BIOINFORMATICS ORIGINAL PAPER Vol. 24 no. 4 2008, pages 521–528
doi:10.1093/bioinformatics/btm625

Structural bioinformatics

Structural flexibility in proteins: impact of the crystal environment

Konrad Hinsen[1,2]

[1]Centre de Biophysique Moléculaire (CNRS), Rue Charles Sadron, 45071 Orléans Cedex 2, France and
[2]Synchrotron Soleil, Saint Aubin, B.P. 48, 91192 Gif sur Yvette Cedex, France

Received on October 4, 2007; revised on November 19, 2007; accepted on December 15, 2007

Advance Access publication December 18, 2007

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: In the study of the structural flexibility of proteins, crystallographic Debye-Waller factors are the most important experimental information used in the calibration and validation of computational models, such as the very successful elastic network models (ENMs). However, these models are applied to single protein molecules, whereas the experiments are performed on crystals. Moreover, the energy scale in standard ENMs is undefined and must be obtained by fitting to the same data that the ENM is trying to predict, reducing the predictive power of the model.

Results: We develop an elastic network model for the whole protein crystal in order to study the influence of crystal packing and lattice vibrations on the thermal fluctuations of the atom positions. We use experimental values for the compressibility of the crystal to establish the energy scale of our model. We predict the elastic constants of the crystal and compare with experimental data. Our main findings are (1) crystal packing modifies the atomic fluctuations considerably and (2) thermal fluctuations are not the dominant contribution to crystallographic Debye-Waller factors.

Availability: The programs developed for this work are available as supplementary material at Bioinformatics Online.

Contact: hinsen@cnrs-orleans.fr

Supplementary information: (1) A full derivation of the normal mode equations in a crystal and in a continuous medium. (2) A movie illustrating the lattice vibrations. Both supplements are available at *Bioinformatics* Online.

1 INTRODUCTION

Protein X-ray crystallography provides not only three-dimensional structures of proteins, but also information about their flexibility in the form of Debye-Waller factors (DWFs). The DWF not only describes the spread of the electron density of an atom due to thermal motion and crystal disorder, but also contains contributions that are due to experimental artefact. For most protein structures in the Protein Data Bank (Berman *et al.*, 2000), the DWFs have been assumed to be isotropic in the refinement process, leading to a single number per atom known as the B factor. More recent structures, in particular high-resolution structures, have been refined using anisotropic DWFs, which are described by a symmetric tensor per atom, whose six independent elements are called anisotropic displacement parameters (ADPs).

It is important to understand that B factors and ADPs are not experimentally observable quantities, but parameters in a theoretical model that is fitted to the experimental diffraction intensities. The refinement procedure typically involves restraints on the DWFs, whose impact can be detected in the fitted ADPs (Kondrashov *et al.*, 2007). Some structures were refined using a theoretical model for collective motions from which the ADPs were derived, resulting in a much smaller number of independent fit parameters. The most popular model for collective motions is the TLS model (Schomaker and Trueblood, 1968; Winn *et al.*, 2001) that describes the protein as an assembly of rigid subunits. Another approach is the use of low-frequency normal mode coordinates to describe the large-amplitude motions of the protein (Diamond, 1990; Kidera and Go, 1990; Poon *et al.*, 2007).

The interpretation of DWFs requires theoretical models as well, because the individual atomic position fluctuations are of little biological interest. It is well known that the largest contributions to thermal fluctuations in macromolecules come from collective motions, which are conveniently described by the normal modes of harmonic potential models or by principal component analysis of molecular dynamics trajectories. Some aspects of static crystal disorder, in particular the co-existence of multiple conformers of the protein in the crystal, are also well described by collective motions. Moreover, collective motions have been demonstrated to be related to the biological function of the protein in a large number of cases (Bahar and Rader, 2005; Ma, 2005). The main utility of crystallographic DWFs in the study of protein flexibility lies thus in their use for calibrating and verifying theoretical models for large-amplitude collective motions.

A family of models for collective motions that has met with considerable success is the elastic network model (ENM), which has first been proposed by Tirion (1996) and then been applied to the identification of collective motions by Hinsen (1998). Since then, a large number of applications has demonstrated its utility (Bahar and Rader, 2005). In its most commonly used form, the ENM consists of point masses located at the positions of the C_α atoms of the protein and interacting through harmonic springs. The potential energy takes the form

$$U(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{\text{all pairs } i,j} U_{ij}(\mathbf{r}_i - \mathbf{r}_j) \qquad (1)$$

with an harmonic pair potential

$$U_{ij}(\mathbf{r}) = k\left(\left|\mathbf{r}_{ij}^{(0)}\right|\right)\left(|\mathbf{r}| - \left|\mathbf{r}_{ij}^{(0)}\right|\right)^2, \quad (2)$$

where $\mathbf{r}_i$ denotes the position of the ith C_α atom and $\mathbf{r}_{ij}^{(0)}$ is the pair distance vector $\mathbf{r}_i - \mathbf{r}_j$ in the given configuration, which becomes by construction the minimum of the potential energy. The function $k(r)$ represents the force constant of the springs as a function of the pair distance at equilibrium. Various functional forms have been used: a Gaussian (Hinsen, 1998), a function fitted to an atomic-level force field (Hinsen *et al*., 2000), and a step function (Atilgan *et al*., 2001). In all these cases, the potential is only determined up to a global scale factor, which is usually obtained by fitting to experimental DWFs. Many studies have compared the isotropic B factors predicted by ENMs to experimental values, usually finding reasonable but not excellent agreement, and two recent studies have extended the comparison to anisotropic displacement parameters (Eyal *et al*., 2007; Kondrashov *et al*., 2007). There have also been validation approaches using molecular dynamics trajectories (Rueda *et al*., 2007) and NMR structures (Yang *et al*., 2007).

2 APPROACH

Until now, ENMs have always been applied to single proteins, modelling the biological reality of proteins in solution. However, this environment is very different from the one of a protein crystal. A crystal is a large assembly of densely packed proteins, meaning that each molecule is in close contact with several other molecules. These contacts considerably reduce the configuration space accessible to each protein, thus modifying its flexibility. Moreover, the crystal as a whole exhibits collective motions that contribute to the thermal fluctuations.

We apply an ENM to a protein crystal (the tetragonal crystal form of lysozyme) in order to study the influence of crystal contacts and of lattice dynamics on the atomic fluctuations in proteins. We also consider the large wavelength limit of the lattice modes in which the protein crystal can be described as an elastic material. We derive the elastic constants of the crystal and compare to recent experimental measurements (Fourme *et al*., 2001; Koizumi *et al*., 2006; Speziale *et al*., 2003). This comparison allows us to establish an absolute energy scale for the ENM using experimental data independent from the DWFs themselves, and thus to predict the thermal contribution to atomic fluctuations in proteins on an absolute scale as well.

3 METHODS

3.1 Elastic network model

We use the ENM proposed by Hinsen *et al*. (2000), which has the form given by Eqs. (1) and (2) with

$$k(r) = \begin{cases} 8.6 \cdot 10^5 \frac{\text{kJ}}{\text{mol nm}^3} \cdot r - 2.39 \cdot 10^5 \frac{\text{kJ}}{\text{mol nm}^2} & \text{for } r < 0.4 \text{ nm} \\ \frac{128 \text{ kJ nm}^4/\text{mol}}{r^6} & \text{for } r \geq 0.4 \text{ nm.} \end{cases} \quad (3)$$

The form for $r < 0.4$ applies to nearest neighbours along the peptide chain, taking into account the rigidity of the peptide plane that links the two C_α atoms. This choice for $k(r)$ was found to best reproduce the normal modes of the all-atom Amber 94 force field (Cornell *et al*., 1995). We apply a cut-off radius of 2.5 nm, at which the potential is negligibly small. The absolute force constant values correspond to a local energy minimum in the Amber 94 force field. For applications of ENMs to large-amplitude motions, the interactions are generally scaled by a global factor that is determined by fitting to experimental data.

For the calculations on protein crystals, we apply the ENM to the unit cell of the crystal, constructed from the lattice parameters and crystallographic symmetry operations contained in the PDB file. We implement periodic boundary conditions using the minimum-image convention, in the same way as is habitually done in molecular dynamics simulations. The atomic fluctuations are calculated from the normal modes of the ENM, excluding the modes describing the global motions of the system (for a crystal, the three translational modes). All calculations are performed using the Molecular Modelling Toolkit (Hinsen, 2000).

3.2 Normal modes of a crystal

The calculation of the normal modes of a crystal is treated in most textbooks on crystal theory (Born and Huang, 1998). However, these derivations do not include the calculation of the amplitudes of thermal motion, which are important for our application. We give a complete derivation in the supplementary material and summarize the results here.

Given a crystal with lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, we denote the equilibrium positions of the atoms by $\mathbf{r}_k^{(j)}$, where $k = 1, \ldots, N$ enumerates the atoms in the unit cell and $j = (j_1, j_2, j_3)$ with integer j_1, j_2, j_3 indicates a specific image of the unit cell. We consider finite-size crystals that consist of n_k copies along lattice vector $\mathbf{a}_k$, with periodic boundary conditions at the surface; ultimately we will extrapolate $n_k \to \infty$. We thus have $0 \leq j_k < n_k$ for $k = 1, 2, 3$. Normal mode analysis assumes small-amplitude vibrations of the atoms around their equilibrium positions; we denote the displacement vectors by $\delta\mathbf{r}_k^{(j)}$.

A complete set of normal modes is given by standing-wave solutions of the equations of motion, which have the form

$$\delta\mathbf{r}_i^{(j)}(\mathbf{q}, t) = \mathbf{u}_i \exp(i\mathbf{q} \cdot \mathbf{r}_i^{(j)}) \cos(\omega t). \quad (4)$$

We will ultimately use the real and imaginary parts of $\delta\mathbf{r}_k^{(j)}$ for the physical displacements. For a finite crystal with periodic boundary conditions, the wave vectors $\mathbf{q}$ must be of the form $\mathbf{q} = 2\pi(q_1\mathbf{b}_1 + q_2\mathbf{b}_2 + q_3\mathbf{b}_3)$, where $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are the reciprocal lattice vectors defined through $\mathbf{a}_i \cdot \mathbf{b}_j = \delta_{ij}$ and the coefficients q_k are integer quotients l_k/n_k with $-n_k/2 < l_k \leq n_k/2$. This yields a set of n_c wave vectors, where $n_c = n_1 n_2 n_3$ is the number of unit cells in the crystal. These wave vectors lie on a uniform grid in reciprocal space.

For each wave vector $\mathbf{q}$ in this set, we obtain $3N$ normal mode sets $(\mathbf{u}, \omega)$ from the equations

$$m_i\omega^2\mathbf{u}_i = \sum_{l,m} \mathbf{K}_{(ij),(lm)} \cdot \mathbf{u}_l \exp\left[i\mathbf{q} \cdot (\mathbf{r}_l^{(m)} - \mathbf{r}_i^{(j)})\right], \quad (5)$$

where $\mathbf{K}$ is the Hessian matrix of the second derivatives of the potential energy and m_i is the mass of particle i. The three lowest-frequency modes for each wave vector are called 'acoustic modes' and describe the collective motions of the atoms in the unit cell relative to the other unit cells. In the remaining $3N-3$ modes, called 'optical modes', the atoms of a unit cell move relative to each other. The acoustic modes for $\mathbf{q} = 0$ have zero frequency and describe the translational degrees of freedom of the crystal. The optical modes for $\mathbf{q} = 0$ are the normal modes of a single unit cell with periodic boundary conditions. The amplitudes of the thermal fluctuations are obtained by requiring that the average kinetic energy of each mode is $k_B T/2$.

3.3 Elastic medium approximation

In elasticity theory, the deformation of an elastic body is described by a deformation field $\mathbf{u}(\mathbf{r})$. Its derivative, the (symmetric) strain tensor

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i}\right),$$

determines the (symmetric) stress tensor σ_{ij} through Hooke's law,

$$\sigma_{ij} = \sum_{kl} \lambda_{ij,kl} e_{kl}, \tag{6}$$

where $\lambda_{ij,kl}$ is a constant tensor that has 21 independent elements in its most general form. This number is further reduced by symmetries present in the elastic medium.

The derivation of the wave equations in an elastic medium can be found in the literature and is reproduced in the supplementary material together with a derivation of the amplitudes of thermal fluctuations. Like the atomic-scale normal mode calculation, it is based on an *ansatz* of standing wave solutions to the equations of motions for the displacement field,

$$\mathbf{u}(\mathbf{r}, t) = \mathbf{a} \exp(i\mathbf{q} \cdot \mathbf{r}) \cos(\omega t), \tag{7}$$

which leads to the wave equations

$$\rho\omega^2 \mathbf{a} = \mathbf{M} \cdot \mathbf{a}. \tag{8}$$

where $\mathbf{M}$ is a symmetric 3×3 matrix that depends on the $\lambda_{ij,kl}$ and on $\mathbf{q}$ and ρ is the mass density of the material. There are three normal mode sets $(\mathbf{a}, \omega)$ for every $\mathbf{q}$.

3.4 Elastic medium limit of the atomic model

In the limit of large wavelengths, i.e. small $\mathbf{q}$, the three acoustic modes of the atomic-scale model correspond to the elastic waves described by Eq. (8). The displacement field can be identified with a collective motion coordinate of the unit cell, such as the center of mass. With that choice, we can identify the matrix $\mathbf{M}$ from Eq. (8) with

$$\mathbf{M}(\mathbf{q}) = \frac{N}{V} \sum_{\lambda=1,2,3} \omega_\lambda^2 \hat{\mathbf{u}}_\lambda \otimes \hat{\mathbf{u}}_\lambda \tag{9}$$

with

$$\hat{\mathbf{u}}_\lambda = \frac{\sum_i m_i \mathbf{u}_i^{(\lambda)}}{\sqrt{\sum_i m_i^2 \left|\mathbf{u}_i^{(\lambda)}\right|^2}}, \tag{10}$$

where $\mathbf{u}_i^{(\lambda)}$ for $\lambda = 1, 2, 3$ are the three lowest-frequency modes obtained from Equation (5), N is the number of atoms in the unit cell, and V is the volume of the unit cell. Calculating $\mathbf{M}(\mathbf{q})$ for several small $\mathbf{q}$ vectors makes it possible to obtain all elastic constants by comparing with the solutions of Equation (8).

3.5 Compressibility of an elastic medium

An elastic medium subject to an isotropic pressure p is characterized by the stress tensor $\sigma_{ij} = -p\,\delta_{ij}$. The displacement field induced by the pressure is described by a strain tensor e_{ij}, which is determined by Equation (6). The volume change due to this strain field is

$$\frac{V + \Delta V}{V} = \sum_i e_{ii}, \tag{11}$$

the compressibility is then given by

$$\beta = -\frac{1}{V}\frac{\partial V}{\partial p} = -\frac{1}{p}\sum_i e_{ii}. \tag{12}$$

An explicit derivation of β for a tetragonal crystal is given in the supplementary material.

4 RESULTS AND DISCUSSION

We have chosen the tetragonal crystal form of hen egg-white lysozyme as the subject for this study, because of the large number of available data for this system. In particular, it is, at the moment, the only protein crystal for which measurements of the elastic constants are available (Koizumi *et al.*, 2006; Speziale *et al.*, 2003). We used two different crystal structures for our calculations: PDB code 1IEE (Sauter *et al.*, 2001), which is the highest-resolution structure available and contains ADPs, and PDB code 2LYM (Kundrot and Richards, 1987), for which the compressibility has been determined experimentally. The first structure has been obtained at a temperature of 110 K, the second one at ambient temperature. The two structures thus permit an estimation of the relative contributions of thermal motion and crystal disorder to the crystallographic DWFs. The RMS difference between the two lysozyme structures is 0.33 Å. The 2LYM unit cell volume is larger than the one of 1IEE by 7.6%, which can be explained by the temperature difference (Kurinov and Harrison, 1995).

In order to clarify the impact of different crystal contacts on the atomic fluctuations, we also look at the structure of the triclinic form of lyzosyme (Walsh *et al.*, 1998, PDB code 3LZT), obtained at a temperature of 120 K. Its structure remains close to that of tetragonal lysozyme (the RMS difference from 1IEE is 0.79 Å, from 2LYM it is 0.72 Å).

Figure 1 compares the magnitude of the atomic fluctuations of the C_α atoms in the three structures 1IEE, 2LYM, and 3LZT. The atomic fluctuation magnitude $<\Delta R^2>$ is related to the crystallographic B factor by $B = 8\pi^2/3 <\Delta R^2>$. The figure shows clearly that the atomic fluctuations are of very similar magnitude in the two structures for the tetragonal form, in spite of the significant temperature difference. Since the thermal fluctuation contribution must decrease with temperature, the contribution of static disorder must be higher in the low-temperature crystal. This effect is considered to be a

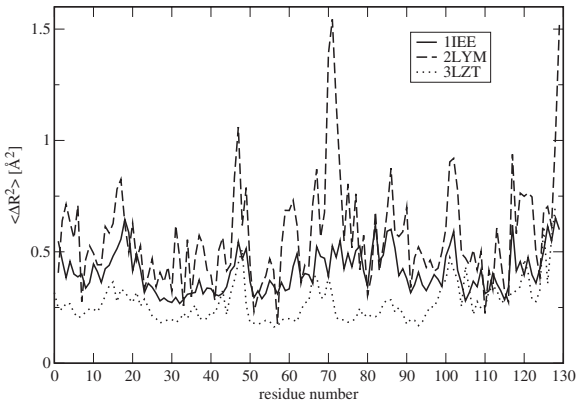

Fig. 1. The magnitude of atomic fluctuations in two different structures for tetragonal lysozyme. Structure 1IEE was recorded at 110 K, structure 2LYM at ambient temperature. The small difference between the two sets of fluctuations indicates a significant contribution from static disorder at least in structure 1IEE.

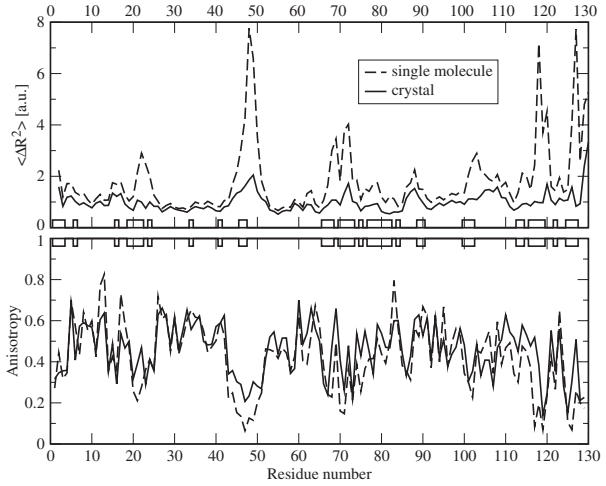

Fig. 2. The influence of crystal packing on the atomic fluctuations for triclinic lysozyme (PDB code 3LZT). The boxes on the x-axis indicate the residues that are in close contact with other molecules in the crystal. Both the magnitude of the atomic fluctuations (*top*) and their anisotropy (*bottom*) are modified by crystal contacts. In the absence of experimental information that permits the establishment of an absolute energy scale for the elastic network model, we have chosen arbitrary units for the magnitudes, defined such that the average magnitude in the crystal is 1.

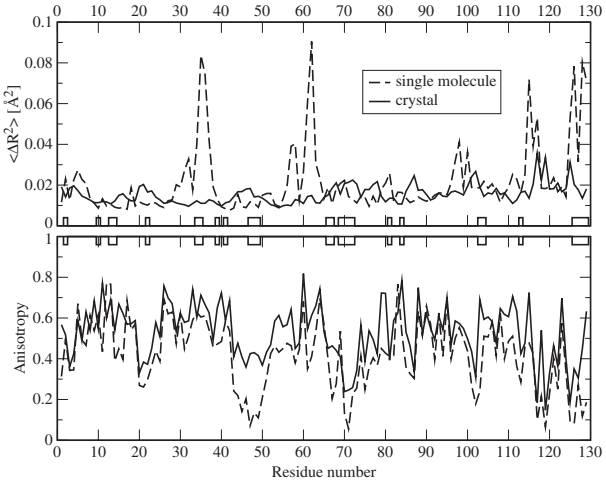

Fig. 3. The influence of crystal packing on the atomic fluctuations for tetragonal lysozyme (PDB code 1IEE). The boxes on the x-axis indicate the residues that are in close contact with other molecules in the crystal. Both the magnitude of the atomic fluctuations (*top*) and their anisotropy (*bottom*) are modified by crystal contacts. For this crystal, the absolute energy scale (and thus the scale for the fluctuations) has been derived from the experimental compressibility, as explained in Section 4.2.

consequence of shock freezing, which freezes dynamic disorder (Kurinov and Harrison, 1995). Comparing 1IEE to 3LZT, we see a more important difference, although the temperatures are almost the same.

4.1 Influence of crystal packing

To study the influence of crystal packing on the atomic fluctuations, we calculate them from (a) an ENM for a single lysozyme molecule and (b) an ENM for the unit cell of the crystal, using periodic boundary conditions. We first look at triclinic lysozyme (PDB code 3LZT), because it has only one lysozyme molecule in the unit cell. Any differences in the atomic fluctuations are thus due to crystal contacts. The comparison is shown in Figure 2. In the top graph, we see that the magnitude of the atomic fluctuations is significantly altered by the crystal contacts. In particular, we note that the residues that exhibit the largest fluctuations in the single molecule (residues 22, 48, 72, 103, 118, and their surroundings) have much smaller fluctuations in the crystal. However, these are not all residues that have crystal contacts themselves (the crystal contacts, defined as residues whose C_α atoms are at <7 Å from any C_α atom in another chain, are indicated in the figure by boxes on the x-axis). A closer look at the normal modes of the ENM shows that the large fluctuations stem from the first three non-zero modes that describe the domain motions in lysozyme. These motions are hindered by the crystal contacts, leading to a reduction of atomic fluctuations even in regions that are not directly affected by crystal contacts themselves.

It should be noted that the magnitudes of all atomic fluctuations are slightly smaller in the crystal than in the single molecule. This is in fact to be expected, given that the crystal contacts add interactions to those already present in the single molecule. Any deformation of the protein thus entails a higher energetic cost in the crystal, leading to a reduction of all atomic fluctuations as long as the temperature remains constant.

The anisotropy of the atomic fluctuations is defined as the ratio of the smallest eigenvalue of the ADP tensor to the largest one. A value of 1 thus indicates an isotropic distribution. The bottom graph of Figure 2 shows the anisotropies for the single molecule and the crystal ENM. The anisotropies are in general very similar, deviating most for those residues that also show the biggest changes in magnitude.

Figure 3 shows the same comparison for tetragonal lysozyme (PDB code 1IEE). As for the triclinic crystal, the strong attenuation of the peaks in the magnitude plot (residues 35, 62, 115, 128 and their surroundings) by the crystal contacts is the most striking feature. Again we note that the residues that are most affected do not necessarily have crystal contacts themselves. An important difference between the two crystals is that some of the atomic fluctuations are larger in the crystal than in the single molecule, and that the fluctuations become more isotropic. This is due to the fact that the unit cell of the tetragonal crystal contains eight lysozyme molecules. The motions of these molecules relative to each other contribute to the atomic fluctuations.

4.2 Lattice vibrations and elastic constants

In the last section, we have looked at the atomic fluctuations inside a unit cell of a protein crystal, taking into account interactions with the neighbouring unit cells, but neglecting the relative motions of molecules in different unit cells. These motions are of a far more collective nature than the collective

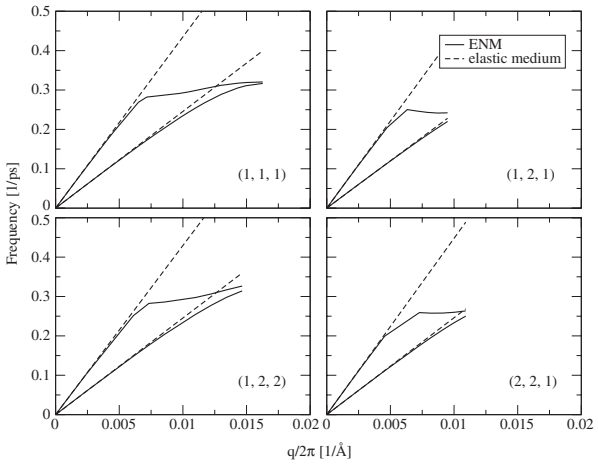

Fig. 4. The vibrational frequency $\omega/2\pi$ of the acoustic modes of the 1IEE crystal as a function of the wave vector length. Both the exact normal modes of the elastic model (full lines) and the modes of a continuum approximation (dashed lines) are shown. The high-frequency branch corresponds to the longitudinal mode and the low-frequency branch to the two degenerate transverse modes. Each graph shows the modes for a specific direction of **q** indicated by the (q_1, q_2, q_3) values.

modes of a protein; their amplitudes of thermal motion are thus very large, and they can be expected to contribute substantially to the atomic fluctuations.

We have calculated the atomic fluctuations for tetragonal lysozyme (PDB code 1IEE) for several crystal sizes using the procedure described in Section 3.2. The optical modes depend very little on **q**, both the frequencies and the motion patterns are very similar to the normal modes of the unit cell (see Section 3 of the supplementary material). It is the acoustic modes that add qualitatively new features to the dynamics of the protein crystal. A movie showing acoustic modes in tetragonal lysozyme crystals is provided as supplementary material.

Figure 4 shows the vibrational frequencies of the acoustic modes of a 1IEE crystal as a function of the length of the wave vectors **q** for four directions of **q**, both for the ENM and for a continuum approximation (Sections 3.3 and 3.4). For each **q** there are two frequencies, the lower one corresponding to the two transverse modes, in which the particle displacements are perpendicular to **q**. These two modes are degenerate due to the symmetry of the crystal. The higher frequency corresponds to the longitudinal mode, in which the particles move along the direction of **q**. A longitudinal mode implies density fluctuations, which in a solid have a higher energetic cost than the shear deformations that characterize transverse modes. In the continuum approximation, the frequency is a linear function of $|\mathbf{q}|$ and depends only weakly on the direction of **q**. For small wave vectors, corresponding to length scales of 200 Å and higher, this is a very good approximation to the ENM frequencies. For larger **q** (shorter length scales), the atomic-scale structure of the crystal leads to marked deviations from the continuum behaviour, in particular for the longitudinal modes.

The frequency scale in Figure 4 is directly related to the energy scale of the ENM. In this work, we use the compressibility of the 2LYM crystal, obtained by comparing to the structure of the same crystal at a pressure of 1000 atm (Kundrot and Richards, 1987, PDB code 3LYM). Its value of 0.105 GPa^{-1} is close to the more recent value of 0.098 GPa^{-1} (Fourme *et al.*, 2001). We use the same compressibility value for the 1IEE crystal, neglecting a probable temperature dependence. The scale factors that need to be applied to the force constants of the ENM in Equation (3) are 2.99 for 1IEE and 3.89 for 2LYM, the difference being due to the different size of the unit cell.

Table 1. The compressibility β and the elastic constants C_{ij} for tetragonal lysozyme

		ENM 1IEE	ENM 2LYM	Dehydrated crystal	Hydrated crystal
β	(GPa^{-1})	0.109	0.109	0.105	0.238
$C_{11}=C_{22}$	(GPa)	16.38	16.20	12.44	5.52
C_{12}	(GPa)	5.99	5.95	7.03	3.12
$C_{13}=C_{23}$	(GPa)	5.87	5.69	8.36	3.71
C_{33}	(GPa)	14.74	15.54	12.79	5.68
$C_{44}=C_{55}$	(GPa)	5.76	5.84	2.97	1.32
C_{66}	(GPa)	6.54	7.08	2.63	1.16

Calculated from elastic network models for structures 1IEE and 2LYM, and obtained experimentally through sound velocity measurements. For the elastic network models, the compressibility is the one obtained crystallographically; this value was used to define the absolute energy scale of the model.

The use of an experimental reference value that is independent of the atomic fluctuations allows us to make a quantitative prediction of the contribution of thermal motions to the experimentally observed DWFs. In fact, the fluctuation scale in Figure 3 is also derived from the scale factor fitted to reproduce the compressibility. In contrast, the fluctuations in Figure 2 are given in arbitrary units, because no compressibility value is available for the triclinic lysozyme crystal. We will discuss the amplitudes of thermal motions later, after adding the contributions of lattice vibrations.

The elastic constants C_{ij} obtained from the ENMs for 1IEE and 2LYM are given in Table 1, together with experimental values for tetragonal lysozyme obtained by measuring sound velocities (Koizumi *et al.*, 2006). While the agreement is not perfect, we find the same order of magnitude in our models as in the dehydrated crystal, which has a similar compressibility. The simplicity of our model is certainly one reason for the discrepancies, but there are other factors as well. For example, the calculation of the crystal compressibility from high-pressure structures is based on the assumption that no solvent enters or leaves the crystal when it is put under pressure. This assumption is likely to be wrong (Kundrot and Richards, 1988). Moreover, the description of a protein crystal as an elastic medium obeying Hooke's law, which underlies both the experimental and the computational determination of the elastic constants, is an approximation whose validity is not yet known.

4.3 Contribution of lattice vibrations to the atomic fluctuations

We have seen in the Section 3.2 that the normal modes of a crystal can be grouped into optical modes and acoustic modes.

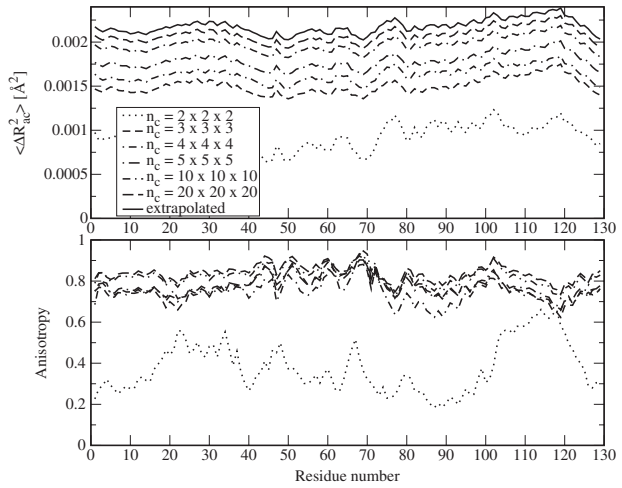

Fig. 5. *Top*: The contribution of the acoustic modes to the magnitudes of the atomic fluctuations in a 1IEE crystal, for different crystal sizes, and the extrapolation to an infinite crystal. *Bottom*: The anisotropy of the acoustic mode contribution.

The optical modes contain all of the motions that are of biological interest. However, the acoustic modes are the lowest-frequency and thus largest-amplitude motions in the crystal. Their contribution to the atomic fluctuations can therefore not be neglected.

Although the optical modes for different **q** are not exactly identical, it turns out that their contributions to the atomic fluctuations are very similar, to the point that on a plot showing $<\Delta R^2_{opt}>(\mathbf{q})$ for various **q** the curves would be nearly indistinguishable. The total contribution of all optical modes to the atomic fluctuations is thus independent of the size of the crystal and equal to the drawn-out curve in Figure 3.

The contribution of the acoustic modes to the atomic fluctuations requires a more careful analysis because the acoustic modes depend strongly on the size of the crystal. At this time, it is not possible to calculate the lattice modes for realistic crystal sizes ($\approx 10^{15}$ copies of the unit cell). However, given the atomic fluctuations for several small crystal sizes, it is possible to extrapolate to infinite crystal size. A description of this extrapolation is provided in the supplementary material.

Figure 5 shows the magnitude of the acoustic mode contribution to the atomic fluctuations for different sizes of a 1IEE crystal, as well as the extrapolation to infinite crystal size. It is much more homogeneous than the contribution of the optical modes (Fig. 3), and also more isotropic. The magnitude of the acoustic contribution is smaller, its average over all residues represents 14% of the average of the optical contribution.

4.4 Comparison to crystallographic Debye-Waller factors

Figure 6 shows the crystallographic isotropic DWFs in comparison to the thermal fluctuation amplitude predicted by the ENM for protein crystals, taking both optical and acoustic modes into account. It is evident that the theoretical amplitudes are much smaller: their average over all residues makes up for

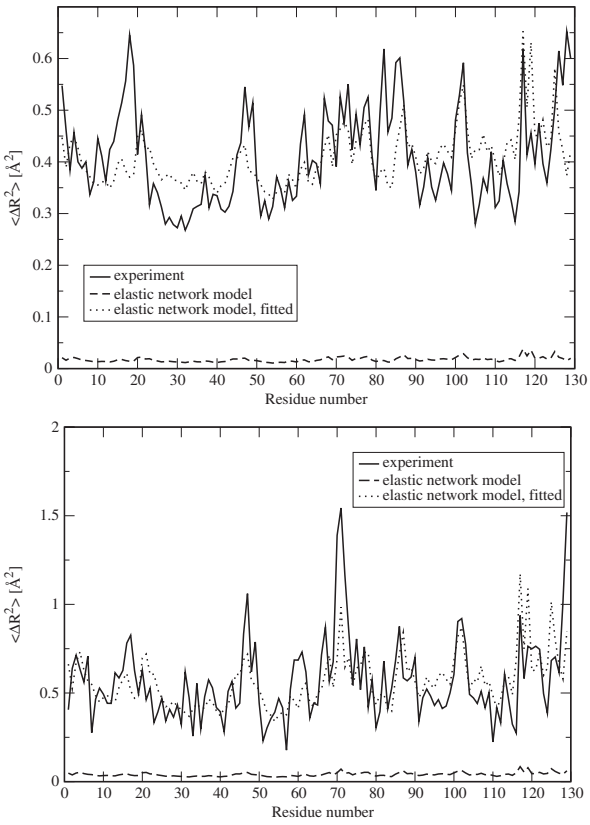

Fig. 6. A comparison between the crystallographic isotropic DWFs and the predictions of the elastic network model for protein crystals for 1IEE (*top*) and 2LYM (*bottom*).

4.3% (1IEE) or 7.2% (2LYM) of the average B factor. Considering that the 2LYM structure was obtained at a higher temperature, it should be expected that thermal fluctuations represent a higher percentage than for 1IEE. The difference between the two percentages should probably be higher, considering that the 1IEE crystal at 110 K is likely to have a lower compressibility, leading to a larger energy scale factor and a smaller thermal fluctuation amplitude.

Although our results indicate that thermal fluctuations are only a minor contribution to the experimentallly observed DWFs, a large number of studies have shown that ENMs are successful at predicting the magnitude and also the shape of crystallographic DWFs if the global energy scale factor is suitably adjusted. This suggests that static crystal disorder, the main non-thermal contribution to DWFs, is also well described by these models. We have therefore added a third curve to Figure 6, representing a fit of our ENM results to the experimental data. The model that we have fitted is $B_{exp} = s_{optical} B_{optical} + s_{acoustic} B_{acoustic}$, i.e. we use different scale factors for the contributions of optical and acoustic modes. The idea behind this model is that optical and acoustic modes could describe different kinds of crystal disorder. The fit parameters for 1IEE are $s_{optical} = 11.0$ and $s_{acoustic} = 115$, the values for 2LYM are $s_{optical} = 13.9$ and $s_{acoustic} = 13.5$. We thus find that the acoustic contribution seems much more important for 1IEE than for 2LYM.

5 CONCLUSION

Our first main conclusion is that the influence of crystal packing on the atomic fluctuations in a protein is very important. Its physical cause is easy to understand. Moreover, taking into account crystal contacts hardly increases the complexity of the model. We therefore consider that all ENM calculations performed with the aim of comparing to crystallographic data should use periodic boundary conditions in order to model the crystal environment. Kundu *et al*. (2002) reach a similar conclusion for the Gaussian Network Model.

There remains the question of how the interactions between the molecules in the crystal should be included in the model. In this work, we have applied an ENM designed for single proteins without modification to a protein crystal. This assumes that the average residue–residue interactions across crystal contacts are of the same strength as those that stabilize protein folds and multimeric proteins. That may not be the case, as there is no evolutionary pressure on proteins to stabilize the unnatural crystal state.

Another aspect of protein crystal dynamics that requires further studies is the role of the solvent. In ENM studies on single proteins, solvent is usually ignored because it has no impact on the free energy of the protein as long as only small fluctuations are considered. In crystals, however, the solvent is partially trapped. Fast compression, as in the propagation of sound waves, involves compression of the solvent as well. Slower deformations of the crystal, such as the ones it undergoes in high-pressure studies, can lead to an increase or decrease of the solvent content of the crystal. It may be necessary to account for such effects in future protein crystal models.

For our study, the most direct impact of the solvent-related uncertainties comes from the experimental compressibilities which we have chosen as the reference data for establishing the absolute energy scale of our ENM. The published values for the compressibility of a tetragonal lysozyme crystal at room temperature vary from 0.098 (Fourme *et al*., 2001) to 0.238 GPa^{-1} (Koizumi *et al*., 2006, extrapolated data), i.e. by a factor of 2.4. The atomic fluctuations obtained from our ENM are directly proportional to the compressibility used for calibration. Choosing the highest published value would increase our prediction of the thermal contribution to the DWFs to 9.7% for 1IEE and 16.3% for 2LYM.

The question of the magnitude of the various contributions to the DWFs (thermal motion, static disorder, lattice defects, experimental setup) is a particular important one, as it is fundamental to the use of DWFs in understanding the structural flexibility of proteins. Our results suggest that thermal motion is a small contribution, in particular for low-temperature structures. molecular dynamics simulations (Burden and Oakley, 2007) also yield thermal fluctuations that are much smaller than crystallographic DWFs. On the other hand, Kurinov and Harrison (1995) estimate by comparison with Mössbauer spectroscopy that thermal fluctuations in tetragonal lysozyme make up for 50% of the DWFs at 300 K.

The validation of theoretical models for protein crystals relies on the availability of experimental data on their mechanical properties. At the moment, such data is scarce, limiting the validation of our model to tetragonal lysozyme crystals. We hope that this study will stimulate more experimental work in this field.

Finally, the success of ENMs in explaining the patterns of crystallographic B factors, in spite of the importance of contributions that are not related to thermal fluctuations, suggests that all these models capture some aspects of static disorder and lattice defects. Improving our understanding of these aspects is likely to lead to important advances in the modelling of structural flexibility in proteins.

ACKNOWLEDGEMENTS

The author would like to thank Pierre Legrand and Roger Fourme for helpful discussions about crystallographic structure refinement and protein crystals under pressure.

Conflict of Interest: none declared.

REFERENCES

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

Bahar,I. and Rader,A. (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.

Born,M. and Huang,K. (1998) *Dynamical Theory of Crystal Lattices*. Clarendon Press, Oxford.

Burden,C. and Oakley,A. (2007) Anisotropic atomic motions in high-resolution protein crystallography molecular dynamics simulations. *Phys. Biol*., **4**, 79–90.

Cornell,W. *et al*. (1995) A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc*, **117**, 5179.

Diamond,R. (1990) On the use of normal modes in thermal parameter refinement: theory and application to the bovine tripsin inhibitor. *Acta Cryst. A*, **46**, 425–435.

Eyal,E. *et al*. (2007) Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. *Bioinformatics*, **23**, i175–i184.

Fourme,R. *et al*. (2001) High-pressure protein crystallography (hppx): instrumentation, methodology and results on lysozyme crystals. *J. Synchrotron Rad*., **8**, 1149–1156.

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

Hinsen,K. (2000) The Molecular Modeling Toolkit: a new approach to molecular simulations. *J. Comp. Chem*., **21**, 79–85. http://dirac.cnrs-orleans.fr/MMTK/

Hinsen,K. *et al*. (2000) Harmonicity in slow protein dynamics. *Chem. Phys*., **261**, 25–37.

Kidera,A. and Go,N. (1990) Refinement of protein dynamic structure: normal mode refinement. *PNAS*, **87**, 3718–3722.

Koizumi,H. *et al*. (2006) Observation of all the components of elastic constants using tetragonal hen egg-white lysozyme crystals dehydrated at 42% relative humidity. *Phys. Rev. E*, **73**, 041910.

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

Kundrot,C. and Richards,F. (1987) Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *J. Mol. Biol*., **193**, 157–170.

Kundrot,C. and Richards,F. (1988) Effect of hydrostatic pressure on the solvent in crystals of hen egg-white lysozyme. *J. Mol. Biol*., **200**, 401–410.

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

Kurinov,I.V. and Harrison,R.W. (1995) The influence of temperature on lysozyme crystals. structure and dynamics of protein and water. *Acta Cryst. D*, **51**, 98–109.

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

Poon,B. *et al*. (2007) Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42 Å crystallographic resolution. *PNAS*, **104**, 7869–7874.

Rueda,M. *et al*. (2007) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure*, **15**, 565–575.

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. D*, **57**, 1119–1126.

Schomaker,V. and Trueblood,K. (1968) On the rigid-body motion of molecules in crystals. *Acta Cryst. B*, **24**, 63–76.

Speziale,S. *et al*. (2003) Sound velocity and elasticity of tetragonal lysozyme crystals by brillouin spectroscopy. *Biophys. J*., **85**, 3202–3213.

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

Walsh,M.A. *et al*. (1998) Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Cryst. D*, **54**, 522–546.

Winn,M. *et al*. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst. D*, **57**, 122–133.

Yang,L.-W. *et al*. (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure*, **15**, 741–749.