THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

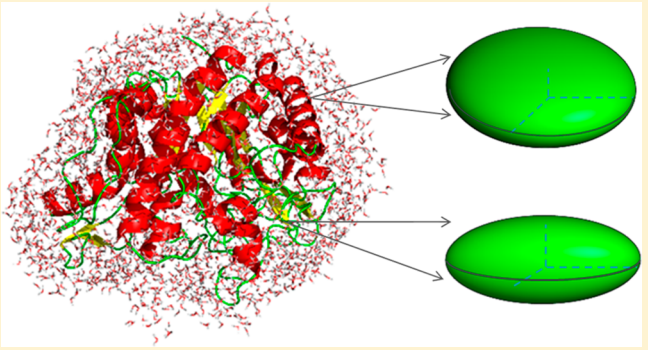
# Aligning Experimental and Theoretical Anisotropic B-Factors: Water Models, Normal-Mode Analysis Methods, and Metrics

Lei Zhou* and Qinglian Liu

Department of Physiology and Biophysics, School of Medicine, Virginia Commonwealth University, Richmond, Virginia 23298, United States

Supporting Information

**ABSTRACT:** The strength of X-ray crystallography in providing the information for protein dynamics has been under appreciated. The anisotropic B-factors (ADPs) from high-resolution structures are invaluable in studying the relationship among structure, dynamics, and function. Here, starting from an in-depth evaluation of the metrics used for comparing the overlap between two ellipsoids, we applied normal-mode analysis (NMA) to predict the theoretical ADPs and then align them with experimental results. Adding an extra layer of explicitly treated water on protein surface significantly improved the energy minimization results and better reproduced the anisotropy of experimental ADPs. In comparing experimental and theoretical ADPs, we focused on the overlap in shape, the alignment of dominant directions, and the similarity in magnitude. The choices of water molecules, NMA methods, and the metrics for evaluating the overlap of ADPs determined final results. This study provides useful information for exploring the physical basis and the application potential of experimental ADPs.



## INTRODUCTION

X-ray crystallography sets up the gold standard for the determination of 3D atomic positions, but its potential in providing the information for protein dynamics has received less attention. In a typical protein PDB file, each atom occupies one line that contains five entries, with three for the *XYZ* coordinates, one for the occupancy, and one for the temperature or B-factor. The value of B-factor represents the magnitude of the electron density in a spherical and isotropic shape. However, in high-resolution protein structures, each atom occupies two lines and the additional line contains six integers that define a symmetric tensor for the anisotropic B-factor, which is also called anisotropic displacement parameter (ADP).[1−3] ADP represents an asymmetrical and multimodal distribution of electron density in the shape of an ellipsoid, which according to the Born−Oppenheimer approximation reflects the anisotropic movement of the atom nucleus.[4,5] In the current PDB database, the high-resolution structures (<1.2 Å) compose about 2.5% of the total deposits. The ADPs from those structures embed a trove of invaluable information for studying the intriguing relationship between protein structure and dynamics.

The notion that structure, dynamics, and function are inseparable in the study of protein biophysics is being increasingly recognized. According to the vibration frequency and amplitude, protein dynamics can be arranged from the high-frequency local movements such as side chain tumbling, to the low-frequency correlated motions that involves every element in the whole molecule.[6−9] It is believed that the motions of the lowest vibrational frequency and also of the largest vibrational amplitude carry the most significance for protein function.[10−13] These molecular motions actually define the pathway that the protein molecule traces in fulfilling the biological function. The lowest frequency and the largest amplitude can be translated into maximal conformation changes with minimal energy costs.

Two computational approaches have been used to study protein dynamics: normal-mode analysis (NMA), an analytical approach based on a harmonic approximation to the protein energy surface, and principal component analysis (PCA), a statistical approach based on the sampling of the conformation space by molecular dynamics (MD) or Monte Carlo simulations.[14−17] Classical NMA starts from a protein structure that is assumed to be at the local minimum on energy surface. The first step is to construct the Hessian Matrix, which contains the second derivative of the system potential energy or the effective force constant between each pair of atoms. Diagonalization of the Hessian matrix yields eigenvalues, corresponding to the vibration frequency (or the amplitude) of the collective molecular motions, and the corresponding eigenvectors, representing the direction of the motions.[18] For a nonlinear system containing N particles, the degree of freedom

**Received:** December 19, 2013
**Revised:** March 26, 2014
**Published:** March 27, 2014

   dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

(DOF) related with vibration is $3N - 6$, with the six DOF of the translation and rotation of the whole molecule removed. NMA is mathematically rigorous and produces $3N - 6$ orthogonal eigenvectors paired with positive eigenvalues. However, the PCA based on MD simulation involves the construction of the covariance matrix based on an ensemble of protein conformations. Compared to NMA, major advantages of PCA include the consideration of the anharmonic behavior of proteins, which could be attributed to the proper treatment of temperature, pressure, and more importantly, solvent, by the MD simulation.

Classical NMA is based on the presentation of the system by all-atom or united-atom force fields. Because of expensive costs of the physical memory to store the all-atom Hessian matrix ($O(N^2)$) and the CPU time to diagonalize the matrix ($O(N^3)$), the application of all-atom NMA (AANM) has been limited. Currently, the storage of Hessian matrix is not the major issue given technical advances in the handling of sparse matrixes, but the matrix diagonalization is still challenging. Iterative algorithms developed for this purpose include the diagonalization in a mixed basis method (DIMB)[19,20] and the Lanczos/Arnoldi factorization method[21] adopted in computational packages of CHARMM[22] and GROMACS,[23] respectively. These iterative methods are still very time-consuming and only yield a small fraction of the total eigenvectors. Therefore, AANM has been mainly applied to protein systems containing at most a few hundred residues and without explicitly treated water molecules.

Various coarse-grained NMA approaches have been pursued to improve the computational efficiency and thus the applicability of NMA to large systems.[24−27] Methods rooted in the all-atom presentation of the system include the rotation translation block (RTB) (also called block normal mode or BNM), which presents the system with a series of rigid blocks,[28,29] and the method that partitions the matrix to relevant and nonrelevant parts (CGNM).[30−32] In addition, coarse-gained NMA methods based on much simplified force-fields have been developed, represented by the anisotropic network model (ANM).[33,34] ANM is the NMA method based on the elastic network model (ENM) (or Gaussian network model, GNM), which only considers C-$\alpha$ atoms and applies a unified force constant (1 kcal/mol/Å$^2$).[35,36] Despite such dramatic simplifications, ENM effectively captures the essentials of the intrinsic connections between structure and dynamics and have been successfully applied to large macromolecules and assemblies that have been beyond the capability of traditional methods.[37−41]

A major advantage of the NMA methods based on all-atom Hessian matrix, including the methods of AANM, RTB, and CGNM, is the incorporation of detailed chemical information embedded in the structure, including structural and surface water molecules.[30] However, because of the expensive computational cost, only a few studies have included explicitly treated water molecular and addressed their effects on protein dynamics.[30,42−46] These studies revealed that surface water, or the hydration layer, has significant impacts on protein structure and dynamics. Especially, having drastically different physical−chemical properties from the bulk water, the surface water contributes to the total atomic fluctuations, reduces the amplitude of protein fluctuations, and shifts the spectrum of molecular motions toward higher vibration frequency. These pieces of information, in conjunction with the MD simulations that explicitly treat both surface and bulk solvent molecules, provide more complete understanding of the unique contribution by surface water.

For both AANM and coarse-grained NMA methods, it is essential to study the feature and applicability of each method. The information on protein dynamics provided by experimental approaches, such as the X-ray crystallography, NMR, and other spectroscopic methods, are useful references. Although the harmonic approximation adopted by NMA obliviates the anharmonic behavior of protein molecules, NMA has been confirmed to be able to capture the essentials of protein dynamics in the vicinity of a local minimum on protein energy surface.[32] Notably, the majority of X-ray data sets were obtained under the cryogenic temperature of 100 K, well below the "glass transition" temperature proposed for protein molecules (160−220 K), when the protein behavior starts to show more signs of anharmonicity.[47] Thus, NMA is especially suitable to reproduce the protein dynamics under cryogenic conditions, when the protein molecule behaves more harmonically and the bulk solvent has less influence over the intrinsic motions.

At atomic level, protein dynamics is reflected in the amplitude and the direction of the thermal fluctuations of each atom. The isotropic thermal factors from X-ray crystal structures used to be the major criteria in evaluating the NMA results. Comparing the absolute amplitude of the B-factor of each atom is not very meaningful, as it is known that NMA tends to yield smaller atomic fluctuations. Most studies use the linear correlation coefficient ($cc_{iso}$) as the metric for the overall alignment between experimental and theoretical B-factors for a single protein, which varies between 1 (perfect correlation), 0 (no correlation), and −1 (perfect anticorrelation) (eq 11).[48] Extensive efforts and considerations have been devoted to improve the prediction results, including the choice between all-atom and coarse-grained methods, the inclusion of solvent molecules, the different setting of cutoff radius for ENM, the location of the residue relative to the whole molecule or certain structural features, the consideration of crystalline environment, etc. However, it has been puzzling that most of the $cc_{iso}$ results fluctuate around 0.6.[37,49,50] Interestingly, ENM-based methods yielded better alignments with NMR data than with X-ray data, with the $cc_{iso}$ above 0.7.[51,52] These facts manifest the limitation of the isotropic B-factor from X-ray data set as the reference in evaluating the NMA results, which could be attributed to the fact that the only useful information provided by isotropic B-factor is the relative flexibility along the primary sequence within each protein.

When the resolution of the X-ray data set reached the level of 1.2 Å or above, it becomes possible to fit the atomic thermal fluctuation with more sophisticated multimodal distributions. In recent years, the ADPs from high resolution X-ray structures have increasingly been used as the reference to measure the NMA results. Compared to the conventional isotropic B-factor, ADP is superior because ADP carries not only the information about the amplitude of fluctuations but more importantly the spatially distributed fluctuations in electron density. Important information about protein dynamics, more specifically the directionality of collective molecular motions, is embedded in these anisotropic fluctuations at atomic level. Therefore, ADP, as defined by a symmetric $3 \times 3$ matrix for each atom, provides clearer and richer information than a single scalar value and thus is more useful in evaluating the NMA results.

The topic of aligning the experimental and theoretical ADPs has been attempted numerous times, but a major technical

hurdle still exists: the suitable metrics to compare the overlap in shape between two ADPs.[37,51,53,54] On the basis of the definition in X-ray crystallography for the correlation between two electron-density maps, Merritt developed the formulation for the correlation coefficient (cc) to quantify the overlap between two ellipsoids in real space (eq 7).[55] However, as noticed from the very beginning, cc is heavily influenced by the degree of anisotropy and biased toward the ADPs that are less anisotropic or close to a sphere in shape (Figure 1). Moreover, cc is sensitive to the amplitude of the isotropic components of two ellipsoids. These factors prompted the development of another two metrics, modified cc (eq 8)[37,50] and normalized cc (eq 10).[55] Modified cc involves the construction of a perfectly misaligned spheroid (keep the eigenvectors but using the eigenvalues, with the smallest and the largest switched, from the other spheroid). Modified cc varies between 1 (perfectly aligned) and 0 (perfectly misaligned). For normalized cc (cc normalized to the overlaps with isotropic sphere of both spheroids), the value will be greater than 1 if two ellipsoids are more similar to each other than the isotropic sphere and less than 1 if otherwise. Another two metrics, Kullback−Leibler (KL) distance and Pearson correlation, have also been used.[53,56,57] Although these different metrics have been repeatedly used in the literature, a detailed investigation of these metrics, especially with regard to the degree of anisotropy, the type, and the orientation of spheroids, is lacking.

In this study, we started from an in-depth evaluation of the metrics used for evaluating the overlap between two ellipsoids. We focused on the influence on the results by the degree of anisotropy, which could be affected by the choice of computational methods and the selection of the pool of eigenvectors in calculating the theoretical ADP. We compared different NMA methods in predicting ADP and used both all-atom force field based and coarse-grained methods. Moreover, we studied the effects by surface water and found that adding a layer of explicitly treated water molecules dramatically improves the energy minimization (EM) results and has certain impacts on the results.

## THEORETICAL METHODS

**Treatment of Experimental ADP.** The ADPs in high-resolution structures are listed as six integers in the order of $U^{11}$, $U^{22}$, $U^{33}$, $U^{12}$, $U^{13}$, and $U^{23}$. Each of these factors has been scaled up by a factor of 10 000, and the unit is in Å$^2$. These six ADPs define a symmetric 3 × 3 tensor:

$$\begin{pmatrix} U^{11} & U^{12} & U^{13} \\ U^{12} & U^{22} & U^{23} \\ U^{13} & U^{23} & U^{33} \end{pmatrix} \quad (1)$$

which defines the shape of an ellipsoid for each atom. Each tensor/ellipsoid has three prominent or principal axes that represent the peaks of three Gaussian distributions. Similar to the isotropic B-factor, the length of each of the axes is inversely related with the electron density. The three columns (or rows) of the tensor represent the projections of three principal axes of an ellipsoid to the orthogonal *XYZ* coordinate system used to define the atomic positions.

Using a simple matrix manipulation (diagonalization), the above 3 × 3 tensor can be transformed into a diagonal matrix:

$$U = R \cdot \begin{pmatrix} W^{11} & 0 & 0 \\ 0 & W^{22} & 0 \\ 0 & 0 & W^{33} \end{pmatrix} \cdot R^T \quad (2)$$

which contains three eigenvalues ($W^{ii}$). The corresponding eigenvectors stored in the matrix $R$ (or $R^T$) define the direction of three prominent axes of the ADP ellipsoid in the orthogonal *XYZ* coordinate system, and $W^{ii}$ corresponds to the distribution of the electron density along each of those prominent axes. The degree of anisotropy is defined as the ratio between $W^{11}$ and $W^{33}$ and varies between 0 and 1 (sphere).

On the basis of the value of $W^{ii}$, the equivalent isotropic B-factor can be calculated based on the following equation:

$$B = \frac{8\pi^2}{3}\text{trace}(W) \times 10^{-4} = \frac{8\pi^2}{3}(W^{11} + W^{22} + W^{33}) \times 10^{-4} \quad (3)$$

**Energy Minimization of the Crystal Structure and Normal-Mode Analysis (NMA).** High resolution structures that contain experimental anisotropic B-factors were downloaded from PDB (see Table S1, Supporting Information, for the list). Water molecules within 3 Å of the protein molecule from the crystal structure were kept during the calculations. First, Modeler was used to fix the missing side chains and residues within the structure.[58] The changes to the overall structure through the Modeler step were minimal (RMSD of C-$\alpha$ atoms, 0.09 ± 0.007Å). The initial crystal structure was energy minimized at double precision. The methods used are steepest-descent (SD), conjugate-gradient (CG), and limited-memory Broyden−Fletcher−Goldfarb−Shanno (L-BFGS).[23] The force field GROMOS96 (53a6) was used. During the SD energy minimization step, oxygen atoms of the structural water and heavy atoms of the protein were position-restrained. The electrostatic energy was described by a switch function with the distance for normal treatment set at 15 Å and the cutoff distance set at 18 Å.[59] Two programs from the GROMACS 4.6.3 package, mdrun and g_nmeig,[23] were used to produce and diagonalize the Hessian matrix, respectively. The eigenvalues and eigenvectors were saved as the results of the NMA based on all-atom (or united-atom) force fields (AANM).

We also tested three different coarse-grained NMA: the NMA based on the elastic network model (ENM),[34] the block NMA (BNM) based on a representation of the whole molecule by rigid blocks,[29] and the NMA based on partitioning the all-atom Hessian matrix (CGNM).[30−32] For ENM, the coordinates of the C-$\alpha$ atoms from the original crystal structure were used. The force constant and the cutoff distance set as 1 kcal/mol/Å$^2$ and 13 Å, respectively. For BNM, we used the Fortran code DIAGRTB (v2.52) with a minor modification of LRWORK, from 32,000,000 to 200,000,000, to accommodate large systems.[28,60] Details of Hessian matrix partition and the related coarse-grained NMA (CGNM) has been described previously in detail.[30] Briefly, the all-atom Hessian matrix was partitioned into four sections representing the interactions among relevant to relevant atoms, nonrelevant to nonrelevant atoms, relevant to nonrelevant atoms, and nonrelevant to relevant atoms:


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

$$H_{\text{all}} = \begin{vmatrix} H_{xx} & H_{xy} \\ H_{yx} & H_{yy} \end{vmatrix} \tag{4}$$

Then the effective force constant matrix for C-$\alpha$ atoms were extracted based on the following equation:[32]

$$H'_{xx} = H_{xx} - H_{xy}\cdot {H_{yy}}^{-1}\cdot H_{yx} \tag{5}$$

Diagonalization of the effective Hessian matrix for C-$\alpha$ atoms yielded the eigenvectors and the corresponding eigenvalues.

On the basis of the eigenvectors and eigenvalues from the above NMA, the theoretical ADP of each atom was calculated using the following equation:

$$V_{ij} = \langle \Delta x_i \cdot \Delta x_j \rangle = \sum_{k=1}^{N} \frac{V_{ki}\cdot V_{kj}}{{\varpi_k}^2} \tag{6}$$

where $i$ and $j$ are indexes for three orthogonal axes and $k$ is the index for eigenvectors/eigenvalues. $N$ represents the number of eigenvectors involved in the analysis.

**Comparing Experimental and Theoretical ADPs.** The correlation coefficient (cc) between two ADPs is calculated according to the following equation:

$$\text{cc}(U, V) = \frac{(\det U^{-1}\cdot \det V^{-1})^{1/4}}{\left[\frac{1}{8}\det(U^{-1} + V^{-1})\right]^{1/2}} \tag{7}$$

where det represents the matrix determinant.[55] Determinant of the matrix can be calculated as the product of eigenvalues or directly from entries in the matrix.

Other than the overlap factor or correlation coefficient (cc), several other metrics have been introduced to quatify the similaries in shape between two ellipsoids. Two representative metrics are

1. Modified cc.

$$\text{cc}_{\text{M}} = \frac{\text{cc} - \text{cc}_{\text{min}}}{1 - \text{cc}_{\text{min}}} \tag{8}$$

where $\text{cc}_{\text{min}}$ corresponds to the minimal cc between $U$ and $V$. $V'$ represents the perfect misalignment defined by

$$V' = ({R_U}^3, {R_U}^2, {R_U}^1)\cdot \begin{pmatrix} {W_V}^{11} & 0 & 0 \\ 0 & {W_V}^{22} & 0 \\ 0 & 0 & {W_V}^{33} \end{pmatrix} \cdot ({R_U}^3, {R_U}^2, {R_U}^1)^T \tag{9}$$

with the prominent $({R_U}^1)$ and the least prominent $({R_U}^3)$ directions being switched and the three eigenvalues being replaced by the values from tensor $V$.[50]

2. Normalized cc.

$$\text{cc}_{\text{N}} = \frac{\text{cc}(U, V)}{\text{cc}(U, \text{Iso})\text{cc}(V, \text{Iso})} \tag{10}$$

where cc(U,V) is divided by the product of the cc between $U$ or $V$ with isotropic sphere.[51]

To compare the distribution of the magnitude of the ADPs, we calculated the isotropic B-factor based on eq 3, normalized the value by the sum of all isotropic factors in that protein, and used the following equation to obtain the linear correlation coefficient ($\text{cc}_{\text{iso}}$) between experimental and theoretical B-factors for each protein:[50]

$$\frac{\sum_{k=1}^{N}(B_k^{\text{exp}} - \langle B^{\text{exp}}\rangle)\cdot(B_k^{\text{pred}} - \langle B^{\text{pred}}\rangle)}{\sqrt{\sum_{k=1}^{N}(B_k^{\text{exp}} - \langle B^{\text{exp}}\rangle)^2 \cdot \sum_{k=1}^{N}(B_k^{\text{pred}} - \langle B^{\text{pred}}\rangle)^2}} \tag{11}$$

**Notice about the Conversion of Units.** In a previous publication, we have listed the conversion of units across different computational systems.[30] For NMA, the GROMACS program g_nmeig generated eigenvalue and eigenfrequency for each eigenvector. The unit of eigenfrequency is listed as wavenumber with the unit in $\text{cm}^{-1}$. Eigenvalue corresponds to the square of the angular speed ($\phi^2$), with the unit in $\text{s}^{-2}$, and requires a conversion factor ($10^{24}$).[30] To convert the eigenvalue into eigenfrequency, the following factor is needed:

$$\begin{aligned} \text{eigenfreq} &= \frac{\varpi}{2\pi c} = \frac{\sqrt{\varpi^2}}{2\pi\cdot 2.998 \times 10^{10}\cdot \text{cm}\cdot \text{s}^{-1}} \\ &= 5.309 \times 10^{-12}\cdot\sqrt{\varpi^2}\cdot \text{cm}^{-1}\cdot \text{s} = 5.309\sqrt{\frac{\varpi^2}{10^{24}}\cdot \text{s}^2}\cdot \text{cm}^{-1} \\ &= 5.309\sqrt{\text{eigenvalue}}\cdot \text{cm}^{-1} \end{aligned} \tag{12}$$

where $c$ represents the speed of light.

**Systematic Survey of Spheroid Orientations in Euler Angle Space.** To study the overlap between randomly positioned ellipsoid, we relied on the Haar measure for Euler angle to approach a uniform coverage of the angle space ($\alpha$, $\beta$, $\gamma$), $\sin(\beta)\Delta\alpha\Delta\beta\Delta\gamma$. The range of $\alpha$, $\beta$, and $\gamma$ are 0 to $2\pi$, $\cos^{-1}(-1.0)$ to $\cos^{-1}(1.0)$, and 0 to $2\pi$, respectively. We used the following Euler's theorem to rotate the ellipsoid:

$$\begin{aligned} R(1,1) &= \cos(\alpha)\cos(\gamma) - \cos(\beta)\sin(\gamma)\sin(\alpha) \\ R(1,2) &= \cos(\alpha)\sin(\gamma) + \cos(\beta)\cos(\gamma)\sin(\alpha) \\ R(1,3) &= \sin(\alpha)\sin(\beta) \\ R(2,1) &= -\sin(\alpha)\cos(\gamma) - \cos(\beta)\sin(\gamma)\cos(\alpha) \\ R(2,2) &= -\sin(\alpha)\sin(\gamma) + \cos(\beta)\cos(\gamma)\cos(\alpha) \\ R(2,3) &= \cos(\alpha)\sin(\beta) \\ R(3,1) &= \sin(\beta)\sin(\gamma) \\ R(3,2) &= -\sin(\beta)\cos(\gamma) \\ R(3,3) &= \cos(\beta) \end{aligned} \tag{13}$$

The following equation was used to obtain the mean value of cc:

$$\begin{aligned} \overline{cc} &= \frac{\iiint \text{cc}_{\alpha\beta\gamma}\cdot\sin(\beta)\Delta\alpha\Delta\beta\Delta\gamma}{\iiint \sin(\beta)\Delta\alpha\Delta\beta\Delta\gamma} \\ &= \frac{\iiint \text{cc}_{\alpha\beta\gamma}\cdot\sin(\beta)\Delta\alpha\Delta\beta\Delta\gamma}{8\pi^2} \end{aligned} \tag{14}$$

## RESULTS

**Metrics for Comparing the Shape of ADP Ellipsoids.** To compare the experimental and computational ADP tensors, it is necessary to study both the shape, the orientation, and the magnitude.[51] The orientation is represented by three eigenvectors of the tensor, which define the three dominant directions of the ADP ellipsoid. The three eigenvalues can be


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

converted to the isotropic B-factor and used to compare the magnitude. In contrast, comparing the shape of two ellipsoids is complicated. The value of cc varies between 0 for misaligned and extremely anisotropic ellipsoids and 1 for perfectly aligned ellipsoids or two ellipsoids that are close to the sphere in shape. The degree of anisotropy is defined by the ratio between the shortest and longest radii of the ellipsoids (Figure 1A). The

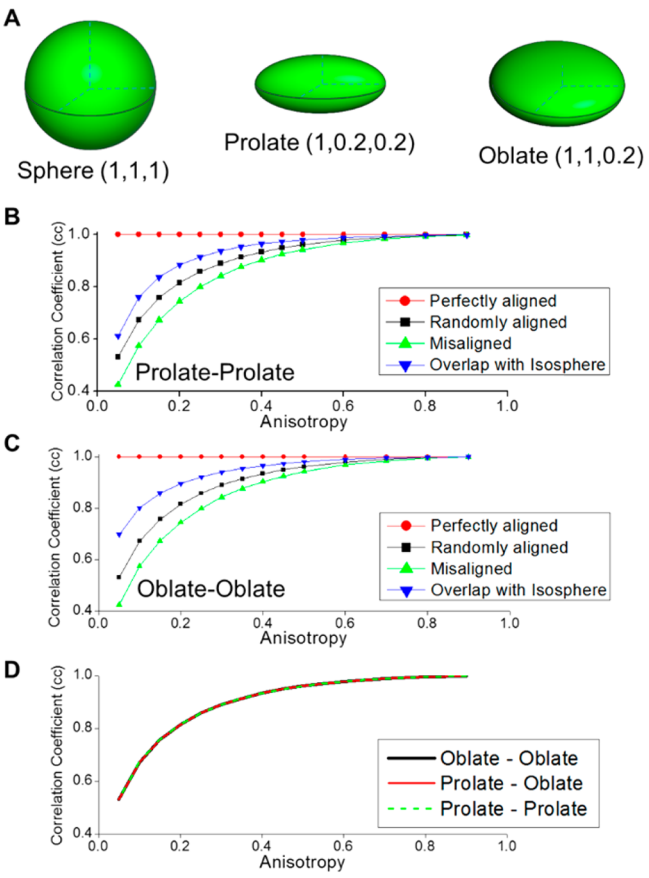



**Figure 1.** Correlation coefficient (cc) for measuring the overlap between two spheroids. (A) Schematic drawings of a sphere, a prolate (cigar-shaped spheroid), and an oblate (pancake-shaped spheroid). The anisotropies for the three objects are 1.0, 0.2, and 0.2, respectively. (B) cc for the overlap between two prolates of the same anisotropy. Red, perfectly aligned prolates; black, averaged cc of randomly orientated prolates; green, perfectly misaligned prolates. Blue trace shows the overlap between a prolate and an isotropic sphere. (C) cc for the overlap between two oblates of the same anisotropy. Red, perfectly aligned oblates; black, averaged cc of randomly orientated oblates; green, perfectly misaligned oblates. Blue trace shows the overlap between an oblate and an isotropic sphere. (D) Averaged cc for randomly oriented prolate and oblate of the same anisotropy (red trace). The traces for prolate−prolate (green) and oblate−oblate (black) are from B and C, respectively.

change in cc value correctly reflects the nature that the ellipsoids of higher anisotropy (close to sphere), tends to have larger overlap (Figure 1B). However, this close correlation between cc and anisotropy has a caveat: the cc results have a bias toward ADPs of high anisotropy. This is an important issue for theoretical ADP because the anisotropy is directly related with the computational model (all-atom force fields or coarse-grained methods, with or without surface water) and, more importantly, the number of eigenvectors involved in the analysis.

On the basis of the definition of cc, several other metrics were proposed to quantify the overlap between two ellipsoids including the normalized cc and the modified cc. Which metrics provide more straightforward and realistic comparison of two ADPs? To this end, we used the prolate or oblate (pancake- or cigar-shaped spheroid) as an example. First, we calculated the overlap between the spheroids of different anisotropy with isotropic sphere, cc($U$,ISO).[55] Clearly, cc($U$,ISO) depends on the type of spheroid because the cc trace of prolate is different from that of oblate, especially at low range of anisotropy (Figure 1B,C, blue traces). Next we calculated the cc of two randomly oriented spheroids with the same anisotropy through a systematic sampling of the Euler angle space (Figure 1B,C, black traces). As expected, the cc between two randomly positioned spheroids is lower than the overlap with the isotropic sphere and also increases along the increase in anisotropy. Noticeably, the overlap between two randomly oriented spheroids is shape-independent: the trace for two prolates and the trace for two oblates are identical (Figure 1D). To further confirm this, we tested the overlap between a prolate with a randomly orientated oblate and obtained the same results. Thus, cc is strongly influenced by the degree of anisotropy and largely insensitive to the specific shape of the ADPs (prolate or oblate). The correlation between cc and anisotropy is an import factor in evaluating experimental and theoretical ADPs.

Next we studied the modified cc and the normalized cc as a function of anisotropy. Our results showed modified cc is not sensitive to anisotropy, which makes it very useful when the degree of anisotropy needs to be occluded from the analysis (Figure 2A,B, left). Furthermore, we calculated the overlap between two randomly oriented spheroids and found that modified cc is sensitive to the shape of the spheroid because the overlap between a prolate and an oblate (red trace) is drastically different from that of prolate−prolate (green, dashed trace) or oblate−oblate (black trace) (Figure 2C, left). However, the almost flat trace of modified cc (black traces in Figure 2A,B, left), even when the anisotropy approaches to 1, is at odds with the expectation that the overlap between two spheroids similar to a sphere should be close to 1. For the normalized cc, it is similar to cc and is also very sensitive to the degree of anisotropy (Figure 2A,B, right). However, to a certain extent, normalized cc is also sensitive to the shape of the spheroids. The traces for the pancake−pancake, cigar−cigar, and cigar−pancake are all different, especially when the anisotropy was much less than 1 (Figure 2C, right). A recent study reported that normalized cc is less useful when evaluating the theoretical ADP based on the segmented TLS model.[61] Taken together, these analyses provided useful information for applying three metrics to compare the overlap between two ADPs.

**Adding Surface Water Makes the Energy Minimized Structure Closer to Native Structure.** EM of the crystal structure is a critical step for classical AANM because NMA is an analytical approach and based on the harmonic approximation of the protein energy surface. To remove the improper contacts in the structure and more importantly ensure the modeled conformation by the force field corresponding to a local minimum on the energy surface, it is essential to use different EM methods and perform the calculations at double precision. To study the effects of surface water, we separately


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

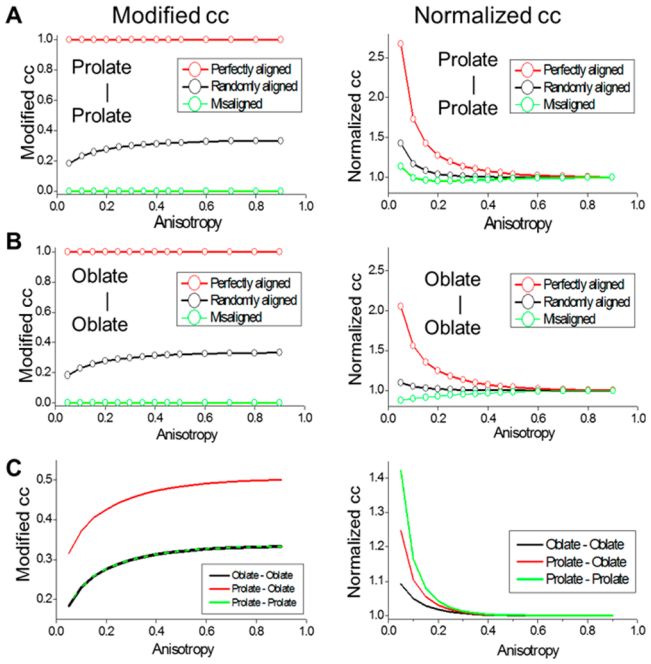



**Figure 2.** Modified cc and normalized cc for the overlap between a pair of spheroids. (A) Modified cc (left) and normalized cc (right) for measuring the overlap between two prolates of the same anisotropy. Red, perfectly aligned prolates; black, mean value for randomly oriented prolates; green, perfectly misaligned prolates. (B) Modified cc (left) and normalized cc (right) for measuring the overlap between two oblates of the same anisotropy. (C) Modified cc (left) and normalized cc (right) for randomly oriented prolate and oblate of the same anisotropy (red trace). The traces for prolate–prolate (green) and oblate–oblate (black) are from A and B, respectively. For modified cc (left), both traces are numerically identical.

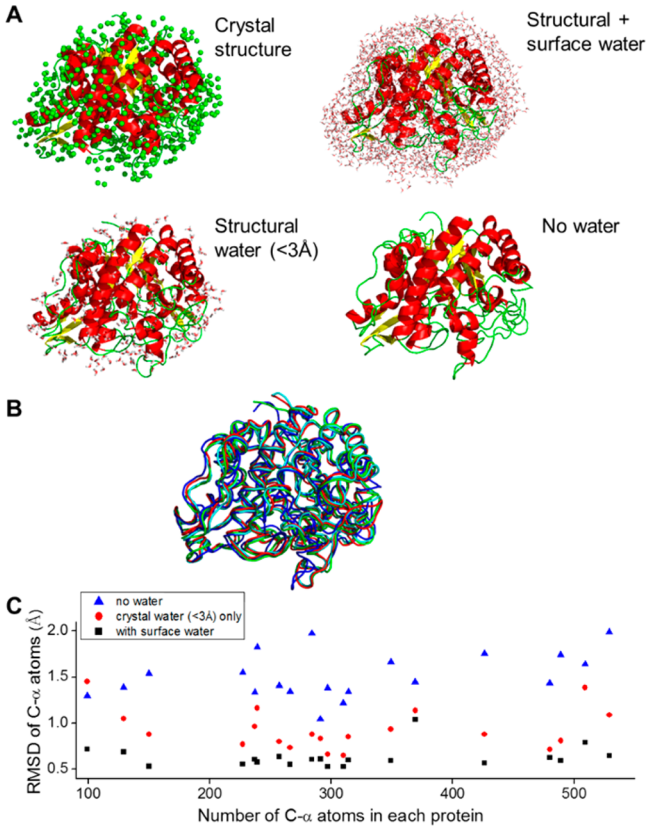



**Figure 3.** Including crystal water and adding surface water improves EM results. (A) Structures of 1UG6 as an example. (clockwise from upper left corner) The original crystal structure including oxygen atoms of water molecules, energy minimized systems including surface and crystal water, no water, and crystal water only. (B) Overlay of all four structures shown in A. Cyan, crystal structure; red, surface + crystal water; green, crystal water only; blue, no water. (C) rmsd of C-$\alpha$ atoms vs the number of C-$\alpha$ atoms for the 20 proteins used in this study (see Table S1, Supporting Information, for the list). For each protein, three water models were tested: systems including surface and crystal water (black), crystal water only (red), and no water (blue).

tested three conditions: (1) protein only and without water; (2) protein plus the crystal water within 3 Å from protein, and (3) protein, the crystal water, and an extra layer (5 Å) of explicitly treated surface water (Figure 3A). After the EM steps, we calculated the root-mean-square deviation (rmsd) with respect to the original crystal structure (Figure 3B,C; Table S1, Supporting Information). The results were impressively consistent: among the three conditions, adding an extra layer of surface water always gave the lowest rmsd (0.65 ± 0.03 Å; $n$ = 20). Oppositely, removing all water from the structure resulted in the highest rmsd, 1.56 ± 0.06 Å, and in the middle was the case with only crystal water (0.94 ± 0.04). To test whether this observation is force-field relevant, we repeated the above calculations using another force field, AMBER99, and obtained the same results.

**Comparing the Overlap in Shape between Experimental and Theoretical ADPs.** Next we asked how the treatment of water molecules and the choice of NMA methods affected the ADP results. In theory, for a system containing $N$ particles, the number of the normal modes available for analysis is $3N - 6$. The selection of eigenvectors in the calculation of theoretical ADP is crucial for the results.[37] Involving too many eigenvectors would increase the computational cost and diminish the meaning of NMA in terms of reducing the number of free parameters. However, involving too few eigenvectors would apparently affect the accuracy of the results. Since the motions of the lowest vibration frequency and of the highest vibration amplitude are of the most functional significance, we calculated the theoretical ADP based on the first 3, 6, 10, 30, 100, or 1000 (if applicable) eigenvectors of the lowest frequencies.

The anisotropy of the theoretical ADP critically depends on the number of eigenvectors (Figure 4A). As a reference, the averaged anisotropy of the experimental ADPs is 0.496 ± 0.099 ($n$ = 20; protein based; filled diamonds) or 0.509 ± 0.168 ($n$ = 6351; residue based; standard deviation). For AANM, the model with explicitly treated surface water apparently outperformed the other two models in approaching the level of experimental anisotropy but with less eigenvectors (black trace in Figure 4A). With only 30 eigenvectors involved (AANM with surface water), the averaged anisotropy reached the level of 0.4, whereas the other two models required hundreds of eigenvectors to reach a similar level. For coarse-grained NMA, both BNM and CGNM methods outperformed the ENM method, which might be related with the all-atom treatment and the consideration of detailed chemical information embedded in the structure.

Does adding surface water lead to an improved overlap between experimental and theoretical ADPs? On the basis of the averaged cc, the answer seemed to be yes. We calculated the cc based on different pools of eigenvectors (Figure 4B). Corresponding to the increase in the number of involved


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069–4079

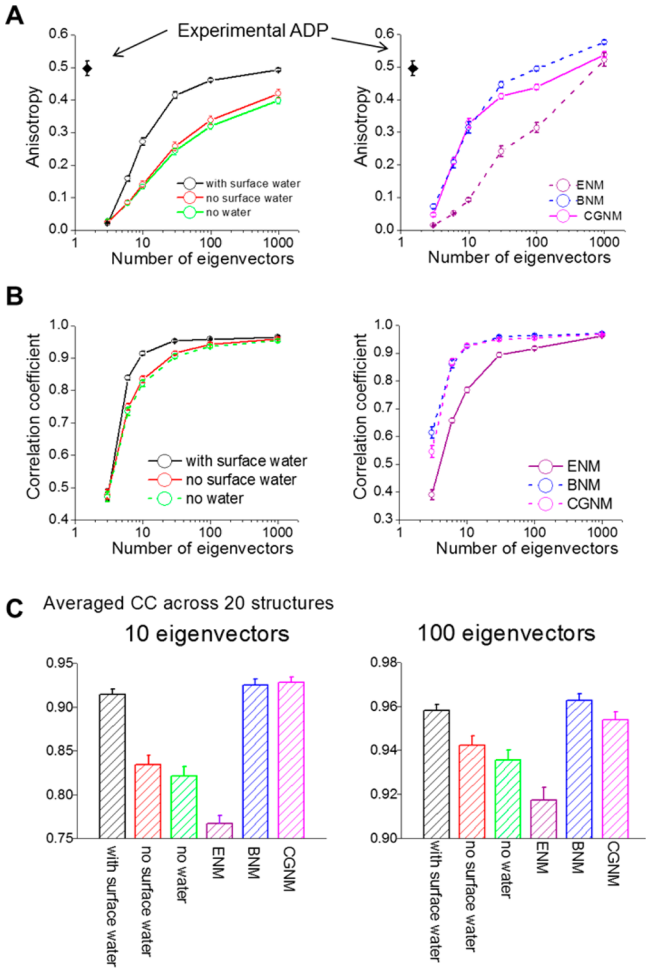



**Figure 4.** Averaged correlation coefficient (cc) based on the results of 20 proteins. (A) Averaged anisotropy of the theoretical ADP vs the number of eigenvectors involved in the calculation. Left, results of NMA based on all-atom force fields. Right, results of coarse-grained NMA. Averaged anisotropy of experimental ADPs is shown in the upper left corner (0.496 ± 0.22, $n$ = 20). (B) Averaged cc vs the number of eigenvectors involved in the analysis. Left, results of NMA based on all-atom force fields. Right, results of coarse-grained NMA. (C) Left, averaged cc based on the first 10 eigenvectors. Right, averaged cc based on the first 100 eigenvectors.

eigenvectors, the cc increases steeply from the value around 0.5 to very close to 1. A replot of the cc values involving 10 or 100 eigenvectors in column format is shown in Figure 4C. It appears that for the theoretical ADP based on the model including surface water, the overlap with experimental ADP is apparently higher. Similarly, the BNM and CGNM methods outperformed the ENM method.

**Negative Controls Are Critical in Evaluating the Overlap in Shape.** To ensure the above conclusion was correct, we asked whether the increase in cc was due to the increase in the anisotropy of theoretical ADPs. To this end, we plotted the cc values that were based on 100 eigenvectors for all protein residues ($n$ = 6351) as a function of the anisotropy of the experiment ADP (Figure S1, Supporting Information). Moreover, we scrambled the input of experimental ADPs for each protein by randomly choosing experimental ADPs from all other 19 proteins (Figure S2, Supporting Information). Choosing atoms from other proteins instead of shuffling the ADPs within each protein increases the randomness because subpopulation of atoms, such as the residues belonging to the same domain, share similarities in the atomic motions. Then we introduced the traces of negative controls to the cc plots shown in Figure 4B (Figure 5A). The close resemblance of the cc

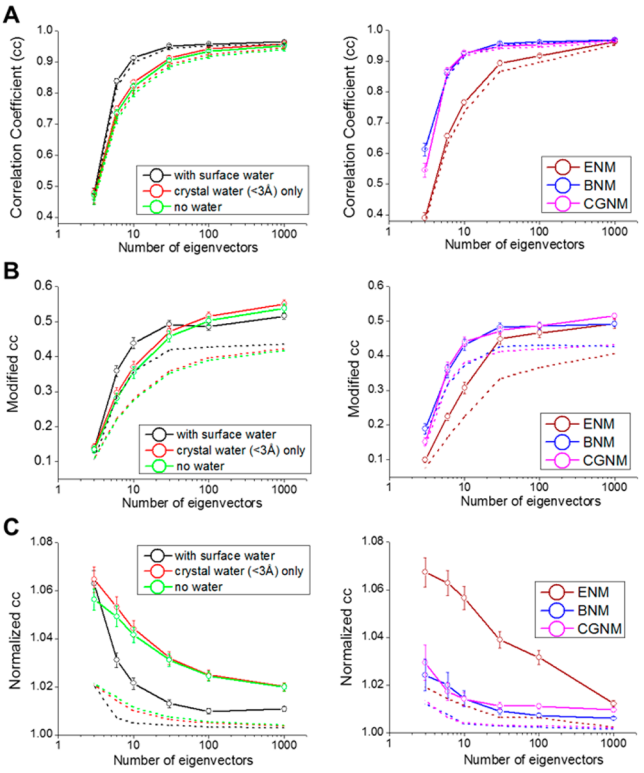



**Figure 5.** Averaged overlap results based on 20 proteins and the corresponding negative controls. (A) Averaged cc based on 20 proteins. Dashed lines represent the negative controls based on scrambled experimental ADPs from the other 19 proteins as input. Left, AANM results. Right, coarse-grained results. (B) Modified cc results. Dashed lines represent the negative controls based on scrambled experimental ADPs as input. (C) Normalized cc results. Dashed lines represent the negative controls based on scrambled experimental ADPs as input.

traces (solid lines with standard error bars) with the corresponding negative control traces (color-matched dashed lines) reinforced the necessity to involve other metrics to help evaluate the overlap between two ADPs.

We calculated the modified cc and the normalized cc. In both cases, the clear separations from the negative control traces demonstrated a better separation between signal and noise than cc. However, two metrics gave opposite evaluations. For modified cc, the AANM with surface water produces better overlap, especially in the range with less than 30 eigenvectors, and the BNM and CGNM methods outperformed ENM (Figure 5B). In contrast, normalized cc revealed totally opposite trends (Figure 5C). This might be related with the fact that the normalized cc biases toward the ADPs of low anisotropy values (Figure 2). Again, these analyses showcase the complexity in interpreting the overlap results.

**Comparing the Distribution of the Magnitude and the Orientation of Experimental and Theoretical ADPs.** We started from comparing the magnitude of experimental and theoretical ADPs, using a linear correlation factor ($cc_{iso}$) for the isotropic B-factors within each protein (eq 11; Figure 6A).[50] The clear separation between the results (solid trace with


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

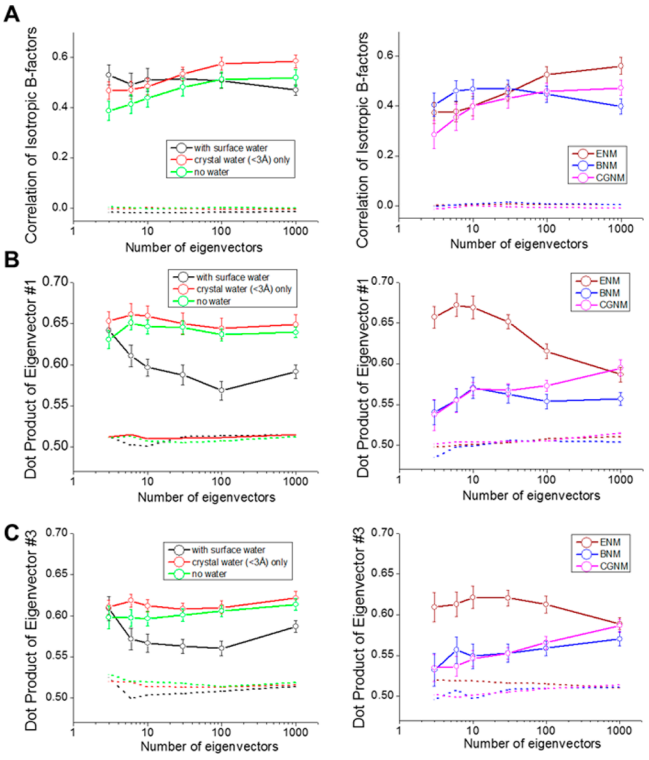



**Figure 6.** Comparing the magnitude and the orientation of theoretical and experimental ADPs. (A) Averaged correlation ($cc_{iso}$) between experimental and theoretical isotropic B-factors for 20 proteins. Dashed lines represent the negative controls based on scrambled experimental ADPs as input. Left, AANM results. Right, coarse-grained results. (B) Averaged dot product (absolute value) of the eigenvectors with the smallest eigenvalues. Results are averaged based on 20 proteins. Dashed lines represent the negative controls based on scrambled experimental ADPs as input. (C) Averaged dot product (absolute value) of the eigenvectors with the largest eigenvalue. Results are averaged based on 20 proteins. Dashed lines represent the negative controls based on scrambled experimental ADPs as input.

standard error bar) from the negative control traces (dashed traces) demonstrated the strength of NMA in predicting the relative flexibility within each protein. With only 3 eigenvectors involved, the AANM method with surface water slightly outperformed other AANM methods and coarse-grained methods. Noticeably, regardless of the computational models, the results by all six methods are very similar, indicating that the isotropic B-factor might not be an effective metric in differentiating NMA results. To compare the orientation of ellipsoids represented by ADP tensors, we studied the eigenvector of the lowest eigenvalue (#1), corresponding to the highest electron density, and the eigenvector of the highest eigenvalue (#3), corresponding to the highest flexibility (Figure 6B,C). The results revealed two pieces of information: First, involving more eigenvectors does not lead to better results, especially for the model involving surface water, and second, ENM significantly outperforms the BNM and CGNM in predicting the orientation of the ADPs.

## DISCUSSION

The ADPs from high-resolution X-ray crystal structures embed rich information for protein dynamics; at the same time, they present an excellent test case for different computational approaches, mainly the NMA and the MD simulation. Apparently MD simulation has the advantage of incorporating a more realistic description of the protein close to the native environment. However, NMA, as an elegant analytical approach for studying protein dynamics, is still being widely used. Since NMA is based on a harmonic approximation of protein energy surface, it is perfectly suitable to reproduce the dynamic information embedded in the crystal structures because most of the X-ray data sets for solving the structure were collected at extremely low temperature (100 K). In this study, using the experimental ADPs from high-resolution structure as the reference, we compared different NMA models in predicting the theoretical ADP and investigated the effects of surface water.

We found that including surface water significantly improved the EM results and outperformed other models in predicting the anisotropy of the ADP. It is well-known that the water near the protein surface has very different physical−chemical properties from the bulk water, such as the 5% increase in the density of water in hydration layer.[62,63] Both experimental studies, including the measurements by terahertz spectroscopy and neutron scattering,[64−66] and theoretical investigations[30,67,68] supported these differences between surface and bulk water. How does adding surface water improve the simulation results? Most likely this is due to the "cage"-like structure formed by surface water molecules, mediated by an extensive network of hydrogen bonds and intimate interactions with polar residues on protein surface.[30,42−46] Thus, including explicitly treated surface water should more faithfully reproduce the intimate interactions between the protein molecule and the solvent. Previously, we showed that including surface water shifts the spectrum of molecular motion generated by NMA toward higher frequencies, which had been reported by spectroscopic studies.[69−71] Here our results showed that the model involving surface water outperformed other models in reproducing the experimentally measured anisotropy. With surface water, much less eigenvectors (30−50) are needed to approach the averaged anisotropy of the experimental ADP. Since 50 eigenvectors only represent a very small fraction of the total eigenvectors ($3N - 6$) of the whole system, the computational cost could be drastically reduced,

The metric for measuring the overlap between two ellipsoids is critical for interpreting the results. Using the prolate and oblate as examples we systematically surveyed the metrics proposed in the literature, including the original cc, modified cc, and normalized cc. We found that cc faithfully reflects the nature that the overlap between two ellipsoids depends on the degree of anisotropy. However, the close correlation between cc and anisotropy makes it difficult to distinguish the real overlaps from the contaminations with the overlap with ellipsoids similar to the sphere. To this end, negative control data sets, based on randomly selected ADPs from other proteins, should be very useful in judging the quality of the overlap results. In contrast, modified cc and normalized cc clearly separate signal from noise, but to different extent both show sensitivities to the shape and the anisotropy of the ellipsoid. Thus, our results show that to compare the overlap between ADPs, it is essential to consider anisotropy, which is related with the computational models and the number of eigenvectors, and more importantly, use different metrics in conjunction with properly designed negative controls to evaluate the results. In practical application, ADPs as negative control could be chosen from a pool of proteins that are totally irrelevant in both function and structure. Noticeably, alignment


dx.doi.org/10.1021/jp4124327 | J. Phys. Chem. B 2014, 118, 4069−4079

of the isotropic B-factors leads to the results around 0.5, regardless of the computational methods and force fields (Figure 6A). In the field, there is a transition from relying on ADP instead of isotropic B-factor as the metric to evaluate NMA results. This should be related with the fact that isotropic B-factors mainly report the relative flexibility within each protein, but however, ADPs provide much richer information for protein dynamics at atomic detail. Importantly, the collective molecular motions, which critically underlies the fulfillment of protein function, are embedded in the anisotropic thermal fluctuations.

A comparison of different models and methods showed that the NMA based on ENM produced results comparable to the results by all-atom force field based methods, including AANM, BNM, and CGNM. ENM, related to GNM, is a knowledge-based model with a drastically simplified presentation of the system: only C-$\alpha$ atoms are being considered, and they are connected through a network of elastic springs with a unified force constant, 1 kcal/mol/Å$^2$. Purely based on the positional definition of C-$\alpha$ atoms, ENM faithfully reproduces the atomic fluctuations determined by NMR and X-ray crystallography[52,72,73] and has been applied to large systems that are beyond the reach of canonical NMA methods.[37−41] Here our results confirmed the effectiveness of the NMA based on ENM, especially in predicting the orientation of the ADP. However, the NMA methods rooted in the all-atom presentation are capable of incorporating a more complex chemical nature and thus have the potential to better reproduce the experimental ADP. Among all the methods, the MD simulation at 300 K better reproduced the absolute amplitude of crystallographic B-factors. Previously, we investigated the effect of surface water by developing the CGNM method.[30] In that study, we reported that CGNM outperformed ENM-based method because the results of AANM instead of the experimental results were used as the reference for the purpose of methodology development.

Here in the current study, we mainly used the experimentally determined ADP as the reference to compare different computational methods. Moreover, we included 20 proteins in the calculation instead of just one in a previous study. For isotropic fluctuations, the correlation coefficient by ENM is comparable to the results by other methods (Figure 6A). However, in terms of predicting the directionality of the ADP ellipsoids, ENM indeed outperformed other coarse-grained methods and matched AANM (Figure 6B). We thought this might be related with the fact that ENM is knowledge-based and rooted in the experimental observation of thermal fluctuations, especially by X-ray crystallography, which might explain the effectiveness of ENM in reproducing empirical results.

## CONCLUSIONS

In summary, we compared the theoretical ADP based on NMA with the experimentally determined ADP embedded in high resolution structures. Incorporating surface water improved the EM results and the alignment with experimental ADP in certain aspects. With the number of the high resolution structures deposited in the PDB databank steadily increasing every year, the ADPs from those structures represent a trove of treasure for the study of protein dynamics but have not been intensively investigated. Further improving the alignment between experimental theoretical ADPs could be accomplished by incorporating the anharmonic behavior of the protein and the crystalline environment, although both only have weak influences over the intrinsic dynamics of the protein under cryogenic conditions.[53,56,74] From the aspect of experimental ADP, the quality of the data could be improved through careful exclusion of the whole-body movement of the crystal and standardization of the refinement procedure, as the methods of SHELX and REFMAC gives slightly different results.[37,75,76] Finally, it would be interesting to ask whether the ADPs within a protein follow certain patterns and to investigate the patterns in the context of structure and protein. Taken together, continuing study of the ADP will strengthen our understanding of the general relationship among structure, dynamics, and function and, retrogradely, would benefit the development of the methodology for refining medium-to-low resolution structures.[77,78]

## ASSOCIATED CONTENT

### Supporting Information

Table S1 provides the list of 20 high resolution structures used in the study. Figures S1 and S2 illustrate the overlap results for all C-$\alpha$ atoms from 20 proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

*(L.Z.) Phone: 804-628-4890. E-mail: lzhou@vcu.edu

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Q.L. and L.Z. are supported by the startup funds from Virginia Commonwealth University and a grant from the NIH (GM098592 to Q.L). We are grateful for the support from Dr. Steven Siegelbaum during the pilot study.

## REFERENCES

(1) Sheriff, S.; Hendrickson, W. A. Description of Overall Anisotropy in Diffraction from Macromolecular Crystals. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1987**, *43*, 118−121.

(2) Schneider, T. R. What Can We Learn from Anisotropic Temperature Factors ? *Proc. CCP4 Study Weekend* **1996**, 133−144.

(3) Trueblood, K. N.; Burgi, H. B.; Burzlaff, H.; Dunitz, J. D.; Gramaccioli, C. M.; Schulz, H. H.; Shmueli, U.; Abrahams, S. C. Atomic Displacement Parameter Nomenclature - Report of a Subcommittee on Atomic Displacement Parameter Nomenclature. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1996**, *52*, 770−781.

(4) Berendsen, H. J. C. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*; Cambridge University Press: Cambridge, U.K., 2007; p xxvii.

(5) Lubich, C. *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*; European Mathematical Society: Zärich, Switzerland, 2008; p ix.

(6) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585−590.

(7) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679−6685.

(8) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A Hierarchy of Timescales in Protein Dynamics Is Linked to Enzyme Catalysis. *Nature* **2007**, *450*, 913−916.

(9) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.

(10) Tsai, C. J.; Del Sol, A.; Nussinov, R. Protein Allostery, Signal Transmission and Dynamics: a Classification Scheme of Allosteric Mechanisms. *Mol. BioSyst.* **2009**, *5*, 207−216.

(11) Kern, D.; Zuiderweg, E. R. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748−757.

(12) Berendsen, H. J.; Hayward, S. Collective Protein Dynamics in Relation to Function. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165−169.
(13) Kitao, A.; Go, N. Investigating Protein Dynamics in Collective Coordinate Space. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164−169.
(14) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373−380.
(15) Bahar, I.; Rader, A. J. Coarse-Grained Normal Mode Analysis in Structural Biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(16) Tama, F. Normal Mode Analysis with Simplified Models to Investigate the Global Dynamics of Biological Systems. *Protein Pept. Lett.* **2003**, *10*, 119−132.
(17) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412−425.
(18) Brooks, B. R.; Janezic, D.; Karplus, M. Harmonic-Analysis of Large Systems 0.1. Methodology. *J. Comput. Chem.* **1995**, *16*, 1522−1542.
(19) Durand, P.; Trinquier, G.; Sanejouand, Y. H. A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. *Biopolymers* **1994**, *34*, 759−771.
(20) Perahia, D.; Mouawad, L. Computation of Low-Frequency Normal Modes in Macromolecules: Improvements to the Method of Diagonalization in a Mixed Basis and Application to Hemoglobin. *Comput. Chem.* **1995**, *19*, 241−246.
(21) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users' Guide*, 3rd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1999.
(22) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(23) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(24) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a Small Globular Protein in Terms of Low-Frequency Vibrational Modes. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696−3700.
(25) Brooks, B.; Karplus, M. Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-Bending Mode of Lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995−4999.
(26) Levitt, M.; Sander, C.; Stern, P. S. Protein Normal-Mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *J. Mol. Biol.* **1985**, *181*, 423−447.
(27) Krebs, W. G.; Alexandrov, V.; Wilson, C. A.; Echols, N.; Yu, H.; Gerstein, M. Normal Mode Analysis of Macromolecular Motions in a Database Framework: Developing Mode Concentration As a Useful Classifying Statistic. *Proteins* **2002**, *48*, 682−695.
(28) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins* **2000**, *41*, 1−7.
(29) Li, G.; Cui, Q. Analysis of Functional Motions in Brownian Molecular Machines with an Efficient Block Normal Mode Approach: Myosin-II and $Ca^{2+}$-ATPase. *Biophys. J.* **2004**, *86*, 743−763.
(30) Zhou, L.; Siegelbaum, S. A. Effects of Surface Water on Protein Dynamics Studied by a Novel Coarse-Grained Normal Mode Approach. *Biophys. J.* **2008**, *94*, 3461−3474.
(31) Ghysels, A.; Miller, B. T.; Pickard, F. C. T.; Brooks, B. R. Comparing Normal Modes Across Different Models and Scales: Hessian Reduction versus Coarse-Graining. *J. Comput. Chem.* **2012**, *33*, 2250−2275.
(32) Hinsen, K.; Petrescu, A.-J.; Dellerue, S.; Bellissent-Funel, M.-C.; Kneller, G. R. Harmonicity in Slow Protein Dynamics. *Chem. Phys.* **2000**, *261*, 25−37.
(33) Doruker, P.; Atilgan, A. R.; Bahar, I. Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to Alpha-Amylase Inhibitor. *Proteins* **2000**, *40*, 512−524.
(34) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80*, 505−515.
(35) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(36) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Fold. Des.* **1997**, *2*, 173−181.
(37) Kondrashov, D. A.; Van Wynsberghe, A. W.; Bannen, R. M.; Cui, Q.; Phillips, G. N., Jr. Protein Structural Variation in Computational Models and Crystallographic Data. *Structure* **2007**, *15*, 169−177.
(38) Rueda, M.; Chacon, P.; Orozco, M. Thorough Validation of Protein Normal Mode Analysis: a Comparative Study with Essential Dynamics. *Structure* **2007**, *15*, 565−575.
(39) Tama, F.; Brooks, C. L., III. Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis. *J. Mol. Biol.* **2005**, *345*, 299−314.
(40) Tama, F.; Sanejouand, Y. H. Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng.* **2001**, *14*, 1−6.
(41) Van Wynsberghe, A. W.; Cui, Q. Comparison of Mode Analyses at Different Resolutions Applied to Nucleic Acid Systems. *Biophys. J.* **2005**, *89*, 2939−2949.
(42) 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.
(43) Yu, X.; Park, J.; Leitner, D. M. Thermodynamics of Protein Hydration Computed by Molecular Dynamics and Normal Modes. *J. Phys. Chem. B* **2003**, *107*, 12820−12828.
(44) Balog, E.; Smith, J. C.; Perahia, D. Conformational Heterogeneity and Low-Frequency Vibrational Modes of Proteins. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5543−5548.
(45) Cui, Q.; Li, G.; Ma, J.; Karplus, M. A Normal Mode Analysis of Structural Plasticity in the Biomolecular Motor F(1)-ATPase. *J. Mol. Biol.* **2004**, *340*, 345−372.
(46) Teeter, M. M.; Case, D. A. Harmonic and Quasiharmonic Descriptions of Crambin. *J. Phys. Chem.* **1990**, *94*, 8091−8097.
(47) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. Solvent Mobility and the Protein 'Glass' Transition. *Nat. Struct. Biol.* **2000**, *7*, 34−38.
(48) Kundu, S.; Melton, J. S.; Sorensen, D. C.; Phillips, G. N., Jr. Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophys. J.* **2002**, *83*, 723−732.
(49) Sen, T. Z.; Feng, Y.; Garcia, J. V.; Kloczkowski, A.; Jernigan, R. L. The Extent of Cooperativity of Protein Motions Observed with Elastic Network Models Is Similar for Atomic and Coarser-Grained Models. *J. Chem. Theory Comput.* **2006**, *2*, 696−704.
(50) Riccardi, D.; Cui, Q.; Phillips, G. N., Jr. Application of Elastic Network Models to Proteins in the Crystalline State. *Biophys. J.* **2009**, *96*, 464−475.
(51) Yang, L.; Song, G.; Jernigan, R. L. Comparisons of Experimental and Computed Protein Anisotropic Temperature Factors. *Proteins* **2009**, *76*, 164−175.
(52) Yang, L. W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A. M.; Bahar, I. Insights into Equilibrium Dynamics of Proteins from Comparison of NMR and X-ray Data with Computational Predictions. *Structure* **2007**, *15*, 741−749.
(53) Hafner, J.; Zheng, W. All-Atom Modeling of Anisotropic Atomic Fluctuations in Protein Crystal Structures. *J. Chem. Phys.* **2011**, *135*, 144114.
(54) Burden, C. J.; Oakley, A. J. Anisotropic Atomic Motions in High-Resolution Protein Crystallography Molecular Dynamics Simulations. *Phys. Biol.* **2007**, *4*, 79−90.
(55) Merritt, E. A. Comparing Anisotropic Displacement Parameters in Protein Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 1997−2004.

(56) Lu, M.; Ma, J. PIM: Phase Integrated Method for Normal Mode Analysis of Biomolecules in a Crystalline Environment. *J. Mol. Biol.* **2013**, *425*, 1082−1098.
(57) Eyal, E.; Chennubhotla, C.; Yang, L. W.; Bahar, I. Anisotropic Fluctuations of Amino Acids in Protein Structures: Insights from X-ray Crystallography and Elastic Network Models. *Bioinformatics* **2007**, *23*, i175−184.
(58) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. Evaluation of Comparative Protein Modeling by MODELLER. *Proteins* **1995**, *23*, 318−326.
(59) van der Spoel, D.; van Maaren, P. J. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. *J. Chem. Theory Comput.* **2006**, *2*, 1−11.
(60) 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.
(61) Zucker, F.; Champ, P. C.; Merritt, E. A. Validation of crystallographic models containing TLS or other descriptions of anisotropy. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 889−900.
(62) Smith, J. C.; Merzel, F.; Bondar, A. N.; Tournier, A.; Fischer, S. Structure, Dynamics and Reactions of Protein Hydration Water. *Philos. Trans. R. Soc. London B, Biol. Sci.* **2004**, *359*, 1181−1189 ; discussion 1189−1190.
(63) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378−5383.
(64) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An Extended Dynamical Hydration Shell around Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20749−20752.
(65) Qvist, J.; Persson, E.; Mattea, C.; Halle, B. Time Scales of Water Dynamics at Biological Interfaces: Peptides, Proteins and Cells. *Faraday Discuss.* **2009**, *141*, 131−144.
(66) Heugen, U.; Schwaab, G.; Brundermann, E.; Heyden, M.; Yu, X.; Leitner, D. M.; Havenith, M. Solute-Induced Retardation of Water Dynamics Probed Directly by Terahertz Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12301−12306.
(67) Zhang, L. Y.; Wang, L. J.; Kao, Y. T.; Qiu, W. H.; Yang, Y.; Okobiah, O.; Zhong, D. P. Mapping Hydration Dynamics around a Protein Surface. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18461−18466.
(68) Tarek, M.; Tobias, D. J. The Dynamics of Protein Hydration Water: a Quantitative Comparison of Molecular Dynamics Simulations and Neutron-Scattering Experiments. *Biophys. J.* **2000**, *79*, 3244−3257.
(69) Xu, J.; Plaxco, K. W.; Allen, S. J. Collective Dynamics of Lysozyme in Water: Terahertz Absorption Spectroscopy and Comparison with Theory. *J. Phys. Chem. B* **2006**, *110*, 24255−24259.
(70) Roh, J. H.; Curtis, J. E.; Azzam, S.; Novikov, V. N.; Peral, I.; Chowdhuri, Z.; Gregory, R. B.; Sokolov, A. P. Influence of Hydration on the Dynamics of Lysozyme. *Biophys. J.* **2006**, *91*, 2573−2588.
(71) Moritsugu, K.; Smith, J. C. Langevin Model of the Temperature and Hydration Dependence of Protein Vibrational Dynamics. *J. Phys. Chem. B* **2005**, *109*, 12182−12194.
(72) Temiz, N. A.; Meirovitch, E.; Bahar, I. *Escherichia coli* Adenylate Kinase Dynamics: Comparison of Elastic Network Model Modes with Mode-Coupling (15)N-NMR Relaxation Data. *Proteins* **2004**, *57*, 468−480.
(73) Tama, F.; Brooks, C. L. Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115−133.
(74) van Gunsteren, W. F.; Karplus, M. Protein Dynamics in Solution and in a Crystalline Environment: a Molecular Dynamics Study. *Biochemistry* **1982**, *21*, 2259−2274.
(75) Herbst-Irmer, R.; Henn, J.; Holstein, J. J.; Hubschle, C. B.; Dittrich, B.; Stern, D.; Kratzert, D.; Stalke, D. Anharmonic Motion in Experimental Charge Density Investigations. *J. Phys. Chem. A* **2013**, *117*, 633−641.
(76) Merritt, E. A. To B or not to B: a Question of Resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 468−477.
(77) Poon, B. K.; Chen, X.; Lu, M.; Vyas, N. K.; Quiocho, F. A.; Wang, Q.; Ma, J. Normal Mode Refinement of Anisotropic Thermal Parameters for a Supramolecular Complex at 3.42-A Crystallographic Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7869−7874.
(78) Chen, X.; Poon, B. K.; Dousis, A.; Wang, Q.; Ma, J. Normal-Mode Refinement of Anisotropic Thermal Parameters for Potassium Channel KcsA at 3.2 A Crystallographic Resolution. *Structure* **2007**, *15*, 955−962.