Use Symmetry to Elucidate the Roles of Global Shape and Local Interactions in Protein Dynamics and Cooperativity

Guang Song

1Department of Computer Science,
2Program of Bioinformatics and Computational Biology,
Iowa State University, Ames, IA 50011, USA

*Correspondence to: Guang Song, Tel: 515-294-1696;
Fax: 515-294-0258; E-mail: gsong@iastate.edu

February 12, 2019

Abstract

Shape had been intuitively recognized to play a dominant role in determining the global motion patterns of bio-molecular assemblies. However, it is not clear exactly how shape determines the motion patterns. What about the local interactions that hold a structure together to a certain shape? The contributions of global shape and local interactions usually mix together and are difficult to tease part. In this work,
we use symmetry to elucidate the distinct roles of global shape and local interactions in protein dynamics. Symmetric complexes provide an ideal platform for this task since in them the effects of local interactions and global shape are separable, allowing their distinct roles to be identified. Our key findings based on symmetric assemblies are: (i) the motion patterns of each subunit are determined primarily by intra-subunit interactions (IRSi), and secondarily by inter-subunit interactions (IESi); (ii) the motion patterns of the whole assembly are fully dictated by the global symmetry/shape and have nothing to do with local iESi or IRSi. This is followed by a discussion on how the findings may be generalized to complexes in any shape, with or without symmetry.

Keywords: symmetry; shape; degeneracy; cooperativity; local interactions; phase shift; vibrations; network models

1 Introduction

Understanding the mechanism by which a molecular system functions is not only of great scientific interests but also has significant practical implications such as in drug design. For example, in the development of HIV-1 anti-viral drugs, modern drug design has become largely mechanistic-based\textsuperscript{1}. Structural and mechanistic understandings gained from basic research have stimulated much hope in drug development\textsuperscript{2}.

The technological advances in structure determination and the increasing availability of atomic or near-atomic structures of large molecular systems as a result have been instrumental in our understanding of the functional mechanisms of bio-molecular systems. Since bio-molecular systems are constantly in motion, another important piece of information that is often needed in uncovering the mechanism is dynamics. Combined together, structure and dynamics provide a fuller understanding of the mechanism. It is intuitively known that there is an intrinsic relationship among structure, dynamics, and function. However, the
relationship is not well understood. It is not fully clear why a given molecular system is structured in a certain way and what dictates its dynamics or motion patterns.

Our objective in the work is to elucidate what determines the global motion patterns of a biomolecular assembly at around its equilibrium state. We will employ normal mode analysis (NMA) as the approach, since NMA is best suited for studying the equilibrium dynamics and global motion patterns, and for understanding the relationship between structure and dynamics.

Regarding the question what determines the global motion patterns of a bio-molecular system, shape had been intuitively recognized to play a dominant role. However, it is not fully clear exactly how shape influences the motion patterns. For one thing, even shape itself is not well defined. For another, what about the local interactions that define and hold the structure together to a certain shape? What is their contribution or what difference do they make to the dynamics?

Much of our understanding that shape determines the overall dynamic patterns was based on empirical evidence, as observed by Tama and Brooks, who noted that “The evolution of the techniques and models for NMA, from an all-atom representation of the biological molecule to multi-resolution EN models, has shed light on and new insights into the role of shape and form in controlling the motions of biological molecules and their assemblies.” In another work, the authors applied randomization to the Hessian matrix elements while maintaining the structure of the matrix (i.e., keeping the zero elements zero and non-zero elements non-zero). The low-frequency normal modes were found to be robust to the randomization. They then concluded that the global shape of molecules played a more dominant role in determining molecular motions than local interactions. However, the approach used only two proteins and exploited randomization only within a small range of spread of the parameters.

In this work, we aim to characterize precisely the motion patterns of an assembly as
a whole and the motion patterns of subunits as components, and exactly what are the underlying determinants of these motion patterns. What is the role of shape? What is the role of local interactions, specifically inter-subunit interactions (IESi) or intra-subunit interactions (IRSi), and how do they affect the motions of each subunit individually and the assembly as a whole? Lastly, we address the issue of cooperativity, i.e., to what extent do the subunits correlate or synchronize their motions? How much does cooperativity among the subunits depend on the strength of inter-subunit interactions?

To answer these questions, we first turn to assemblies with symmetry, for the following reasons. Though it is generally accepted that shape determines the motion patterns of single-chain structures or structure assemblies\(^6\)–\(^10\), unlike volume or weight, shape is generally not well defined geometrically\(^11\)–\(^12\). According to wikipedia (https://en.wikipedia.org/wiki/Shape), “A shape is the form of an object or its external boundary, outline, or external surface, as opposed to other properties such as color, texture or material composition.” The definition does not exactly quantify what a shape it.

Fortunately, there are some special cases where shape is precisely defined. One such case is when shape is associated with symmetry. To say an assembly has a symmetric shape or it has a certain point group symmetry means the same thing. One may thus say a certain symmetry rigorously defines a certain shape or a certain attribute of the shape. For example, the cyclic symmetry defines a ring shape. the dihedral symmetry define a double-ring shape. Symmetry can be thought of as a special kind of “shape” that has a precise definition.

It is helpful to note that shape in biological systems is often associated with symmetry. In protein data bank (PDB)\(^13\), nearly 40% of the entries are complexes with global symmetry. Symmetry is useful structurally in constructing complexes with certain structural shapes, such as rings, double-rings, or spherical containers that are often used to store viral genomes or build bacterial micro-compartments\(^14\)–\(^15\), etc. Symmetry allows large complexes to be assembled out of a small number of protein types. Symmetry is important functionally.
as some functional processes rely heavily on the symmetry. Take the AAA+ family for example, most of the proteins in this family are ring-shaped and have symmetric motions that are used to thread proteins through their central pore\(^{16,17}\). Computationally, symmetry is useful as it can be taken advantage of to significantly accelerate the normal mode computations\(^{17–25}\). Moreover, as will be shown next, symmetric complexes provide an ideal platform to elucidate the roles of global shape and local interactions. These roles mix together in non-symmetric systems and are difficult to tease apart. In symmetric complexes, the effects of local interactions and global shape become separable, allowing their roles to be clearly identified.

## Methods and Results

### The distinct roles of global shape and local interactions in the motions of symmetric complexes

In this section, we will show that the global shape and local interactions have distinct roles in determining the motions of symmetric complexes.

Given a symmetric complex of any symmetry group that has \(h\) asymmetric subunits (or ASUs), its Hessian matrix in each ASU’s local coordinates\(^{22,23}\) is given as follows:

\[
H_{local} = \begin{bmatrix}
H_0 & H_1 & \cdots & H_{h-2} & H_{h-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots
\end{bmatrix},
\]  

(1)

where \(H_0\) represents intra-subunit interactions and \(H_j\) inter-subunit interactions between the current subunit (or the 0\(^{th}\) subunit) and the \(j^{th}\) subunit. Note that due to symmetry, the block elements on the other rows of \(H_{local}\) (not listed) are the same as those on the first row except the order is different. *This is true for all symmetry groups.*
What also is true for all symmetry groups is that there exists a symmetry transformation $T$, under which the Hessian matrix $H^{\text{local}}$ becomes block diagonal. That is,

$$H^{\text{sym}} = T^\top H^{\text{local}} T,$$

where $H^{\text{sym}}$ is the Hessian matrix in symmetry coordinates and is block diagonal. $\top$ denotes transpose. $T$ is unique for each symmetry group and is solely determined by representation matrices of the symmetry group. In the following, we will show how this works out particularly for assemblies with cyclic symmetry or icosahedral symmetry.

**Assemblies with the cyclic symmetry.** Consider a ring-like assembly that has $h$ identical subunits and cyclic symmetry. Its Hessian matrix in local coordinates is a circulant matrix:

$$H^{\text{local}} = \begin{bmatrix}
H_0 & H_1 & \ldots & H_{h-2} & H_{h-1} \\
H_{h-1} & H_0 & \ldots & H_{h-3} & H_{h-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
H_2 & H_3 & \ldots & H_0 & H_1 \\
H_1 & H_2 & \ldots & H_{h} & H_0
\end{bmatrix},$$

where $H_0$ represents intra-subunit interactions and $H_j$ inter-subunit interactions between the current subunit and the $j^{\text{th}}$ subunit down the ring. $H_{h-j} = H_j^\top$ (the transpose of $H_j$) since the matrix has to be symmetric.

The inter-subunit interactions (IESi), $H_j$ ($1 \leq j \leq h - 1$), generally are substantially weaker than intra-subunit interactions (IRSi), or $H_0$.

Based on the character table of cyclic symmetry, one can write down a transformation
matrix $T^{22}$,

$$
T = \frac{1}{\sqrt{h}} \begin{bmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & \omega_h & \omega_h^2 & \ldots & \omega_h^{h-1} \\
1 & \omega_h^2 & \omega_h^4 & \ldots & \omega_h^{2(h-1)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \omega_h^{h-1} & \omega_h^{2(h-1)} & \ldots & \omega_h^{(h-1)^2}
\end{bmatrix}
$$

(4)

where $\omega_h = e^{\frac{2\pi i}{h}}$ and $h$ is the number of subunits, such that when applied to the original Hessian matrix, produce the Hessian matrix in the symmetry coordinates that is block diagonal, i.e.,

$$
H_{sym} = T^\top H_{local} T = \text{diag}(M_0, M_1, \ldots, M_{h-1}),
$$

(5)

where $M_j = \sum_{k=0}^{h-1} \omega_h^{kj} H_k$.

The motion pattern of a subunit is thus dictated by $M_j$'s ($0 \leq j \leq h - 1$), which can be rewritten as:

$$
M_j = H_0 + \sum_{k=1}^{h-1} \omega_h^{kj} H_k.
$$

(6)

In the above equation, the first term $H_0$ represents intra-subunit interactions (or IRSi), while the second term inter-subunit interactions (or IESi). Since the second term consisting of non-bonded interactions is generally much weaker than the first term that is composed mostly of bonded interactions, the motion pattern of each subunit, $M_j$, is thus determined by the intra-subunit interests (IRSi) primarily and inter-subunit interactions (IESi) secondarily.

Assume $u_j$ is one of the eigenvectors of $M_j$, then $w_j = (0_{3N}, 0_{3N}, \ldots, u_j^T, \ldots, 0_{3N})^T$ is an eigenvector of $H_{sym}$ in eq. 5, where $0_{3N}$ represents a row vector with $3N$ zeros and $N$ is the number of atoms in each subunit. The normal mode of the whole assembly in the local
coordinates is $22$,

$$
v = (T \otimes I_{3N}) w_j = 
\begin{bmatrix}
  u_j \\
  u_j \omega_h^j \\
  \vdots \\
  u_j \omega_h^{j(h-1)}
\end{bmatrix}
$$

(7)

Eq. (7) is the general form of the normal modes of assemblies with cyclic symmetry. It clearly shows that different subunits have exactly the same motions except for a constant phase shift (or phase lag). Thus the motion patterns of the whole assembly are standing waves, in which the motions of the subunits are the same but there is a phase shift/lag between neighboring subunits. It is worth noting that this global motion pattern is dictated solely by $T$ (Eq. 4), the same transformation matrix that is shared by all complexes with cyclic symmetry and $h$ subunits (or $C_h$ in short), having nothing to do with IESi or IRSi. For different complexes with the same cyclic symmetry, the difference in their subunit compositions affects only the motions patterns of individual subunits, not the motion patterns of the assembly as a whole.

**Summary.** The above results regarding complexes of cyclic symmetry can be summarized as follows:

- The motion patterns of each subunit are determined primarily by intra-subunit interactions (IRSi), and secondarily by inter-subunit interactions (IESi), which serve mostly as a perturbation.

- The motion patterns of the whole assembly are fully dictated by the symmetry. Symmetry dictates that the overall motion patterns are standing waves. Surprisingly, neither iESi nor IRSi affects the global motion patterns.

A third issue not included above is the extent of cooperativity: how stable are the motion patterns? How strongly do subunits maintain their cooperativity with one another?
Some motion patterns may be highly vulnerable to degeneracy and may disappear at slight disturbance in structure. We will address the issue of cooperativity later in the text.

**Assemblies with the icosahedral symmetry.** For complexes with icosahedral symmetry, the Hessian matrix in local coordinates takes the same form as Eq. (1) with $h$ being 60. The order of the blocks in other rows is known and can be found in Ref. 25.

Similar to complexes with cyclic symmetry, there exists a transformation matrix $T$ (also given in Ref. 25) under which the Hessian matrix becomes block diagonal. $T$ is solely determined by representation matrices of the icosahedral point group and has nothing to do with the structural composition of the subunits nor their interactions.

That is

$$H_{\text{sym}} = T^\top H_{\text{local}} T = \begin{bmatrix}
H_E & 0 & 0 & 0 & 0 \\
0 & I_3 \otimes H_{T1} & 0 & 0 & 0 \\
0 & 0 & I_3 \otimes H_{T2} & 0 & 0 \\
0 & 0 & 0 & I_4 \otimes H_G & 0 \\
0 & 0 & 0 & 0 & I_5 \otimes H_H
\end{bmatrix},$$

where $H_{\text{sym}}$ is the Hessian matrix of icosahedral complexes in symmetry coordinates. The diagonal blocks correspond to the five irreps of the icosahedral point group.

Again, the motions of each subunit are dictated by the diagonal blocks $H_E$, $H_{T1}$, $H_{T2}$, $H_G$, and $H_H$ (whose subscripts correspond to the five irreps of icosahedral point groups, namely $E$, $T_1$, $T_2$, $G$, and $H$), all of which can be expressed as a summation of an intra-ASU interaction term and an inter-ASU interaction term. For example,

$$H_E = H_0 + \sum_{j=1}^{h-1} H_j$$
where $H_0$ (Eq. 1) represents interactions within the first ASU while the second term represents the interactions between the first ASU and the other ASUs.

On the other hand, the transformation matrix $T$ of icosahedral symmetry, which depends solely on the symmetry and has nothing to with IESi nor IRSi, determines the global motion patterns in a similar fashion as in Eq. (7). $T$ determines how the motion patterns of individual subunits are to be pieced together to form the normal mode of the whole assembly. To put this in an analogy, if the motion patterns of a subunit were a tile and the motion patterns of the whole assembly were a floor of tiles, $T$ contains the rule for laying out the tiles in order to achieve a global “floor pattern” that is consistent with the symmetry of the assembly.

The same conclusion found for complexes with cyclic or icosahedral symmetry clearly holds on complexes with other symmetries as well. As shown in Eq. (2), for every symmetry group, there exist a set of symmetry coordinates under which the Hessian matrix, $H^{sym}$, is block diagonal, and a transformation matrix $T$ that relates $H^{sym}$ with $H^{local}$. As a result, the local motion patterns of each subunit is determined primarily by IRSi within the subunit and secondarily by IESi, while the global motion patterns of the whole complex are pieced together from component (or subunit) motions according to a rule expressed in $T$ that is determined solely by the global symmetry of the complex.

### 2.2 Use symmetry to understand cooperativity

The functional processes of many biological systems require cooperativity. Understanding cooperativity is thus highly important\textsuperscript{25}. The range of cooperativity has been studied using elastic network models and normal mode analysis\textsuperscript{20}. In this work, when speaking of cooperativity, we focus on the following specific aspect of it: How strongly correlated are the motions of subunits in a large assembly? What are key factors that influence this cooperativity? One of our findings is that cooperativity among subunits is frequency-dependent,
i.e., cooperativity among subunits is stronger at some frequencies and weaker at others.

**Inter-subunit interactions effect cooperativity among subunits.** In the following, we show how the extent of cooperativity among subunits is influenced by inter-subunit interactions (IESi).

Recall that the motions of a subunit are dictated by intra-subunit interactions primarily and inter-subunit interactions secondarily. Without IESi, all subunits have the same vibrational spectrum and the vibrations of the whole assembly is fully degenerate, and there is no cooperativity among the subunits. Inter-subunit interactions serve as a perturbation and cause the otherwise degenerate modes of the subunits to mix together. As a result, the otherwise degenerate vibrational frequency lines are split and the modes become mostly non-degenerate. Apparently, the extent of split depends on the magnitude of the perturbation, or the strength of inter-subunit interactions. If the magnitude of perturbation is very small, the split will be narrow and under slight structure variations, these modes become effectively degenerate again\cite{27} and cooperativity is lost as a result.

To illustrate this quantitatively, we consider the simplest assembly with two subunits: a dimer. The Hessian matrix of a symmetric dimer when written in local coordinates is\cite{22,23,24,25}:

$$
H_{\text{local}} = \begin{bmatrix}
H_0 & H_1 \\
H_1 & H_0
\end{bmatrix},
$$

(10)

where \(H_1\) represents inter-subunit interactions. \(H_{\text{local}}\) becomes block diagonal when transformed to symmetry coordinates\cite{23,26}:

$$
H_{\text{sym}} = \begin{bmatrix}
H_0 + H_1 & 0 \\
0 & H_0 - H_1
\end{bmatrix}.
$$

(11)
Now let $v_i^{(0)}$ and $\lambda_i$ (where $1 \leq i \leq 3N$ and $N$ is the number of atoms in each subunit) represent respectively the eigenvalues and eigenvectors of $H_0$. If $H_1 = 0$, i.e., if there is no inter-subunit interactions, $H^{sym}$ becomes,

$$H^{sym} = \begin{bmatrix} H_0 & 0 \\ 0 & H_0 \end{bmatrix}. \quad (12)$$

In such a case, all the normal modes of the dimer are obviously doubly-degenerate: the two subunits have identical vibrational spectra.

Inter-subunit interactions $H_1$ remove this degeneracy. Indeed, inter-subunit interactions $H_1$ in Eq. (11) cause the otherwise degenerate frequency lines to split. The extent of split can be well estimated by the zero-order approximation according to the perturbation theory\cite{30,31} by:

$$\Delta \lambda_i = v_i^{(0)\top} H_1 v_i^{(0)} \quad (13)$$

That is, instead of having two degenerate modes corresponding to the eigenvalue $\lambda_i$, there are two split modes of eigenvalues $\lambda_i + \Delta \lambda_i$ and $\lambda_i - \Delta \lambda_i$.

$\lambda_i + \Delta \lambda_i$ ($1 \leq i \leq 3N$) are the eigenvalues of $H_0 + H_1$ and correspond to fully symmetric modes in which the two subunits in the dimer have identical motions in their local coordinate frames. $\lambda_i - \Delta \lambda_i$ ($1 \leq i \leq 3N$), on the other hand, are the eigenvalues of $H_0 - H_1$ and correspond to fully anti-symmetric modes in which the two subunits in the dimer have exactly opposite motions\cite{23}. In either case, the motions of the two subunits are fully correlated. Therefore, the inter-subunit interactions that cause the splitting in the spectrum cause also the motions of the two subunits in the dimer to change from fully uncorrelated to fully correlated. This is illustrated in Fig. [1] Intuitively we know this is true: that interactions should cause the motions of the subunits to correlate.

In reality, the transition from fully uncorrelated to fully correlated is often not black and
white, but has a degree of correlatedness that depends on the extent of the split, which in turn depends on the magnitude of inter-subunit interactions. This again is in line with our intuition that weak interactions among subunits should incur a weak correlation and strong interactions a strong correlation. The following explains why this is the case.

Since it is evident that the extent of the split $\Delta E$, which amounts to $2\Delta \lambda$, is linearly proportional to the magnitude of inter-subunit interactions $H_1$, i.e.,

$$\Delta E = 2\Delta \lambda_i \propto ||H_1||,$$

(14)

weaker inter-subunit interactions $H_1$ should therefore lead to a smaller split. When the split is small, the two split modes will have nearly identical vibrational frequencies. According to our previous work on the effective degeneracy of proteins$^{27}$, normal modes with nearly identical frequencies are effectively degenerate. The degeneracy implies that each subunit can move on its own, independent of the other subunit. A larger split on the other hand ensures a stronger entanglement of the vibrations of the subunits and their cooperativity. The width of the split is thus an indicator of the extent of cooperativity (see Figure 1).

The extent of cooperativity is frequency-dependent. Figure 2 shows quantitatively how inter-subunit interactions (IESi) influence the frequencies of the modes, using the crystal structure of tyrosinase from Bacillus megaterium (pdb-id: 5OAE, a dimer)$^{32}$ as an example. To this end, we apply sbNMA$^{33}$ and first compute $\Delta \lambda$ using eq. (13). Since the split in eigenvalues is $2\Delta \lambda$ and the vibrational frequency $\omega = \frac{1}{2\pi c} \sqrt{\lambda}$ (in cm$^{-1}$) where $c = 2.997925 \times 10^{10}$ cm/sec is the speed of light, the extent of split in the frequency spectrum $\Delta \omega$ is thus:

$$\Delta \omega_i = \frac{1}{2\pi c} \frac{\Delta \lambda_i}{\omega_i} = \frac{1}{2\pi c} \frac{v_i^{(0)^T} H_1 v_i^{(0)}}{\omega_i}$$

(15)

$\Delta \omega_i$ represents the zero-order correction to the frequency spectrum due to the existence
of inter-subunits interactions. $\psi_i^{(0)}$ represents the eigen-modes of the systems without inter-subunit interactions.

On one hand, it is clear from Eq. (13) that the extent of split is linearly proportional to the magnitude of inter-subunit interactions $H_i$. On the other hand, the extent of split is inversely proportional to the vibrational frequency $\omega$. This frequency dependence is evident also in Fig. 2(B), where it is seen that on average the split $\Delta \omega$ is most pronounced at the the low frequency range and becomes much smaller (or sparser) at high frequencies, except for a few spikes.

The low value of $\Delta \omega$ at the high frequency end can be understood intuitively as follows. At high frequencies, most vibrations are highly localized, and as a result, there may be no vibration at the boundary between subunits. In such a case, inter-subunit interactions have little or no effect on the dynamics of the subunits. The extent of split is consequently very small or even none.

Secondly, $H_1$ represents contributions of the non-bonded interactions between subunits. These contributions were shown to influence vibrations mostly in the frequency range of 0-500 cm$^{-1}$. At this frequency range intra-subunit and inter-subunit vibrations take a similar frequency and consequently the subunits have the strongest correlations. That is, subunits resonate with one another the strongest when the inter-subunit interactions are able to transmit their vibrations to one another most effectively. Currently charge-charge interactions or hydrogen bonds are not included in the inter-subunit interactions in sbNMA. Their inclusion might increase the strength of IESi somewhat and consequently the split in frequency shown in Figure 1.

**Summary.** In summary, inter-subunit interactions determine the extent of cooperativity. No/weak inter-subunit interactions means no/low cooperativity. Another new finding is that the extent of cooperativity varies with frequency. The cooperativity should take place mostly
within the frequency range of $[0, 500 \text{ cm}^{-1}]$. Vibrations of subunits at high frequencies should be mostly uncorrelated. It is postulated that key function-related motion patterns should take place at the frequency range where cooperativity is the highest.

3 Discussion

In this work, we have used symmetry to elucidate the roles of global shape and local interactions in protein dynamics and cooperativity. Our key findings from studying symmetric complexes are: (i) the motion patterns of each subunit are determined primarily by intra-subunit interactions (IRSi) and secondarily by inter-subunit interactions (IESi); (ii) the motion patterns of the whole assembly are fully dictated by the global symmetry and have nothing to do with iESi or IRSi.

From symmetry to shape. As aforementioned, shape in general is not well defined mathematically while symmetry is. Symmetry represents a special kind of shape (or an aspect of shape) that has a precise mathematical definition. A key result from the current study is that the overall motion patterns of symmetric assemblies are dictated solely by global symmetry, or global shape, while local symmetry or interactions influence only the motions patterns of individual subunits. Can we generalize this result about complexes with symmetry to complexes without symmetry? Though symmetry is most often represented by a character table in group theory while a general shape cannot, fortunately both symmetry and shape can be represented by a network model.

Represent symmetry using a network of nodes instead of the character table. Symmetry in structures is usually represented by a character table and representation matrices. However, the same symmetry, for instance the icosahedral symmetry possessed by a viral capsid shown in Figure 3(A), can be fully captured also by a simple network, as
shown in Figure 3(B) where each node represents the mass center of a corresponding protein chain of the capsid shown in Figure 3(A). There are 60 nodes in Figure 3(B) and their connectivity represents a shape. If we are to use this network and solve for its normal modes, we will obtain the same motion patterns as we get using the character table and the representation matrices (Eq. 8). That is to say, the network in Figure 3(B) is an alternative way to the group theory in capturing precisely the overall symmetry and consequently motion patterns of the assembly. Figure 4 shows the motion patterns of three modes (out of 180 modes) based on the network model in Figure 3(B). All three modes, and only these three, are singly-degenerate and fully symmetric, as is expected from the group theory that 1/60 of the modes must be singly-degenerate and fully symmetric. Now what is advantageous about this network representation (Figure 3(B)) is that it can be easily generalized to other shapes with less or no symmetry. This is what has been done intuitively already, for example through elastic network models.

Figure 5 shows the atomic structure of HIV-1 capsid on the left and its network representation on the right, with nodes placed at the mass centers of the 1,356 capsid proteins. By extrapolation it is reasonable to conjecture that just as the network in Figure 3(B) fully captures the global motion patterns of the capsid in Figure 3(A), the global motion patterns of the HIV-1 capsid in Figure 5(A) should be captured by the network model given in Figure 5(B). Since the shape of HIV-1 capsid represented by the network of nodes in Figure 5(B) is ellipsoidal, the vibration of the whole capsid should be similar to a standing wave on an ellipsoidal membrane. Movies of some of these modes are given in Supplemental Information. Indeed, these modes show motion patterns that resemble standing waves on a membrane.

The roles of shape and local interactions in determining the local or global motion patterns of a structural assembly. Consider the motion pattern of a large assembly that
vibrates at frequency $\omega$, all its subunits are vibrating at the same frequency $\omega$. As shown early, the vibration pattern of each subunit is determined primarily by the IRSi (intra-subunit interactions) and secondarily by IESi (inter-subunit interactions). Since the vibration is primarily determined by IRSi, it is expected to resemble closely to its vibration at frequency $\omega$ in isolation (i.e., without IESi). This is precisely the physical basis of resonance that was observed in an earlier work.\(^{39,40}\) Note that the motions of individual subunits represent only local motion patterns, determined by IRSi primarily and IESi secondarily.

On the other hand, the global motion patterns of a whole assembly, which are pieced together by the motions of individual subunits, are determined solely by global symmetry for symmetric complexes. This is most evident in complexes with cyclic symmetry, where the net effect of global symmetry is that each subunit is assigned a different phase shift (see Eq. 7 and the related text). The relative phase shift among the subunits essentially captures the global motion patterns. This observation is possibly extendable to any shape in general, i.e., the global symmetry or shape of an assembly causes a phase shift to be assigned to each subunit and the relative phase shift among the subunits determines the global motion patterns.

Besides the local motion patterns of individual subunits and the relative phase shifts assigned to them, another aspect of the global motion patterns is the relative magnitudes of motions of the subunits. The magnitude of fluctuations was shown to be inversely proportional to the packing density.\(^{41}\) In our terms, the relative magnitude of motions of a subunit should be roughly proportional to the magnitude of its IESi: the weaker the IESi, the larger the magnitude. IESi determines also the extent of the cooperativity between neighboring capsomeres.

In summary, for subunit $i$ that is vibrating at frequency $\omega$ within an assembly and whose vibration $\mathbf{v}_i$ is denoted as

$$\mathbf{v}_i = u_i \cos(\omega t + \phi_i),$$

(16)
it is reasonable to conjecture that the influence of local interactions and global symmetry or shape should come into play as follows: i) $u_i$, which represents the normal mode (or the direction of motion) of the $i^{th}$ subunit as in Eq. [7] is determined primarily by IRS$i$ and secondarily by IES$i$; ii) the phase shift $\phi_i$ is determined by global symmetry or shape; iii) the magnitude of $u_i$ is determined mostly by IES$i$. This conjecture is known to be true for symmetric assemblies based on the results shown earlier, but remains to be tested for other assemblies.

**Elastic network models.** Insights gained in this work regarding the roles of shape and local interactions and where they come into play in a subunit’s vibration expressed in Eq. (16) can help NMA developers or users develop or select the right models when studying protein dynamics. For example, if one cares to know only the global motion patterns of a large assembly, Eq. (16) indicates that a simplified network that captures its global symmetry or shape should be sufficient. This is the underlying basis for the success of coarse-grained elastic network models. It also reveals the models’ weakness: the local motion patterns $u_i$ (Eq. 16) are often inaccurate. Their frequency spectra (or $\omega$ in Eq. 16) are inaccurate either, as they were found to deviate significantly from the universal spectrum of globular proteins\cite{34,42,43}. On the other hand, if accurate local motion patterns are desired, e.g., when studying how the pore of a capsomere is regulated, all-atom ENM models and all-atom force fields would be necessary\cite{33,39,40}, since local dynamics are mostly determined by local interactions. Mixed-grained models\cite{44,45} also seem plausible but the challenge is that an accurate force field for mixed-grained models is hard to find.

Elastic network models were developed as entropy-based models and their success was often ascribed to a good modeling of the entropy\cite{29}. Here we see that the success of network models originates also from their accurate representation of the shape of the structures being studied, which, as the major determinant of the global motion patterns as is shown in this
work, leads to a fairly accurate reproduction of global dynamics, as was observed also by other authors.\[6–10\]

**Supplemental Information**

**hivExpand.mpg, hivRotate.mpg, hivContract.mpg**: movies of the normal modes of HIV-1 capsid. The modes are computed through an extremely simplified network model that represents each capsid protein as a node (thus there are 1,356 nodes in the network model). The movies show three typical motions: the radial expansion of the capsid, the rotations of capsomeres, and the contraction and expansion of the capsomeres. These motions are similar to standing waves on an ellipsoidal membrane. One of them resembles a transverse wave, while the other two longitudinal waves.

**References**

[1] E. J. Arts and D. J. Hazuda. Hiv-1 antiretroviral drug therapy. *Cold Spring Harbor Perspectives in Medicine*, 2(4), 2012.

[2] S. K. Carnes, J. H. Sheehan, and C. Aiken. Inhibitors of the hiv-1 capsid, a target of opportunity. *Curr Opin HIV AIDS*, 13(4):359–365, 2018.

[3] B. Brooks and M. Karplus. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, 80(21):6571–6575, November 1983.

[4] M. Levitt, C. Sander, and P. S. Stern. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int. J. Quant. Chem.*, 10:181–199, 1983.
[5] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA*, 80(12):3696–3700, June 1983.

[6] D. Ming, Y. Kong, M. A. Lambert, Z. Huang, and J. Ma. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. USA*, 99(13):8620–8625, 2002.

[7] M. Lu and J. Ma. The role of shape in determining molecular motions. *Biophys. J.*, 89(4):2395–2401, 2005.

[8] F. Tama and C. L. Brooks III. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.*, 35:115–133, 2006.

[9] P. L. Freddolino, A. Arkhipov, A. Y. Shih, Y. Yin, Z. Chen, and K. Schulten. Application of residue-based and shape-based coarse graining to biomolecular simulations. In G. A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 20, pages 299–315. Chapman and Hall/CRC Press, 2008.

[10] O. Kurkcuoglu, P. Doruker, T. Z. Sen, A. Kloczkowski, and R. L. Jernigan. The ribosome structure controls and directs mrna entry, translocation and exit dynamics. *Phys. Biol.*, 5(4):046005, Nov 2008.

[11] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam. Analytical shape computation of macromolecules: I. molecular area and volume through alpha shape. *Proteins*, 33(1):1–17, 1998.

[12] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam. Analytical shape computation of macromolecules: II. inaccessible cavities in proteins. *Proteins*, 33(1):18–29, 1998.
[13] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–42, 2000.

[14] T. Klose, D. G. Reteno, S. Benamar, A. Hollerbach, P. Colson, B. La Scola, and M. G. Rossmann. Structure of faustovirus, a large dsdna virus. *Proc. Natl. Acad. Sci. USA*, 113(22):6206–6211, 2016.

[15] M. Sutter, B. Greber, C. Aussignargues, and C. A. Kerfeld. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science*, 356(6344):1293–1297, 2017.

[16] P. I. Hanson and S. W. Whiteheart. AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.*, 6:519–529, July 2005.

[17] H. Na and G. Song. Predicting the functional motions of p97 using symmetric normal modes. *Proteins*, 84(12):1823–1835, 2016.

[18] T. Simonson and D. Perahia. Normal modes of symmetric protein assemblies. application to the tobacco mosaic virus protein disk. *Biophys J*, 61(2):410–27, 1992.

[19] H. W. van Vlijmen and M. Karplus. Normal mode analysis of large systems with icosahedral symmetry: application to (Dialanine)60 in full and reduced basis set. *J. Chem. Phys.*, 115(2):691698, 2001.

[20] H. W. van Vlijmen and M. Karplus. Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry. *J Mol Biol*, 350(3):528–42, 2005.

[21] Y. Matsunaga, R. Koike, M. Ota, J. R. Tame, and A. Kidera. Influence of structural symmetry on protein dynamics. *PLoS One*, 7(11):e50011, 2012.
[22] G. Song. Symmetry in normal modes and its strong dependence on symmetry in structure. *J. Mol. Graphics Modell.*, 75:32 – 41, 2017.

[23] G. Song. The finite number of global motion patterns available to symmetric protein complexes. *Proteins*, 85(9):1741–1758, 2017.

[24] G. Song. A comparative study of viral capsids and bacterial compartments reveals an enriched understanding of shell dynamics. *Proteins*, 86(2):152–163, 2018.

[25] G. Song. A time and memory efficient recipe for fast normal mode computations of complexes with icosahedral symmetry. *J. Mol. Graphics Modell.*, 87:30 – 49, 2019.

[26] F.A. Cotton. *Chemical Applications of Group Theory*. Wiley, New York, 3rd edition, 1990.

[27] H. Na and G. Song. The effective degeneracy of protein normal modes. *Phys Biol*, 13(3):036002, 2016.

[28] Q. Cui and M. Karplus. Allostery and cooperativity revisited. *Prot. Sci.*, 17(8):1295–1307.

[29] L. Yang, G. Song, and R. L. Jernigan. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA*, 106(30):12347–12352, July 2009.

[30] L. D. Landau and E. M. Lifschitz. *Quantum Mechanics: Non-relativistic Theory*. Elsevier, 3rd edition, 1977.

[31] Wikipedia: Perturbation theory (quantum mechanics). [https://en.wikipedia.org/wiki/Perturbation_theory_(quantum_mechanics)](https://en.wikipedia.org/wiki/Perturbation_theory_(quantum_mechanics))

[32] S. Ferro, B. Deri, M. P. Germanò, R. Gitto, L. Ielo, M. R. Buemi, G. Certo, S. Vittorio, A. Rapisarda, Y. Pazy, A. Fishman, and L. De Luca. Targeting tyrosinase: Development
and structural insights of novel inhibitors bearing arylpiperidine and arylpiperazine fragments. *Journal of Medicinal Chemistry*, 61(9):3908–3917, 2018. PMID: 29634898.

[33] H. Na and G. Song. Bridging between normal mode analysis and elastic network models. *Proteins*, 82:2157–2168, September 2014.

[34] H. Na, G. Song, and D. Ben-Avraham. Universality of vibrational spectra of globular proteins. *Phys. Biol.*, 13(1):016008, February 2016.

[35] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77:1905–1908, 1996.

[36] T. Haliloglu, I. Bahar, and B. Erman. Gaussian dynamics of folded proteins. *Phys Rev Lett.*, 79(16):3090–3093, 1997.

[37] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding & Design*, 2(3):173–181, 1997.

[38] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins*, 33(3):417–429, March 1998.

[39] H. Na and G. Song. Fast normal mode computations of capsid dynamics inspired by resonance. *Phys Biol*, 15(4):046003, 2018.

[40] H. Na and G. Song. All-atom normal mode dynamics of hiv-1 capsid. *PLoS Comput Biol*, 14(9):e1006456, 2018.

[41] Bertil Halle. Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. USA*, 99(3):1274–1279, 2002.
[42] D. ben-Avraham. Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B*, 47(21):14559–60, 1993.

[43] K. Hinsen and G. R. Kneller. Projection methods for the analysis of complex motions in macromolecules. *Mol. Sim.*, 23:275–292, 2000.

[44] O. Kurkuoglu, R. L. Jernigan, and P. Doruker. Collective dynamics of large proteins from mixed coarse-grained elastic network model. *QSAR Comb. Sci.*, 24:443–448, June 2005.

[45] O. Kurkuoglu, O. T. Turgut, S. Cansu, R. L. Jernigan, and P. Doruker. Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.*, 97:1178–1187, August 2009.

[46] S. W. Lane, C. A. Dennis, C. L. Lane, C. H. Trinh, P. J. Rizkallah, P. G. Stockley, and S. E.V. Phillips. Construction and crystal structure of recombinant stnv capsids. *Journal of Molecular Biology*, 413(1):41 – 50, 2011.
Figure Legends

Figure 1.

The split in spectrum lines caused by inter-subunit interactions in turn causes the motions of the subunits in a dimer to change from fully uncorrelated to fully correlated. The degree of correlatedness depends on the width of the split.

Figure 2.

(A) The vibrational spectrum of a dimer (the crystal structure of tyrosinase from Bacillus megaterium (pdb-id: 5OAE) without taking into account the effect of inter-subunit interactions, and (B) the extent of split $\Delta \omega$ at different frequencies.

Figure 3.

The structure of STNV viral capsid (pdb-id: 4V4M) in (A) atomic details (image taken from PDB), and (B) the same structure in a much simplified network model.

Figure 4.

Porcupine plots of a few selected global motion patterns of STNV viral capsid as revealed by the simple 60-node network model shown in Figure 3(B): (A) radial expansion of the whole capsid, (B) rotation along the 5-fold axes, and (C) expansion or contraction of each pentamer. These motion patterns are in full agreement with the results that one would get using the group theory.
Figure 5.

The structure of HIV-1 capsid in (A) atomic details (image taken from PDB\textsuperscript{163}), and (B) the same structure in a network model, whose nodes are placed at the mass centers of the 1,356 capsid proteins. For clarity, only edges within each hexamer or pentamer are shown.
Figure 1:

Degree of correlatedness
Figure 2:
Figure 3:
Figure 4:
Figure 5: