BIOINFORMATICS

Vol. 23 ISMB/ECCB 2007, pages i175–i184
doi:10.1093/bioinformatics/btm186

Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models

Eran Eyal[1], Chakra Chennubhotla[1], Lee-Wei Yang[1,2] and Ivet Bahar[1,*]

[1]Department of Computational Biology, School of Medicine, University of Pittsburgh. Suite 3064, Biomedical Science Tower 3, 3051 Fifth Ave., Pittsburgh, PA 15213, USA and [2]Institute of Molecular and Cellular Biosciences, University of Tokyo, R107, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

ABSTRACT

Motivation: A common practice in X-ray crystallographic structure refinement has been to model atomic displacements or thermal fluctuations as isotropic motions. Recent high-resolution data reveal, however, significant departures from isotropy, described by anisotropic displacement parameters (ADPs) modeled for individual atoms. Yet, ADPs are currently reported for a limited set of structures, only.

Results: We present a comparative analysis of the experimentally reported ADPs and those theoretically predicted by the anisotropic network model (ANM) for a representative set of structures. The relative sizes of fluctuations along different directions are shown to agree well between experiments and theory, while the cross-correlations between the (*x*-, *y*- and *z*-) components of the fluctuations show considerable deviations. Secondary structure elements and protein cores exhibit more robust anisotropic characteristics compared to disordered or flexible regions. The deviations between experimental and theoretical data are comparable to those between sets of experimental ADPs reported for the same protein in different crystal forms. These results draw attention to the effects of crystal form and refinement procedure on experimental ADPs and highlight the potential utility of ANM calculations for consolidating experimental data or assessing ADPs in the absence of experimental data.

Availability: The ANM server at http://www.ccbb.pitt.edu/anm is upgraded to permit users to compute and visualize the theoretical ADPs for any PDB structure, thus providing insights into the anisotropic motions intrinsically preferred by equilibrium structures.

Contact: bahar@ccbb.pitt.edu

Supplementary information: Two Supplementary Material files can be accessed at the journal website. The first presents the tabulated results from computations (Pearson correlations and KL distances with respect to experimental ADPs) reported for each of the 93 proteins in *Set I* (the averages over all proteins are presented above in Table 3). The second file consists of three sections: (A) detailed derivation of Equation (7), (B) analysis of the effect of ANM parameters on computed ADPs and identification of parameters that achieve optimal correlation with experiments and (C) description of the method for computing the tangential and radial components of equilibrium fluctuations.

*To whom correspondence should be addressed.

1 INTRODUCTION AND THEORY

1.1 Anisotropic displacement parameters (ADPs) from X-ray crystallography

An essential step in the resolution and refinement of X-ray structures is the determination of model parameters that optimally describe the uncertainties and/or displacements in atomic positions (Willis and Pryor, 1975). These displacements are usually described by Debye-Waller or B-factors, for each atom, assuming the atomic fluctuations about their mean positions to be Gaussianly distributed and isotropic. In cases where sufficiently high-resolution diffraction data are available, on the other hand, a set of six anisotropic displacement parameters (ADPs) have been reported per atom with the structural data deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000).

The ADPs describe the mean-square displacements of atoms along three directions as well as their cross-correlations (Dunitz *et al.*, 1988; Merritt, 1999). As such, they provide a much more detailed description than the single isotropic parameters, B-factors, assigned to each atom. However, the extraction of such 6-fold more detailed data also requires the determination of the structure at sufficiently high resolution (e.g. higher than 1.2 Å), which has not been readily achievable for biological molecules, or large asymmetrical macromolecules until recently. More than 1600 such structures can now be found in the PDB, compared to only about 30 in 1998 (Merritt, 1999). Yet, these structures currently amount to only 5% of those collected in the PDB.

The ADPs represent the six distinctive elements—three diagonal and three off-diagonal—of the 3×3 covariance matrix $\boldsymbol{C}$ associated with the Gaussian probability distribution of atomic fluctuations in space (see Methods section). Apart from describing the atomic fluctuations, the ADPs also convey information on the collective motions of proteins (Rosenfield *et al.*, 1978; Schomaker and Trueblood, 1968). In view of the importance of understanding collective dynamics for inferring biomolecular function, methods for improving the accuracy of ADPs in macromolecular refinements have been developed. A useful approach has been to resort to translation-libration-screw (TLS) model parameters (Painter and Merrit, 2006; Schomaker and Trueblood, 1968), and representing the structure as a pseudo-rigid body composed of multiple TLS groups, instead of modeling the six parameters for each

individual atom separately (Painter and Merritt, 2006; Winn *et al*., 2001).

How anisotropic are atomic fluctuations? A simple measure is the mean anisotropy A, i.e. the ratio of the shortest axis to the longest one, when representing the volume swept during atomic fluctuations as an ellipsoid. Examination of about 30 PDB structures for which the ADPs were available in 1999 yielded an average anisotropy A of 0.4–0.5 (Merritt, 1999; Merritt, *et al*., 1998). This is a significant departure from isotropic fluctuations ($A=1$). The significance of the anisotropic nature of atomic motions has also been evidenced by the improvement in R factors achieved upon adoption of ADPs (Schneider, 1996; Winn, *et al*., 2001).

Despite recognizing the anisotropy of atomic fluctuations, in view of the difficulties in assessing ADPs, B-factors have been widely examined and compared with predictions from theoretical approaches as a measure of the mobility of atoms in folded structures. Agreement with B-factors has indeed been gauged as a criterion for benchmarking theoretical models, including in particular the elastic network (EN) models used in conjunction with normal mode analysis (NMA) (Eyal *et al*., 2006; Kundu *et al*., 2002; Yang *et al*., 2006). A more stringent approach would, however, be to consider the ADPs. Examination of ADPs and comparison with theoretical predictions is particularly timely and important, given the current availability of a sizeable set of such refined structures in the PDB, and the recent success of EN models for describing functional motions (Alexandrov *et al*., 2005; Bahar and Rader, 2005; Cui and Bahar, 2006; Krebs *et al*., 2002; Ma, 2005; Nicholay and Sanejouand, 2006; Tama and Brooks, 2006; Yang *et al*., 2006).

The article is organized as follows: First we present an analysis of existing experimental data. We considered, to this aim, the reproducibility of the ADP data for the same protein in different crystal forms, or in the same crystal form resolved independently in different experiments, which showed that the parameterized displacements are affected by the crystal geometry. Next, we computed the theoretical predictions based on the anisotropic network model (ANM) (Atilgan *et al*., 2001; Doruker *et al*., 2000). The level of agreement between theory and experiments is shown to be comparable to that between the two sets of experimental data for the same proteins under different crystal forms. The theoretical results thus present an effective means of consolidating experimental data when available. We have now implemented our algorithm and its visualization features in the ANM web server, so that users may have access to theoretical ADPs and corresponding graphical interfaces, for any query structure.

1.2 Covariance matrix and ADPs

The covariance matrix, $\mathbf{C}$, is a $3N \times 3N$ dimensional symmetric matrix, for a structure of N atoms, the elements of which are 3×3 submatrices of the form

$$C_{ij} = \begin{bmatrix} <\Delta X_i \Delta X_j> & <\Delta X_i \Delta Y_j> & <\Delta X_i \Delta Z_j> \\ <\Delta Y_i \Delta X_j> & <\Delta Y_i \Delta Y_j> & <\Delta Y_i \Delta Z_j> \\ <\Delta Z_i \Delta X_j> & <\Delta Z_i \Delta Y_j> & <\Delta Z_i \Delta Z_j> \end{bmatrix} \quad (1)$$

where i and j refer to the atom indices $1 \le i, j \le N$, ΔX_i, ΔY_i and ΔZ_i are the components of the fluctuation vector $\Delta \mathbf{R}_i$ associated with the displacements of atom i away from its mean position, and the angular brackets refer to expected (or average) values. The sum over the diagonal elements of $\mathbf{C}_{ij}$, or the trace (tr) of $\mathbf{C}_{ij}$ describe the cross-correlation between the fluctuations of atoms i and j, i.e.

$$tr(\mathbf{C}_{ij}) = \langle \Delta X_i \Delta X_j \rangle + \langle \Delta Y_i \Delta Y_j \rangle + \langle \Delta Z_i \Delta Z_j \rangle = \langle \Delta \mathbf{R}_i \Delta \mathbf{R}_j \rangle \quad (2)$$

while the off-diagonal elements describe the cross-correlations between the components of the two fluctuation vectors $\Delta \mathbf{R}_i$ and $\Delta \mathbf{R}_j$. The *mean-square fluctuations* $\langle (\Delta X_i)^2 \rangle, \langle (\Delta Y_i)^2 \rangle$ and $\langle (\Delta Z_i)^2 \rangle$, and the *self-correlations* between the components of $\Delta \mathbf{R}_i$, on the other hand, are given by the respective diagonal and off-diagonal elements of the $\mathbf{C}_{ii}$ [i.e. submatrices of $\boldsymbol{C}$ where $i=j$ in Equation (1)], such that

$$tr(\mathbf{C}_{ii}) = \langle (\Delta X_i)^2 \rangle + \langle (\Delta Y_i)^2 \rangle + \langle (\Delta Z_{i.})^2 \rangle = \langle (\Delta R_i)^2 \rangle \quad (3)$$

$\mathbf{C}_{ii}$ has six distinct elements, three diagonal ($\langle (\Delta X_i)^2 \rangle$, $\langle (\Delta Y_i)^2 \rangle$ and $\langle (\Delta Z_i)^2 \rangle$) and three off-diagonal ($\langle \Delta X_i \Delta Y_i \rangle$, $\langle \Delta X_i \Delta Z_i \rangle$ and $\langle \Delta Y_i \Delta Z_i \rangle$), which define the six ADPs that characterize the anisotropy of the motions of atom i. $\mathbf{C}_{ii}$ is referred to as the *ADP matrix* for atom i.

The anisotropic displacements, or the space sampled by a fluctuating atom, can be represented by an ellipsoid assuming that the fluctuations along each direction are Gaussianly distributed. The size of the ellipsoids depends on the selected probability/confidence levels and the relative sizes of the three axes are found by diagonalizing the ADP matrix $\mathbf{C}_{ii}$. The resulting eigenvectors define the three principal axes of the ellipsoids and the eigenvalues ($d_{i1} > d_{i2} > d_{i3}$) describe the mean-square fluctuations (sizes) along these directions. Ellipsoids of axial lengths equal to $d_{i1}^{1/2}$, $d_{i2}^{1/2}$ and $d_{i3}^{1/2}$ enclose the space in which atom i is found with a probability of 0.683 (i.e. one SD away from the mean of the Gaussian distribution). The diagrams presented in this study refer to this size of ellipsoids. Doubling the axes' lengths encapsulates 95.4% of the distribution. The ratio between the smallest and largest eigenvalues of $\mathbf{C}_{ii}$ defines the anisotropy (Merritt, 1999; Trueblood, *et al*., 1996) $A_i = d_{i3}/d_{i1}$. A_i·($0 \le A_i \le 1$) serves as a quantitative measure of anisotropy for the displacements of atom i, the lower and upper limits corresponding to planar and isotropic motions, respectively. The size of atomic fluctuations, on the other hand, are described by the corresponding ellipsoid volume $\mathrm{V}_i = 4\pi\,(d_{i1}\,d_{i2}\,d_{i3})^{1/2}/3$.

1.3 Anisotropic network model (ANM)

The concept of an *elastic network* (EN) for modeling the equilibrium dynamics of proteins was introduced a decade ago (Bahar *et al*., 1997; Haliloglu *et al*., 1997), inspired by a NMA with uniform harmonic potentials performed by Tirion (1996). The applicability of an EN–NMA at residue level was first shown by Hinsen (1998), followed by others (Atilgan *et al*., 2001; Doruker *et al*., 2000; Tama and Sanejouand, 2001), as recently reviewed (Bahar and Rader, 2005; Chennubhotla *et al*., 2005; Ma, 2005; Tama and Brooks, 2006). In the ANM (Atilgan *et al*., 2001; Doruker *et al*., 2000), the network nodes are identified with the positions of the α-carbons, and uniform

elastic springs with force constant γ connect the nodes located within a cutoff distance of r_c such that the underlying potential is

$$V = \frac{\gamma}{2}\sum_i \sum_{j>i} (\Gamma_{ij})(R_{ij} - R_{ij}^0)^2 \quad (4)$$

Here R_{ij} and R_{ij}^0 are the respective instantaneous and equilibrium distances between residues i and j, Γ_{ij} is the ijth element of the Kirchhoff matrix Γ of inter-residue contacts equal to 1 if nodes i and j are connected by a spring, zero otherwise. The NMA of the ANM requires the eigenvalue decomposition of the Hessian $\boldsymbol{H}$, a $3N \times 3N$ matrix composed of $N \times N$ super elements, $\boldsymbol{H}_{ij}$ ($i \neq j$) of the form

$$\mathbf{H}_{ij} = \frac{\gamma\,\Gamma_{ij}}{(R_{ij}^0)^2}\begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix} \quad (5)$$

obtained from the 2nd derivative of V with respect to residue positions. The diagonal super elements of $\mathbf{H}$ are given by $\mathbf{H}_{ii} = -\sum_{j,\ j\neq i} \mathbf{H}_{ij}$. Here X_{ij}, Y_{ij}, and Z_{ij} are the components of the distance vector $\mathbf{R}_{ij}^0$. The pseudo-inverse of $\mathbf{H}$ is the $3N \times 3N$ covariance matrix, $\mathbf{C}^{\mathrm{ANM}}$, predicted by the ANM. The diagonal super-elements of $\mathbf{C}^{\mathrm{ANM}}$ (3×3 matrices $\mathbf{C}_{ii}^{\mathrm{ANM}}$) will be compared below to the experimental ADPs $\mathbf{C}_{ii}$ for each node/residue i, represented by the corresponding α-carbon, consistent with the ANM. $\mathbf{C}^{\mathrm{ANM}}$ can be expressed in terms of the $3N$-6 non-zero eigenvalues λ_k and corresponding eigenvectors $\mathbf{u}_k$ of $\mathbf{H}$ as

$$\mathbf{C}^{\mathrm{ANM}} = \sum_{k=1}^{3N-6} \frac{1}{\lambda_k}\mathbf{u}_k\mathbf{u}_k^T \quad (6)$$

The B-factor predicted by the ANM for residue i is calculated from the trace of $\mathbf{C}_{ii}^{\mathrm{ANM}}$ using $B_i^{\mathrm{ANM}} = (8\pi^2 k_B T/3\gamma)\ tr(\mathbf{C}_{ii}^{\mathrm{ANM}})$ where k_B is the Bolzmann constant and T is the absolute temperature. The value of γ is determined *a posteriori* if experimental data are available, and does not affect the fluctuation profile of residues.

2 METHODS

2.1 Datasets

We extracted 1432 PDB entries for which the ADPs were reported as of January 2006, eliminated the redundant structures in this set using PISCES (Wang and Dunbrack, 2005) such that, in the final set, no protein pair had more than 15% sequence identity. Furthermore, only proteins with resolution higher than 1.5 Å and R-factor smaller than 0.3 Å were retained. Finally, PDB files for which ADPs were assigned to the majority of the atoms were selected. The final set includes 93 proteins listed in the Supplementary Material Table S1. This representative set of non-redundant high-resolution structures with available ADPs will be referred to as *Set I*. Two additional datasets were compiled for proteins having at least two independent X-ray structures with ADPs. The first (19 pairs) is comprised of pairs of structures having the similar crystal form, i.e. it includes proteins with two crystal structures based on the same space group and unit cell dimensions. The second (8 pairs) is comprised of pairs of structures (same protein in each pair) resolved in different crystal form (different space group or unit cell dimensions). These two sets are listed in the respective Tables 1 and 2.

2.2 Measuring similarities between ADPs

The similarities between ADPs were measured in terms of (i) the Pearson correlation $\rho(\boldsymbol{s_A}, \boldsymbol{s_B})$ between the N (or $6N$)-dimensional arrays $\boldsymbol{s}_{\mathbf{A}}$ and $\boldsymbol{s}_{\mathbf{B}}$ that describe the ADPs of the N amino acids (or atoms) in the structures A and B, and (ii) the *Kullback–Leibler (*KL*) distances D* between the ADP matrices corresponding to the structures A and B. The arrays $\boldsymbol{s}_{\mathbf{A}}$ and $\boldsymbol{s}_{\mathbf{B}}$ include the anisotropies $\{A_i\}$, volumes $\{V_i\}$, diagonal $\{\langle(\Delta X_i)^2\rangle, \langle(\Delta Y_i)^2\rangle, \langle(\Delta Z_i)^2\rangle\}$ and off-diagonal $\{\langle\Delta X_i \Delta Y_i\rangle, \langle\Delta X_i \Delta Z_i\rangle, \langle\Delta Y_i \Delta Z_i\rangle\}$ elements of $\mathbf{C}_{ii}$ considered both separately and together, and the traces $\{\langle(\Delta R_i)^2\rangle\}$ of $\mathbf{C}_{ii}$, for the structures A and B and $1 \le i \le N$. For simplicity, we adopt the notation $\rho(\boldsymbol{s})$, the argument designating the array.

In the following, we will compare the ADPs for α-carbons at the *same* sequence position (e.g. i) for a pair of structures (that are otherwise structurally aligned). Two types of comparisons will be performed: (i) pairs of PDB structures A and B determined for the same protein in different settings, termed *experiments versus experiments* comparisons, and (ii) pairs of experimental and computational ADP data for the same PDB entry, termed *experiments versus theory*.

The KL distance between the trivariate (x-, y- and z-) Gaussian probability distributions a and b, associated with the anisotropic fluctuations of atom i, is expressed in terms of the respective eigenvalues (d_{ak}, d_{bk}, $1 \le k \le 3$) and eigenvectors ($\boldsymbol{v}_{ak}, \boldsymbol{v}_{bk}, 1 \le k \le 3$) of the corresponding ADP matrices $\mathbf{C}_{ii}^{\mathrm{a}}$ and $\mathbf{C}_{ii}^{\mathrm{b}}$ as

$$D_i^{ab} = -\frac{3}{2} + \frac{1}{2}\sum_{k=1}^{3} ln\frac{d_{bk}}{d_{ak}} + \frac{1}{2}\sum_{k=1}^{3}\sum_{l=1}^{3}\frac{d_{ak}}{d_{bl}}\left|\mathbf{v}_{ak}^{\mathrm{T}}\mathbf{v}_{bl}\right|^2 \quad (7)$$

See Supplementary Material for a complete derivation. As D_i^{ab} is asymmetrical, we use the arithmetic average $\frac{1}{2}(D_i^{ab} + D_i^{ba})$ for evaluating the distance between $\mathbf{C}_{ii}^{\mathrm{a}}$ and $\mathbf{C}_{ii}^{\mathrm{b}}$. $D_i^{ab} = 0$ for two identical distributions, and increases with the divergence between them.

2.3 Visualization and implementation in ANM server

The ANM website (http://ccbb.pitt.edu/anm) (Eyal *et al.*, 2006) has been further developed to compute, from knowledge of structural coordinates, the C^{α} ADPs for PDB entries or user-supplied structures submitted in PDB format, and display the ellipsoids created by Rastep (Merritt and Bacon, 1997; Merritt and Murphy, 1994) or Povscript (Fenn *et al.*, 2003). If isotropic temperature factors are available in the input file, the server offers an option to use them for rescaling the absolute sizes of atomic fluctuations and evaluating the corresponding ADPs.

2.4 Solvent accessibilities and secondary structures

Solvent accessibilities were computed using Voronoi polyhedra method (McConkey *et al.*, 2002). The accessibility of residue X is defined as the ratio of its total solvent-accessible surface in the native structure to that in the peptide GGXGG with the same backbone conformation. Residues with accessibility of less than 0.1 are accepted as buried, those between 0.1 and 0.5 are assume to be partially exposed and those above 0.5 are taken as exposed. Secondary structures are assigned using DSSP (Kabsch and Sander, 1983).

3 RESULTS AND DISCUSSION

3.1 Experiments versus experiments comparisons of ADPs

We first examined the extent of anisotropy observed in different structural elements, Figure 1 displays the distribution of

Table 1. Correlations between the ADPs reported in 20 pairs of PDB structures for the same protein in identical crystal forms*

PDB structures A and B		Pearson correlation coefficient ρ between							KLD[b]
		N	ADP volumes[a]	Anisotropy[a]	$tr(\mathbf{C}_{ii})$[a]	All ADP[a]	Diagonal ADP[a]	Off-diagonal ADP[a]	
1nwczA	1uwnX	120	0.930	0.062	0.951	0.925	0.857	0.717	0.231
1pq7A	1gdnA	51	0.801	0.176	0.830	0.902	0.638	0.474	0.183
1kmvA	1kmsA	183	0.900	0.710	0.924	0.937	0.856	0.876	0.135
1f9yA	1q0nA	158	0.864	0.400	0.865	0.895	0.771	0.622	0.308
1lugA	1oq5A	130	0.847	0.613	0.840	0.902	0.778	0.692	0.222
1lokA	1rtqA	291	0.832	0.535	0.795	0.846	0.719	0.467	0.200
1m40A	1nymA	263	0.806	0.367	0.806	0.933	0.638	0.534	0.129
1oc7A	1oc6A	364	0.933	0.007	0.939	0.970	0.914	0.664	0.058
1sx7A	1swzA	164	0.990	0.896	0.963	0.962	0.948	0.779	0.070
1i1wA	1i1xA	299	0.818	0.443	0.812	0.941	0.784	0.615	0.114
1a6m	1bzpA	151	0.664	0.532	0.638	0.858	0.613	0.482	0.134
1pwmA	1z8aA	315	0.921	0.843	0.933	0.928	0.903	0.837	0.120
1me4A	1me3A	208	0.939	0.478	0.949	0.957	0.888	0.711	0.121
1rgzA	1q2qA	359	0.926	0.347	0.930	0.912	0.827	0.672	0.230
1m1rA	1m1qA	90	0.881	0.798	0.900	0.962	0.872	0.754	0.066
1g6xA	1k6uA	58	0.774	0.860	0.828	0.930	0.797	0.802	0.043
3lzt	4lzt	129	0.899	0.345	0.845	0.870	0.731	0.525	0.166
1g66A	1bs9	207	0.896	0.351	0.897	0.918	0.799	0.687	0.157
1kt5A	1kt7A	175	0.855	0.084	0.872	0.872	0.752	0.606	0.230
Mean		195	0.867	0.464	0.870	0.916	0.793	0.658	0.154

*The two crystals share the same symmetry (defined by the space group) and have the same dimensions of the unit cell.
[a]The reported values are the mean of the correlation coefficient values obtained over all residues.
[b]For each protein the mean KL distance of all residues is calculated.

Table 2. Correlations between the ADPs reported for eight pairs of PDB structures for the same protein in dissimilar crystal forms*

PDB structures A and B		Pearson correlation coefficient ρ between							*KLD*[b]
		N	ADP volumes[a]	Anisotropy[a]	$tr(\mathbf{C}_{ii})$[a]	All ADP[a]	Diagonal ADP[a]	Off-diagonal ADP[a]	
1nwzA	1kouA	119	0.682	0.294	0.644	0.812	0.480	0.482	0.287
1pq7A	1ppzA	223	0.522	0.206	0.557	0.817	0.433	0.366	0.376
1f9yA	1rb0A	151	0.219	0.109	0.372	0.676	0.253	0.183	0.968
2a6zA	2a70B	222	0.854	0.506	0.853	0.920	0.797	0.675	0.163
1a6m	1u7sA	151	0.448	0.326	0.436	0.799	0.381	0.386	0.540
1pwmA	1t41A	315	0.747	0.042	0.756	0.836	0.656	0.540	0.392
3lzt	1ieeA	129	0.587	0.128	0.560	0.800	0.506	0.281	0.317
1iqzA	1ir0A	81	0.295	0.308	0.356	0.716	0.339	0.426	0.256
Mean		173	0.544	0.240	0.569	0.797	0.480	0.417	0.412

*The two crystals differ by their symmetry (defined by the space group) or by the dimensions of the unit cell.
[a]The reported values are the mean of the correlation coefficient values obtained over all residues.
[b]For each protein the mean KL distance of all residues is calculated.

anisotropies A_i for α-carbons, evaluated for buried residues (panel A) and solvent-exposed residues (panel B), obtained from the statistical examination of the *Set I* of 93 high-resolution structures. Clearly a broad spectrum of anisotropies is observed, consistent with the bell-shaped distribution obtained earlier with a smaller set (Merritt, 1999; Merritt *et al.*, 1998). The anisotropies are more pronounced (skewed towards smaller A_i values) in the case of exposed residues. The average anisotropies are $\langle A_i \rangle = 0.59$ and 0.49 for the two respective subsets of residues. Panels A′ and B′ display the corresponding ellipsoid volumes V_i that encapsulate the fluctuations within one SD away from the mean position. The mean volume shifts to a lower value in the case of buried residues compared to exposed ones ($0.039\,\text{Å}^3$ versus $0.082\,\text{Å}^3$), as expected from the tighter packing of amino acids in the core.

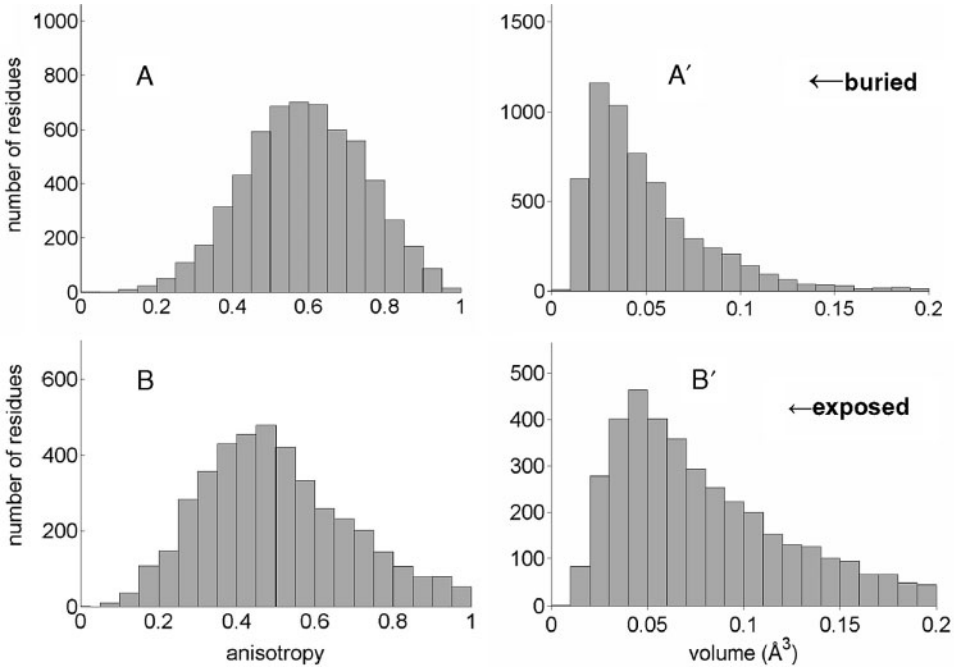

Fig. 1. Histograms of anisotropies (A) experimentally reported for (**A**) buried, and (**B**) solvent-exposed Cα atoms in 93 high-resolution protein structures (Supplementary Material Table S1). The ordinate represents the observed counts for each interval of size $\Delta A = 0.05$, in the range $0 \leq A \leq 1$ ($A = 1$ for *isotropic* fluctuations). Panels **A′** and **B′** are the respective histograms of the ellipsoid volumes encapsulating the fluctuations within one SD away from mean positions.

Next, we proceed to a quantitative assessment of the degree of correlations between the ADPs reported *for the same proteins* under similar or different conditions/crystal forms. Table 1 refers to pairs of structures **A** and **B** determined in isomorphous crystals. Columns 1 and 2 list the PDB codes for each pair, and column 3 gives the size (N) of the corresponding protein. Columns 4–9 list the Pearson correlations $\rho(\boldsymbol{s})$ corresponding to $\boldsymbol{s} = \{V_i\}$, $\{A_i\}$, $\{tr(\mathbf{C}_{ii})\}$, all six ADPs $\{\langle(\Delta X_i)^2\rangle, \langle(\Delta Y_i)^2\rangle, \langle(\Delta Z_i)^2\rangle, \langle\Delta X_i \Delta Y_i\rangle, \langle\Delta X_i \Delta Z_i\rangle, \langle\Delta Y_i \Delta Z_i\rangle\}$, the diagonal and off-diagonal elements of ADP matrices, where $1 \leq i \leq N$. The last column lists the average KL distance $\langle D\rangle$ between the $\mathbf{C}_{ii}$ matrices over all residues $1 \leq i \leq N$ of the two proteins. The values $\langle\rho(\mathbf{s})\rangle$ and $\langle D\rangle$ averaged over the entire set of 19 pairs are listed in the last row. Ideally, pairs of 'identical' proteins/structures resolved under isomorphic crystal conditions, would be expected to yield correlation coefficients equal to unity, and the KL distances equal to zero. Obviously, Table 1 shows departures from these 'ideal' values. The correlation coefficients are close to unity at columns 4, 6 and 7, i.e. the V_i, and $tr(\mathbf{C}_{ii})$ values are in good agreement between the two sets. All these three quantities are dominated, if not fully defined, by the mean-square fluctuations $\langle(\Delta R_i)^2\rangle$ of residues, and the corresponding high correlation coefficients indicate the accord between the two structures' dynamics in so far as mean-square fluctuation profiles of residues (or even the components of $\langle(\Delta \mathbf{R}_i)^2\rangle$) are concerned. The off-diagonal terms of the ADP matrices (*col* 8) show reasonable correlation, while the anisotropies are weakly correlated (0.464), and the average KL distance $\langle D\rangle$ is 0.154. The mean correlations and KL distance $\langle D\rangle$ in the last row of Table 1 may be viewed as the optimal levels of accuracy that can be achieved between two sets of ADPs, given the experimental uncertainties. The interpretation of the ensuing results will be made with reference to these optimal/control values.

We now proceed to the set of structures resolved for the same protein, but under different crystal space group and unit cell dimensions. Table 2 lists the results for eight such pairs retrieved from the PDB. There is a significant decrease in the correlation coefficients, and an increase in KL distance, revealing differences between the ADPs reported for the same protein resolved/modeled with different structure determination/refinement protocols. The correlation $\langle\rho\{tr(\mathbf{C}_{ii})\}\rangle$ between $tr(\mathbf{C}_{ii}) = B_i$ values in this case is 0.569, that between the diagonal elements of $\mathbf{C}_{ij}$ is 0.48 and the correlation of off-diagonal elements 0.42. The mean correlation between the ellipsoid volumes is $\langle\rho\{V_i\}\rangle = 0.54$ and that between anisotropies is $\langle\rho\{A_i\}\rangle = 0.24$. The mean KL distance increases to 0.43. These results provide us with estimates for the levels of agreement one might expect to reach between theory and experiments. In particular, the low $\langle\rho\{A_i\}\rangle$ implies that the anisotropy is a very sensitive measure highly dependent on experimental conditions and techniques.

3.2 Experiments versus theory comparisons

ANM calculations have been performed for the 93 proteins (*Set I*) extracted from the PDB as explained above. Figure 2 illustrates the types of calculations and comparative analysis performed, for an example protein, antifungal protein EAFP2

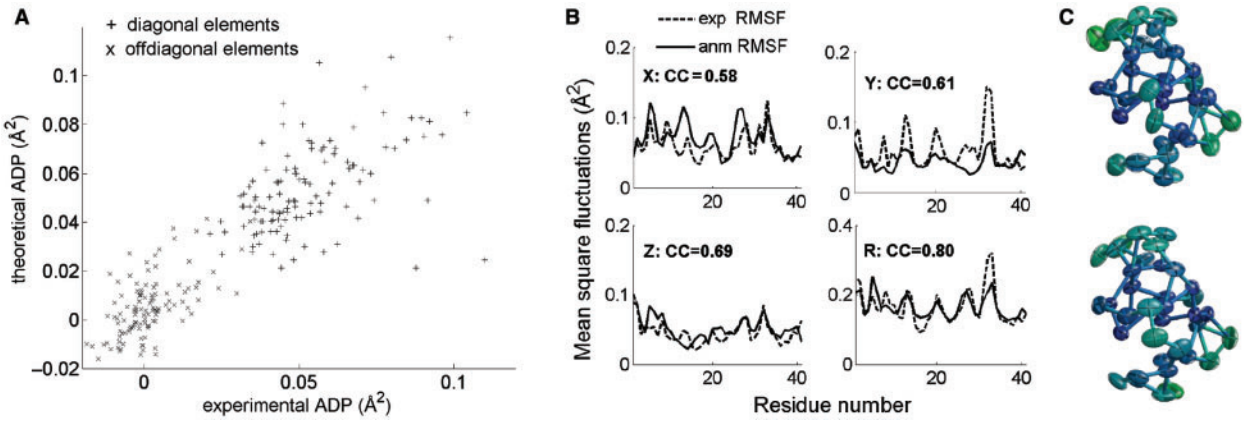

Fig. 2. Comparison of experimental and theoretical anisotropic fluctuation behavior, illustrated for antifungal protein EAFP2 from Oliver tree (PDB code 1p9g, Xiang *et al*., 2004). Theoretical results are computed using the anisotropic network model (ANM). (**A**) Experimental ADPs plotted against their theoretical counterparts for all α-carbons. The ADPs corresponding to the diagonal and off-diagonal elements of the covariance matrix are shown by different symbols. The correlation coefficient between the two sets is 0.85; those evaluated for the subsets of diagonal and off-diagonal elements are 0.61 and 0.51, respectively. (**B**) Comparison between theoretical and experimental mean square fluctuations of Cα-atoms along each axis $\left(\langle(\Delta X_i)^2\rangle, \langle(\Delta Y_i)^2\rangle \text{ and } \langle(\Delta Z_i)^2\rangle\right)$ separately and between the overall fluctuations $\langle(\Delta R_i)^2\rangle$, plotted as a function of residue index i. Solid and dotted curves refer to experimental and theoretical results, respectively. The correlation coefficients (CC) between two sets of data are indicated in each plot. (**C**) Visualization of experimental (top) and theoretically predicted (bottom) ADPs displayed using Rastep (Merritt and Bacon, 1997) The size and direction of the residue displacements are indicated by the orientation and color of the ellipsoids, darker blue indicating smaller fluctuations.

from *Eucommia ulmoides* Oliver tree. Panel A displays all the predicted (by ANM) and the experimental (deposited in the PDB) ADPs, grouped in two subsets, consisting of the diagonal $\left(\langle(\Delta X_i)^2\rangle, \langle(\Delta Y_i)^2\rangle \text{ and } \langle(\Delta Z_i)^2\rangle\right)$ and off-diagonal $(\langle\Delta X_i \Delta Y_i\rangle, \langle\Delta X_i \Delta Z_i\rangle \text{ and } \langle\Delta Y_i \Delta Z_i\rangle)$ elements of $\mathbf{C}_{ii}$ for $1 \le i \le N$. The correlation between the two sets of ADPs is 0.85 in this case. The correlations for the subsets of diagonal and off-diagonal elements, computed separately, are 0.61 and 0.51, respectively. Panel B compares the predicted and the experimental mean-square fluctuations $\langle(\Delta X_i)^2\rangle$, $\langle(\Delta Y_i)^2\rangle$ and $\langle(\Delta Z_i)^2\rangle$, as well as the overall mean-square fluctuations $\langle(\Delta R_i)^2\rangle$. In panel C, we display the respective experimental (*upper*) and the theoretical (*lower*) ADPs as color-coded ellipsoids, blue and red referring to the smallest and largest size fluctuations, and the orientation of the beads indicating the anisotropic directions of the fluctuations.

Similar calculations performed for the complete set of 93 proteins in *Set I* led to the average results presented in Table 3. For detailed results, similar to Tables 1 and 2, the reader is referred to the Table S1 in the Supplementary Material. The first row shows the correlation (0.77) between the two sets of $6N$ ADPs, from the PDB and ANM predictions, averaged over all proteins (i.e. by repeating for all proteins the analysis illustrated in Fig. 2A for EAFP2). This is a good agreement in view of the uncertainties in experiments and approximations in the theory. The corresponding average correlations for the diagonal and off-diagonal elements of the ADP matrices, considered separately, are listed in the 2nd and 3rd rows. Separation of the ADPs into these two populations lowers the overall correlation, because the relative sizes of auto- and cross-terms of ADP matrices, which are presumably in good agreement between theory and experiments, are not being taken into consideration upon treating the two subsets as independent entities.

Table 3. Summary* of the average correlations between experimental ADPs and those computed using different elastic network models

	ANM	Tirion's model	GNM
All ADP	0.771	0.712	
Diagonal ADP	0.486	0.455	0.481
Off-diagonal ADP	0.343	0.305	
B-factors (ADP trace)	0.571	0.567	0.587
Anisotropy	0.260	0.230	
Volume	0.547	0.560	0.538
KL distance	0.429	1.065	0.340

*The reported values are the mean correlation coefficients averaged over the results obtained for 93 proteins. See the Table in Supplementary Material I for the values obtained for each protein.

The fourth row in Table 3 lists the correlations between the traces of the ADP matrices. The average correlation obtained, 0.57, is similar to that reported in previous studies of B-factors (Hamacher and McCammon, 2006). The better agreement between theoretical and experimental $\langle(\Delta R_i)^2\rangle$, compared to that obtained for the X-, Y- and Z- components is a common feature in the examined set, suggesting that the *sizes* of the motions are more accurately measurable/predictable, than their *directions*, in general. The correlations $\langle\rho\{A_i\}\rangle$ and $\langle\rho\{V_i\}\rangle$ in Table 3 are generally similar to values observed in Table 2.

The low $\langle\rho\{A_i\}\rangle$ further strengthens the view that A_i values should be interpreted with caution; whereas V_i provides a reasonable description of the fluctuation volumes, consistent

with B_i. Finally, the KL distance averaged over all proteins (last row) is also comparable to the one obtained in Table 2, overall lending support to the view that the two sets of data, experimental and theory, do not differ from each other more than those obtained experimentally for the same proteins under different settings.

3.3 How do results depend on the coarse-graining of the theoretical model?

The ANM predictions presented above are obtained with a residue-level representation of the proteins, each residue position on the network being identified by that of its α-carbons. Of interest is to assess the effect of the selected coarse-grained model on theoretical results. Does the absence of atomic details in the model give rise to inaccuracies in predicting directional fluctuations? How would theory and experiments compare if fluctuations at atomic scale were considered in details?

To answer these questions, we repeated the computations using an atomic level EN model (Tirion, 1996). A cutoff interaction distance of 5 Å between atoms (Sen *et al.*, 2006) was adopted for defining the interacting pairs of atoms. The calculations were done for 60 proteins in the set, the sizes of which are smaller than 250 residues. While the calculations yield data on the fluctuations of all atoms, for direct comparison with our residue-level results we examined the fluctuation dynamics of C^{α} atoms, derived from this full atomic description. The entries in the middle column of Table 3 list the results obtained in this case. As can be clearly seen, the results are comparable to those found with the residue-level model. The coarse-graining is clearly not a major source for inducing additional differences between residue level $\mathbf{C}_{ii}^{\mathrm{ANM}}$ and experimentally refined ADPs.

We also examined the fluctuations obtained using the GNM (Bahar *et al.*, 1997), an isotropic network model. The C^{α} ADP matrices based on the GNM are simply diagonal matrices with identical mean-square fluctuations along all three directions. Strikingly, we found (right column of Table 3) that the agreement between GNM results and experimental data is comparable to that obtained with the ANM. This is presumably due to the more accurate prediction of fluctuation sizes by the GNM compared to the ANM (Bahar *et al.*, 2007; Chennubhotla *et al.*, 2005), which more than compensates the discrepancies arising from the lack of anisotropy.

3.4 KL distance: a measure of assessing the robustness of ADPs

We now concentrate on KL divergence as a measure of the probability distributions of residue displacements in 3D space, or the entire shape of ellipsoids (rather than a single parameter per ADP matrix like the anisotropy which is shown to be highly sensitive to model parameters, see Supplementary Material). The KL divergences are proposed here to serve as a metric for estimating the robustness of the ADPs.

As mentioned above, the KL distance between two distributions vary in the range of $0 \leq D_i \leq \infty$, with the lower limit corresponding to identical distributions; and D_i increases with

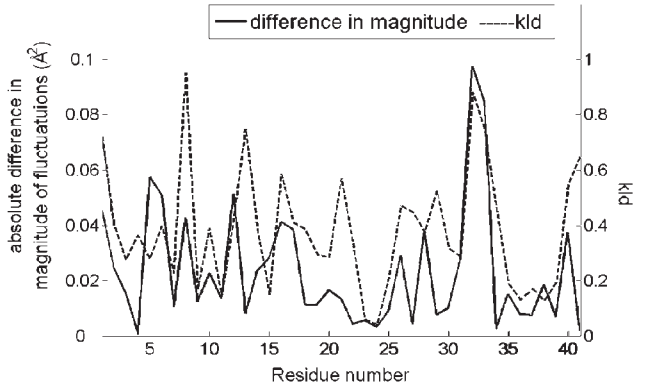

Fig. 3. Difference in the magnitude of fluctuations (absolute difference $\Delta\langle(\Delta\mathbf{R}_i)^2\rangle$ between traces of the experimental and theoretical ADP matrices) and *KL* distance values for each residue shown for protein EAFP2 (PDB code 1p9g).

the dissimilarity in the shape of the ADPs corresponding to the ith atom in the two sets of data. To view how KL divergences incorporate information different from (and additional to) those typically conveyed by isotropic fluctuations, let us first consider two quantities: (i) the KL divergence, D_i between the ADP matrix $\mathbf{C}_{ii}^{\mathrm{ANM}}$ computed by ANM for atom i, and its experimental counterpart, $\mathbf{C}_{ii}$, and (ii) the information derived from the diagonal elements alone of the ADP matrices, i.e. using $\Delta\langle(\Delta R_i)^2\rangle \equiv \{tr(\mathbf{C}_{ii}^{\mathrm{ANM}}) - tr(\mathbf{C}_{ii})\}$. Figure 3 compares these two quantities for the α-carbons in the protein EAFP2 analyzed above (Fig. 2). D_i and $\Delta\langle(\Delta\mathbf{R}_i)^2\rangle$ profiles are presented as dotted and solid curves respectively, with the respective scales shown along the right and left ordinates. Significant differences between the two quantities are observed at particular residues. These originate from the shape of the ellipsoids and the cross-correlations between the fluctuations along different directions at those particular residues, which are not accounted for by $\Delta\langle(\Delta\mathbf{R}_i)^2\rangle$, but included in D_i. For example, the C-terminus appears to diverge between the two sets of data according to D_i values while $\Delta\langle(\Delta\mathbf{R}_i)^2\rangle$ is negligibly small, i.e. the sizes of the motions are comparable, while their spatial distributions differ. Therefore, the differences (between the theoretical and experimental data) overlooked by $\Delta\langle(\Delta\mathbf{R}_i)^2\rangle$ are captured by the relatively high D_i values.

Notably, minima in the D_i profile indicate the structural regions whose experimental and theoretical ADPs are in agreement. The KL divergence between theoretical and experimental data thus provides information on the regions whose ADPs are more robustly defined. D_i values may even serve as a metric for assessing the confidence levels of the ADPs. Figure 4 shows such results for different types of secondary structural elements, or different extents of solvent exposure. The KL distances are lower at the protein interior, and at ordered regions (α-helix, β-strand). This difference between regular/ordered and disordered regions (such as loops, coils and turns) becomes more pronounced when the residues are solvent exposed.

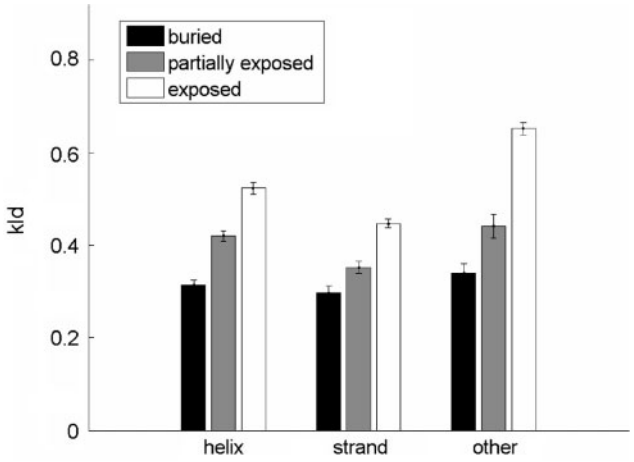

Fig. 4. KL distance between experimental and theoretical distributions of atomic fluctuations (covariance matrices) for different types of secondary structures and different levels of solvent accessibilities. Buried β-strands exhibit the lowest KL distances, i.e. the best agreement between theory and experiments.

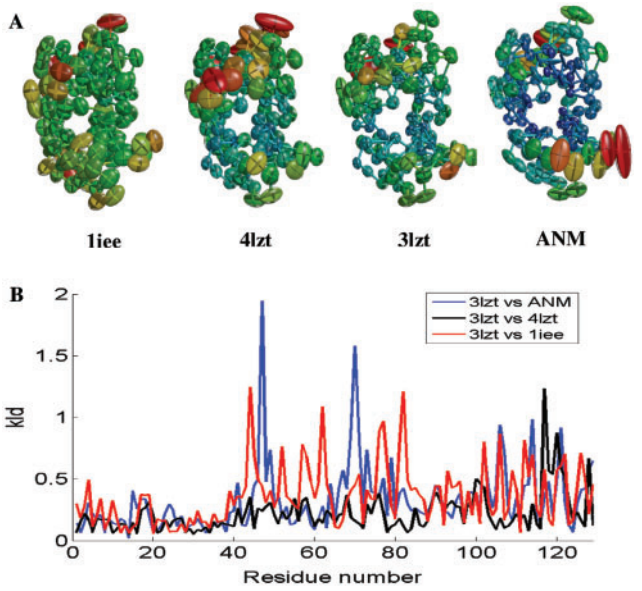

Fig. 5. Comparison of ADPs obtained from different crystal structures and from theoretical model for hen egg white lysozyme. (**A**) Diagrams displaying the ADP values (with the ellipsoid volumes corresponding to one SD displacements away from the mean positions) shown for 1iee, 4lzt, 3lzt (experimental) and using the theoretical values calculated by ANM for 3lzt, created using Rastep (Merritt and Bacon, 1997). The orientation of the color-coded ellipsoids (red to blue: largest to smallest size fluctuations) display the direction of fluctuations. (**B**) The KL divergence for individual residues ADPs, evaluated for the indicated pairs: the structures resolved from isomorphous crystals of HEWL (3lzt, 4lzt: space group P1), the HEWL structures resolved from different crystal forms (3lzt, 1iee with space groups P1 and P 43 21 2, respectively) and the pair of experimental and theoretical dataset for 3lzt.

Figure 5 illustrates, for hen egg white lysozyme (HEWL), how the KL distances may be used for identifying structural regions or residues distinguished by their well-defined (robust) anisotropic fluctuations. Panel A displays the experimentally assigned ADPs for α-carbons in three high-resolution structures (PDB files 3lzt, 4lzt and 1iee (Sauter *et al.*, 2001; Walsh *et al.*, 1998) of HEWL, and those predicted by the ANM for 3lzt, as color/size-coded ellipsoids. Here 3lzt and 4lzt correspond to isomorphous crystals of HEWL, and the pair 3lzt and 1iee refer to different crystal forms. Panel B displays the KL distances between the ADPs for the pairs listed in the inset. Maxima point to regions where the two sets of data exhibit the strongest departure, and minima indicate the regions where the two sets of data concur. Interestingly, the N-terminal segment of about 40 residues is observed in all cases to exhibit the lowest divergence, pointing to the robustness of the fluctuation dynamics at this region. Additionally, the helical segment 80–100 appears to have well-defined preferred motions. If we focus in particular on the structure (3lzt) for which ANM calculations have been performed (the blue curve in Figure 5A), we also distinguish the β-strand region (residues 50–65) to exhibit minimal divergence.

3.5 Tangential versus radial displacements

The good agreement between the ADPs corresponding to pairs of isomorphous structures and the departures between the ADPs from structures based on different crystal forms were originally noted by Merrit (Merritt, 1999). Rigid body motions were suggested to be responsible for these differences. If rigid body rotations of proteins in the crystal are the main contributors to ADPs, the atoms located far away from the mass center would be expected to exhibit larger tangential fluctuations compared to the atoms that are closer to the mass center. For the radial fluctuations, this trend would not be expected (Schneider, 1996). Analysis of the experimental fluctuations for many proteins in our set reveals a clear trend in which tangential displacements indeed become larger with the increasing distance from the protein center, and the radial displacements are almost not affected. However, a very similar trend is obtained for the theoretical fluctuations calculated by the ANM, as well, although the rigid body motions are by definition eliminated in the ANM fluctuations. Figure 6 illustrates this for oxy-myoglobin (Vojtechovsky *et al.*, 1999). Having this trend detected by a model which is solely based on the internal motions of the protein, suggests that it is also internal phenomenon (probably a by-product of the large scale slow mode motions) and should not be explained only by rigid body motions in the crystal. It is likely that other effects such as crystal contacts also make crystal-dependent contributions to the ADPs. Differences in the refinement protocols also seem to contribute, as suggested by the small differences between the pairs of isomorphous structures. The agreement between the experimental and theoretical values is found to be better for tangential fluctuations as shown for oxi-myoglobin in Figure 6 panel C.

4 CONCLUSION

The degree to which experimental ADPs represent internal dynamics, rigid body motions or static disorders is still under

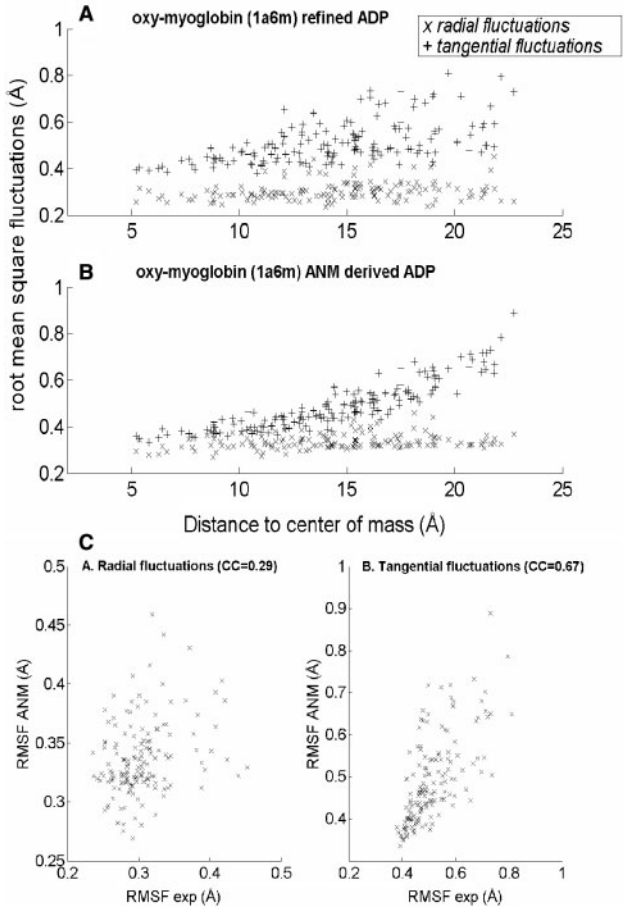

Fig. 6. Radial and tangential components of the mean-square fluctuations of C^{α}-atoms as a function of the distance from the mass center, shown for oxy-myoglobin [PDB code: 1a6m (Vojtechovsky *et al.*, 1999)]. (**A**) Experimental fluctuations. (**B**) Theoretical (ANM) fluctuations. (**C**) Comparison of theoretical and experimental results for radial (left) and tangential (right) fluctuations.

long debate for both proteins and small molecules (Dunitz *et al.*, 1988; Harata *et al.*, 1998; Merritt, 1999). It is also not clear how much crystal contacts affect the internal dynamics of proteins in crystals, but such effects surely exist (Kundu *et al.*, 2002). Comparative analysis of the experimental data for a given protein structurally resolved in different crystal forms has been shown earlier (Diamond, 1990; Kidera and Go, 1992; Winn *et al.*, 2001), and confirmed here with a larger set (Table 2), to display poor-to-moderate agreement with regard to the anisotropic nature of residue fluctuations, raising questions about the physical meaning of the ADPs, or the criteria for assessing the levels of agreement between different datasets. Anisotropies (A_i) of individual amino acids emerge as a poor metric, even for pairs of structures resolved in isomorphic crystals ($\langle\rho\{A_i\}\rangle = 0.464$; see Table 1). On the other hand, direct examination of the complete set of $6N$ distinctive ADPs for the N amino acids in a given protein point to some conservation/reproducibility of the anisotropic fluctuation data despite all perturbing effects.

Comparison of theoretical ANM results with experimental data yielded overall correlations, or KL divergences, comparable to those observed in Table 2. Such an agreement between theory and experiments is satisfactory in view of several possible sources of discrepancy between the two sets: approximations in the ANM, experimental artifacts due to static disorder induced by the slightly different location of atoms in different unit cells, crystal contacts or inaccuracies and errors in the refinement process such as interpretation of alternative conformations as a single structure with large displacement parameters.

Our analysis suggests that there are subsets of residues that tend to have more robust preferences with regard to the spatial distributions of their fluctuations/displacements than others. Here, to identify and characterize such amino acids, we propose to use the residue-based KL divergences between $\mathbf{C}_{ii}$ and $\mathbf{C}_{ii}^{\mathrm{ANM}}$. KL divergences vary in the range $0 \leq D_i \leq \infty$. The D_i profiles for individual proteins, illustrated in Figure 5B provide an assessment of the structural regions that exhibit consistent fluctuations between theory and experiments (e.g. $D_i \leq 0.3$, which corresponds to the 58% of all 17 308 residues examined in our entire dataset). In this context, the ADPs at the protein core or at regular structures appear to be more accurately measured/predicted than other regions in general (Fig. 4).

Finally, the examination of residue displacements along radial and tangential directions shows that the amplitude of tangential motions increase with the distance from the mass center, while radial motions remain unaffected. This feature in close accord with experimental data has been largely attributed to rigid body rotational motions of the proteins in the crystal, resulting in larger tangential displacements of exterior residues. Yet, the same behavior obtained here with the ANM purely internal conformational fluctuations in the absence of any rigid body motions suggests that this interpretation may not be complete. In fact, internal motions also lead to larger tangential displacements at distant positions, and this is a natural consequence of the dominant effect of slow modes that drive global/domain motions with respect to central hinge centers or domain interfaces. We also note that an extensive study of X-ray crystallographic B-factors by Phillips and coworkers (Kundu *et al.*, 2002) using the TLS model and an elastic network model demonstrated that the elastic network models yield better agreement with experiments, compared to TLS models. This is also consistent with the proven success of normal mode analysis in assisting structural refinement (Delarue and Dumas, 2004; Diamond, 1990; Kidera and Go, 1990; Suhre and Sanejouand, 2004).

Therefore we conclude that the ADPs reported for highly refined PDB structures do convey information on the directional preferences of individual residues. While they are subject to several perturbing effects, a reasonable correlation is observed with the ANM predictions that are based solely on internal conformational motions, particularly at core regions and regular structural elements. Combined examination of experimental and computed ADPs may thus assist in identifying the amino acids that possess strong structure-encoded preferences to fluctuate in specific directions. Notably, in the absence of experimental data, theoretical ADPs can be resorted to for a first estimation of the anisotropic displacements of

amino acids. The features of the ANM server (www.ccbb.pitt.edu/anm) have now been augmented to permit users to calculate and visualize the ADPs for query structures.

ACKNOWLEDGEMENTS

Support by NIH R33 GM068400-01A2 is gratefully acknowledged by I.B. While publishing this work, we have been alerted to a new related paper on comparing anisotropic temperature factors with theoretical models (Kondrashov *et al*., 2007).

Conflict of Interest: none declared.

REFERENCES

Alexandrov,V. *et al*. (2005) Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. *Proteins Sci*., **14**, 633–643.

Atilgan,A. *et al*. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J*., **80**, 505–515.

Bahar,I. *et al*. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des*., **2**, 173–181.

Bahar,I. *et al*. (2007) On the theoretical foundations of the Gaussian network model and its applications to proteins. *Phys. Biol*., **4**, 64–65.

Bahar,I. *et al*. (2005) Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol*., **15**, 586–592.

Berman,H. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Chennubhotla,C. *et al*. (2005) Elastic network models for understanding biomolecular machinary: from enzymes to supramolecular assemblies. *Phys. Biol*., **2**, S173–S180.

Cui,Q. *et al*. (2006) *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Chapman & Hall/CRC, Boca Raton, FL, USA.

Delarue,M. *et al*. (2004) On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl Acad. Sci. USA*, **101**, 6957–6962.

Diamond,R. (1990) On the use of normal modes in thermal parameter refinement: theory and application in the bovine pancreatic trypsin inhibitor. *Acta. Cryst*., **A46**, 425–435.

Doruker,P. *et al*. (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to amylase inhibitor. *Proteins*, **40**, 512–524.

Dunitz,J. *et al*. (1988) Atomic motions in molecular crystals from diffraction measurements. *Angew. Chem. Int. Ed. Engl*., **27**, 880–895.

Eyal,E. *et al*. (2006) Anisotropic Network Model: systematic evaluation and a new web interface. *Bioinformatics*, **22**, 2619–2627.

Fenn,T. *et al*. (2003) POVScript+: a program for model and data visualization using persistence of vision ray-tracing. *J. Appl. Cryst*., **36**, 944–947.

Haliloglu,T. *et al*. (1997) Gaussian dynamics of folded proteins. *Phys. Rev. Lett*., **79**, 3090–3093.

Hamacher,K. *et al*. (2006) Computing the amino acid specificity of fluctuations in biomolecular systems. *J. Chem. Theory Comput*., **2**, 873–878.

Harata,K. *et al*. (1998) Full-matrix least-square refinement of lysozymes and analysis of anisotropic thermal motion. *Proteins*, **30**, 232–243.

Hinsen,K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33**, 417–429.

Kabsch,W. *et al*. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kidera,A. *et al*. (1990) Refinement of protein dynamics structure: normal mode refinement. *Proc. Natl Acad. Sci. USA*, **87**, 3718–3722.

Kidera,A. *et al*. (1992) Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J. Mol. Biol*., **225**, 457–475.

Kondrashov,D.A. *et al*. (2007) Protein structural variation in computational models and crystallographic data. *Structure*, **15**, 169–177.

Krebs,W. *et al*. (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins*, **48**, 682–695.

Kundu,S. *et al*. (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophys. J*., **83**, 723–732.

Ma,J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, **13**, 373–380.

McConkey,B.J. *et al*. (2002) Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics*, **18**, 1365–1373.

Merritt,E. (1999) Comparing anisotropic displacement parameters in protein structures. *Acta. Cryst*., **D55**, 1997–2004.

Merritt,E. (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta. Cryst*., **D55**, 1109–1117.

Merritt,E. *et al*. (1997) Raster3D Photorealistic Molecular Graphics. *Meth. Enzymol*., **277**, 505–524.

Merritt,E. *et al*. (1998) The 1.25 Å resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site. *J. Mol. Biol*., **282**, 1043–1059.

Merritt,E. *et al*. (1994) Raster3D Version 2.0 – a program for photorealistic molecular graphics. *Acta. Cryst*., **D50**, 869–873.

Nicholay,S. *et al*. (2006) Functional modes of proteins are among the most robust. *Phys. Rev. Let*., **96**, 078104.

Painter,J. *et al*. (2006) Optimal description of protein structure in terms of multiple groups undergoing TLS motion. *Acta. Cryst*., **D62**, 439–450.

Rosenfield,R. *et al*. (1978) A test for rigid vibrations, based on generalisation of Hirshfeld's "rigid bond" postulate. *Acta. Cryst*., **A34**, 828–829.

Sauter,C. *et al*. (2001) Structure of tetragonal hen egg-white lysozyme at 0.94 Å from crystals grown by the counter-diffusion method. *Acta. Cryst*., **D57**, 1119–1126.

Schneider,T. (1996) What can we learn from anisotropic temperature factors? In: *Proceedings of the CCP4 study weekend,* SERC Darsbury Laboratory, Daresbury UK, pp. 133–144.

Schomaker,V. *et al*. (1968) On the rigid-body motion of molecules in crystals. *Acta. Cryst*., **B24**, 63–76.

Sen,TZ. *et al*. (2006) The extent of cooperativity of protein motionsobserved with elastic network models is similar for atomic and coarser-grained models. *J. Chem. Theory Comput*., **2**, 696–704.

Suhre,K. *et al*. (2004) On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta. Cryst*., **D60**, 796–799.

Tama,F. *et al*. (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct*., **35**, 115–133.

Tama,F. *et al*. (2001) Conformational changes of proteins arising from normal mode calculations. *Protein Eng*., **14**, 1–6.

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

Trueblood,K. *et al*. (1996) Atomic displacement parameters nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. *Acta. Cryst*., **A52**, 770–781.

Vojtechovsky,J. *et al*. (1999) Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J*., **77**, 2153–2174.

Walsh,M. *et al*. (1998) Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta. Cryst*., **D54**, 522–546.

Wang,G. *et al*. (2005) PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res*., **33**, W94–W98.

Willis,B. *et al*. (1975) *Thermal Vibrations in Crystallography*. University press, London.

Winn,M. *et al*. (2001) Use of TLS parameters to model anisotropic displacement in macromolecular refinement. *Acta. Cryst*., **D57**, 122–133.

Xiang,Y. *et al*. (2004) Crystal structure of a novel antifungal protein distinct with five disulfide bridges from Eucommia ulmoides Oliver at an atomic resolution. *J. Struct. Biol*., **148**, 86–97.

Yang,L.W. *et al*. (2006) oGNM: online computation of structural dynamics using the gaussian network model. *Nucleic Acids Res*., **34**, W24–W31.