar significance, we compared the corresponding eigenvectors of the Hessians, i.e. the directions of those normal modes. Fig. 5 illustrates their similarities and differences.

A normal mode k captures the collective motion of the atoms of a system. This information is given in the form of a vector $\mathbf{V}_k$, i.e. an eigenvector of the Hessian of the potential considered, such that $\mathbf{V}_k(i)$ defines the relative contribution of atom i to the collective motion. Two different models for the coarse-grained potential lead to two different Hessians, and therefore to two different sets of eigenvectors, $\mathbf{V}_k^E$ and $\mathbf{V}_k^G$, for the Elastic Tirion potential, and for the Gō-like potential, respectively. We first compare those two sets of eigenvectors by computing all pairwise dot products between them; results are shown in Fig. 5A and B for one E protein in DENV3-37C, and for the corresponding full virus envelope, respectively. For the E protein alone, only a fraction of the eigenvectors are conserved between the two coarse-grained models, while for the full viral envelope, the similarities between the two models extend to a much larger number of eigenvectors. To further quantify the differences between the two different coarse grained models, we computed the effective number N_{eff} of Tirion modes that are needed to describe one Gō mode k using Equation (51). In effect, N_{eff} provides the number of non-zero overlaps between the Gō mode k and all Tirion modes. A small number for N_{eff} means that this mode k is conserved between the two models, or, using the terminology of Nicolay and Sanejouand (2006), that k is a robust mode, while a large number for N_{eff} indicates that mode k is only diffusely present in the Tirion modes. For the E protein alone, the first 8 non-zero modes are nearly perfectly conserved between the two models, with N_{eff} below 1.5, and 19 non-zero modes have a N_{eff} value below 5. The diffusion of the Gō modes within the Tirion modes however grows really fast, with N_{eff} larger than 50 for Gō modes with indices larger than 50. In contrast, the modes for the full envelope are much better conserved between the two models, with 260 modes among the first 1000 Gō modes having an N_{eff} below 5, and all those first 1000 modes with N_{eff} below 50. We note however that the different behaviors observed on the E protein alone, and on the full viral envelope are not inconsistent with each other: the first 20 non zero modes of the E protein represent approximately 2% of the full spectrum of modes, while the first 261 modes for the full envelope also represent approximately 1% of the corresponding full spectrum.

4.3.3. Correlated dynamics of E proteins under two coarse grained normal mode models

The comparisons between the Tirion and the Gō normal modes presented above focus on the eigenvalues and eigenvectors of the corresponding Hessian matrices. However, ultimately, these comparisons should be focused on directly checking for differences in the corresponding motions observed in the molecular system under study. This is done in Fig. 6, in which we compare the cross correlation matrices for the dynamics of one isolated E protein, and of one E protein as part of the full envelope of DENV3-37C, for the two coarse grained normal mode models.

The cross correlation matrices (CCMs) computed from the Tirion and the Gō normal modes are remarkably similar for both systems: for the E protein alone, the max-distance between the CCMs for the Tirion and the Gō is 0.4, while for the full envelope, the same max-distance is 0.2. In the previous section, we had shown that there are only 20 Gō modes that are robust for the E protein alone; the fact that the corresponding CCMs are well conserved between the two models re-emphasize the importance of the low frequency modes for describing the dominant, correlated motions in a protein. These matrices resemble those obtained on DENV1, and on the Zika virus (Hsieh et al. (2016)). Briefly, the CCMs for the E protein alone reveal 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 full envelope, the same positive correlations are observed within each of the three domains. However, the interactions between the domains change significantly. In the full envelope, the dynamics of domain I is strongly correlated to the dynamics of domain III, as opposed to being anti-correlated in the E protein alone.

4.3.4. The Gō model leads to less deformation of the structure during normal mode dynamics

The previous three subsections have shown that the Tirion model and the Gō model lead to very similar low-frequency normal modes for a molecular system, which then leads to similar correlated motions within this system. The question arises then as to which of these two models should be used, if they are so similar. To answer this question, we applied normal mode mapping to the structural changes between two known conformations of a protein, following the model originally proposed by Petrone and Pande (2006). We repeated their analyses for DENV3, and studied the structural changes in its envelope between its mature form at 28 °C, DENV3-28C, and its mature form bound with FAB fragments, DENV3-FAB (see Fig. 2 above). We computed the first 4000 modes of DENV3-28C, using both the Tirion model and the Gō model. We then assessed how those modes can be used to map the conformational changes. The mapping was measured using a coordinate RMS between the conformation generated by combining the normal modes, and the target conformation (see section 3.3.4 above). In parallel, we monitored the resulting deformations in the lengths of the Cα-Cα pseudo bonds within the intermediate structures that are generated. Results are shown in Fig. 7.

As we keep adding modes to the computed collective motions mapped onto the conformational changes between the structures of DENV3-28C and DENV3-FAB, the cRMS between the mapped conformation and the target conformation decreases. Overall, this cRMS decreases from 2.79 to 1.43 Å, for both the Tirion model and the Gō model. We note that we do not expect these cRMS values to drop to 0, even if we were to include all normal modes. Normal modes only capture collective, harmonic motions. In addition, Cartesian normal modes based on a simple potential such as the elastic potential, when weighted with possibly large amplitudes, are likely to stretch the structures in non-physical ways. This is where the two coarse-grained normal mode models were found to differ the most. The first 1000 modes lead to the largest drop in the cRMS value. Over these 1000 modes, the Cα-Cα pseudo bond length is distorted on average by 0.2 Å with the Tirion model, and by only 0.025 Å for the Gō model. For even larger number of modes, the distortion increases significantly for the Tirion model, reaching on average 0.6 Å, while the same distortion never goes above 0.3 Å for the Gō model. As the latter explicitly contains terms that ensure the

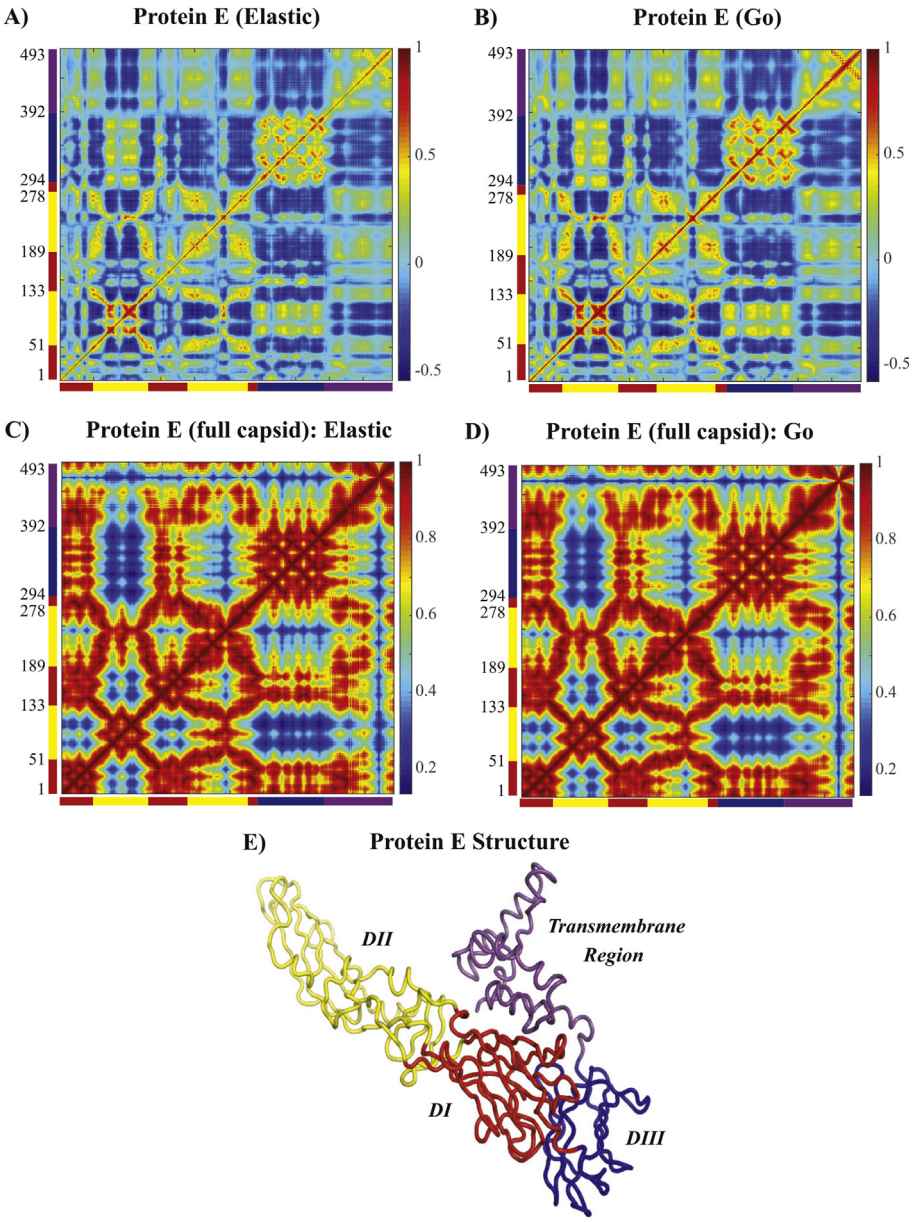

Fig. 6. Correlated motions in the DENV3-37C E protein. Cross Correlation Matrices (CCM) obtained from the first 1000 modes for the E protein alone and the E protein in the whole envelope for two coarse-grained normal mode models, the Tirion model, and the Gō-like model. Those plots show correlations between the motions of Cα atoms. Both axes of a matrix are the residue index within the E protein. Each cell in a matrix shows the correlation between the motions of two residues (Cα atoms) in the protein on a range from −1 (anticorrelated, blue) to 1 (correlated, red), with 0 conferring no correlation. In panel E, the E protein is shown in cartoon mode. The color code for the structure in E) as well as for the X and Y axes of the CCM plots follows the standard designation of the E protein domains I (red), II (yellow) and III (blue). The transmembrane domain is shown in purple. Panel E was generated using Pymol.

preservation of local, pseudo-bond properties of the molecule, it is not surprising that it does better in maintaining the local geometry. We will therefore only use the Gō model in the remaining part of this study.

4.4. Structural transitions in mature forms of DENV3 induced by binding of FAB fragments

DENV, a member of the Flaviridae family, is one of the most important viruses that target humans. It is estimated that 400 million people worldwide are infected by DENV every year, with approximately 100 millions of those showing direct signs of infections, with diverse levels of severity (Bhatt et al. (2013)). As mentioned above, the nucleocapsid of DENV is surrounded by a lipid bilayer membrane, itself surrounded by an external envelope consisting of 180 E proteins and 180 M proteins. E proteins play an important role in the entry of the virus into the host cell as it facilitates the fusion of the virus with the host cell membrane (Modis et al. (2004)). Mapping of potential epitopes, i.e. antibody bind sites, on the E protein is therefore crucial for vaccine development. Here we propose a dual approach to understanding antibody binding to E proteins. Instead of directly searching for residues that may bind to an antibody, we identify those residues in the E protein that are indirectly important for antibody binding by studying the structural transition between a FAB-free conformation of the DENV envelope, and a FAB-bound conformation of the same virus. Those residues are essential for allowing structural changes. We use DENV3, as the structure of the mature form of its envelope is

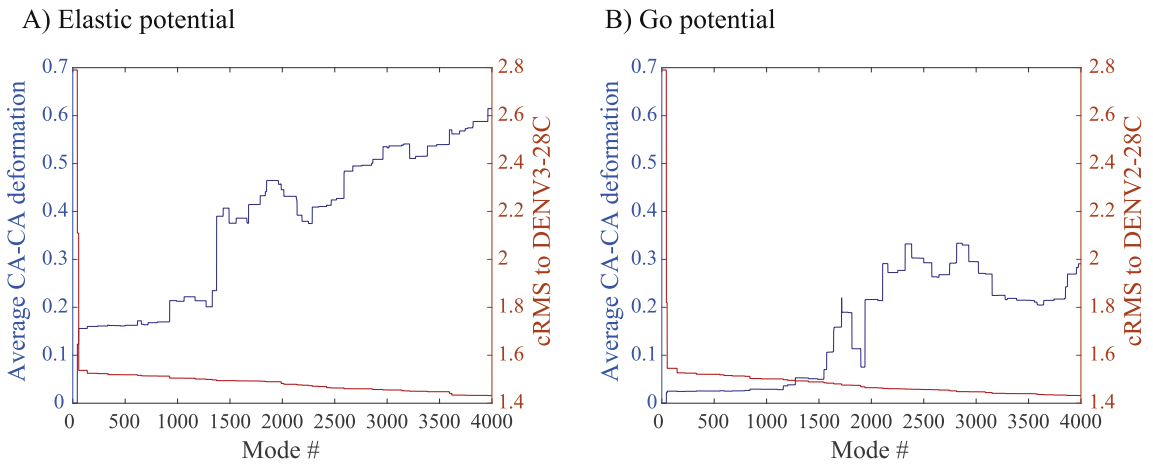

Fig. 7. cRMS between the best mapped conformation and the target conformation (right axis) for the transition between DENV3-37C and DENV3-FAB (which has FAB fragments bound to it, see Fig. 2) (right axis, red), and average distortion of the Cα-Cα pseudo bond length (left axis, blue) as a function of the normal mode number, for the elastic, Tirion normal mode model (A), and the Gō normal mode model (B).

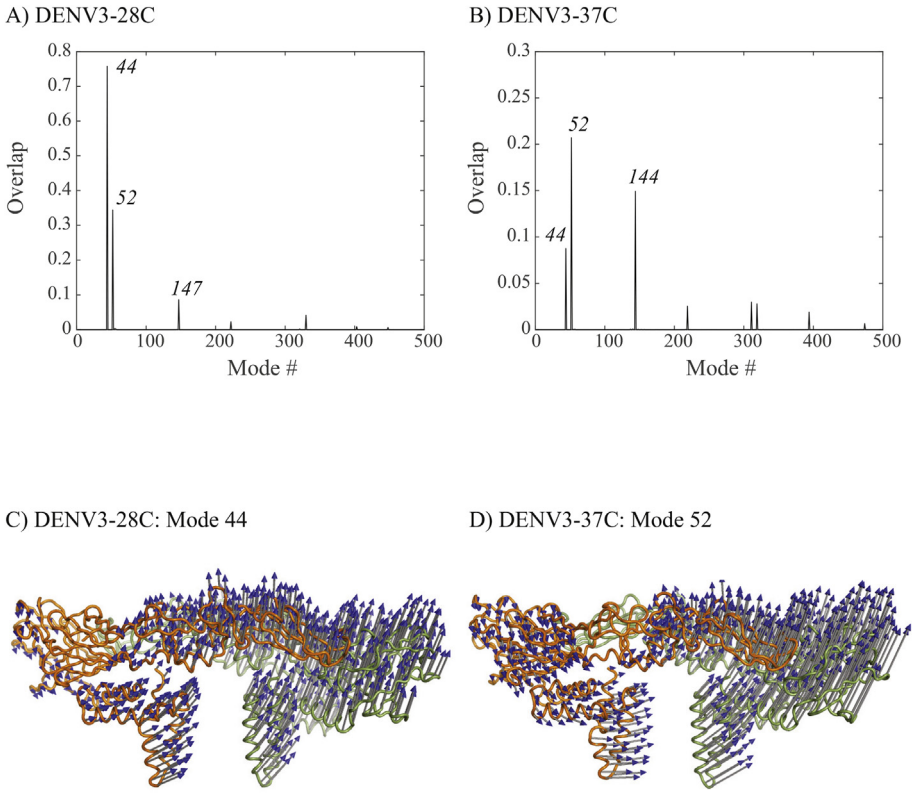

Fig. 8. Overlaps between normal modes and conformational changes for the transition between the mature form of DENV3 at 28 °C and the same mature form with FAB fragments bound to it (panel A), and for the transition between the mature form of DENV3 at 37 °C and the same FAB-bound conformation (panel B). The normal modes were computed based on the FAB-free structures, using the Gō coarse-grained model. Modes are indexed based on increasing frequencies. The indices of the key modes (i.e. those with contribution higher than 5%) are provided. Panels C and D show "porcupine plots" of the large-scale collective motions associated with mode 44 for DENV3 at 28 °C (panel C) and with mode 52 for DENV3 at 37 °C, shown on one E-protein dimer. The orientation and length of the blue cones indicate the direction and the relative amplitude of motion of the Ca atoms. The porcupine plots were generated using Pymol.

known at two different temperatures, 28 °C (DENV3-28C) and 37 °C (DENV3-37C), and the structure of its envelope is also known when it is complexed with FAB fragments (DENV3-FAB).

We computed the first 4000 normal modes of the two mature forms using the Gō model, as described above, and assessed how those modes can be used to map the conformational changes with the two other forms of the envelope. We note first that those changes of conformation are relatively modest: the cRMS between DENV3-28C and DENV3-FAB is 2.79 Å, while the cRMS between DENV3-37C and DENV3-FAB is 1.64 Å (in comparison, the cRMS between DENV3-28C and DENV3-37 is 2.74 Å). The mapping between normal modes and conformational changes is measured

using an overlap value (see section 3.3.4). Results are shown in Fig. 8. We find that for the transition DENV3-28C to DENV3-FAB, the main conformational changes occur in the lowest frequency modes, i.e. within the first 200 modes. These modes contribute 84% of the transition; this number increases only to 86% when all 4000 modes are included. This transition is therefore dominated by a few highly collective modes. The transition DENV3-37C to DENV3-FAB is more diffuse and requires more modes. In this case, the first 200 modes contribute 27% to the transition; this number increases to 44% when the 4000 modes are included. It is therefore expected that many more localized motions (i.e. those with higher frequencies) drive this transition.

From Fig. 8, we identify mode 44 as the dominant mode for the DENV3-28C to DENV3-FAB transition (representing 76% of the transition), and mode 52 as the dominant mode for the DENV3-37C to DENV3-FAB transition (representing 21% of the transition). Those modes lead to similar motions within one E protein and its closest neighbour, as illustrated with the porcupine plots in panels C and D of Fig. 8. From those modes it is possible to identify the important residues, also called "hot-spot residues" using a structural perturbation method (SPM) (Zheng et al. (2005, 2007)). The basic premise of this technique is that, for a given mode, the dynamic importance of the i-th atom can be assessed by computing the response to a local perturbation in the potential at i, using Equation (49). We computed those response scores over a whole raft of the DENV3 envelope, which contains 6 E proteins, for the two dominant modes identified above. Results for all residues in the unit are shown in Fig. 9, while the hot-spot residues, i.e. those with the highest response scores, are listed in Table 1.

The E proteins of DENV3 form local structures within the virus envelope, referred to as raft. A raft includes six E proteins forming 3 dimers arranged in a parallel manner, resulting from the combination of two asymmetrical units. The whole envelope contains 30 such rafts. In mode 44 of DENV3-28C, the SPM response is

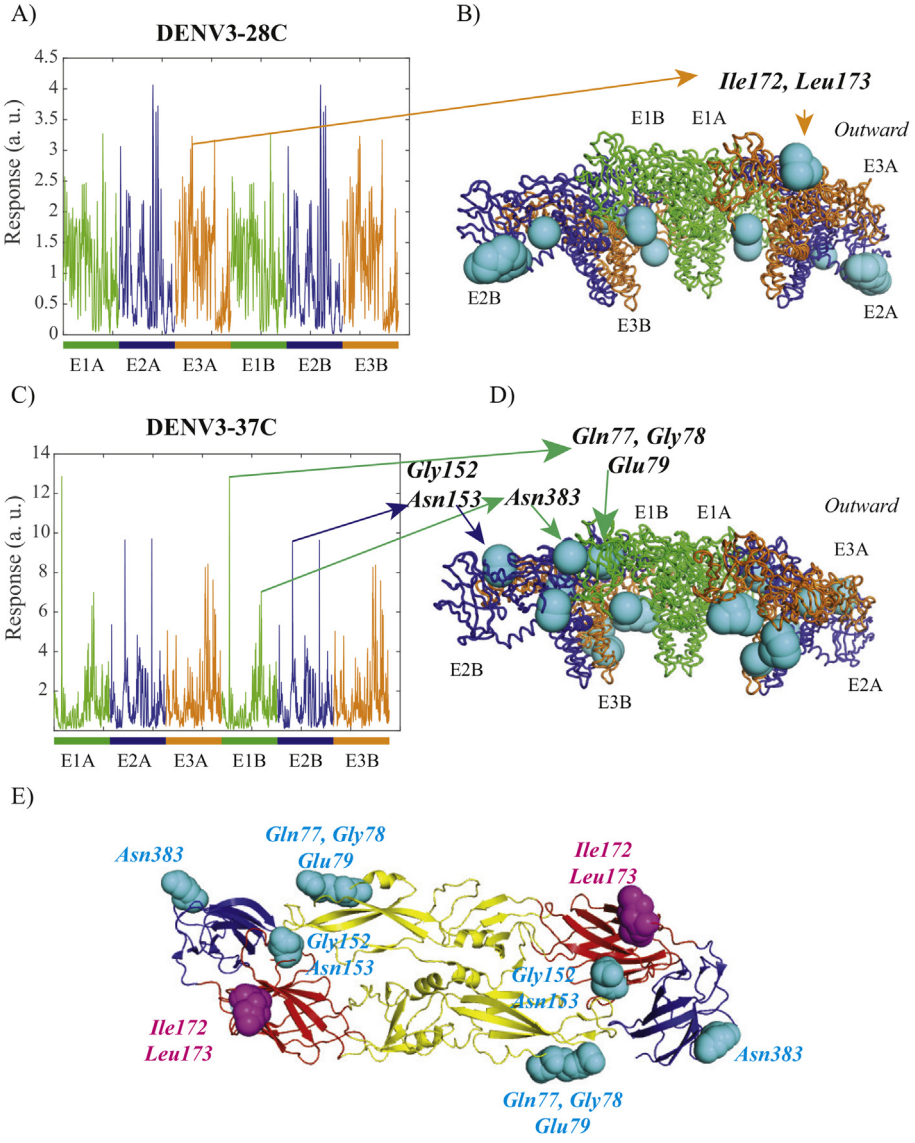

Fig. 9. SPM response scores for all residues in a raft of DENV3 (that contains 6 E-proteins) for normal mode 44 of DENV3-28C (panel A), and for mode 52 of DENV3-37C (panel C). The higher those response scores, the more important the residues are for the motion captured by the normal modes. Structures of the assymmetric unit of DENV3-28C (panel B) and DENV3-37C (panel D), with the residues with the highest SPM response scores, referred to as host-spot residues, that point outward highlighted using a sphere representation. Panel E: cartoon representation of the E protein dimer (from PDB code 1UZG), with the outward pointing hot-spot residues represented as spheres, in magenta for those from DENV3-28C, and in cyan for those from DENV3-37C. The molecular images were generated using Pymol.

Table 1
Important residues in the transition from FAB-free to FAB-bound conformations of DENV3.

E Protein domain [a]	DENV3-28C [b]	DENV3-37C [c]
Domain I	**Ile172**, **Leu173**	**Gly152**, **Asn153**
Domain II		**Gln77**, **Gly78**, **Glu79**
Domain III	Asp339, Asp340, Gln341 Gly342, Lys343, Ala344 Asn346, Gly372, Glu373 Lys392, Gly393	**Asn383**
Trans Membrane	Phe400	Ile396, Gly397, Lys398 Met399, Phe400, Thr403 Ala404, Val423, Gly424 Gly425, Val426, Leu427 Cys477, Ile478

[a] See Fig. 6 for a definition of the domains. Important residues that faces outward, i.e. away from the lipid membrane of the nucleocapsid, are highlighted in bold.
[b] Only those residues whose SPM scores are above 3.1.
[c] Only those residues whose SPM scores are above 6.3.

conserved between the two subunits of a raft (see Fig. 9A). Defining hot-spot residues as those whose response scores are higher than 3.1, we find that most of those residues points towards the inner membrane that sits on top of the nucleocapsid, as illustrated in Fig. 9B, and listed in Table 1. There is however a pocket of residues within domain I of E protein that points outward, with high response scores, namely Ile172, Leu173 (and to a lesser extent their immediate neighbors in the sequence). It is interesting to notice that none of those residues directly interact with the FAB fragments (see Table 1 of Fibriansah et al. (2015)). However, they do play a key role in allowing these FAB fragments to bind to the viral envelope.

In mode 52 of DENV3-37C, the SPM response is again conserved between the two subunits of a raft (see Fig. 9C). Setting a cutoff of 6.3 to define hot-spot residues, we find that most of those residues point towards the inner membrane, just like for DENV3-28C (see Fig. 9C, and Table 1). However, in the case of DENC3-37C we observe three pockets of residues that point outward and that have high response scores. Those pockets are centered around residues Gly152 and Asn153 of domain I, residues Gln77, Gly78, and Glu79 of domain II, and residue Asn383 of domain III. As observed with DENV3-28C, none of those residues directly interact with the FAB fragments (Fibriansah et al. (2015)).

5. Concluding remarks

Our interests in this paper are three fold. First, we proposed a new formalism for characterizing the Hessian of a coarse-grained quadratic potential using tensor products. This formalism can be used for computing normal modes of very large viral systems. This formalism is developed for the pairwise potential originally proposed by Tirion (1996), as well as for a Gō like potential (Clementi et al. (2000); Lin and Song (2010)). When combined with a fast method for computing some eigenpairs of the Hessian, this new formalism enables the computation of up to 1000 normal modes of a full viral envelope with more than 100,000 atoms in less than 1 h and 30 min (clock time) on a standard desktop computer. Second, we compared the two coarse-grained potentials mentioned above with respect to the eigenvalues and eigenvectors of their Hessians, as well as the collective motions they generate. We have shown that, despite significant differences in their formulations, the Tirion and the Gō like potentials capture very similar dynamics characteristics of the molecules under study. However, the Gō like potential should be preferred as it leads to significantly less local deformations in the structure of the molecule during normal mode dynamics. This is probably due to the fact that the Go-like potential has bond, angle, and torsion angle terms that are designed to preserve the geometry of the molecule. Finally, we used coarse-grained normal mode analysis based on the Gō like potential to characterize the structural transitions that occur when FAB fragments bind to the envelope of serotype 3 of the Dengue virus. Using the concept of Structural Perturbation Method (SPM), we have identified residues at the surface of the envelope that are important for the transition between the FAB-free and FAB-bound conformations, and therefore important for the binding of antibodies to Dengue viruses.

The tensor formalism for computing the Hessians of the Tirion and Gō like potential allows for simplified representations of those Hessians, and more importantly for better parallelization of Hessian - vector multiplications. However, the most important improvements for computing some eigen pairs of those Hessians came from the implementation of the Block Chebyshev Jacobi Davidson (BlockChebDav) algorithm (Zhou (2010)). We have recently tested other recent algorithms for computing those eigen pairs and found BlockChebDav to be the most efficient (P. Koehl, submitted). We do not exclude however that there are other algorithms available that may result in even further improvements. In addition, the current implementation of BlockChebDav is found to be limited to computing the first 5000 eigen pairs of a (very large) Hessian. We will need new implementations if more eigen pairs are needed; we are currently looking at solutions to this problem, such as the spectrum slicing methods (Lin et al. (2016)).

The new representation of the Hessian of a quadratic potential, combined with the fast method BlockChebDav for computing a subset of the normal modes associated with the Hessian, is a general formalism that can be applied to any large molecular system. In this paper, we have applied it to the outer shells of viral systems. In those specific cases, the computations of normal modes could have been simplified by using either the symmetry properties of the outer shells (Simonson and Perahia (1992); van Vlijmen and Karplus (2005); Peeters and Taormina (2009)), or by looking for specific relationships between the normal modes of the full viral shells and those of its components (Na and Song (2018)). We argue here that such simplifications are not necessary, opening the doors to studying the dynamics of general large molecular systems using normal modes.

The Gō like potential considered here is identical to the potential introduced by Lin and Song (2010), with two differences. First, we rewrote completely the computation of its Hessian, as it is significantly simplified using the tensor representation. Second, we introduced a cutoff for the non-bonded terms in the potential, that was not used in the original implementation. The value given to that cutoff, namely 25 Å, is somewhat arbitrary and should be studied in details. This Gō potential was found to better preserve local geometry of the molecules during normal mode dynamics. However, in this paper we have only looked at modest ranges of motion for those normal modes. Internal coordinate normal modes (Noguti and Go (1982); Levitt et al. (1985); Lopez-Blanco et al. (2011)) combined with Gō potential would certainly be even more robust to deformation. It is unclear at this stage if they can be adapted to studying very large molecular systems.

Using the coarse-grained normal mode analysis based on the Gō like potential, we were able to identify hot-spot residues of the envelope of DENV3, namely residues that are important for the conformational transition involved in the binding of FAB fragments to the surface of the envelope. Further analysis is needed to ensure that these “dynamically important” residues are conserved among all DENV epitopes, but in principle this prediction of epitopes should be useful for the design of neutralizing antibodies in general. We plan to make this tool available on a web page in the near future.

Acknowledgement

Patrice Koehl acknowledges support from the Ministry of Education - Singapore through Grant Number: MOE2012-T3-1-008. Some of the ideas discussed here originated from a workshop organized by the Institute for Mathematical Sciences, National University of Singapore during the summer in 2017. We thank their hospitality and financial support.

References

Abe, H., Gō, N., 1981. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Applications to two-dimensional lattice proteins. Biopolymers 20, 1013–1031.
Arnoldi, W., 1951. The principle of minimized interations in the solution of the matrix eigenvalue problem. Q. Appl. Math. 9, 17–29.
Atilgan, A., Durell, S., Jernigan, R., Demirel, M., Keskin, O., Bahar, I., 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. 80, 505–515.
Bahar, I., Atilgan, A., Erman, B., 1997. Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. Folding Des. 2, 173–181.
Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P., 2000. The protein Data Bank. Nucleic Acids Res. 28, 235–242.
Bhatt, S., Gething, P., Brady, O., Messina, J., Farlow, A., Moyes, C., Drake, J., Brownstein, J., Hoean, A., Sankoh, O., Myers, M., George, D., Jaenisch, T., Wint, G., Simmons, C., Pan, J.J., Farrar, T.W.S., Hay, S., 2013. The global distribution and burden of Dengue. Nature 496, 504–507.
Blondel, A., Karplus, M., 1996. New formulation for elimination of singularities. J. Comput. Chem. 17, 1132–1141.
Bond, P., Cuthbertson, J., Deol, S., Sansom, M., 2004. MD simulations of spontaneous membrane protein/detergent micelle formation. J. Am. Chem. Soc. 126, 15948–15949.
Brooks, B., Bruccoleri, R., Olafson, B., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187–217.
Brooks, B., Karplus, M., 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. Unit. States Am. 80, 6571–6575.
Brüschweiler, R., 1995. Collective protein dynamics and nuclear spin relaxation. J. Chem. Phys. 102, 3396–3403.
Chennubotla, C., Rader, A., Yang, L., Bahar, I., 2005. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. Phys. Biol. 2, S173–S180.
Clementi, C., Nymeyer, H., Onuchic, J., 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. J. Mol. Biol. 298, 937–953.
Dill, K., Ozkan, S., Shell, M., Weikl, T., 2008. The protein folding problem. Annu. Rev. Biophys. 37, 289–316.
Dykeman, E., Sankey, O., 2008. Low frequency mechanical modes of viral capsids: an atomistic approach. Phys. Rev. Lett. 100, 028101.
Dykeman, E., Sankey, O., 2010. Normal mode analysis and applications in biological physics. J. Phys. Condens. Matter 22, 423202.
Eyal, E., Yang, L., Bahar, I., 2006. Anisotropic network model: systematic evaluation and a new web interface. Bioinformatics 22, 2619–2627.
Fibriansah, G., Tan, J., Smith, S., de Alwis, R., Ng, T., Kostyuchenko, V., Jadi, R., Kukkaro, P., de Silva, A., Crowe, J., Lok, S., 2015. A highly potent human antibody neutralizes dengue virus serotype 3 by binding across three surface proteins. Nat. Commun. 6, 6341.
Gō, N., Abe, H., 1981. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. Biopolymers 20, 991–1011.
Hagan, M., 2014. Modeling viral capsid assembly. Adv. Chem. Phys. 155, 1–68.
Haliloglu, T., Bahar, I., Erman, B., 1997. Gaussian dynamics of folded proteins. Phys. Rev. Lett. 79, 3090–3093.
Hochstenbach, M., Notay, Y., 2006. The Jacobi-Davidson method. GAMM-Mitteilungen 29, 368–382.
Hsieh, Y.-C., Poitevin, F., Delarue, M., Koehl, P., 2016. Comparative normal mode analysis of the dynamics of DENV and ZIKV capsids. Frontiers Bio. Sci. 3, 85.
Ichiye, T., Karplus, M., 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins Struct. Funct. Genet. 11, 205–217.
Kar, P., Feig, M., 2014. Recent advances in transferable coarse-grained modeling of proteins. Adv. Protein Chem. Struct. Biol. 96, 143–180.
Karplus, M., 2014. Development of multiscale models for complex chemical systems: from H+H_2 to biomolecules (nobel lecture). Angew. Chem., Int. Ed. Engl. 53, 9992–10005.
Kim, M., Jernigan, R., Chirikjian, G., 2003. An elastic network model of HK97 capsid maturation. J. Struct. Biol. 143, 107–117.
Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A., Kolinski, A., 2016. Coarse-grained protein models and their applications. Chem. Rev. 116, 7898–7936.
Kondrashov, D., Cui, Q., Phillips Jr., G., 2006. Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. Biophys. J. 91, 2760–2767.
Kostyuchenko, V., Chew, P., Ng, T., Lok, S., 2014. Near-atomic resolution cryo-electron microscopic structure of dengue serotype 4 virus. J. Virol. 88, 477–482.
Kostyuchenko, V., Lim, E., Zhang, S., Fibriansah, G., Ng, T.-S., Ooi, S., Shi, J., Lok, S.-M., 2016. Structure of the thermally stable Zika virus. Nature 533, 425–428.
Kostyuchenko, V., Zhang, Q., Tan, J., Ng, T., Lok, S., 2013. Immature and mature dengue serotype 1 virus structures provide insight into the maturation process. J. Virol. 83, 7700–7707.
Levitt, M., 2014. Birth and future of multiscale modeling for macromolecular systems (nobel lecture). Angew. Chem., Int. Ed. Engl. 53, 10006–10018.
Levitt, M., Sander, C., Stern, P., 1985. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. 181, 423–447.
Levitt, M., Warshel, A., 1976. Computer simulation of protein folding. Nature 253, 694–698.
Lezon, T., Shrivastava, I., Yan, Z., Bahar, I., 2010. Elastic network models for biomolecular dynamics: theory and application to membrane proteins and viruses. In: Boccaletti, S., Latora, V., Moreno, Y. (Eds.), Handbook on Biological Networks. World Scientific Publishing Co, Singapore, pp. 129–158.
Li, G., Cui, Q., 2002. A coarse-grained normal mode approach for macromolecules: an efficient implementation and applications to CA2+-ATPase. Biophys. J. 86, 743–763.
Li, M., Zhang, J., Xia, F., 2016. A new algorithm for construction of coarse-grained sites of large biomolecules. J. Comput. Chem. 37, 795–804.
Lin, L., Saad, Y., Yang, C., 2016. Approximating spectral densities of large matrices. SIAM Rev. 58, 34–65.
Lin, T.-L., Song, G., 2010. Generalized spring tensor models for protein fluctuation dynamics and conformation changes. BMC Struct. Biol. 10, S3.
Lopez-Blanco, J., Chacon, P., 2016. New generation of elastic network models. Curr. Opin. Struct. Biol. 37, 46–53.
Lopez-Blanco, J., Garzin, J., Chacon, P., 2011. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics 27, 2843–2850.
Mahajan, S., 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.
Marrink, S., Risselada, H., Yefimov, S., Tieleman, D., Vries, A.D., 2007. The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B 111, 7812–7824.
Marzinek, J., Holdbrook, D., Huber, R., Verma, C., Bond, P., 2016. Pushing the envelope: dengue viral membrane coaxed into shape by molecular simulations. Structure 24, 1410–1420.
Ming, D., Wall, M., 2005. Allostery in a coarse-grained model of protein dynamics. Phys. Rev. Lett. 95, 198103.
Modis, Y., Ogata, S., Clements, D., Harrison, S., 2004. Structure of the Dengue virus envelope protein after membrane fusion. Nature 427, 313–319.
Na, H., Lin, T.-L., Song, G., 2014. Generalized spring tensor models for protein fluctuation dynamics and conformation changes. Adv. Exp. Med. Biol. 805, 107–135.
Na, H., Song, G., 2018. Fast normal mode computations of capsid dynamics inspired by resonance. Phys. Biol. 15, 046003.
Nerenberg, P., Head-Gordon, T., 2018. New developments in force fields for biomolecular simulations. Curr. Opin. Struct. Biol. 49, 129–138.
Nicolay, S., Sanejouand, Y.-H., 2006. Functional modes of proteins are among the most robust. Phys. Rev. Lett. 96, 078104.
Nielsen, S., Lopez, C., Srinivas, G., Klein, M., 2004. Coarse grain models and the computer simulation of soft materials. J. Phys. Condens. Matter 16, R481–R512.
Noguti, T., Go, N., 1982. Collective variable description of small-amplitude conformational fluctuations in a globular protein. Nature 296, 776–778.
Peeters, K., Taormina, A., 2009. Group theory of icosahedral virus capsid vibrations: a top-down approach. J. Theor. Biol. 256, 607–624.
Petrone, P., Pande, V., 2006. Can conformational change be described by only a few normal modes? Biophys. J. 90, 1583–1593.
Polles, G., Indelicato, G., Potestio, R., Cermelli, P., Twarock, R., Micheletti, C., 2013. Mechanical and assembly units of viral capsids identified via quasi-rigid domain decomposition. PLoS Comput. Biol. 9, e1003331.
Quentrec, B., Brot, C., 1973. New method for searching for neighbors in molecular dynamics computations. J. Comput. Phys. 13, 430–432.
Rader, A., Vlad, D., Bahar, I., 2005. Maturation dynamics of bacteriophage HK97 capsid. Structure 13, 413–421.
Reddy, T., Sansom, M., 2016. The role of the membrane in the structure and biophysical robustness of the Dengue virion envelope. Structure 24, 375–382.
Riniker, S., Allison, J., vanGunsteren, W., 2012. On developing coarse-grained models for biomolecular simulation: a review. Phys. Chem. Chem. Phys. 14, 12423–12430.
Sanejouand, Y., 2013. Elastic network models: theoretical and empirical foundations. Methods Mol. Biol. 914, 601–616.
Saunders, M., Voth, G., 2013. Coarse-graining methods for computational biology. Annu. Rev. Biophys. 42, 73–93.
Scott, K., Bond, P., Ivetac, A., Chetwynd, A., Khalid, S., Sansom, M., 2008. Coarse--grained MD simulations of membrane protein–bilayer self–assembly. Structure 16, 621–630.
Shelley, J., Shelley, M., Reeder, R., Bandyopadhyay, S., Klein, M., 2001. A coarse grain model for phospholipid simulations. J. Phys. Chem. B 105, 4464–4470.
Shih, A., Freddolino, P., Arkhipov, A., Schulten, K., 2007. Assembly of lipoprotein particles revealed by coarse–grained molecular dynamics simulations. J. Struct.

Biol. 157, 579–592.
Shillcock, J., Lipowsky, R., 2005. Tension–induced fusion of bilayer membranes and vesicles. Nat. Mater. 4, 225–228.
Simonson, T., Perahia, D., 1992. Normal modes of symmetric protein assemblies. application to the tobacco mosaic virus protein disk. Biophys. J. 61, 410–427.
Sinitskiy, A., Voth, G., 2013. Coarse-graining of proteins based on elastic network models. Chem. Phys. 422.
Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T., Rossmann, M., Kuhn, R., 2016. The 3.8 å resolution cryo-EM structure of Zika virus. Science 352, 467–470.
Sleijpen, G., van der Vorst, H., 1996. A Jacobi Davidson iteration method for linear eigenvalue problems. SIAM J. Matrix Anal. Appl. 17, 401–425.
Smit, B., Esselink, K., Hilbers, P., van Os, N., Rupert, L., Szleifer, I., 1993. Computer simulations of surfactant self–assembly. Langmuir 9, 9–11.
Tama, F., Brooks III, C., 2002. The mechanism and pathway of ph induced swelling in cowpea chlorotic mottle virus. J. Mol. Biol. 318, 733–747.
Tama, F., Brooks III, C., 2005. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. J. Mol. Biol. 345, 299–314.
Tama, F., Brooks III, C., 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu. Rev. Biophys. Biomol. Struct. 35, 115–133.
Tama, F., Gadea, F., Marques, O., Sanejouand, Y.-H., 2000. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins Struct. Funct. Genet. 41, 1–7.
Tama, F., Sanejouand, Y.-H., 2001. Conformational change of proteins arising from normal mode calculations. Protein Eng. 14, 1–6.
Tirion, M., 1996. Large amplitude elastic motions in proteins from a single parameter, atomic analysis. Phys. Rev. Lett. 77, 1905–1908.
Tozzini, V., 2005. Coarse-grained models for proteins. Curr. Opin. Struct. Biol. 15, 144–150.
Ueda, Y., Taketomi, H., Gō, N., 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effects of specific amino acid sequence represented by specific inter-unit interactions. Int. J. Pept. Res. 7, 445–459.
van Schaik, R., Berendsen, H., Torda, A., van Gunsteren, W., 1993. A structure refinement method based on molecular dynamics in four spatial dimensions. J. Mol. Biol. 234, 751–762.
van Vlijmen, H., Karplus, M., 2005. Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry. J. Mol. Biol. 350, 528–542.
Voth, G., 2017. A multiscale description of biological active matter: the chemistry underlying many life processes. Acc. Chem. Res. 50, 594–598.
Warshel, A., 2014. Multiscale modeling of biological functions: from enzymes to molecular machines (nobel lecture). Angew. Chem., Int. Ed. Engl. 53, 10020–10031.
Zhang, X., Ge, P., Yu, X., Brannan, J., Bi, G., Zhang, Q., Schein, S., Zhou, Z., 2012. Cryo-EM structure of the mature dengue virus at 3.5 å resolution. Nat. Struct. Mol. Biol. 20, 105–110.
Zhang, Z., 2015. Systematic methods for defining coarse-grained maps in large biomolecules. Adv. Exp. Med. Biol. 827, 33–48.
Zhang, Z., Lu, L., Noid, W., Krishna, V., Pfaendtner, J., Voth, G., 2008. A systematic methodology for defining coarse-grained sites in large biomolecules. Biophys. J. 95, 5073–5083.
Zhang, Z., Sanbonmatsu, K., Voth, G., 2011. Key intermolecular interactions in the e. coli 70s ribosome revealed by coarse-grained analysis. J. Am. Chem. Soc. 133, 16828–16838.
Zheng, W., Brooks, B., Doniach, S., Thirumalai, D., 2005. Network of dynamically important residues in the open/closed transition in polymerase is strongly conserved. Structure 13, 565–577.
Zheng, W., Brooks, B., Thirumalai, D., 2007. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. Biophys. J. 93, 2289–2299.
Zheng, W., Wen, H., 2017. A survey of coarse-grained methods for modeling protein conformational transitions. Curr. Opin. Struct. Biol. 42, 24–30.
Zhou, Y., 2010. A block Chebishev-Davidson method with inner-outer restart for large eigenvalue problems. J. Comput. Phys. 229, 9188–9200.
Zhou, Y., Saad, Y., 2006. A Chebishev-Davidson algorithm for large symmetric eigenvalue problems. SIAM J. Matrix Anal. Appl. 29, 341–359.