J Comput Aided Mol Des
DOI 10.1007/s10822-010-9410-0

Applying forces to elastic network models of large biomolecules using a haptic feedback device

M. B. Stocks · S. D. Laycock · S. Hayward

Received: 24 September 2010 / Accepted: 30 December 2010

Abstract Elastic network models of biomolecules have proved to be relatively good at predicting global conformational changes particularly in large systems. Software that facilitates rapid and intuitive exploration of conformational change in elastic network models of large biomolecules in response to externally applied forces would therefore be of considerable use, particularly if the forces mimic those that arise in the interaction with a functional ligand. We have developed software that enables a user to apply forces to individual atoms of an elastic network model of a biomolecule through a haptic feedback device or a mouse. With a haptic feedback device the user feels the response to the applied force whilst seeing the biomolecule deform on the screen. Prior to the interactive session normal mode analysis is performed, or pre-calculated normal mode eigenvalues and eigenvectors are loaded. For large molecules this allows the memory and number of calculations to be reduced by employing the idea of the important subspace, a relatively small space of the first M lowest frequency normal mode eigenvectors within which a large proportion of the total fluctuation occurs. Using this approach it was possible to study GroEL on a standard PC as even though only 2.3% of the total number of eigenvectors could be used, they accounted for 50% of the total fluctuation. User testing has shown that the haptic version allows for much more rapid and intuitive exploration of the molecule than the mouse version.

Keywords Force feedback · Protein flexibility · Normal mode analysis

M. B. Stocks · S. D. Laycock (✉) · S. Hayward
School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK
e-mail: sdl@cmp.uea.ac.uk

Introduction

Biomolecules undergo conformational change in response to the binding of their natural ligands and one of the aims of biomolecular modelling is to understand and predict these functional motions. One simple, but successful model employed for this task is the Elastic Network Model (ENM) introduced into protein research by Tirion [1]. This is a coarse-grained model where a subset of atoms are connected to each other with elastic bonds. The dynamical behaviour of an ENM can be determined by normal mode analysis. Originally, normal mode analysis of biomolecules used the same accurate force-fields that are employed for Molecular Dynamics (MD) simulation and required considerable computational effort, but ENM normal mode analysis has largely taken over due to its relative ease of implementation, its lighter computational load, and due to the fact that results are comparable to results of accurate force-field normal mode analysis [1]. The lighter computational load occurs because usually a subset of atoms is used (for proteins normally the C_α atoms, for DNA and RNA the P atoms) and energy minimization is not carried out. As a consequence ENM normal mode analysis has been performed on a large number of biomolecules some of which are of considerable size [2–7]. Given the assumptions that are inherent in normal mode analysis, the small difference between the results from these two methods is probably of little consequence. Indeed, comparisons between movements in the low frequency modes of ENMs with functional movements derived from pairs of X-ray structures [8] suggest the same level of correspondence seen in similar studies using accurate force-field normal mode analysis. The binding of a ligand to a biomolecule gives rise to specific forces that induce the conformational change [9–11], and a pertinent question is whether these

Published online: 15 January 2011

forces applied to a harmonic model of the biomolecule would produce a movement that corresponds well to the experimentally determined functional movement. Ikeguchi et al. [10] have already performed such an analysis on harmonic models derived from MD simulations using quasi-harmonic analysis and also to an ENM. Their results are very encouraging. For an ENM of F1-ATPase the movement in the β_E subunit correlated well (correlation coefficient of 0.84) with the experimentally determined functional movement. External forces have also been applied to ENMs of proteins to determine residues that play a key role in eliciting a known functional movement [12] and to understand the response of proteins to forces such as those that might arise in atomic force microscopy [13]. Therefore, a tool that allows a user to explore the response of an ENM of a biomolecule to applied forces would be of considerable value.

In this paper we describe the implementation and testing of such a tool. We have developed two versions, one where an atom is selected and then a force is applied using a mouse, and the other where a haptic feedback device is used to select and apply forces to atoms. The application of haptics in molecular simulation has quite a long history with the first such project, GROPE1, starting in 1967 at the University of North Carolina [14]. More recent applications using personal computers allow the user to feel interaction forces between a probe molecule and the biomolecule [15–20]. These applications assume that both the protein and the ligand are completely rigid. A method for including molecular flexibility based on minimizing the potential energy of a protein-ligand system has also been described [21], but with this system problems would arise from the delay in calculating the global response of large molecules to changes in position of the ligand. "Interactive Molecular Dynamics" includes molecular flexibility by allowing the user to apply forces through a haptic device during an MD simulation [22]. A drawback with this approach, however, is that the user will react to events that occur on a very short timescale. This is avoided in an ENM, although it is known that some key aspects of dynamical behaviour of biomolecules at physiological temperature are not well described by the harmonic model. Bolopion et al. [23] have recently developed a tool that allows the user to interact with a molecular simulation based on quasi-statics [24], via a haptic feedback device. They give details on the coupling between the haptic device and the molecular simulation.

In this work we use ENMs and make use of the concept of the important subspace [25–27] in normal mode analysis which is related to the concept of the essential subspace as determined by performing a principal component analysis of MD trajectories [28–32]. The important subspace, is a space of the first M lowest frequency normal mode eigenvectors within which a large proportion of the total fluctuation occurs, and is used here in order to make it possible to use the software on standard computers even with very large biomolecules.

Methods

Theory

The ENM used in this work has the potential energy function [1]:

$$V = \frac{\gamma}{2} \sum_{|r_{ij}^0| < R_C} (r_{ij} - r_{ij}^0)^2 \tag{1}$$

Where r_{ij} is the distance between atoms i and j and r_{ij}^0 the distance between the atoms in the reference structure, e.g. a crystallographic structure. This summation is only performed over atom pairs separated by a distance less than a cut-off, R_C. γ is the force constant for the elastic bond between each pair of atoms, and is the same for all atom pairs. The second derivative matrix, the Hessian, of this potential energy function has been given previously [33]. Diagonalization of the $3N \times 3N$ Hessian, where N is the number of atoms, gives a set of eigenvalues and associated eigenvectors. For $\gamma = 1$, let $\mathbf{w}_j$ represent the eigenvector from the jth column of the $3N \times 3N$ eigenvector matrix $\mathbf{W}$, and ω_j^2 its corresponding eigenvalue at the jth position in the $3N \times 3N$ diagonal eigenvalue matrix, $\mathbf{\Omega}$. For $\gamma \neq 1$ the eigenvectors are unchanged and the eigenvalues are $\gamma\omega_j^2$. If the eigenvalues and eigenvectors are in ascending order of the eigenvalues, then the first 6 eigenvalues are zero. These correspond to the external degrees of freedom (non shape-changing modes). These 6 eigenvalues and eigenvectors are removed from $\mathbf{\Omega}$ and $\mathbf{W}$ to give a $(3N - 6) \times (3N - 6)$ matrix $\mathbf{\Omega_I}$ and a $3N \times (3N - 6)$ matrix $\mathbf{W_I}$ corresponding to the internal modes of motion (shape-changing modes).

Let $\mathbf{f}$ represent a $3N \times 1$ column matrix, where its elements f_{3i-2}, f_{3i-1} and f_{3i} are the x, y and z components, respectively, of an external force acting on atom, i. It can be shown that in static equilibrium the displacement vector $\mathbf{\Delta r}$, where its elements $\Delta r_{3i-2}, \Delta r_{3i-1}$ and Δr_{3i} are the displacements $\Delta x_i, \Delta y_i$ and Δz_i, respectively, of atom i, due to $\mathbf{f}$ is given by:

$$\mathbf{\Delta r} = \frac{1}{\gamma}(\mathbf{W_I \Omega_I^{-1} W_I^t})\mathbf{f} = \frac{1}{\gamma}\mathbf{Y f} \tag{2}$$

Here $\mathbf{\Omega_I^{-1}}$ is the inverse of $\mathbf{\Omega_I}$ (a diagonal matrix of elements, $1/\omega_j^2$), and t denotes the transpose. The matrix $\mathbf{Y}$ is a $3N \times 3N$ matrix which multiplies the $3N \times 1$ matrix $\mathbf{f}$. Note that as the modes for the external degrees of freedom have been removed the applied force will not produce any

overall translation or rotation of the molecule (effectively the molecule is held by infinitely stiff bonds on the external degrees of freedom). $\mathbf{Y}$ can be calculated in the pre-processing stage, i.e. prior to the interactive session during which the user applies forces to the molecule. As forces are applied to one atom at a time, $\mathbf{f}$ is sparse with all elements equal to zero except the x, y and z components of the atom to which the force is applied. Therefore one need only perform $9N$ multiplications to evaluate Eq. 2 rather than $9N^2$ necessary were $\mathbf{f}$ to have no zero-valued elements. This gain in speed is important for the haptic application, since a 1000 Hz update rate is required to ensure smooth and stable forces are perceived. The direct use of the pre-calculated matrix, $\mathbf{Y}$, in Eq. 2 to calculate the response will be referred to as "Method 1".

For large biomolecules the matrix $\mathbf{Y}$ may become too large to store in memory. In order to overcome this we exploit a well-known feature of the normal mode analysis of biomolecules, namely that a large proportion of the total fluctuation occurs in a relatively small subspace, the "important subspace", of the first M lowest frequency normal mode eigenvectors. We are able to exploit the important subspace in order to reduce the memory by calculating the atomic displacements as:

$$\mathbf{\Delta r} \approx \mathbf{\Delta r_M} = \frac{1}{\gamma}\sum_{j=1}^{M}\frac{1}{\omega_j^2}\mathbf{w}_j\mathbf{w}_j^t\mathbf{f} = \frac{1}{\gamma}\mathbf{W_M}\mathbf{\Omega_M^{-1}}\mathbf{W_M^t}\mathbf{f} \tag{3}$$

where $\mathbf{\Omega_M}$ is the $M \times M$ submatrix of $\mathbf{\Omega_I}$ (rows $1-M$, columns $1-M$), $\mathbf{W_M}$ is the $3N \times M$ matrix of M corresponding eigenvectors from $\mathbf{W_I}$ and $\mathbf{w}_j$ is the jth eigenvector in $\mathbf{W_M}$. Equation 3 above is equivalent to Eq. 4 in Ikeguchi et al. [10]. Equation 3 shows that in general the lowest frequency eigenvalues contribute most to the evaluation of $\mathbf{\Delta r}$. The far right hand side of Eq. 3 shows that instead of storing a $3N \times 3N$ matrix, the $3N \times M$ matrix $\mathbf{W_M}$ (and the M eigenvalues in $\mathbf{\Omega_M}$) is stored, where M is chosen so that $\mathbf{W_M}$ is small enough to fit in memory for the machine being used. Multiplying these matrices from right to left starting with $\mathbf{W_M^t f}$ and exploiting the fact that $\mathbf{f}$ has only 3 non-zero elements and $\mathbf{\Omega_M^{-1}}$ is diagonal, results in $M(3N+4)$ multiplications (multiplying these matrices in any other order results in more calculations) and during the process no matrix larger than $3N \times M$ occurs. *All three matrix multiplications in Eq. 3 are performed at runtime* as the saving on memory and time necessitates that $\mathbf{W_M^t f}$ is evaluated first, requiring knowledge of $\mathbf{f}$. If $M(3N+4)$ multiplications is too many to ensure smooth and stable forces are perceived then M is reduced further. The reduction in the number of multiplications and memory requirement through the appropriate choice of M is key to being able to use the software on large molecules. The use of Eq. 3 will be referred to as "Method 2".

Method 2 is used, irrespective of memory considerations, when the user opts to use a subset of modes during the session. In this case $\mathbf{\Omega_M}$ and $\mathbf{W_M}$ comprise the subset of eigenvalues and eigenvectors corresponding to the modes selected.

F_M, the percentage of the total fluctuation due to the modes used is:

$$F_M = \frac{100\sum_{j\in S}\frac{1}{\omega_j^2}}{\sum_{j=1}^{3N-6}\frac{1}{\omega_j^2}} \tag{4}$$

where the sum in the numerator is performed over the elements of the set S which is either $\{1 \le j \le M\}$ or the set of user selected modes. In the "Results" section, F_M is used to quantify the approximation required to run the software on a standard personal computer for biomolecules of various sizes due to memory and time limitations.

As in previous studies [34], if the structures are solved by X-ray crystallography the isotropic B-factors can be used to estimate a value for γ as both the ENM and the B-factors give a value for the total fluctuation as:

$$\sum_{i=1}^{N}\langle\Delta r_i^2\rangle = \frac{3}{8\pi^2}\sum_{i=1}^{N}B_i = \frac{k_bT_{exp}}{\gamma}\sum_{j=1}^{3N-6}\frac{1}{\omega_j^2} \tag{5}$$

where T_{exp} is the temperature in Kelvin at which the X-ray diffraction data was collected, B_i is the isotropic B-factor of atom i as given in the Protein Data Bank (PDB) file of the biomolecule concerned, and k_B is Boltzmann's constant. If all of these quantities are known γ can be determined.

The force applied through the haptic device, f_{haptic}, obviously needs to be scaled so that the force applied to the biomolecule is realistic. If f_{haptic} is the force applied to the haptic device, then let $s_{atom}f_{haptic}$ be the force applied to the atom where s_{atom} is the "atomic scaling-factor". If the scaled force is applied to coordinate k, then the displacement of atom $i = \lceil k/3 \rceil$ in the direction of the force will be given by:

$$\Delta r_k = \frac{1}{\gamma}\mathbf{Y}_{kk}s_{atom}f_{haptic} \tag{6}$$

Note we could set $\gamma = 1.0$ and deal with s_{atom} only in Eq. 6. However, using Eq. 5 to estimate γ means our scaled force $s_{atom}f_{haptic}$ will have a value that is of the right magnitude within the limits of the model, for the displacement it causes. Note that this displacement is for a biomolecule in the crystal environment.

Our value of s_{atom} could be determined such that when the force applied to the haptic device is equivalent to 1N, which is $1.4396 \times 10^{10}\text{kcalmol}^{-1}\text{Å}^{-1}$, it produces a displacement in the direction of the force on the atom that it is applied to, that is equal to the root mean-square fluctuation

of the atom at 300K as determined by its isotropic B-factor. That is:

$$\frac{1}{\gamma}\mathbf{Y}_{kk}s_{atom}\cdot 1.4396\times 10^{10}=\sqrt{\frac{3B_i}{8\pi^2}\times\frac{300}{T_{exp}}} \tag{7}$$

However, rather than base this on an individual atom we set the average over all terms of the left-hand side and right-hand side of Eq. 7 equal to determine the scaling factor as:

$$s_{atom}=\frac{\gamma\sqrt{\frac{3B_i}{8\pi^2}\times\frac{300}{T_{exp}}}}{\overline{\mathbf{Y}_{kk}}1.4396\times 10^{10}}=\frac{3N\gamma\sqrt{\frac{3B_i}{8\pi^2}\times\frac{300}{T_{exp}}}}{1.4396\times 10^{10}\sum_{j=1}^{3N-6}\frac{1}{\omega_j^2}} \tag{8}$$

The force applied to the haptic device, f_{haptic}, and the force applied to the atom, $s_{atom}f_{haptic}$, are both displayed to the user, the former in Newtons, the latter in kcalmol^{-1}Å^{-1}. In the knowledge that a force of 1 N will cause realistic displacement (at least for a biomolecule in the crystal environment), this allows the user to judge a realistic force to apply.

Let f_{hapmax} be the maximum force that can be applied to the haptic device without causing it damage. Then $s_{atom}f_{hapmax}$ would be the maximum force that can be applied to an atom. There may be occasions when the user will want to exceed this force. In order to allow this, another scaling factor, the "haptic scaling-factor", s_{haptic}, is defined. The force applied to an atom will now be given by $s_{atom}s_{haptic}f_{haptic}$. Thus one can vary s_{haptic} in order to increase the force applied to an atom such that $f_{haptic}\leq f_{hapmax}$.

To summarise the atomic displacements are given by:

$$\mathbf{\Delta r}=\frac{1}{\gamma}s_{atom}s_{haptic}\mathbf{Y}\mathbf{f}_{\mathbf{haptic}},\qquad \text{Method 1} \tag{9}$$

$$\mathbf{\Delta r}_{\mathbf{M}}=\frac{1}{\gamma}s_{atom}s_{haptic}\mathbf{W}_{\mathbf{M}}\mathbf{\Omega}_{\mathbf{M}}^{-1}\mathbf{W}_{\mathbf{M}}^{t}\mathbf{f}_{\mathbf{haptic}},\qquad \text{Method 2} \tag{10}$$

Implementation

Initially a PDB file is read to determine the atomic coordinates. If the structure was solved using NMR then only "Model 1" coordinates are used. The nodes of the ENM are C_α atoms, for protein, and P atoms for DNA or RNA. The default value for R_c is 11Å, however, this may be adjusted by the user. Normal mode analysis is performed to calculate the eigenvalues and eigenvectors, which can be saved to avoid recalculation in future sessions. Matrix $\mathbf{Y}$, shown in Eq. 2, can now be computed and also saved. However, for large biomolecules there may be insufficient memory or it may take a signficant amount of time to compute, requiring Method 2 to be used. The software may also read the eigenvalues and eigenvectors computed by other ENM normal mode analysis software, providing the input file contains eigenvalues followed by eigenvectors sorted in ascending order. Furthermore, output from the ENM webserver, ElNemo [35] may be used. ElNemo provides up to 25 modes in the form of PDB files containing displaced coordinates and a list of eigenvalues that can be used in Method 2, if the original PDB file is also loaded.

B-factors and T_{exp} are required to calculate γ and s_{atom}, as shown in Eqs. 5 and 8. If no B-factor information is available, as for structures solved by NMR, then γ can no longer be determined using Eq. 5 and it is set equal to 1.0. In addition, the B-factors used in Eq. 8 are also all set to 1.0. If T_{exp} is unspecified it will be set to 300 K as a default value. These default values for the B-factors and T_{exp} are also used when only a subset of normal mode eigenvalues and eigenvectors are loaded.

When Method 2 is used M is calculated as the minimum of two values, M_M and M_T, where M_M is restricted by the available memory and M_T is restricted by the processor speed. These values are determined before loading any eigenvalues and eigenvectors. M_M is calculated in two ways depending on whether Methods 1 and 2 are used together in the same session or whether only Method 2 is used. If Method 1 is used then the amount of memory required to store $\mathbf{Y}$ is computed and the remaining available memory determines M_M. If Method 2 is used exclusively then the total amount of selectable modes, M_M, is computed based on the entire amount of available memory. M_T is calculated based on the total amount of time taken to evaluate Eq. 10. An estimate could be made based on the number of double precision floating point operations allowed within 60 ms, which ensures interactive performance. However, a more accurate value for M_T is determined by timing the entire calculation. To compute the maximum value for M_T such that Eq. 10 can be computed within 60 ms a binary search is carried out. A flow chart describing the calculation of M is shown in Fig. 1. During the session the user may select individual modes or subsets of the available modes.

Use of the software poses two main issues; how to navigate the large biomolecular structures and how to apply desired forces to the atoms. Two types of input devices are considered in the software implementation, either full mouse control or enhanced haptic control with optional mouse commands.

Navigation and exploration

Haptics

To facilitate efficient navigation and exploration of a biomolecule we have implemented our Navigation Cube technique [20]. However, when visualising the effects of the application of forces, it is important to be able to see both

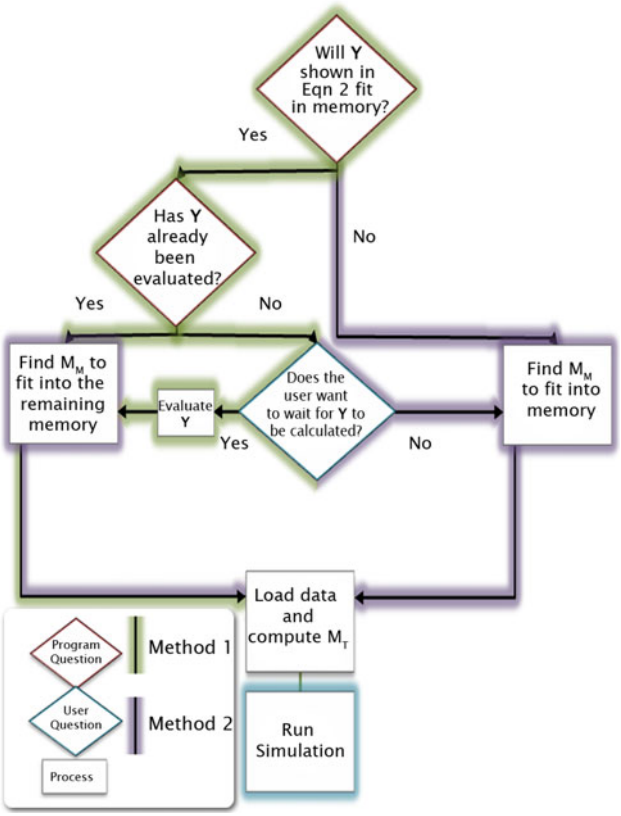

Fig. 1 A flow chart describing the processes involved before the interactive session commences. It is presumed that all eigenvalues and eigenvectors are available

local and global deformations. To this end, a separate mouse controlled camera has been employed allowing the user to translate and rotate the molecule without affecting the region that can be reached by the haptic probe. To aid in depth perception the haptic probe is represented by a sphere which has the same radius as the spheres depicting the atoms.

As the user navigates the structure further visual cues are required to aid the user in exploring and selecting atoms. Naturally, stereoscopic displays capable of generating three-dimensional views of the biomolecule present one strategy for enabling the user to gauge the probe's proximity to the atoms. An approach for quad-buffer stereo has been implemented to enable stereoscopic viewing. We also include our "haptic light" technique to enhance the user's perception of the position of the haptic probe [36]. When the haptic probe makes contact with an atom a semi-transparent sphere is rendered around the selected atom and the user "feels" the atom as a hard sphere.

Mouse

To explore the biomolecule using the mouse the user must first select "Navigation Mode" from the interface. Translation and rotation of the molecule can then be achieved by dragging the mouse with either the left or right mouse buttons depressed. Atoms are selected and forces are applied only during the "Simulate" mode.

The application of forces using both the haptic device and the mouse is discussed in the following section.

Applying forces

Haptics

Forces can be applied to the atoms and Eqs. 9 and 10 are used to determine the deformation of the structure. The magnitude and direction of a force vector is controlled using a three dimensional force feedback device, such as the Phantom Omni or Novint Falcon. When the haptic probe comes into contact with an atom the user can press the stylus button and drag the device in the desired direction to generate a force proportional to the distance dragged. To assist in maintaining the desired direction a stick-slip friction model is employed [37]. During the drag motion an arrow emanating from the surface of the atom is rendered to depict the magnitude and direction of the force. It is coloured in a colour range from blue to red based on the ratio of the magnitude of the applied force to f_{hapmax}. Once the magnitude of the force becomes equal to f_{hapmax}, the arrow is coloured red and no larger force vectors are permitted to prevent damage to the haptic device. The haptic device is not just a convenient input device but is constantly providing feedback to the user regarding the forces they are applying to the structure. One could continue to allow larger force vectors to be created but this would break the relationship between the forces input to the ENM and the forces returned to the user via the haptic device. The intention is to enable the user to experience a continuous range of forces from $0 - f_{hapmax}$ whilst the force range of $0 - s_{atom}f_{hapmax}$ is applied to the biomolecule. As discussed in the theory section, the user may apply forces larger than f_{hapmax} by manipulating a scale factor, s_{haptic} (which can be varied between 0 and 100). Once adjusted, the user may apply a force, f_{haptic}, in the range from $0 - f_{hapmax}$, whilst visualising the effect of the range $0 - s_{atom}s_{haptic}f_{hapmax}$ on the biomolecule. As the user applies a force the values for f_{haptic} and $s_{atom}s_{haptic}f_{haptic}$ are displayed. This approach enables devices capable of a maximum force of 3N, such as the Phantom Omni, to be used as effectively as devices capable of exerting a maximum force of 37.5 N, such as the Phantom Premium 1.5 High Force. Figure 2 illustrates a maximum force being applied to GroEL. An additional option is to save the current force allowing multiple forces to be applied.

Mouse

Applying forces with a mouse is more complicated due to the nature of performing a three dimensional task with a two dimensional input device. In this case a force is applied to an atom by first selecting "Simulate Mode" from the user interface. This will then allow a user to select an atom on which to apply a force. The selected atom is encompassed by a semi-transparent sphere. After selection dragging the mouse with the left button down adjusts the angle of the force and holding the right button down adjusts the magnitude. Removal of the force can be achieved by clicking on an atom.

After forces have been applied using either mouse or haptic input devices the deformed molecular structure can be output as a new PDB file. Furthermore, entire user sessions can be saved that will allow the program to restart at a later date with all applied force data and user interface settings intact.

User testing

To assess the advantages gained from using a haptic device over a mouse, user testing was undertaken. The small protein crambin was selected for the test. Typically a user will colour an atom of interest and then navigate the structure to obtain an appropriate view. After navigation, the user will then apply a force to that atom to investigate the flexibility of the molecule. To replicate this in the experiment, five atoms were randomly highlighted individually for the user. On each atom an arrow, of random length and direction, is depicted for the user to match within a threshold, see Fig. 3. This is performed five times for each of the five atoms. The times are then recorded for the mouse and haptic interface to enable an assessment of the interaction modes.

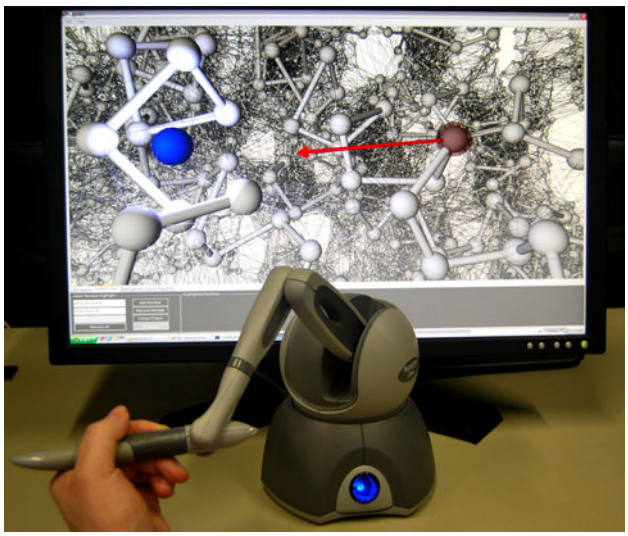

Fig. 2 A force being applied to GroEL. The *blue sphere* represents the haptic probe and the *red semi-transparent sphere* represents the atom on which the force is being applied. The haptic light effect can be seen to illuminate the atoms and network in close proximity to the haptic probe, which is useful for navigation

Results

User testing

User tests have been conducted among 20 volunteers. An identical test was taken by all volunteers. The time for selecting each highlighted atom was averaged per user as was the time for applying all five forces. Figure 4 shows the averaged times for each user.

The results show that on average locating the atom was 1.6 times faster using a haptic device rather than a mouse. The time taken to apply the correct force is 6 times faster using a haptic device than a mouse.

Testing on individual proteins

The software has been tested on seven proteins spanning a considerable size range: crambin, liver alcohol dehydrogenase, *S*-adenosylhomocysteine hydrolase, aspartate transcarbamylase, ferritin, glutamine synthetase and GroEL. The software has been designed to run with large molecules on a standard PC. To demonstrate this we have determined the general performance on a machine with

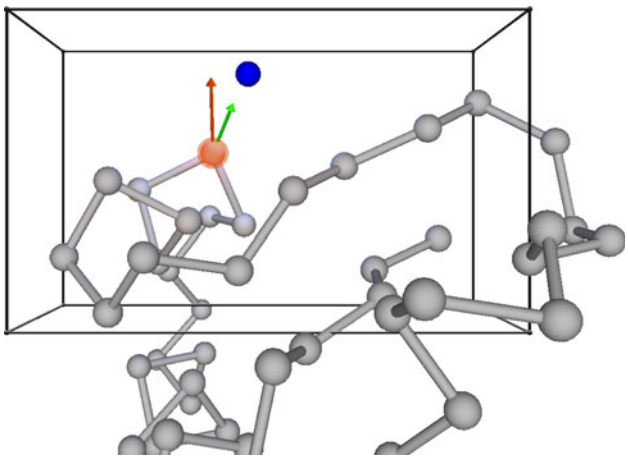

Fig. 3 A screen capture of the user testing program for the haptic device, where the user is interacting with crambin. The user is first instructed to select the flashing atom, depicted in red, with the input device. The navigation cube used for translating the structure is illustrated. An *arrow* is displayed depicting a force (shown in *red* in the image) and the user must then match this force by moving the probe sphere (the user's current force is shown in *green* in the image and the probe sphere is shown in *blue*). If the user succeeds a *new arrow* is then displayed and the test continues

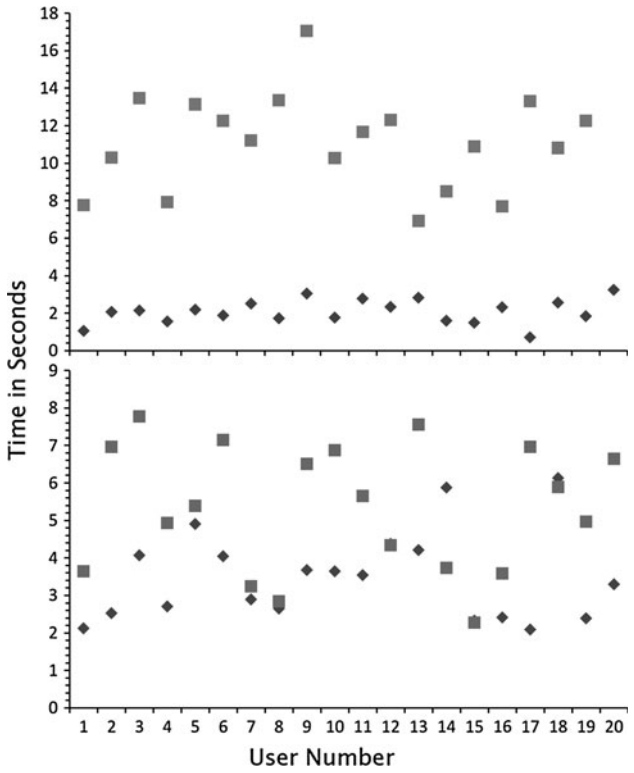

Fig. 4 Graphs showing the average times recorded for all 20 users. The *bottom graph* shows the average times recorded to locate the atom. The *top graph* shows the average time to apply all forces. The times recorded for the haptic device are depicted with *diamonds* and the mouse times with *squares*

2 GB of available memory and able to perform 11 million operations per cycle. We believe this is more informative than describing the results from running the software on a particular PC where performance is dependent on many factors. Table 1 gives the details for each of the seven biomolecules. Apart for glutamine synthetase and GroEL there is sufficient memory to load the pre-calculated matrix, **Y**, and for these molecules Method 1 can be used which is sufficiently fast not to exceed time constraints. For glutamine synthetase and GroEL Method 2 must be used. For glutamine synthetase and GroEL, M has been reduced so that the matrices of Method 2 can fit into memory. For glutamine synthetase $M_M = 15{,}371$ which is 88% of the total number of modes and for GroEL $M_M = 12{,}194$ which is 55% of the total number of modes. Despite this reduction, F_M, the percentage of the total fluctuation, is 97% for glutamine synthetase and 88% for GroEL showing that most of the fluctuation occurs in this subset of modes.

Method 2 is considered for all molecules even though Method 1 could be used for all but glutamine synthetase and GroEL. For Method 2, M_T will often be smaller than M_M, due to the $M(3N + 4)$ multiplications required to evaluate Eq. 10, although for crambin and liver alcohol dehydrogenase no further reduction in M is necessary. However, for the remaining proteins, M has to be reduced considerably, and for ferritin, glutamine synthetase and GroEL below 10% of the total number of modes. However, this drastic reduction in M does not necessarily mean a drastic reduction in F_M as seen in the last column of Table 1. For GroEL, even though only 2.3% of the total number of modes can be used, these modes still account for 50% of the total fluctuation. The general reason for this is shown in Fig. 5. Figure 5 plots F_M against the percentage of modes used ($M/3N - 6$) for each of the seven biomolecules. As can be seen, even though they span a considerable size range they all have the same general

Table 1 Performance results for computation time and memory for seven proteins

Protein (PDB code, No. of residues)	Method 1 Memory requirement (GB)	Method 2			
		M_M restricted by 2 GB of memory		M_T restricted by time[a]	
		$M_M/(3N - 6)(\%)$	F_M percentage of total fluctuation	$M_T/(3N - 6)(\%)$	F_M percentage of total fluctuation
Crambin (1CRN, 46)	0.0001	132/132(100%)	100%	132/132(100%)	100%
Liver Alcohol Dehydrogenase (1ADG, 748)	0.04	2238/2238(100%)	100%	2238/2238(100%)	100%
S-Adenosylhomocysteine Hydrolase (1B3R, 1712)	0.2	5130/5130(100%)	100%	2143/5130(41.8%)	82%
Aspartate Transcarbamylase (1AT1, 2736)	0.5	8202/8202(100%)	100%	1340/8202(16.3%)	77%
Ferritin (1FHA, 4128)	1.1	12378/12378(100%)	100%	888/12378(7.2%)	47%
Glutamine Synthetase (1FPY, 5820)	2.3	15371/17454(88.1%)	97%	630/17454(3.6%)	48%
GroEL (3FBH, 7336)	3.6	12194/22002(55.4%)	88%	499/22002(2.3%)	50%

[a] Based on 11 million operations per CPU cycle, utilising a 2.4 GHz processor

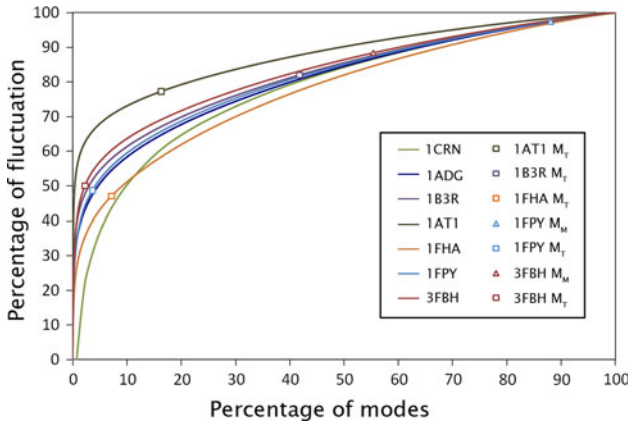

Fig. 5 A graph of F_M against the percentage of the total number of modes for the seven biomolecules tested. The values of M_M and M_T are indicated, precise values are given in Table 1. The results are based on a PC capable of 11 million operations per CPU cycle and 2 GB of available memory

trend, namely a steep rise followed by a much shallower approach to the 100% maximum. The steep region indicates the important subspace. The points on the curves show the value of M and F_M at the memory and time cut-offs as given in Table 1. These points are located near the top, or beyond the region, of steep rise.

Discussion

Elastic Network Models have proved useful for understanding the large-scale motions that occur in biomolecules. One particular application is to apply forces to specific regions of the molecule and to see its response. This can be done either to mimic forces that may arise in the interaction with a ligand or as part of an investigation to understand the mechanical properties of a molecule. ENMs are of particular use in the investigation of the properties of large biomolecules rather than small ones [2]. There are two reasons for this. The first reason is that small molecules do not have domains and their functional movements involve protein side chains or loops which are badly modelled by an ENM. The second reason is that the simulation of very large systems is computationally expensive using more accurate methods such as MD. The advantage of using an ENM is that it separates the computationally expensive calculation of the eigenvalues and eigenvectors from the calculations required during the interactive session enabling it to run on standard computers. Haptic interaction during a standard MD simulation, however, would require access to a powerful supercomputer for large molecules. In the development of the system, it has been clear that unless it were able to deal with large systems it would be of limited usefulness. Despite the separation of the calculation of the eigenvalues and eigenvectors from the calculations required in the interactive session it is apparent that for large systems such as GroEL the basic calculation (as given in Eq. 9) would require too much memory for an average PC. In order to deal with this we have utilized the concept of the important subspace from normal mode analysis. This has allowed us to develop the software to deal with large systems with a quantifiable approximation. The approximation allows us to reduce the memory required for large systems and involves the effective removal of the higher frequency modes which are not thought to be related to global functional movements. If memory reduction is necessary the same approximation can be used to speed up the calculation. Speed is particularly important for the haptic application due to the sensitivity of the human sense of touch.

Existing systems, such as the one by Bolopian et al. [23], strongly couple the haptic device to the molecular simulation in a continuous control loop with scaling factors for the force and position information. In these systems transparency and stability can present difficulties. In our approach the force is calculated based on the magnitude and direction of a vector the user creates by manipulating the haptic device. Only this force vector needs to be scaled down to the molecular level for it to be used in the ENM. The resulting conformation is displayed but no information is fed back to the haptic device, other than a force based on the original force vector the user created. In this sense our approach presents a linear pipeline to the force application and simulation process where the haptic device provides an intuitive interface for inputting forces in three dimensions and for comparing their effect on the flexibility of the biomolecule. For example, the user may apply a force close to f_{hapmax} in one direction and see little change to the structure whilst applying the same force in another direction could cause significant conformational change. Use of the important subspace means that an update rate of 1000Hz can be achieved, even for large biomolecules. This rapid rate means that as the user varies the force the molecule's structure is perceived to update instantly. This approach avoids stability issues which might result when simulating dynamics as it simply calculates the response of the molecule in static equilibrium.

As an example of possible usage consider the famous allosteric enzyme aspartate transcarbamylase. The binding of the regulatory molecule cytosine triphosphate (CTP) to the regulatory dimers influences catalysis by causing the two catalytic trimers to counter rotate. This mechanical coupling between the regulatory dimers and the catalytic trimers can be clearly seen using the software. Applying forces at the region where CTP binds causes the expected counter rotation of the catalytic trimers.

Although we have developed two versions, one for haptic devices, the other for a mouse, it is clear from user

testing that the haptic version is by far the most convenient. Our testing showed that even on a very small biomolecule the haptic version allowed for much more rapid and intuitive exploration than the mouse version.

Availability

Both haptic feedback and mouse versions of the software are freely available for non-commercial use and can be downloaded from http://www.haptimol.co.uk/enm/mgi.

References

1. Tirion M (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77:1905–1908
2. Chacon P, Tama F, Wriggers W (2003) Mega-dalton biomolecular motion captured from electron microscopy reconstructions. J Mol Biol 326:485–492
3. Wang Y, Rader AJ, Bahar I, Jernigan RL (2004) Global ribosome motions revealed with elastic network model. J Struct Biol 147:302–314
4. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15:586–592
5. Lezon TR, Sali A, Bahar I (2009) Global motions of the nuclear pore complex: insights from elastic network models. PLoS Comput Biol 5
6. Atsushi M, Ishida H (2009) Global conformational changes of ribosome observed by normal mode fitting for 3D cryo-EM structures. Structure 17:1605–1613
7. Bahar I, Lezon TR, Bakan A, Shrivastava IH (2009) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 110:1605–1613
8. Tama F, Sanejouand Y-H (2001) Conformational change of proteins arising from normal mode calculations. Prot Eng 14:1–6
9. Hayward S (2004) Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. J Mol Biol 339:1001–1021
10. Ikeguchi M, Ueno J, Sato M, Kidera A (2005) Protein structural change upon ligand binding: linear response theory. Phys Rev Lett 94:078102-1-4
11. Hayward S, Kitao A (2006) Molecular dynamics simulations of NAD+-induced domain closure in horse liver alcohol dehydrogenase. Biophys J 91:1823–1831
12. Atilgan C, Atilgan AR (2009) Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput Biol 5
13. Eyal E, Bahar I (2008) Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. Biophys J 94:3424–3435
14. Brooks FP, Ouh-Young M, Battert JJ, Kilpatrick PJ (1990) Project GROPE-haptic displays for scientific visualization. Comput Graph 24:177–185
15. Comai S, Mazza D (2010) A haptic-based framework for chemistry education: experiencing molecular interactions with touch, technology enhanced learning. Quality of teaching and educational reform. Springer, Berlin Heidelberg, pp 338–344
16. Nagata H, Mizushima H, Tanaka H (2002) Concept and prototype of protein-ligand docking simulator with force feedback technology. Bioinformatics 18:140–146
17. Sankaranarayanan G, Weghorst S, Sanner M, Gillet A, Olson A (2003) Role of haptics in teaching structural molecular biology. HAPTICS '03, 363
18. Wollacott AM, Merz KM Jr. (2006) Haptic applications for molecular structure manipulation. J Mol Graph Model 25:801–805
19. Subasi E, Basdogan C (2008) A new haptic interaction and visualization approach for rigid molecular docking in virtual environments. Presence Teleoper Virtual Environ 17:73–90
20. Stocks M, Hayward S, Laycock S (2009) Interacting with the biomolecular solvent accessible surface via a haptic feedback device. BMC Struct Biol 9:69
21. Daunay B, Régnier S (2009) Stable six degrees of freedom haptic feedback for flexible ligand-protein docking. Computer-Aided Design, pp 886–895
22. Stone J, Gullingsrud J, Grayson P, Schulten K (2001) A system for interactive molecular dynamics simulation. ACM Siggraph, pp 191–194
23. Bolopion A, Cagneau B, Redon S, Régnier S (2010) Comparing position and force control for interactive molecular simulators with haptic feedback. J Mol Graph Model 29(2):280–289
24. Rossi R, Isorce M, Morin S, Flocard J, Arumugan K, Crouzy S, Vivaudou M, Redon S (2007) Adaptive torsion-angle quasi-statics: a general simulation method with applications to protein structure analysis and design. Bioinformatics 23(13):408–417
25. Nishikawa T, Go N (1987) Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. Proteins Struct Funct Genet 2:308–329
26. Go N (1990) A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. Biophys Chem 35:105–112
27. Hayward S, Go N (1995) Collective variable description of native protein dynamics. Ann Rev Phys Chem 46:223–250
28. Kitao A, Hirata F, Go N (1991) The effects of solvent on the conformation and the collective motions of protein—normal mode analysis and molecular-dynamics simulations of melittin in water and in vacuum. Chem Phys 158:447–472
29. García AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699
30. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins Struct Funct Genet 17:412–425
31. Hayward S, Kitao A, Hirata F, Go N (1993) Effect of solvent on collective motions in globular proteins. J Mol Biol 234:1207–1217
32. Kitao A, Go N (1999) Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 9:164–169
33. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80:505–515
34. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding Design 2:173–181
35. Karsten S, Yves-Henri S (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucl Acids Res 32:W610–614
36. Laycock SD, Stocks MB, Hayward S (2010) Navigation and exploration of large data-sets using a haptic feedback device. SIGGRAPH '10: ACM SIGGRAPH 2010 Posters
37. Zilles CB, Salisbury JK (1995) A constraint based God-object method for haptic display. In: Proceedings of the IEEE conference on intelligent robots and systems, pp 146–151