Published in final edited form as:

J Mol Biol. 2013 March 25; 425(6): 1082–1098. doi:10.1016/j.jmb.2012.12.026.

PIM: Phase Integrated Method for Normal Mode Analysis of Biomolecules in Crystalline Environment

Mingyang Lu[†] and Jianpeng Ma^{†,‡,§}

[†]Verna and Marrs McLean Department of Biochemistry and Molecular Biology Baylor College of Medicine One Baylor Plaza, BCM-125 Houston, TX 77030

[‡]Department of Bioengineering Rice University Houston, TX 77005

Abstract

In this study, a normal mode analysis, named phase integrated method (PIM), is developed for computing modes of biomolecules in crystalline environment. PIM can calculate low-frequency modes on one or a few asymmetric units and generate exact modes of a whole unit cell according to space group symmetry, while the translational symmetry between unit cells is maintained via periodic boundary condition. Therefore, the method can dramatically reduce computational cost in mode calculation in the presence of crystal symmetry. PIM also has an option to map modes onto a single asymmetric unit to form an orthonormalized mode set, which can be directly applied to normal-mode-based thermal parameter refinement in X-ray crystallography. The performance of PIM was tested on all 65 space groups available in protein crystals (one protein for each space group) and on another set of 83 ultra-high-resolution X-ray structures. The results showed that considering space group symmetry in mode calculation is crucial for accurately describing vibrational motion in crystalline environment. Moreover, the optimal inter-asymmetric-unit packing stiffness was found to be about 60% of that of intra-asymmetric-unit interactions (nonbonded interaction only).

Keywords

Normal mode analysis; Crystal packing; Space group; Asymmetric unit; X-ray refinement; Phase shift; Phase integrated method; Inter- and Intra-asymmetric unit packing interactions

Introduction

Normal mode analysis (NMA) is an important computational method for studying largescale deformational motions of biomolecules¹⁻⁶. During the recent two decades, there are increasing interests in dealing with deformational motions of supramolecular complexes⁷⁻¹⁸. Consequently, many coarse-grained methods have been developed to reduce computational costs. They can be roughly categorized into two groups. The first group belongs to force-field-based methods, in which the potential function is calculated by a molecular mechanics force field, such as CHARMM¹⁹⁻²¹ and AMBER²²⁻²⁵. Applications of these methods include original rotations-translations of blocks (RTB)²⁶ and all-atom-

[§]To whom the correspondence should be addressed. Mailing address: One Baylor Plaza, BCM-125, Baylor College of Medicine, Houston, TX 77030. jhma@bcm.tmc.edu, Phone: 713-798-8187, Fax: 713-796-9438. .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

derived coarse-grained methods²⁷⁻²⁸. The second group belongs to network-based methods, such as elastic network models (ENMs), in which the potential function is the sum of harmonic terms between interacting nodes defined by a coarse-grained structure model²⁹⁻⁵⁴.

Force-field-based methods utilize detailed molecular interactions, therefore usually producing more accurate modes than network-based methods. However, conventional force-field-based methods require initial energy minimization that distorts structure⁵⁵ and imposes tremendous difficulties in applications such as normal-mode-based X-ray refinement⁵⁶⁻⁶⁶. The network-based methods bypass initial energy minimization step, but the conventional methods such as ENMs sometimes generates abnormally localized motions, known as “tip-effects”, even in the low-frequency eigenvectors⁵⁰. The “tip-effects” can be a big issue, especially in normal-mode-based X-ray thermal parameter refinement, because the modeled thermal parameters can be erroneously large for some surface atoms.

To cope with above issues, over the past years, we have developed a series of methods that not only avoid initial energy minimization, but also deliver considerably more accurate low-frequency eigenvectors^{50-51, 55}. We first developed a C α -based ENM method with a scheme for stiffening internal coordinate degrees of freedom⁵⁰. More recently, we have developed two mode calculation methods, minimalist network model (MNM)⁵⁵ and NMA with molecular geometry restraints (MGR)⁵¹. MNM is a force-field-based method, but bypass energy minimization step by slightly modifying second derivative matrix. Our tests show that MNM can produce very similar eigenvectors to CHARMM normal modes. Because MNM does not need energy minimization, it even outperforms CHARMM in fitting experimental anisotropic displacement parameters (ADPs) of 83 ultra-high resolution crystal structures. MGR, on the other hand, is a network-based method. In addition to non-bonded terms originally used in ENMs, the MGR potential function has harmonic restraining terms on bonded interactions to strengthen local stiffness. Our tests show that MGR produces modes significantly closer to CHARMM modes than the conventional ENMs. Both MNM and MGR do not need energy minimization step, and can eliminate “tip effects”, usually found in the conventional ENMs. The two methods have their own advantages. Because MNM utilizes a force field, it describes molecular interactions more accurately, but the method is effective only when force field is applicable, e.g., in highquality structures. MGR, however, depends less on structure quality, as it can be applied to structures with bad contacts, bad geometry and missing atoms, commonly found in low-resolution structures.

Another important issue in NMA is that mode calculations are normally performed on an isolated molecule. Such a feature may cause problems in applications in X-ray refinement since all crystal structures are solved in a crystal, in which the deformational motions can be affected by crystal contacts. To deal with this problem, some methods have been recently developed to introduce crystal contacts such as work by Hinsen⁶⁷ and Riccardi et al.⁶⁸. In their study, they calculated ENM modes on a whole unit cell of a crystal for a series of wave vectors. However, these methods can be very costly for a large unit cell. For example, a typical ribosome molecule has more than 5000 residues, while some typical ribosome crystals with space group I422 have 16 asymmetric units (AUs) in each unit cell. Therefore, the system is 16 times as large as a ribosome molecule, making it very difficult to handle by current computational resources. Compared to the mode calculation on a single AU, their methods can cost about more 4×10^3 CPU time (considering CPU time for diagonalization is usually $\propto(N^3)$). Besides, even if the modes are successfully computed, it is still very hard to extract deformational motions for a single molecule or an AU in the context of crystalline environment.

Recently, Hafner et. al.⁶⁹ also calculated normal modes of an AU in crystalline environment by using some special boundary conditions. In that study, they argued that crystal packing interactions are very weak (2% of interactions between protein molecules). While in the studies of Hinsén and Riccardi et al.⁶⁷⁻⁶⁸, crystal packing interactions were modeled as same strength as intra-asymmetric-unit interactions. Therefore, it is important to carry out more indepth study on how crystal packing affects low-frequency modes and their interactions strength.

In this study, we designed a normal mode analysis for molecules with crystal packing environment, named phase integrated method (PIM). To simplify the problem, we restrain our normal mode analysis within a unit cell and translational symmetry between unit cells is maintained via periodic boundary condition. In PIM, space group symmetry is employed in the calculation to reduce computational cost. In the literature, some mode calculation methods have been developed by means of point-group theory for some symmetric molecules⁷⁰⁻⁷⁴. Unlike these studies, we designed a method to link deformations among various AUs according to space group. Based on these relations, PIM can calculate Hessian matrix and normal modes on one or a few AUs, and obtain exact normal modes of a whole unit cell. There are many advantages of PIM method. First, PIM can dramatically reduce computational cost for most crystal space groups. Second, PIM focuses more on internal motions inside AU(s), rather than vibration of crystal lattice. Third, PIM is able to map modes onto a single AU to form an orthonormalized low-frequency mode set, which can be directly applied to normal-mode-based thermal parameter refinement in X-ray crystallography⁵⁶⁻⁶³.

In this paper, we first introduced the theory of PIM method, then numerically tested the PIM method with the mode calculation method MGR⁵¹, including accuracy test and mode comparison test. Furthermore, we optimized the stiffness of crystal packing interactions by fitting against ADPs from experimental data. The normal mode method MGR and test methods were explained in the Materials and Methods.

Theory

In this study, we only consider the case that unit cells have fluctuations under periodic boundary conditions. Other than directly calculating low frequency normal modes from the whole unit cell⁶⁷, in the PIM method, we calculate the modes for the structures in a crystal with one or a few asymmetric units. These modes are found to be sufficient to represent the modes of the whole unit cell. We showed in the following how to select the asymmetric unit(s) and how to calculate modes based on them.

Energy Decomposition—For a crystal, the total energy can be written as

$$E = \sum_i E_i + \sum_{i,j} E_{ij}, \text{ where } E_i \text{ is self-energy for the interactions within an asymmetric unit (AU) } i, E_{ij} \text{ is the energy from the interactions between the two AUs } i \text{ and } j$$

All asymmetry units in a crystal lattice are related to each other by space group symmetry within a unit cell and translational symmetry between unit cells. The micro-environment around all AUs are the same and all units are equivalent, therefore, total energy can be re-

expressed as $E = N \left(E_i + 1/2 \sum_j E_{ij} \right)$, where the summation is over all the AUs that have contacts with the AU i , N is the number of AUs in the crystal (i.e. a very large number). Therefore, mode calculations can be performed based on the potential energy,

$$E_0 = E_i + 1/2 \sum_j E_{ij}. \quad (\text{Eq.1})$$

It focuses energy on a single AU.

Multiplication Table—To exploit symmetry relations of AUs, we utilized a multiplication table, also known as Cayley table⁷⁵, for space group symmetry.

The multiplication table for space group contains all possible combinations of how one element (one AU) is related to another by symmetry operations. For instance, the multiplication table for the space group $P2_12_12_1$ is

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

where the element C_{ij} at i th row and j th column represents the element resulted from the i th element after performing the j th space group symmetry operation. In this study, the definition of symmetry elements for any space group follows the convention used in the CCP4 SYMLIB library⁷⁶. For instance, starting from the second element $(\bar{x}+1/2, \bar{y}, z+1/2)$, we can perform the third symmetry operation $(x+1/2, \bar{y}+1/2, \bar{z})$ on it and obtain the fourth element $(\bar{x}, y+1/2, \bar{z}+1/2)$. Note, unit translational invariance is applied based on periodic boundary condition. Therefore, in this example, the element of the 2nd row and the 3rd column has a value 4, representing the 4th element.

Moreover, as we discovered, it is the structure of the multiplication tables, not the specific symmetry operations, that determines the way to calculate modes. We showed in the later sections how to calculate modes based on the multiplication tables in the PIM.

Single AU Case—To begin with, we consider the easiest case, in which only one AU is required in mode calculation, while other AUs share the same mode patterns as the studied unit. In other word, all the AUs share the same set of eigenvectors, but only differ in the phase of the sine functions. This type of case is named as single AU case.

Consider a crystal system in which a unit cell contains n AUs, the same eigenvector is used for all AUs in a specific mode, except that the i th AU has phase θ_i . So, the mode in a unit cell has a set of phases $\{\theta_i\}$. To satisfy space group symmetry, mode of the i th AU should become the mode of C_{ij} AU after the j th symmetry operation according to the multiplication table, i.e.,

$$e^{i\theta_i} e^{i\theta_j} = e^{i\theta_{C_{ij}}}. \quad (\text{Eq.2})$$

There are totally n^2 equations that completely satisfy the multiplication table.

We computationally enumerated all possible phase solutions for all the 65 chiral space groups available to protein crystals. According to the group theory, there always exist one or a few positive integers m that satisfy

$$e^{i,m\theta_i} = 1, m|n.$$

Here, m is an integer less than n , and m is a submultiple of n . Therefore, the smallest sampling grid for the phase set $\{\theta_i\}$ equals to 2π divided by the largest possible m . We found that it is possible to sample each phase θ_i for every $\pi/6$ without missing any solution in protein crystals. According to the enumeration results, each space group has no more than n phase solutions. It was also found that, when the number of phase solutions is exactly n , it covers all possible modes for the whole system. Each of these phase solutions is defined as a branch of the PIM. There are totally 29 out of the 65 space groups in this case, shown in the leftmost column of Table 1. In Protein Data Bank (PDB), 73.3% structures belong to this case. If the number of phase solutions is less than n , we found that, although the solutions are still viable, they are not complete anymore, meaning that it misses some modes of a whole unit cell. These cases were discussed in the later sections on double AU case and triple AU case.

For example, in the case of space group $P2_12_12_1$, all four possible phase solutions, or branches, are:

	(1)	(2)	(3)	(4)
1:	0	0	0	0
2:	0	π	π	0
3:	0	π	0	π
4:	0	0	π	π

The numbers in the parenthesis in the first line denote the indices of AUs. In each line afterwards, the number before colon is the solution index, the four consecutive numbers after colon represent the phases for all the AUs in a unit cell. As the convention of the SYMLIB library, the first element of any space group always corresponds to the unit/identical symmetry operation, *i.e.* $\theta_1 = 0$.

Given a set of phases $\{\theta_j\}$, the Hessian matrix of the PIM can be constructed accordingly. The first term in Eq. 1, the self-energy, is the same as the total energy of an isolated AU. So

the Hessian matrix can be calculated by $\mathbf{H}_{ab} = \frac{\partial^2 E_i}{\partial X_a \partial X_b}$ for any atom pair a and b . The second term in Eq. 1, the packing energy, includes all the interactions between AU i and the other units that are in contact with it. Here, we assume that the packing interactions are all pairwise, which is valid for the elastic network models³¹ and MGR model⁷⁷. For any interaction between the atom a of the AU i and the atom b of the AU j , with energy E_{ab} , the Hessian matrix block can be calculated by

$$\mathbf{H}_{a^{(i)}, b^{(j)}} = \begin{pmatrix} \mathbf{H}_{aa} & H_{ab} \\ \mathbf{H}_{ba} & \mathbf{H}_{bb} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 E_{ab}}{\partial^2 X_a} & \frac{\partial^2 E_{ab}}{\partial X_a \partial X_b} \\ \frac{\partial^2 E_{ab}}{\partial X_b \partial X_a} & \frac{\partial^2 E_{ab}}{\partial^2 X_b} \end{pmatrix},$$

where $\mathbf{x}_a$ and $\mathbf{x}_b$ are the displacement vectors of the atoms a and b , the notation $a^{(i)}$ represents the atom a in the AU i . In order to calculate normal modes on a single AU, the above Hessian matrix needs to be converted to a form based on the atomic displacements of the same unit. In the other word, we replace $\mathbf{x}_b$ by the displacement of the atom that is equivalent to the atom b in the AU i , $\mathbf{x}_b^{(i)}$. Here, $\mathbf{x}_b^{(i)}$ is related to $\mathbf{x}_b$ by the rotation matrix of the symmetry operation $\mathbf{R}_j$ of the AU j and the phase α_j :

$$\mathbf{x}_b = \mathbf{R}_j e^{i\theta_j} \mathbf{x}_b^{(i)}.$$

Therefore, the modified Hessian block for the displacement vector $\begin{pmatrix} X_a \\ X_b^{(i)} \end{pmatrix}$ is

$$\mathbf{H}_{|a^{(i)}, b^{(i)}} = \begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \mathbf{R}_j e^{i\theta_j} \\ \mathbf{R}_j^T \mathbf{H}_{ba} e^{-i\theta_j} & \mathbf{R}_j^T \mathbf{H}_{bb} \mathbf{R}_j \end{pmatrix}. \quad (\text{Eq.3})$$

Finally, the Hessian matrix blocks for all packing terms are integrated into the whole Hessian matrix of the self-energy part. Note that any matrix in the form of Eq.3 is Hermitian, so the normal modes calculated in the single AU case always have real frequencies. Besides, the conjugation of the Hessian matrix equals to the Hessian matrix for the branch of phase solution $\{-\theta_j\}$. So, the phase solutions $\{\theta_j\}$ and $\{-\theta_j\}$ give degenerate modes. Therefore, only one branch needs to be calculated. For single AU case, one needs to calculate no more than n diagonalizations of Hessian matrices on a single AU.

For the case of space group $P1$, since there's only one AU in a unit cell, the normal mode calculation is just performed on the unit cell under periodic boundary condition.

Double AU Case—Some other space groups have less than n phase solutions, but each of the space groups has at least one subgroup of order $n/2$ which belongs to the single AU case. For these space groups, the normal modes can still be calculated based on a single AU. However, the displacement vectors of the AUs are related to each other not only by phases as in the single AU case, but also by conjugation for some of AUs. There are totally 23 space groups in this category, as listed in the middle column in Table.1. These cases were named as double AU case, because, in the calculation of some branches, the Hessian matrices have the dimension of two AUs. In the PDB, 24.0% structures belong to the case.

For example, the space group type $P422$ has eight AUs. It has a subgroup $P4$ for the first four elements $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $(\bar{\mathbf{y}}, \mathbf{x}, \mathbf{z})$, $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \mathbf{z})$ and $(\mathbf{y}, \bar{\mathbf{x}}, \mathbf{z})$. So it belongs to the double AU case. The multiplication table for $P422$ is:

1	2	3	4	5	6	7	8
2	3	4	1	8	5	6	7
3	4	1	2	7	8	5	6
4	1	2	3	6	7	8	5
5	6	7	8	1	2	3	4
6	7	8	5	4	1	2	3
7	8	5	6	3	4	1	2
8	5	6	7	2	3	4	1

The eight branches are:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1:	0	0	0	0	0	0	0	0
2:	0	π	0	π	π	0	π	0
3:	0	π	0	π	0	π	0	π
4:	0	0	0	0	π	π	π	π
5:	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	0*	$\frac{\pi}{2}$ *	π *	$\frac{3\pi}{2}$ *
6:	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	π *	$\frac{3\pi}{2}$ *	0*	$\frac{\pi}{2}$ *
7:	0	$\frac{3\pi}{2}$	π	$\frac{\pi}{2}$	0*	$\frac{3\pi}{2}$ *	π *	$\frac{\pi}{2}$ *
8:	0	$\frac{3\pi}{2}$	π	$\frac{\pi}{2}$	π *	$\frac{\pi}{2}$ *	0*	$\frac{3\pi}{2}$ *

Here, two operations are defined for each branch solution. The first one is phase shift operation θ , already used in the single AU case, representing phase difference θ from the first AU (i.e. the first element). The second one is a new operation θ^* , introduced to represent a θ phase shift after a conjugation operation, i.e. it changes a vector $\mathbf{x}$ to $e^{i\theta} \mathbf{x}^*$. The first four branches only involve phase shifts, so the Hessian matrix can be constructed in the same way as the single AU case by eq.3. The rest four branches use the second operations, so it needs to solve linear equations that involve conjugation. The validity of these solutions is shown in the Results.

Given a set of operations $\{\theta_j, \theta_j^*\}$ in a branch, the Hessian matrix can be constructed accordingly. There are two kinds of interactions. The first is the interactions between two atoms of the same AU or two AUs whose displacement vectors only differ by a phase. The

energy of these interactions is denoted as $E^{(1)}$. Suppose $\mathbf{H}_{ab} = \frac{\partial^2 E^{(1)}}{\partial X_a \partial X_b}$ represents the second derivative matrix for the atom pair a (from AU i) and b (from AU j), the Hessian matrix

block for the vector $\begin{pmatrix} X_a \\ X_b^{(i)} \end{pmatrix}$ is the same as eq.3, or it can be written in a conjugation form for the vector $(\mathbf{x}_a \mathbf{x}_b^{(i)} \mathbf{x}_a^* \mathbf{x}_b^{(i)*})^T$ as

$$H^1|_{a^i, b^{(i)}} = \begin{pmatrix} \mathbf{x}_a \mathbf{x}_b^{(i)} \mathbf{x}_a^* \mathbf{x}_b^{(i)*} \end{pmatrix}^T \text{ as } \begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \mathbf{R}_j e^{i\theta_j} & 0 & 0 \\ \mathbf{R}_j^T \mathbf{H}_{ba} e^{-i\theta_j} & \mathbf{R}_j^T \mathbf{H}_{bb} \mathbf{R}_j & 0 & 0 \\ 0 & 0 & \mathbf{H}_{aa} & \mathbf{H}_{ab} \mathbf{R}_j e^{-\theta_j} \\ 0 & 0 & \mathbf{R}_j^T \mathbf{H}_{ba} e^{i\theta_j} & \mathbf{R}_j^T \mathbf{H}_{bb} \mathbf{R}_j \end{pmatrix} \quad (\text{Eq.4.a})$$

The second type is the interactions between two atoms of two different AUs whose displacement vectors differ by both a conjugation operation and a phase shift operation. The

energy of these interactions is denoted as $E^{(2)}$. Suppose $\mathbf{K}_{ab} = \frac{\partial^2 E^{(2)}}{\partial X_a \partial X_b}$ represents the second derivative matrix for the atom pair a (from AU i) and b (from AU j), the Hessian matrix block for the vector $(\mathbf{x}_a \mathbf{x}_b^{(i)} \mathbf{x}_a^* \mathbf{x}_b^{(i)*})^T$ is

$$H^2|_{a^i, b^{(i)}} = \begin{pmatrix} \mathbf{x}_a \mathbf{x}_b^{(i)} \mathbf{x}_a^* \mathbf{x}_b^{(i)*} \end{pmatrix}^T \text{ is } \begin{pmatrix} \mathbf{K}_{aa} & 0 & 0 & \mathbf{K}_{ab} \mathbf{R}_j e^{i\theta_j} \\ 0 & \mathbf{R}_j^T \mathbf{K}_{bb} \mathbf{R}_j & \mathbf{R}_j^T \mathbf{K}_{ba} e^{i\theta_j} & 0 \\ 0 & \mathbf{K}_{ab} \mathbf{R}_j e^{-i\theta_j} & \mathbf{K}_{aa} & 0 \\ \mathbf{R}_j^T \mathbf{K}_{ba} e^{-i\theta_j} & 0 & 0 & \mathbf{R}_j^T \mathbf{K}_{bb} \mathbf{R}_j \end{pmatrix} \quad (\text{Eq.4.b})$$

Both H^1 and H^2 in eq.4.a and eq.4.b take the form of $\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix}$, where A is a Hermitian matrix and B is a symmetric matrix ($A^+ = A, B^T = B$). Therefore, the whole Hessian matrix, as the variable of displacement vectors and their conjugations, also takes the form of

$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix}$ and is also a Hermitian matrix. Hence, all calculated normal modes also have real frequencies in the double AU case. Besides, the four branches $\{\theta_j, \theta_j^*\}$, $\{\theta_j, -\theta_j^*\}$, $\{-\theta_j, \theta_j^*\}$ and $\{-\theta_j, -\theta_j^*\}$ have the same Hessian matrix or its conjugation, so only one branch $\{\theta_j, \theta_j^*\}$ needs to be computed.

In the double AU case, one needs to calculate no more than $n/2$ diagonalizations of the Hessian matrices on a single AU, and $n/8$ diagonalizations of the Hessian matrices on two AUs.

Triple AU Case—The rest 13 space groups belong to triple AU case (Table 1, rightmost column). In the PDB, only 2.2% structures belong to the case. These space groups have less than n phase solutions, and have at least one subgroup of order $n/3$ that belongs to the single AU case or the double AU case. As the name implies, for some solutions, it requires modes to be calculated on three asymmetric units.

Unfortunately, in the triple AU case, it is not easy to find an algebraic operation to relate the displacement vectors of various AUs in the unit cell. Instead, we generated a structure of three specific AUs, called the triple-unit, as the part in unit cell for normal mode calculation. The triple-unit consists of three AUs of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $(\mathbf{z}, \mathbf{x}, \mathbf{y})$ and $(\mathbf{y}, \mathbf{z}, \mathbf{x})$. In the unit cell, the triple-unit follows subgroup space group symmetry, so the normal modes for the whole crystal can be exactly calculated on the subgroup symmetry by the methods of the two previous cases.

Since the triple-unit is three times as large as an AU, mode calculation on it increases computational cost. But since all these cases have higher order symmetries, many branches produce degenerate modes, so one only needs to calculate modes on a small portion of branches.

The 13 space groups can be further divided into three sub-branches based on the way of mode calculation. The first sub-branch A, as listed in the upper portion of the rightmost column in Table.1 (III.A, a total of 5, 1.4% structures), has a subgroup of order $n/3$ that belongs to the single AU case. For example, space group P 2 3 has subgroup P222, which belongs to the single AU case. So, modes can be calculated with the previous methods by regarding the crystal as a triple-unit of the three AUs $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $(\mathbf{z}, \mathbf{x}, \mathbf{y})$ and $(\mathbf{y}, \mathbf{z}, \mathbf{x})$ with P222 symmetry. But unlike the standard single AU case, it is not necessary to calculate on every branch due to mode degeneracy. For any mode in the phase solution (branch) of $\{0 \ \pi \ \pi 0\}$ with eigenvalue λ and eigenvector $(\mathbf{a} \ \mathbf{b} \ \mathbf{c})^T$, where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ present the displacement vectors for the three AUs respectively, the phase solution $\{0 \ \pi 0 \ \pi\}$ always has a mode with the same eigenvalue λ and eigenvector $(\mathbf{c} \ \mathbf{a} \ \mathbf{b})^T$. Similarly, the solution $\{0 0 \ \pi \ \pi\}$ also has a mode with the same eigenvalue λ and eigenvector $(\mathbf{b} \ \mathbf{c} \ \mathbf{a})^T$. Hence, one just needs to calculate modes on the first phase solution $\{0 \ \pi \ \pi 0\}$ and obtains the three degenerate modes altogether.

The second sub-branch B, as listed in the middle portion of the rightmost column in Table.1 (III.B, a total of 6, 0.5% structures), has a subgroup of order $n/3$ that belongs to the double AU case. Similar to the first sub-branch A, the sub-branch also has degenerate modes. But the corresponding degenerate eigenvectors are $(\mathbf{a} \ \mathbf{b} \ \mathbf{c})^T$, $(\mathbf{b}+i\mathbf{c} \ \mathbf{c}+i\mathbf{a} \ \mathbf{a}+i\mathbf{b})^T$ and $(\mathbf{b}-i\mathbf{c} \ \mathbf{c}-i\mathbf{a} \ \mathbf{a}-i\mathbf{b})^T$ instead.

The third sub-branch C contains two special space groups F 432 and F 4₁ 3 2, as listed in the lower portion of the rightmost column in Table.1 (III.C, 0.3% structures). They have a subgroup I 422 and I 4₁ 2 2 respectively, by changing the axis of the unit cell from $(\hat{x}\hat{y}\hat{z})$ to $((\hat{x}+\hat{y})/2 \ (\hat{x}-\hat{y})/2 \ \hat{z})$ ⁷⁸. The procedure afterwards is the same as that of the sub-branch B.

In the triple AU case A and B, one needs to calculate no more than $n/4$ diagonalizations of the Hessian matrices on a single AU, and $n/4$ diagonalizations of the Hessian matrices on three AUs.

Mode Integration—So far, we have presented a mode calculation method for a unit cell under periodic boundary condition. It calculates modes on one or a few AUs for different branches based on space group. In each branch, the calculated eigenvectors form an

orthonormalized mode set. If necessary, the modes can be further extended to a whole unit cell. But, in many applications, one might be just interested in a single set of normal modes for one asymmetric unit. Therefore, we introduce a method to integrate modes from the various branches into a single set of orthonormalized normal modes based on the displacement variance-covariance matrix.

To achieve this, the effective displacement variance-covariance matrix $\mathbf{H}^{eff}$ is calculated by adding all the displacement variance-covariance matrices, or the inversion of the Hessian matrices, for all branches:

$$\mathbf{H}^{eff} = \sum_j \mathbf{H}_l^{-1}, \quad (\text{Eq.5})$$

where $\mathbf{H}_l$ is the Hessian matrix for branch l , and summation is over all branches. Some $\mathbf{H}_l$ may have zero frequency modes, so it can not be inverted directly. See the next section on how to process the Hessian matrix with zero frequency modes.

In the double AU case, some branches involve conjugation operations. For those branches,

the Hessian matrix has double-sized dimension and takes the form of $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix}$ (see Eq. 4).

The matrix $\mathbf{H}_l^{-1}$ for one asymmetric unit, as required in Eq.5, is the upper diagonal block of the inversion of the Hessian matrix (see Eq.4), or

$$\mathbf{H}_l^{-1} = (\mathbf{A} - \mathbf{B}\mathbf{A}^{*-1}\mathbf{B}^*)^{-1}. \quad (\text{Eq.6})$$

Since the conjugation of $\mathbf{H}_l^{-1}$ is also the positional fluctuation matrix for the conjugated branch, the summation in (Eq.5) needs to include twice the real part of $\mathbf{H}_l^{-1}$ in (Eq.6).

In the triple AU case, the solutions involve mode calculations on a triple-unit. Suppose the

inversion of Hessian matrix on the triple-unit takes the form of $\begin{pmatrix} \mathbf{H}^{-1}_{11} & \dots & \dots \\ \dots & \mathbf{H}^{-1}_{22} & \dots \\ \dots & \dots & \mathbf{H}^{-1}_{33} \end{pmatrix}$,

the displacement variance-covariance matrix $\mathbf{H}_l^{-1}$ for one asymmetric unit, as required in (Eq.5), can be calculated by

$$\mathbf{H}_l^{-1} = \mathbf{H}^{-1}_{11} + \mathbf{R}_2^T \mathbf{H}^{-1}_{22} \mathbf{R}_2 + \mathbf{R}_3^T \mathbf{H}^{-1}_{33} \mathbf{R}_3, \quad (\text{Eq.7})$$

where $\mathbf{R}_2$ and $\mathbf{R}_3$ are the rotational matrix for the second and the third elements in the triple-unit.

Diagonalization on the effective positional fluctuation correlation matrix $\mathbf{H}^{eff}$ gives the integrated normal modes. The modes with large eigenvalues correspond to the low frequency modes. Usually there are no zero-frequency modes in the integrated modes, because an AU in a crystal is not an isolated structure. But in some cases, one might be interested in separating the internal motions from the external motions of an asymmetric unit. To achieve this, $\mathbf{H}^{eff}$ is further modified by the zero modes processing method described in the following.

Zero Modes Processing—Since molecules in a crystal can have zero frequency translational modes, the Hessian matrix may not be strictly positive-definite. So, in order to calculate the matrix inversion, the Hessian matrix should be first modified by

$$\mathbf{H}^{\text{new}} = \mathbf{H} + \alpha \sum_k \mathbf{z}_k \mathbf{z}_k^T, \quad (\text{Eq.8})$$

where α is a very large number (10^6 in this study), the summation includes all zero-frequency $\mathbf{z}_k$ for the Hessian matrix. The method is similar the RAISE method in CHARMM eigenvector package²¹.

In the last step of the mode integration method, we separated the internal modes of an AU from the external modes (six translation-rotational modes) in a similar way, except that the summation in (Eq.8) includes all the six external modes.

Results

Accuracy Test

First, we tested accuracy of MGR-PIM (see Materials and Methods, simply denoted as PIM afterwards). The normal modes calculated in various branches from PIM were compared with those calculated on one whole unit cell under periodic boundary condition. Here, the stiffness of the packing interactions k_p in Eq. (9) was set to be the same as that of the long-range interactions k_r in the MGR (unity). Therefore, the system has exactly the same interactions in both calculations, so they should produce same modes.

We first tested on a CypA crystal structure (PDB ID: 3K0M) with a space group $P 2_1 2_1 2_1$. As shown in the theory section, space group $P 2_1 2_1 2_1$ belongs to the single AU case and has four normal mode branches. For each branch, normal modes are always calculated on a single AU, so the dimension of the eigenvectors is one-fourth of the dimension of those from the whole-unit-cell calculation.

We first examined the eigenvalues produced in both calculations. The normal modes of the whole-unit-cell calculation have no degeneracy modes except for the first three zero-frequency modes that represent translation invariance of a crystal. Unlike the normal modes of a free molecule, there are no zero-frequency rotational modes for a crystal. For each positive vibrational mode of the whole-unit-cell calculation, there is always one corresponding mode with exactly the same eigenvalue from one of the branches in the PIM calculation. Besides, each of branch 2, 3 and 4 also contains one zero-frequency mode, representing the translational modes along z , y and x axes, respectively.

Afterwards, we examined the eigenvectors calculated in both methods. Fig.1 shows two eigenvectors in both calculations. Fig.1(a) shows the fourth low-frequency mode in the whole-unit-cell calculation (on the left), which corresponds to the first low-frequency mode of branch 1 in the PIM calculation (on the right). On the left side of the figure, the whole eigenvector is for a whole unit cell. It can be divided into four different regions, each of which corresponds to a different AU in the unit cell, i.e. the first region is the displacement vector for the first AU, which is exactly same as the displacement vector from the PIM calculation (on the right). The rest three regions on the left side of Fig.1(a) are those for the three other AUs in the unit cell after space group symmetry operations. It is clear that all four AUs have same displacement vectors and the same phase, indicating that the mode belongs to branch 1. For example, the x components for the first and the second region of the displacement differ only by a sign, due to the space group operations between the two AUs. Other modes in branch 1 also follow this property. Besides, in these modes, the displacement vectors of one specific AU are orthogonal to each other, so they form an orthonormalized set of eigenvectors. To summarize, the modes from the whole-unit-cell calculation can be easily converted to those from the PIM calculation, and *vice versa*. Fig.

1(b) shows the sixth low-frequency mode, which corresponds to the second (the first non-zero) low-frequency mode of branch 4 in the PIM calculation. Similar to the previous case, the displacement vectors for all four regions are same, except that those for the last two regions have π phase shift. This indicates that this belongs to a mode of branch 4. For example, in the upper panel (left side), the first and the third region of the displacement vector are same (branch 1), but the same regions in the lower panel differ by a sign (indicating they have a π phase shift in branch 4).

We then tested on an RNase HI crystal structure (PDB ID: 2E4L) with space group P 4 2 2. As shown in the Theory section, space group P 4 2 2 belongs to the double AU case and has eight normal mode branches. For each of the first four branches, modes are still calculated on single AU, with the same procedure as that for the single AU case. For the last four branches, one needs to construct and diagonalize a double-sized Hessian matrix (see Theory). Half of the calculated modes belong to branch 5 and 7 (degenerate modes of branch 5), while the other half belongs to branch 6 and 8 (degenerate modes of branch 6). Similarly, our tests showed that modes from the whole-unit-cell calculation and from the PIM calculation are consistent with each other for both eigenvalues and eigenvectors. For the last four branches, the PIM calculation produces complex eigenvectors, while the whole-unit-cell calculation always produces real eigenvectors. But each real eigenvector can still be expressed as a linear combination of the two degenerate complex eigenvectors of the same eigenvalue in the PIM calculation. Therefore, in these cases, the PIM calculation and the whole-unit-cell calculation are consistent.

We also tested on the other structures in the space group test set in Table 2. Similarly, both calculations produced modes of exactly same frequencies. Hence, we concluded that the PIM methods can accurately calculate modes based on one or a few asymmetric units.

Tests on Branch Modes

After confirming accuracy of the PIM method, we examined the difference among the modes of various branches. Fig.2 shows the eigenvalues for the lowest frequency modes calculated in different branches for the structures 3K0M and 2E4L. It is clear that the profiles of the lowest frequency eigenvalues are similar among different branches. However, larger differences were found for the first 10 lowest frequency modes (inset). Since the magnitude of the eigenvalues for these modes are relatively small, the differences can be as large as two or three folds. As also mentioned in the previous section, zero-frequency modes can only be found in some certain branches, depending on the space group.

Differences can also be found in eigenvectors of various branches. Because of spatial arrangements and phase differences among various AUs, some branches are likely to have modes with lower frequency, indicating less packing conflicts among neighboring AUs. The difference becomes smaller for higher frequency modes. Moreover, the upper half of Fig.3 shows the overlap indices (defined in the Materials and Methods section) among the lowest frequency mode subspaces of a free AU (indicated as free-molecule in the following) and an AU in a crystal in various branches for the structure 3K0M. It was found that the mode differences among various branches (four middle circles in Fig.3) are in general smaller than the differences between one of the many branches and free molecule calculation (top circle in Fig.3), suggesting that packing interactions affect low frequency modes. It was also found that the mode differences among certain branch pairs can be as large as the differences between modes of a branch and a free molecule, suggesting large difference among some specific branches. For example, in this 3K0M case, the overlap index between branch 2 and branch 3 is 0.929 and that between branch 1 and branch 4 is 0.931, while that between any branch and free molecule is 0.927 on average. Similar results were also found for the other space groups. Also as shown in Fig.4, the average overlap index between any branch and

free molecule (dot line) is in general smaller than that between any two branches (small dash line), indicating mode subspaces of free molecule are usually more different from those of branches.

The comparison test also suggests that modes of various branches are different, therefore, it is necessary to calculate modes for all of them. For example, it is not sufficient to capture the packing effects of the low frequency normal modes for a structure in a crystal by only imposing periodic boundary conditions of AUs (i.e. branch 1 in current study). In the next section, we will discuss the benefit of integrating the modes of all branches.

Tests on Integrated Modes

Another important feature of the PIM method is the capability of producing integrated modes for a single AU. In many applications, such as normal-mode-based refinement of crystallographic thermal parameters⁵⁸⁻⁶⁰, orthonormalized modes for single AU is essential. In the PIM method, the mode sets for various branches are integrated into a single mode set by summing up all displacement variance and co-variance matrices. Therefore, it doesn't even need diagonalization procedure for each branch, but only requires a few matrix inversion operations on the calculated Hessian matrices. So the integrated mode calculation is much less computationally intensive. No matter whether the crystal system belongs to the single, double or triple AU cases, one only needs to diagonalize for once on the Hessian matrix with the dimension of a single AU.

In the following, we tested whether the integrated modes can represent the modes of various branches. As shown in Fig.3, for the case of the structure 3K0M, the differences between the integrated modes and the modes of a branch, as indicated by the overlap indices (lines between bottom circle and any of middle four circles), are consistently smaller than or similar to the differences between the modes of two branches (lines between any pair of the middle four circles). Besides, the integrated modes are significantly closer to the branch modes (lines between bottom circle and any of middle four circles) than the free-molecule modes (lines between top circle and any of middle four circles). These suggest that the integrated modes can represent the mode sets of various branches well. Moreover, we also found that the difference of the integrated modes and the free-molecule modes (line between top circle and bottom circle, 0.952) are significantly smaller than the difference between the branch modes and the free-molecule modes (on average 0.927). This finding suggests that each branch can only represent a part of molecular motions in the crystal, while modes after integration can represent motions better.

Similar overlap tests were also performed on all the other space groups except for P 1 (only one AU in a unit cell) and F 4 3 2 / F 4 1 3 2 (same treatments as their subgroup I 4 2 2 / I 4 1 2 2) (Fig.4). For overlap calculations that involve branch modes, we excluded the branches of complex modes in the double AU case. Besides, for the space groups that belong to the triple AU case, we only included the branches that can be treated in the same way as single AU cases. Therefore, in all cases, the mode subspaces are of the same dimension. Along with Fig.4, the average value for the various overlap indices were also calculated and summarized in Table 3. It is clear that we got similar results to the case of 3K0M.

We also calculated the average local character indicator of both the integrated modes and the free-molecule modes for the structures of various space groups (shown in Table 3). As expected, the integrated modes have less local character indicator in magnitude than the free-molecule modes, suggesting that packing interactions affect localized motions of internal motions.

Test against Experimental Data

We tested the effectiveness of the MGR-PIM in fitting experimental anisotropic displacement parameters (ADPs) in crystallographic data. There are three purposes for this test. First, we wanted to see the overall effects of crystal packing on the internal motions of structures in crystalline environment; second, to see how the integrated modes work in fitting; third, to obtain optimal stiffness of the packing interactions k_p (set as unity, or k_p in the previous tests) based on the experimental data.

The procedure is mostly the same as that used in our previous study⁵⁵, except that each fitting test is repeated for 10 times and the average KL distance and Pearson correlation were calculated. The test set (Table 2) is also the same as the one used in the previous studies⁵⁵. The scoring function to be minimized was chosen as the KL distance between experimental ADPs and calculated ADPs. Smaller KL distance indicates better fitting. Besides, the Pearson correlation between experimental ADPs and calculated ADPs were used to cross-validate the fitting results. In the test, we also used the MNM method⁵⁵ to calculate modes without packing interactions in crystal, and performed on the same test. The MNM results were shown here as a reference.

As shown in Fig.5, if no packing interactions are included in the fitting test, or $k_p = 0$, the MGR performs worse than the MNM. However, when packing interactions are included in the calculation, the performance of the MGR-PIM is significantly improved and starts to perform better than the MNM even when the packing stiffness is weak ($k_p = 0.1$). Increasing the packing stiffness lowers the average KL distance and raises the average Pearson correlation, while both quantities reach optimal values when the packing stiffness k_p is about 0.6. Further increasing the packing stiffness does not improve the average KL distance, and even slightly worsens the average Pearson correlation. This seems to suggest that 0.6 is the optimal packing stiffness for the MGR-PIM. Interestingly, the magnitude of packing stiffness is very close to the ratio of Miyazawa-Jernigan contact energy for inter-protein and intra-protein interactions (0.68)⁶⁹. The value of the optimized packing stiffness ($k_p = 0.6$) suggests that the crystal packing interactions between the AUs are weaker than the intra-AU interactions ($k_r = 1.0$), but not substantially weaker. Consequently, it is clear that inter-AU packing interactions have a large impact on the internal motions of structures in crystalline environment. Our finding is in sharp contrast to what reported in references^{69, 79}. It is worth mentioning that it is an approximation that the inter-AU interactions are treated as the packing interactions in crystal. In a small number of cases in which the AU contains multiple copies of molecules, the inter-molecular packing interactions also exist within one AU. See Supplemental Materials for more test results on fitting experiment.

Discussions

In this study, we developed a normal mode analysis, named phase integrated method (PIM), for studying biomolecules in crystalline environment. PIM aims to reduce computational cost of mode calculation for structures in a unit cell according to the space group of crystal. It focuses more on obtaining internal deformation for the structure in an asymmetric unit (AU), rather than the vibration of the entire crystal lattice. Under periodic boundary condition, we are able to obtain several “branches”, each of which defines specific relations among the displacement vectors of various AUs. These relations, in turn, enable PIM to calculate a set of normal modes on one or a few AUs for each branch, therefore dramatically reducing computational cost. From these branch modes, PIM can further obtain a set of integrated modes for a single AU.

PIM can accurately compute normal modes for a unit cell in crystalline environment on one or a few AUs, because of the following three features. First, in each branch, PIM only uses phase shift and conjugation operations to relate displacement vectors for different AUs. Note that these operations are only dependent on space group, rather than mode frequencies. Therefore, PIM is able to obtain an orthogonal complete set of eigenvectors for each branch. Second, there are usually multiple branches (except for P_1). Therefore, PIM calculation can cover all degrees of freedom in a unit cell, rather than those for one AU. Third, any two unit-cell modes (derived from AU modes) calculated from different branches are always orthogonal. Because of the 2nd and 3rd features, modes from all branches, when put together, form an orthogonal complete set of eigenvectors for a whole unit cell.

All 65 chiral space group symmetries for biomolecules belong to three cases: single AU, double AU, and triple AU cases. The single AU case covers 29 space group symmetries and about 73.3% of all available crystal structures. The displacement vectors of any two AUs differ from each other by only phase shift, and the derived PIM Hessian has the size of one AU. The double AU case covers 23 space group symmetries and about 24.0% crystal structures. The displacement vectors of any two AUs relate to each other by both phase shift and conjugation operations, and the derived PIM Hessian has the size of two AUs. The triple AU case covers the rest 13 space groups and only about 2.2% crystal structures. In the triple AU case, we can not find a clear relation among the displacement vectors of some AU pairs. To deal with this situation, a big unit that contains three AUs is used to perform PIM calculation, so the triple AU case is essentially reduced to one of the previous two cases. In addition, the high order symmetry in the triple AU case allows to bypassing mode calculations of some branches, due to mode degeneracy.

In addition to the ability to compute modes for a unit cell, PIM can be used to obtain an orthogonal complete set of integrated normal modes for one AU. PIM achieves this by summing positional variance-covariance matrices for all branches, then diagonalizing the integrated matrix. It's essentially based on equipartition theorem that modes from any branch should have same contribution to thermal fluctuation. If one only needs to obtain the integrated modes, instead of the detailed branch modes, such calculation only involves several matrix inversions and one matrix diagonalization on one AU for all cases. So it can further reduce computational cost. Our tests show that the integrated modes can better represent the molecular deformation in crystalline environment than one individual branch of the branch modes.

Compared to the conventional methods, in which mode calculation was performed on a unit cell, PIM can dramatically reduce computational cost when there are a lot of AUs in a unit

cell. Consider a unit cell with n AUs, the PIM can compute unit cell modes with $\frac{c}{n^2}$ CPU time of the conventional method, where c is a constant determined by the space group symmetry. If only integrated modes are required to be computed, the CPU cost can be

further reduced to $\frac{1}{n^3}$.

PIM scheme was implemented with MGR method⁷⁷, and was found to improve normal-mode-based x-ray crystallographic thermal parameter refinement according to a fitting study on a set of 78 crystal structures. The optimal inter-AU packing stiffness (the long range interactions between AUs) was found to be about 60% in magnitude of the intra-AU stiffness (the long range interactions within one AU). With the optimal packing stiffness, PIM can even outperform force-field-based normal mode analysis (MNM⁵⁵) on a free molecule. Interestingly, the ratio of stiffness between inter-AU and intra-AU interactions discovered in this study is surprisingly close to the estimation by the Miyazawa-Jernigan

contact energy (~68%)⁶⁹, but is dramatically different from 2% ~ 25%, as suggested by Zheng's group⁶⁹. Our results are very different from those by Zheng's group, mainly because the contributions of thermal parameters from overall rigid body motions were directly estimated from theoretical calculation in Zheng's method. While in our fitting procedure, we used the same method as the normal-mode-based refinement, i.e. the contributions from overall rigid body motions were modeled by external modes of AU. Besides, the two methods used different mode calculation methods. We argued that the inter-AU packing interactions, although slightly weaker than the intra-AU molecular interactions, are still strong enough to have substantial impact to internal molecular vibration. PIM will be applied to x-ray crystallographic refinement experiment in future studies.

In this study, we applied unit-cell periodic boundary condition on the eigenvectors among unit cells in mode calculation. Under this condition, each AU has exactly same environment around it. Besides, the same AUs in all unit cells should have same deformation patterns, so we can safely exploit relations of molecular deformation according to space group symmetry. Without periodic boundary condition, it becomes much tougher to utilize space group symmetry, and to reduce computational cost in mode calculation.

Among all solved crystal structures, about three quarter structures belong to the simplest single AU case. In contrast, about one quarter structures belong to the double AU case, and only 2.2% structures belong to the triple AU case. It seems that biomolecules are more likely to be crystallized with low order symmetries⁸⁰⁻⁸². In the most abundant single AU case, deformations from different AUs only differ from each other by a phase shift. An interesting question is if such uniform deformation patterns across all AU help to stabilizing crystal? We believe that some further analyses on the dynamics of crystal structure can help to answer the question.

Materials and Methods

In this section, we mainly describe the procedures to implement and test PIM. We first explain the normal mode calculation method MGR and the way to implement it in PIM. Then we define some test criteria used in this study.

MGR Method

In this study, we implemented the PIM method for crystal packing on the MGR method previously developed in our lab⁵¹, named MGR-PIM. MGR is an all-atom based normal mode analysis with molecular geometry restraints for large biomolecular complexes. In this method, the potential function contains not only long range pair-wise harmonic interactions used in the Elastic Network Models (ENMs), but also short range bond interactions to strengthen the bond stiffness. Similar to ENMs, MGR does not require initial energy minimization. While compared to ENMs, MGR produces closer lowest-frequency modes to those calculated from the force-field-based normal mode methods. Same as what was implemented in our original study, the MGR method was combined with the residue-based RTB coarse-graining scheme to reduce computational cost.

The MGR potential function used in this study is

$$V = \sum k_l (l - l^0)^2 + \sum k_\theta (\theta - \theta^0)^2 + \sum k_\phi (\phi - \phi^0)^2 + \sum_{r_{ij}^0 \leq r_c} k_r (r_{ij} - r_{ij}^0)^2 + \sum_{r_{ij}^0 \leq r_p} k_p (r_{ij} - r_{ij}^0)^2, \quad (\text{Eq.9})$$

where the first four terms represent bond length, bond angle, bond dihedral angle and the long range interactions originally used in the MGR method, and the last term represents the

packing interactions among various AUs. In this study, the cutoff distance for the packing interactions r_p was chosen to be same as the cutoff distance for the long-range interactions r_c , i.e. 6Å. The stiffness of packing interactions k_p is an undetermined parameter. We show how to choose the k_p value of in the Results section.

Test Criteria

Similar to what was used in our previous study^{50-51, 55}, two quantities were mainly calculated in evaluating low-frequency eigenvectors. The first one is the “overlap index”, representing similarity of the lowest frequency eigenvector subspaces between two methods. The overlap index is defined in this study as the average projection of the eigenvectors of one method onto the mode subspace of the other method,

$$P = \frac{1}{n_m} \sum_{i,j=1}^{n_m} \|\mathbf{v}_i \cdot \mathbf{u}_j\|^2, \quad (\text{Eq.10})$$

where $\mathbf{v}_i$ and $\mathbf{u}_j$ are i th and j th eigenvectors of the two methods respectively, n_m is the number of lowest frequency modes included in each subspace (chosen as 100 in this study), and the summation is over all eigenvectors in both subspaces. Unlike the definition in our previous study, the modified quantity is independent to the order of two methods and complex eigenvectors are allowed in the test. A P closer to one indicates higher overlap between two subspaces, therefore higher similarity between the two methods.

The second quantity is the normalized “local character indicator”⁴, used to measure the amount of localized displacement of eigenvectors. It was defined as

$$L = \frac{n}{n_m} \sum_{i=1}^{n_m} \sum_{j=1}^n \mathbf{v}_i(j)^4, \quad (\text{Eq.11})$$

where $\mathbf{v}_i(j)$ is j th element of i th eigenvector, n is the dimension of eigenvectors, n_m is the number of modes (chosen as 100), and the summation includes all the vector elements and all the low frequency modes. Usually, a large L indicates large “tip effect” for the method⁵⁰. The vector dimension n is included in Eq.11 to make the quantity less dependent on the molecular size.

In the fitting test against experimental crystallographic data, same as that in our previous study⁵⁵ we used the average KL distance and Pearson correlation for two sets of anisotropic displacement parameters. The KL distance indicates the difference between the Gaussian probability distributions defined by the experimental and theoretical ADPs⁸³⁻⁸⁴. The KL distance for each atom is

$$D_{KL} = -\frac{3}{2} + \frac{1}{2} \sum_{p=1}^3 \ln \frac{\omega_{(p)}^{\text{mode}}}{\omega_{(p)}^{\text{data}}} + \frac{1}{2} \sum_{p=1}^3 \sum_{q=1}^3 \frac{\omega_{(p)}^{\text{data}}}{\omega_{(q)}^{\text{mode}}} \left| \left(\mathbf{a}_{(p)}^{\text{data}} \right)^T \mathbf{a}_{(q)}^{\text{mode}} \right|^2. \quad (\text{Eq.12})$$

where $\omega_{(p)}^{\text{mode}}$, $\omega_{(p)}^{\text{data}}$ ($p \in 1, 2, 3$) are eigenvalues, $\mathbf{a}_{(p)}^{\text{mode}}$, $\mathbf{a}_{(p)}^{\text{data}}$ ($p \in 1, 2, 3$) are eigenvectors of the theoretical (*mode*) and experimental (*data*) ADP matrices. In this study, the KL distance for each structure is calculated by averaging the atomic KL distance for all heavy atoms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge support from National Institutes of Health (R01-GM067801), National Science Foundation (MCB-0818353), and the Welch Foundation (Q-1512).

Abbreviations

MNM	Minimalist Network Model
RTB	Rations-Translations of Blocks
PIM	Phase Integrated Method
NMA	Normal Mode Analysis
CHARMM	Chemistry at HARvard Molecular Mechanics
AMBER	Assisted Model Building with Energy Refinement
ENM	Elastic Network Model
MGR	NMA with Molecular Geometry Restraint
ADP	Anisotropic Displacement Parameter
AU	Asymmetric Unit
KL	Kullback-Leibler

References

1. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: Trypsin Inhibitor, crambin, ribonuclease and Lysozyme. *J. Mol. Biol.* 1985; 181:423–447. [PubMed: 2580101]
2. McCammon, JA.; Harvey, S. Dynamics of Proteins and Nucleic Acids. Cambridge University Press; Cambridge: 1987.
3. Brooks CL III, Karplus M, Pettitt BM. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. *Adv. Chem. Phys.* 1988; 71:1–249.
4. Brooks BR, Janezic D, Karplus M. Harmonic Analysis of large systems. I. Methodology. *J. Comput. Chem.* 1995; 16:1522–1542.
5. Ma J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure.* 2005; 13:373–380. [PubMed: 15766538]
6. Ma J. New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. *Curr Protein Pept Sci.* 2004; 5:119–123. [PubMed: 15078222]
7. Ma J, Karplus M. The allosteric mechanism of the chaperonin GroEL: A dynamic analysis. *Proc. Natl. Acad. Sci. USA.* 1998; 95:8502–8507. [PubMed: 9671707]
8. Ma J, Karplus M. Ligand-induced conformational changes in *ras* p21: A normal mode and energy minimization analysis. *J. Mol. Biol.* 1997; 274:114–131. [PubMed: 9398520]
9. Seno Y, Go N. Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction. *J Mol Biol.* 1990; 216:95–109. [PubMed: 2231732]
10. Seno Y, N Go. Deoxymyoglobin studied by the conformational normal mode analysis II. The conformational change upon oxygenation. *J. Mol. Biol.* 1990; 216:111–126. [PubMed: 2231726]
11. Thomas A, Field MJ, Perahia D. Analysis of the low-frequency normal modes of the R state of aspartate transcarbamylase and a comparison with the T state modes. *J Mol Biol.* 1996; 261:490–506. [PubMed: 8780788]
12. Thomas A, et al. Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase. *J Mol Biol.* 1996; 257:1070–1087. [PubMed: 8632469]
13. Thomas A, et al. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins.* 1999; 34:96–112. [PubMed: 10336386]

14. Li G, Cui Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca(2+)-ATPase. *Biophys J.* 2002; 83:2457–2474. [PubMed: 12414680]
15. Cui Q, et al. A Normal Mode Analysis of Structural Plasticity in the Biomolecular Motor F(1)-ATPase. *J Mol Biol.* 2004; 340:345–372. [PubMed: 15201057]
16. Krebs WG, et al. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins.* 2002; 48:682–95. [PubMed: 12211036]
17. Lu M, Ma J. The Role of Shape in Determining Molecular Motions. *Biophys. J.* 2005; 89:2395–2401. [PubMed: 16055547]
18. Flynn TC, Ma J. Theoretical analysis of twist/bend ratio and mechanical moduli of bacterial flagellar hook and filament. *Biophys J.* 2004; 86:3204–3210. [PubMed: 15111433]
19. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 1996; 105:1902–1921.
20. MacKerell AD, et al. All-atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem.* 1998; B102:3586–3616.
21. Brooks BR, et al. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry.* 1983; 4:187–217.
22. Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry.* 2000; 21:1049–1074.
23. Ponder JW, Case DA. Force fields for protein simulations. *Adv Protein Chem.* 2003; 66:27–85. [PubMed: 14631816]
24. Wang J, et al. Development and testing of a general amber force field. *J Comput Chem.* 2004; 25:1157–74. [PubMed: 15116359]
25. Case DA, et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005; 26:1668–88. [PubMed: 16200636]
26. Tama F, et al. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins.* 2000; 41:1–7. [PubMed: 10944387]
27. Zhou L, Siegelbaum SA. Effects of surface water on protein dynamics studied by a novel coarse-grained normal mode approach. *Biophys J.* 2008; 94:3461–74. [PubMed: 18212016]
28. Ming D, Wall ME. Allostery in a coarse-grained model of protein dynamics. *Phys Rev Lett.* 2005; 95:198103. [PubMed: 16384030]
29. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 1996; 77:1905–1908. [PubMed: 10063201]
30. Haliloglu T, Bahar I, Erman B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* 1997; 79:3090–3093.
31. Atilgan AR, et al. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J.* 2001; 80:505–515. [PubMed: 11159421]
32. Doruker P, Jernigan RL, Bahar I. Dynamic of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* 2002; 23:119–127. [PubMed: 11913377]
33. Kurkcuoglu O, Jernigan RL, Doruker P. Collective dynamics of large proteins from mixed coarse-grained elastic network model. *Qsar & Combinatorial Science.* 2005; 24:443–448.
34. Eom K, et al. Coarse-graining of protein structures for the normal mode studies. *J Comput Chem.* 2007; 28:1400–10. [PubMed: 17330878]
35. Ming D, et al. How to Describe Protein Motion Without Amino-acid Sequence and Atomic Coordinates. *Proc. Natl. Acad. Sci. USA.* 2002; 99:8620–8625. [PubMed: 12084922]
36. Tama F, Wrighers W, Brooks CL. Exploring Global Distortions of Biological Macromolecules and Assemblies from Low-resolution Structural Information and Elastic Network Theory. *J Mol Biol.* 2002; 321:297–305. [PubMed: 12144786]
37. Hinsen K. The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry.* 2000; 21:79–85.

38. Kondrashov DA, et al. Protein structural variation in computational models and crystallographic data. *Structure*. 2007; 15:169–77. [PubMed: 17292835]
39. Lyman E, Pfaendtner J, Voth GA. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys J*. 2008; 95:4183–92. [PubMed: 18658214]
40. Moritsugu K, Smith JC. Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian. *Biophys J*. 2007; 93:3460–9. [PubMed: 17693469]
41. Woodcock HL, et al. Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J Chem Phys*. 2008; 129:214109. [PubMed: 19063546]
42. Zheng W. Anharmonic Normal Mode Analysis of Elastic Network Model Improves the Modeling of Atomic Fluctuations in Protein Crystal Structures. *Biophys J*. 2010; 98:3025–3034. [PubMed: 20550915]
43. Stember JN, Wriggers W. Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *J Chem Phys*. 2009; 131:074112. [PubMed: 19708737]
44. Mendez R, Bastolla U. Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. *Phys Rev Lett*. 2010; 104:228103. [PubMed: 20867208]
45. Kong Y, et al. Conformational pathways in the gating of Escherichia coli mechanosensitive channel. *Proc Natl Acad Sci U S A*. 2002; 99:5999–6004. [PubMed: 11972047]
46. Kong Y, et al. Conformational Flexibility of Pyruvate Dehydrogenase Complexes: A Computational Analysis by Quantized Elastic Deformational Model. *J. Mol. Biol.* 2003; 330:129–135. [PubMed: 12818207]
47. Ming D, et al. Simulation of F-actin Filaments of Several Microns. *Biophys J*. 2003; 85:27–35. [PubMed: 12829461]
48. Ming D, et al. Substructure Synthesis Method for Simulating Large Molecular Complexes. *Proc. Natl. Acad. Sci. USA*. 2003; 100:104–109.
49. Ming D, et al. Domain Movements in Human Fatty Acid Synthase by Quantized Elastic Deformational Model. *Proc. Natl. Acad. Sci. USA*. 2002; 99:7895–7899. [PubMed: 12060737]
50. Lu M, Poon B, Ma J. A New Method for Coarse-Grained Elastic Normal-Mode Analysis. *Journal of Chemical Theory and Computation*. 2006; 2:464–471. [PubMed: 21760758]
51. Lu MY, Ma JP. Normal mode analysis with molecular geometry restraints: Bridging molecular mechanics and elastic models. *Archives of Biochemistry and Biophysics*. 2011; 508:64–71. [PubMed: 21211510]
52. Kim JIN, Sungsoo, Eom Kilho. Domain decomposition-based structural condensation of large protein structures for understanding their conformational dynamics. *J. Comput. Chem*. 2010; 32:161–169. [PubMed: 20645300]
53. Yang LW. Models with energy penalty on interresidue rotation address insufficiencies of conventional elastic network models. *Biophys J*. 100:1784–93. [PubMed: 21463592]
54. Orellana L, et al. Approaching Elastic Network Models to Molecular Dynamics Flexibility. *Journal of Chemical Theory and Computation*. 6:2910–2923.
55. Lu M, Ma J. A minimalist network model for coarse-grained normal mode analysis and its application to biomolecular x-ray crystallography. *Proc Natl Acad Sci U S A*. 2008; 105:15358–63. [PubMed: 18832168]
56. Chen X, et al. Normal-mode refinement of anisotropic thermal parameters for potassium channel KcsA at 3.2 Å crystallographic resolution. *Structure*. 2007; 15:955–62. [PubMed: 17698000]
57. Ni F, et al. Application of normal-mode refinement to X-ray crystal structures at the lower resolution limit. *Acta Cryst*. 2009; D65:633–643.
58. Poon BK, et al. Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-Å crystallographic resolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:7869–7874. [PubMed: 17470791]
59. Chen X, et al. Structural improvement of unliganded simian immunodeficiency virus gp120 core by normal-mode-based X-ray crystallographic refinement. *Acta Crystallogr D Biol Crystallogr*. 2009; 65:339–47. [PubMed: 19307715]

60. Chen X, et al. Structure of the full-length Shaker potassium channel Kv1.2 by normal-mode-based X-ray crystallographic refinement. *Proc Natl Acad Sci U S A*. 2010; 107:11352–7. [PubMed: 20534430]
61. Diamond R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallographica Section A*. 1990; 46:425–435.
62. Kidera A, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J Mol Biol*. 1992; 225:457–75. [PubMed: 1593630]
63. Kidera A, Go N. Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci U S A*. 1990; 87:3718–22. [PubMed: 2339115]
64. Schroder GF, Levitt M, Brunger AT. Super-resolution biomolecular crystallography with low-resolution data. *Nature*. 2010; 464:1218–22. [PubMed: 20376006]
65. Brunger AT, et al. Application of DEN refinement and automated model building to a difficult case of molecular-replacement phasing: the structure of a putative succinyl-diaminopimelate desuccinylase from *Corynebacterium glutamicum*. *Acta Crystallogr D Biol Crystallogr*. 2012; 68:391–403. [PubMed: 22505259]
66. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci U S A*. 2004; 101:6957–62. [PubMed: 15096585]
67. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics*. 2008; 24:521–8. [PubMed: 18089618]
68. Riccardi D, Cui Q, Phillips GN Jr. Application of elastic network models to proteins in the crystalline state. *Biophys J*. 2009; 96:464–75. [PubMed: 19167297]
69. Hafner J, Zheng W. Optimal modeling of atomic fluctuations in protein crystal structures for weak crystal contact interactions. *J Chem Phys*. 2010; 132:014111. [PubMed: 20078153]
70. van Vlijmen HWT, Karplus M. Normal mode analysis of large systems with icosahedral symmetry: Application to (Dialanine)₆₀ in full and reduced basis set implementations. *J Chem Phys*. 2001; 115:691–698.
71. van Vlijmen HWT, Karplus M. Normal Mode Calculations of Icosahedral Viruses with Full Dihedral Flexibility by Use of Molecular Symmetry. *J Mol Biol*. 2005; 350:528–542. [PubMed: 15922356]
72. Yang Z, Bahar I, Widom M. Vibrational Dynamics of Icosahedrally Symmetric Biomolecular Assemblies Compared with Predictions Based on Continuum Elasticity. *Biophys J*. 2009; 96:4438–4448. [PubMed: 19486668]
73. Simonson T, Peahia D. Normal modes of symmetric protein assemblies. Application to the tobacco mosaic virus protein disk. *Biophys J*. 1992; 61:410–427. [PubMed: 1547329]
74. Dykeman EC, Sankey OF. Low Frequency Mechanical Modes of Viral Capsids: An Atomistic Approach. *Physical Review Letters*. 2008; 100:028101. [PubMed: 18232930]
75. Cayley A. On the theory of groups, as depending on the symbolic equation $\theta^n = 1$. *Philosophical Magazine*. 1854; 7:40–47.
76. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994; 50:760–3. [PubMed: 15299374]
77. Lu M, Ma J. Normal mode analysis with molecular geometry restraints: bridging molecular mechanics and elastic models. *Arch Biochem Biophys*. 2011; 508:64–71. [PubMed: 21211510]
78. Hahn, T.; International Union of Crystallography. Brief teaching edition of volume A, Space-group symmetry. 5th ed. Vol. vi. Published for the International Union of Crystallography by Kluwer Academic Publishers; Dordrecht ; Boston: 2002. International tables for crystallography; p. 154
79. Hafner J, Zheng W. All-atom modeling of anisotropic atomic fluctuations in protein crystal structures. *J Chem Phys*. 2011; 135:144114. [PubMed: 22010705]
80. Andersson KM, Hovmöller S. The protein content in crystals and packing coefficients in different space groups. *Acta Crystallogr D Biol Crystallogr*. 2000; 56:789–90. [PubMed: 10930824]
81. Chruszcz M, et al. Analysis of solvent content and oligomeric states in protein crystals--does symmetry matter? *Protein Sci*. 2008; 17:623–32. [PubMed: 18359856]

82. Stephanie W, Wukovitz TOY. Why protein crystals favour some space-groups over others. *Nature structural biology*. 1995; 2:1062–1067.
83. Ming D, Wall ME. Allostery in a coarse-grained model of protein dynamics. *Physical Review Letters*. 2005; 95:159902.
84. Eyal E, et al. Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. *Bioinformatics*. 2007; 23:I175–I184. [PubMed: 17646294]
85. Kondrashov DA, Cui Q, Phillips GN Jr. Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data. *Biophys J*. 2006; 91:2760–7. [PubMed: 16891367]

Highlight

1. A normal mode analysis, named phase integrated method (PIM), is developed for computing modes of biomolecules in crystalline environment.
2. PIM can efficiently calculate low-frequency modes for a unit cell based on one or a few asymmetric units
3. PIM has an option to map modes onto a single asymmetric unit to form an orthonormalized mode set
4. From fitting test on a set of 83 ultra-high-resolution structures, the optimal inter-asymmetric-unit packing stiffness was found to be about 60% of that of intra-asymmetric-unit interactions (nonbonded interaction only).
5. PIM can dramatically reduce computational cost in mode calculation on structures in crystalline environment. It can be directly applied to normal-mode-based thermal parameter refinement in X-ray crystallography.

Fig.1.

Two unit-cell eigenvectors for the structure 3K0M ($P2_12_12_1$) in crystalline environment. The x-axis is the atom index. The profile is divided into four regions, each of which represents the displacement for an AU in the unit cell. The symbols above the plots denote the space group symmetry operations for the corresponding AUs. The y-axes are the x-, y- and z-components of eigenvectors. (a) the fourth lowest-frequency mode of the unit-cell calculation, which corresponds to the first mode of the branch 1 (phase set $0,0,0,0$); (b) the sixth lowest-frequency mode of the unit-cell calculation, which corresponds to the second mode of the branch 4 (phase set $0,0,\pi,\pi$).

Fig.2.

Eigenvalue profile of the first 100 lowest frequency modes for various branches in the PIM method. Upper panel: 3K0M, lower panel: 2E4L. Inset figures provide fine-scale resolution for the first 10 modes.

Fig.3.

The overlap indices of the 100 lowest frequency mode subspaces for various mode calculations on the structure 3K0M. The top-most circle denotes the mode subspace for the free-molecule calculation; the bottom-most circle denotes that for the integrated calculation; the middle four circles denote those of the four branches. The number notations for branches (e.g. 1:0000) denote the branch number and the corresponding phases. The number along the connection between any two circles is the overlap index of the two corresponding mode subspaces.

Fig.4.

The average overlap indices for the 62 proteins in the space group test set. The x-axis is the structure index. The y-axis is average overlap indices. Solid line is the overlap indices between free molecule and the integrated calculation. Dot line is the average overlap indices between free molecule and any branch. Dash line is that between the integrated calculation and any branch. Small dash line is that between any two branches.

Fig.5.

The performance of PIM in the fitting test for various packing stiffness. The average KL Distance is shown as solid line plus squares (left y-axis). The average Pearson correlation is shown as dash dot line plus hollow squares (right y-axis). The values for the MNM method are shown in the figure as the reference (two horizontal lines, solid and dash dot).

Table 1

The PIM cases for all chiral space group symmetries. Each column lists all space groups that belongs to the single AU case (denoted as Single (I)), the double AU case (denoted as Double (II)), or the triple AU case (denoted as Triple (III)). The triple AU case can be further divided into three groups A, B and C. The percentage digits in the parenthesis following each case category are the abundance of structures in the PDB.

Single (I) (73.3%)	Double (II) (24.0%)	Triple (III)
P 1	P 4 2 2	Group A. (1.4%)
P 1 2 1	P 4 ₂ 2	P 2 3
P 1 2 ₁ 1	P 4 ₁ 2 2	P 2 ₁ 3
C 1 2 1	P 4 ₁ 2 ₁ 2	I 2 3
P 2 2 2	P 4 ₂ 2 2	I 2 ₁ 3
P 2 2 2 ₁	P 4 ₂ 2 ₁ 2	F 2 3
P 2 ₁ 2 ₁ 2	P 4 ₃ 2 2	Group B (0.5%)
P 2 ₁ 2 ₁ 2 ₁	P 4 ₃ 2 ₁ 2	P 4 3 2
C 2 2 2 ₁	I 4 2 2	P 4 ₁ 3 2
C 2 2 2	I 4 ₁ 2 2	P 4 ₂ 3 2
F 2 2 2	H 3 2	P 4 ₃ 3 2
I 2 2 2	P 3 1 2 I	4 3 2
I 2 ₁ 2 ₁ 2 ₁	P 3 2 1	I 4 ₁ 3 2
P 4	P 3 ₁ 1 2	Group C (0.3%)
P 4 ₁	P 3 ₁ 2 1	F 4 3 2
P 4 ₂	P 3 ₂ 1 2	F 4 ₁ 3 2
P 4 ₃	P 3 ₂ 2 1	
I 4	P 6 2 2	
I 4 ₁	P 6 ₁ 2 2	
P 3	P 6 ₂ 2 2	
P 3 ₁	P 6 ₃ 2 2	
P 3 ₂	P 6 ₄ 2 2	
H 3	P 6 ₅ 2 2	
P 6		
P 6 ₁		
P 6 ₂		
P 6 ₃		
P 6 ₄		
P 6 ₅		

Table 2

List of PDB codes for Protein structures tested in this study. There are two test sets: the space group test set and the fitting test set. The space group test set consists of 65 proteins, each of which has unique space group. The set was used to test mode calculations for structures with various space group symmetries. The last three structures listed in the parenthesis were not included in the evaluation of overlap index and local character indicator. (see main text) The fitting test set (see the reference ^{55, 85}) was used in the fitting test against experimental ADP data.

Space group test set (65 proteins)	3HHV, 2WEG, 3IP0, 2Z51, 3HT1, 3KKW, 3K0M, 3FE6, 3GNZ, 3CGN, 3KCS, 3HYF, 2ON8, 3H93, 2QVO, 3H1X, 2W2R, 3IM6, 1O5J, 3ID4, 2J4X, 3GI2, 3H57, 3FAJ, 3GFP, 3I93, 2PL1, 3JU0, 2E4L, 1Y63, 3KOL, 3JYZ, 2YX8, 3FSD, 2VO8, 3A3Q, 3JTE, 2QCB, 3KKF, 1RUT, 2NMU, 2DVP, 3KDW, 2A15, 3HWL, 2J97, 3KB0, 3KKG, 3F4M, 3ES1, 2R6Q, 3HQJ, 2EJB, 3BG8, 3IXC, 3JSC, 1YZV, 2C4V, 1YOA, 1VMJ, 2EEY, 2ZDB, (2WLC, 3GYO, 2IUU)
Fitting test set (83 proteins)	1A6M, 1BYI, 1C75, 1C7K, 1EA7, 1EB6, 1EXR, 1F94, 1F9Y, 1G4I, 1G66, 1G6X, 1GA6, 1GKM, 1GQV, 1GVK, 1GWE, 1HJ9, 1I1W, 1IQZ, 1IUA, 1IX9, 1IXH, 1J0P, 1JFB, 1K5C, 1KTH, 1KWF, 1L9L, 1LKK, 1LNI, 1LUG, 1M1Q, 1M40, 1MC2, 1MJ5, 1MUW, 1MWQ, 1N4W, 1N55, 1NKL, 1NLS, 1NWZ, 1O7J, 1OAI, 1OD3, 1OK0, 1PQ7, 1R2M, 1R6J, 1RB9, 1RTQ, 1SFD, 1SSX, 1TG0, 1TQG, 1TT8, 1U2H, 1UFY, 1UG6, 1UNQ, 1US0, 1V0L, 1V6P, 1VBW, 1VYR, 2CHH, 1W0N, 1X6Z, 1X8Q, 1XMK, 1Y55, 1YLJ, 1ZK4, 1ZZK, 2BT9, 2BW4, 2CWS, 2F01, 2FDN, 2PVB, 3LZT, 7A3H

Table 3

Summary of the tests on the lowest frequency mode subspaces of various mode calculations. Tests were performed on the 62 proteins in the space group test set listed in Table 2.

Mode Sets	Average Overlap with Branches	Average Overlap with Free Molecule	Average Local Character Indicator
Free Molecule	0.895	---	8.62
Branches	0.933	(0.895)	---
Integrated	0.949	0.932	7.26