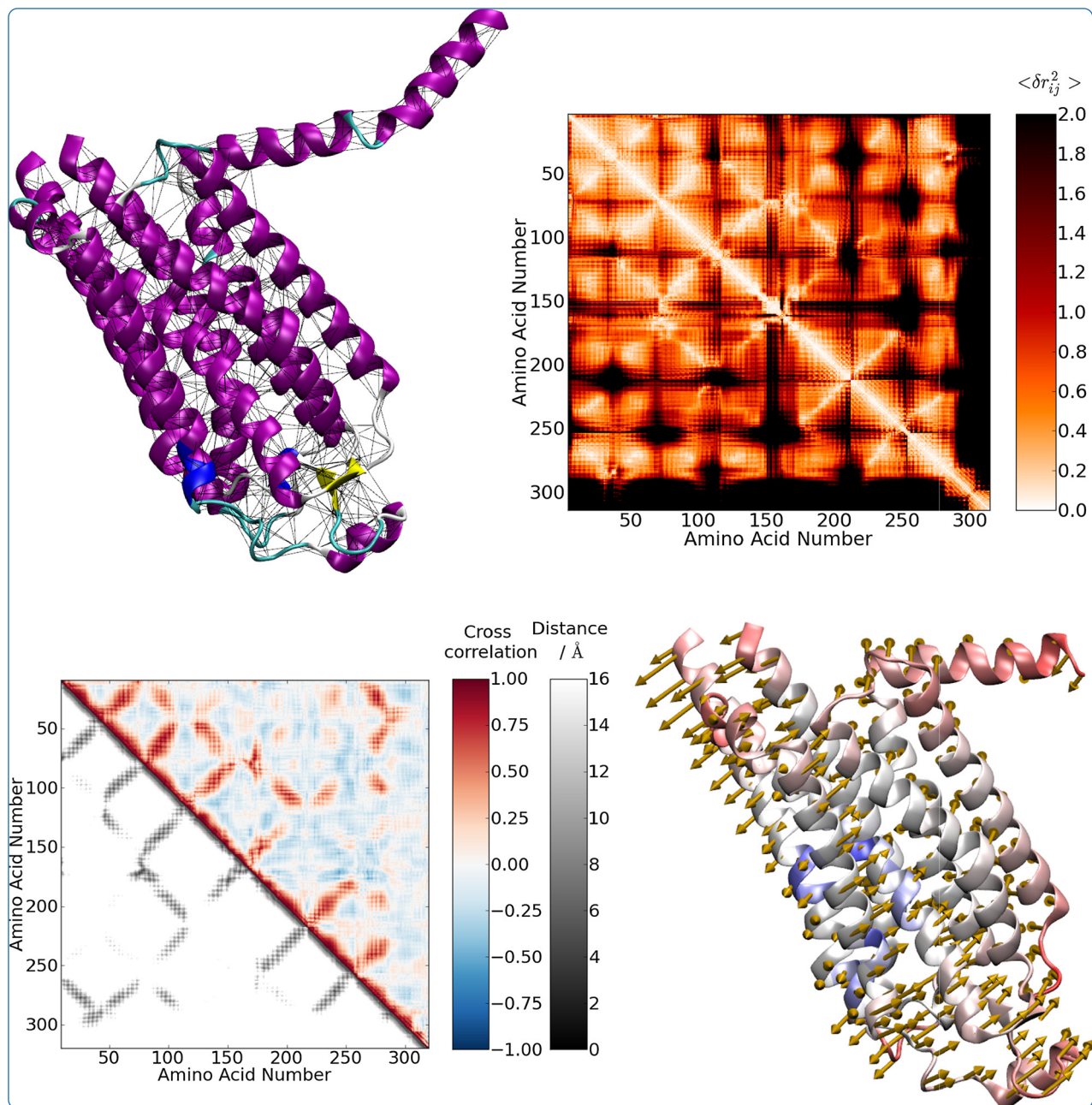




$\Delta\Delta$ PT: a comprehensive toolbox for the analysis of protein motion

Rodgers *et al.*

SOFTWARE

Open Access

$\Delta\Delta$ PT: a comprehensive toolbox for the analysis of protein motion

Thomas L Rodgers^{1,2*}, David Burnell^{1,2}, Phil D Townsend^{1,3}, Ehmke Pohl^{1,2,3}, Martin J Cann^{1,3}, Mark R Wilson^{1,2} and Tom CB McLeish^{1,4}

Abstract

Background: Normal Mode Analysis is one of the most successful techniques for studying motions in proteins and macromolecules. It can provide information on the mechanism of protein functions, used to aid crystallography and NMR data reconstruction, and calculate protein free energies.

Results: $\Delta\Delta$ PT is a toolbox allowing calculation of elastic network models and principle component analysis. It allows the analysis of pdb files or trajectories taken from; Gromacs, Amber, and DL_POLY. As well as calculation of the normal modes it also allows comparison of the modes with experimental protein motion, variation of modes with mutation or ligand binding, and calculation of molecular dynamic entropies.

Conclusions: This toolbox makes the respective tools available to a wide community of potential NMA users, and allows them unrivalled ability to analyse normal modes using a variety of techniques and current software.

Background

Normal mode analysis (NMA) is both one of the most commonly used and best suited theoretical methods for studying motions in proteins and other macromolecules. This produces a collection of collective modes which represent the true protein dynamics [1]. The first normal mode studies were performed in the early 1980s [2-4], and they remained restricted to small-size proteins until the mid 1990s. From this time, methodological advances [5-9], simplified protein descriptions [10-13], and faster computer systems allowed them to address increasingly large macromolecular systems. By the early 2000s, entire protein complexes could be addressed, including the whole ribosome [14-16].

Krebs *et al.* 2002 [17] have analysed more than 3800 experimentally determined protein motions, and have shown that more than half of them can be approximated by applying a perturbation in the direction of at most two low-frequency normal modes of the considered protein; often a single low frequency normal mode is enough, and it is usually one of the three lowest-frequency modes [14,15]. Conformational changes on ligand binding of proteins have also been represented by motion along low frequency normal modes [8,14,15]. This method has also

been used in the study of membrane channel opening [18], the analysis of structural movements of the ribosome [16], viral capsid maturation [19], transconformations of the SERCA1 Ca-ATPase [9,20], tertiary and quaternary conformational changes in aspartate transcarbamylase [21], mapping G-actin crystal form onto the F-actin crystal form highlighting possible transition pathways [22], the regulation of the Kv7.1 Potassium Channel by KCNE1 [23], and the unfolding of Amylosucrases [24].

B-factors calculated from crystallographic data have been predicted and refined using normal mode analysis [25,26]. The residue average B-factors (the average over all the heavy atoms, i.e. not including hydrogens) of alpha lytic protease have been well predicted [27] and extended to examine differences in motion of the S1 binding pocket in either a symmetric or antisymmetric direction. It has been found that the symmetric direction allowed a much large opening of the binding pocket. The diffuse scattering produced by correlated displacements of atoms during X-ray scattering experiments have also been predicted from normal mode analysis for lysozyme [28]. Cryo-EM structures have also been refined using elastic network models [29].

NMA is most often used to predict conformational changes that proteins undergo to fulfil their function, and can be used to check if a conformational change proposed on the basis of non-structural experimental data is likely to occur. These functional motions have led to the determination of domains within the proteins [30].

*Correspondence: thomas.rodders@durham.ac.uk

¹Biophysical Sciences Institute, Durham University, Durham, UK

²Department of Chemistry, Durham University, Durham, UK

Full list of author information is available at the end of the article

For example, Class I major histocompatibility complex molecule fluctuations have been found to be dependant on the conformation of their three domains [31] and it has been shown that each domain motion has a different function within the molecule. Human growth hormone induces dimerization of its binding protein; it has been shown that this is due to a marked decrease in domain motion after binding [32].

NMA can also be used to predict entropy changes on ligand binding as each normal mode has a calculable entropy associated with it. This means that for entropically controlled allosteric binding, it would be possible predict changes in the allosteric binding ratios [33]. The free energy of large functional motions can also be predicted by NMA [34,35]. The vibrational energy of G-actin has been calculated by regarding the molecule as a collection of independent harmonic oscillations (the normal modes) [35].

The major goal of normal mode analysis is to reduce the complexity of the full dynamics of a complex system and to describe them in a few generalised coordinates. However, if the long range hydrodynamics of water and anharmonicity are important variants to the protein motion then a method that is capable of reducing a complex system to a few general components but is not dependant on a harmonic approximation is needed. This method is principle component analysis (PCA) [36], and it is a technique used in a wide variety of fields, e.g. from finance to biology.

PCA computes the second moment of a multivariate distribution and describes the deviations from an average in terms of a set of principle components that represent the collective motions of the largest deviations. These principle components are the eigenvectors of a covariance matrix of the motion, whether the system is harmonic, heavily damped, or does not oscillate at all. Like NMA generally, only a small number of the lowest frequency modes are needed to describe most of the protein motion [37]. The lowest frequency modes tend to describe possible conformational changes in the protein while the slightly higher frequency modes describe vibrational, or breathing, motions around the average structure.

For ubiquitin, with molecular dynamics simulations starting from a variety of different X-ray structures, it was found that the first ten quasi-harmonic analysis modes contributed 78% of all the dynamic movement and that these modes described fluctuations of the structures seen with NMR [38].

PCA need not even be applied to dynamic fluctuations, but can be used to explore a mapping of many different conformers or mutants of a family of proteins. Recent work has explored 40 different X-ray structures of Ras kinase proteins and found that the structural variance can be described by a small number of principal components [39].

NMA and PCA thus represent a powerful tools with a wide range of applications in structural biology. Due to this there are a number of on-line web servers currently available that can calculate elastic network models, e.g. EL-Nemo [15] provides the scaled frequency, fluctuations, and shapes of calculated normal modes; ANM web server [40] provides calculation of the normal modes and allows on-line display of the modes with a Jmol plugin; and FlexServ [41] provides calculation of normal modes, and also allows simulation by discrete molecular dynamics and Brownian dynamics.

There are also programmatic libraries available for analysis of NMA and PCA, however, these mean the user has to write their own code and integrate the subroutines from these libraries manually; e.g. MMTK [42] which provides python subroutines for molecular dynamics, NMA, and structural minimisation; and ProDY [43] which provides python subroutines for PCA and NMA.

We designed $\Delta\Delta$ PT as a comprehensive, but still easy-to-use toolbox for NMA/PCA, with increased functionality for normal mode analysis over currently available methods, and easier to use than current programmatic libraries. Particular emphasis was put on its ability to analyse data from Elastic Network Models, Gromacs simulations [44], Amber simulations [45], and DL_POLY simulations [46,47] in an interchangeable manner with all the post analysis tools available irrespective of the input data. Due to the modular nature of the software it is also easy to produce additional input or analysis programs to adapt to the needs of most researchers; however, this is not required to use the program.

Methodology

Normal mode calculation is based on the harmonic approximation of the potential energy function, V , around a minimum energy conformation, Equation 1, where r is the distance between atoms, R is the equilibrium distance between atoms, u is the difference from equilibrium distance between atoms, i and j refer to the atom number, and α and β refer to the direction of the motion.

$$\mathbf{V}(\mathbf{r}_n) = \frac{1}{2} \sum_i^N \sum_j^N \frac{\partial^2 \mathbf{V}}{\partial r_{i\alpha} \partial r_{j\beta}} \bigg|_R u_{i\alpha} u_{j\beta} \quad (1)$$

This approximation allows an analytic solution of the equations of motion by diagonalising the mass-weighted Hessian matrix, $\mathbf{D}$, (the mass-weighted second derivatives of the potential energy matrix), Equation 2, where $D_{i\alpha,j\beta} = \partial^2 \mathbf{V} / \partial r_{i\alpha} \sqrt{m_i} \partial r_{j\beta} \sqrt{m_j} \big|_R$ and m is the mass.

$$\mathbf{e}^{-1} \mathbf{D} \mathbf{e} = \text{diag} [\omega_{1 \rightarrow n}^2] \quad (2)$$

The eigenvectors of this matrix, e , are the normal modes, and the eigenvalues are the squares of the associated frequencies, ω . The protein movement can then be represented as a superposition of these normal modes, fluctuating around a minimum energy conformation. The normal modes responsible for most of the amplitude of the atomic displacement are associated with the lowest frequencies.

In order to avoid time-consuming energy minimisations, a single-parameter Hookean potential can be used, which is shown to yield low-frequency normal modes as accurate as those obtained with more detailed, empirical, force fields [10]. The spring constant of the Hookean potential, k , is generally assumed to be the same for all interacting pairs within an arbitrary cut-off, R_c , beyond which interactions are not taken into account.

$$V_{ij} = \begin{cases} \frac{k_{ij}}{2} (r_{ij} - R_{ij})^2 & R_{ij}^2 \leq R_c^2 \\ 0 & R_{ij}^2 > R_c^2 \end{cases} \quad (3)$$

$\Delta\Delta$ PPT toolbox has a default cut-off of 12 Å and a Hookean potential of 1 kcal mol⁻¹ Å⁻², these can be changed with the relevant flags (-c and -r respectively) in the GENENMM program. This approximation implies that the reference structure represents the minimum energy conformation. As default, all atom masses are set to the same fixed value in the kinetic energy term, 1 Da, as this approximation was shown to have little influence on the low-frequency modes; however, if desired the true atomic masses can be used (add -mass flag) or, if the model is based on the C_α atoms, only the residue mass can be assigned (add -res flag with -ca flag), Figure 1.

The GENENMM program also allows elastic network models with a varying spring constant, either with an empirical power decay on the interaction (-an flag), with the Hinsen exponential spring constant (-hine) [12], with Hinsen fitted spring constants (-hin) [42], or with individually set values between residues (-f file). GNMPROD also allows the production of the one-dimensional Gaussian network model instead of the three-dimensional elastic network model^a.

The resulting Hessian can be either fully diagonalised using the DIAGSTD program (not recommended for many more than 1000 sites - although in reality a system this size will only take around 10 minutes to solve on a desktop PC - run serial on an AMD Phenom™II 3.2 GHz Quad Core) or diagonalised using the rotation-translation-block (RTB) approach, DIAGRTB program. The RTB approach groups several atoms into a single point, which is generally achieved by division into residue blocks, or multiple residue blocks. The rigid-body rotations and translations of these 'super'-sites are used as the new co-ordinate system instead of Cartesian co-ordinates [6]. When a small number of residues per block are

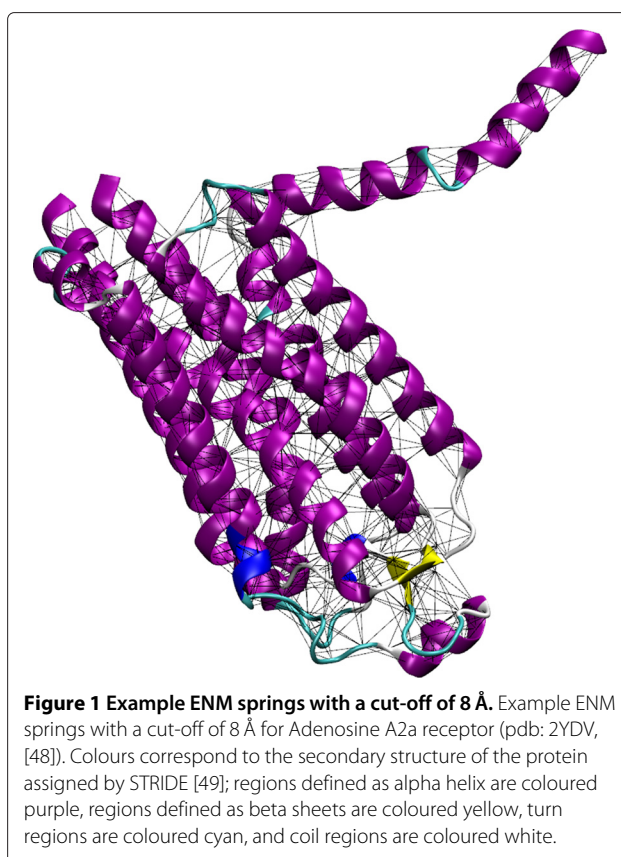


Figure 1 Example ENM springs with a cut-off of 8 Å. Example ENM springs with a cut-off of 8 Å for Adenosine A2a receptor (pdb: 2YDV, [48]). Colours correspond to the secondary structure of the protein assigned by STRIDE [49]; regions defined as alpha helix are coloured purple, regions defined as beta sheets are coloured yellow, turn regions are coloured cyan, and coil regions are coloured white.

used, the approximation has very little effect on the low frequency modes; although the frequencies do increase predictably due to internal block stiffening [8]. Using this approximation, it becomes possible to treat very large proteins, or protein complexes, in an all-atom level of description in reasonable computing time. DIAGRTB can be set to block into groups by a number of residues (-r n), block into the protein secondary structure (-str SECO), or block into custom domains (-str DOMN). The lowest frequency modes mainly depend on the overall shape of the system; they can be captured at extremely high levels of coarse-graining [50] or by using low-resolution structural data [51].

For comparison with atomistic simulations, the COVAR program allows calculation of a mass weighted covariance matrix, $\mathbf{F}$, from trajectories generated with Gromacs, Amber, or DL_POLY, Equation 4, where $\mathbf{x}$ is the atomic position matrix and $\mathbf{m}$ is the mass matrix.

$$\mathbf{F} = \left\langle \mathbf{m}^{1/2} (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{m}^{1/2} (\mathbf{x} - \langle \mathbf{x} \rangle))^T \right\rangle \quad (4)$$

COVAR also corrects the displacements by removing the centre of mass motion and rigid body rotations; this produces more accurate results as the motion is not dominated by the rigid body motions. These displacements can

be expanded into normal modes, principle component analysis, Equation 5, where Q is the eigenvalue matrix.

$$\mathbf{e}^{-1} \mathbf{F} \mathbf{e} = \langle \mathbf{Q}^2 \rangle \quad (5)$$

As we are again approximating the full motion to harmonic style motions, the solution is governed by harmonic oscillatory statistical mechanics. This means that for each eigenvector Equation 6 must hold true [52], where k is the Boltzmann constant, T is the temperature, and ν is the normal mode number.

$$\langle Q_\nu^2 \rangle = \frac{kT}{\omega_\nu^2} \quad (6)$$

The COVAR program also plots the trajectory frames onto the lowest frequency eigenvectors, Figure 2. If the inbuilt principle component analysis tools in Gromacs or Amber are preferred, GroAMED can convert the default outputs respectively for use of the other toolbox programs.

The FREQ/EN program calculates the mode frequencies, the free energy, and the entropy from the calculated eigenvalues. The free energy and entropy are calculated using the full solution, Equations 7 and 8, and the Schlitter approximation [53] for comparison with other programs^b,

where G is the free energy, S is the entropy, and $\hbar$ is the reduced Plank constant.

$$G_\nu = -kT \ln \left(\frac{1}{1 - \exp \left(-\frac{\hbar \omega_\nu}{kT} \right)} \right) \quad (7)$$

$$S_\nu = k \left(\frac{\frac{\hbar \omega_\nu}{kT}}{\exp \left(\frac{\hbar \omega_\nu}{kT} \right) - 1} - \ln \left(1 - \exp \left(-\frac{\hbar \omega_\nu}{kT} \right) \right) \right) \quad (8)$$

The RMS/COLL program calculates the root mean squared displacements of all the atoms for each of the selected modes along with the collectivity, κ , of the modes, Equation 9 [54], where α is the collectivity constant selected so that $\sum_i^N \alpha |\mathbf{e}_i^2(\nu)| = 1$ and N is the number of atoms.

$$\kappa_\nu = \frac{1}{N} \exp \left(- \sum_i^N \alpha |\mathbf{e}_i^2(\nu)| \log (\alpha |\mathbf{e}_i^2(\nu)|) \right) \quad (9)$$

The degree of collectivity indicates the fraction of atoms that are significantly affected by a given mode. For modes involving all the atoms, the degree of collectivity tends to be one, whereas for localised motions the degree of collectivity approaches zero (actually $1/N$). The first 25-50 low-frequency normal modes tend to have a collectivity of above 0.4 meaning a significant part of the protein

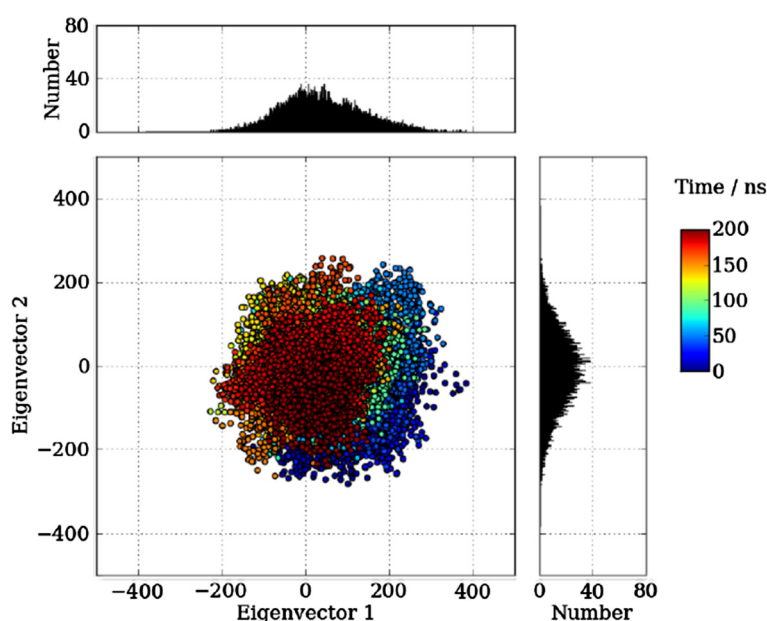


Figure 2 Plot of the trajectory from an Amber simulation. Plot of the trajectory from an Amber simulation of CAP (pdb: 1G6N) onto the two lowest frequency eigenvectors. Each trajectory position is plotted as the dot product of the co-ordinates and the eigenvector, representing the extend of the displacement along each eigenvector from the average position. The distribution of these values are displayed as the adjoining histograms. The colour of the points responds to the simulation time.

is involved in each mode. Low collectivity in the lowest frequency modes is indicative of extended parts of the system, either N- or C- termini or large unstructured loops. These loops cannot be modelled in a meaningful way as they intrinsically adopt multiple conformations, can appear to be invisible in one crystal form but visible in a different crystal form [55], or can even appear ordered due to crystal packing [56].

It is common practise to remove these extended parts prior to the normal mode computation. If a RTB approximation is used, there is some advantage to blocking and representing large unstructured loops by one block so they are included but do not dominate the motion.

The RMS/COLL program also calculates the B-factors, B , Equation 10 [52], from the mean square displacements of the first 25 lowest frequency modes (this can be changed with the `-e n` option), ignoring the six rigid block rotational and translation modes (starting from the seventh mode, `-s 7`, is the default). The B-factors should be calculated with the same mass weighting options as GenENMM. Correlations to crystallographic B-factors are typically found to be greater than 0.5-0.6 [15], and can even be greater than 0.8 [1]^c.

$$B_i = \frac{8kT\pi^2}{3m_i} \sum_v \frac{|\mathbf{e}_i^2(v)|}{\omega_v^2} \quad (10)$$

Adjusting the cutoff value can slightly improve such correlations, and if possible it is recommended that the correlation between the shape of the predicted and the experimental values is iterated upon when setting the cutoff value if no other information is available. The comparison between the shape of the computed and observed crystallographic B-factors provides a measure of how well the protein's flexibility in its crystal environment is described by the normal modes. This motion tends to echo, but is more restricted (by crystal packing) than the motion in solution.

The CROSCOR program calculates the cross-correlation, C , of atoms over the first 25 modes (although this can be changed with the `-b n` and `-e n` flags), Equation 11. The cross-correlation shows which atoms tend to move in the same direction with a correlated motion in the modes, Figure 3(a). A value of 1 implies perfectly correlated motion and -1 perfectly anti-correlated motion. As the numerator is calculated as the dot product between the two vectors, as is a common manner of calculation, the correlation is dependant on the angle of the motion, i.e. fluctuations of the same period and phase but with a difference in orientation of 90° will give a value of 0. Thus, the cross-correlation is useful for identifying which atoms make up a group with correlated motions; however, a spherical breathing mode is difficult to identify from the cross-correlations because they are positive for

atoms on the same side, negative for atoms on opposite sides, and 0 for atoms at 90° [57].

$$C_{ij} = \sum_v \left(\frac{\mathbf{e}_i(v) \cdot \mathbf{e}_j(v)}{(|\mathbf{e}_i(v)|^2 |\mathbf{e}_j(v)|^2)^{0.5}} \right) \quad (11)$$

The extent of this motion, $\langle \delta r_{ij}^2 \rangle$, can be calculated with the MOVEING program. This calculates the change in the distance between atoms between the equilibrium value and the value after applying the eigenvectors, Equation 12, Figure 3(b).

$$\langle \delta r_{ij}^2 \rangle = \sum_v \left(\frac{(\mathbf{x}_i + \mathbf{e}_i(v) - (\mathbf{x}_j + \mathbf{e}_j(v)))^2 - (\mathbf{x}_i - \mathbf{x}_j)^2}{\omega_v^2} \right) \quad (12)$$

The OVERLAP program calculates the overlap of the atomic motion between eigenvectors, v_1 and v_2 , Equation 13. There can be different eigenvectors for the same NMA, eigenvectors produced with ENM to atomistic NMA, NMA eigenvector to difference in two crystal structures, or any combination thereof. A value of 1 indicates that the motions are identical whereas a value of 0 indicates that the motions are completely different.

$$I_{v_1 v_2} = \left(\frac{\sum_i |\mathbf{e}_i(v_1) \cdot \mathbf{e}_i(v_2)|}{\left(\sum_i |\mathbf{e}_i(v_1)|^2 \sum_i |\mathbf{e}_i(v_2)|^2 \right)^{0.5}} \right) \quad (13)$$

Implementation

There are four principal inputs into the toolbox: protein coordinates written in PDB format [58]; NMA output data from Gromacs; PCA output data from Gromacs or Amber; or a trajectory output from Gromacs, Amber, or DL_POLY.

For the ENM implementation, the PDB file where all ATOM records are read by the GENENMM program is all that is needed to determine the interaction matrix (HETATM records can also be included by using the `-het` flag, these are commonly used for ligands and provide an easy method of looking at differences on ligand binding). DNA can also be read into the GENENMM using the `-DNA` flag; this then includes the C4 and C1' carbon atoms if the `-ca` flag (for C_α only) is used. This interaction matrix is output so that it can be solved directly using either DIAGSTG or blocked in RTBs and solved with DIAGRTB. This simple approach will likely produce useful results when using an original (unprocessed) PDB file, but some modifications of the input data are advisable, e.g. removal of water or buffer molecules. To prevent lumping

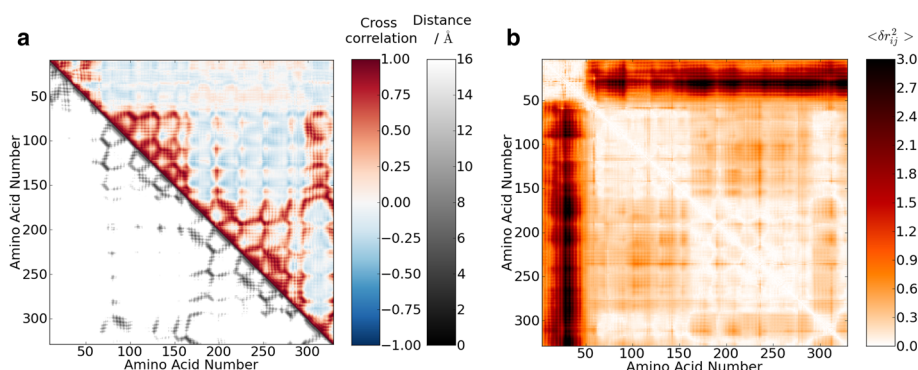


Figure 3 Cross correlation and mean square fluctuations. Plots of (a) the cross correlation of the residue motions and the distance between the C_{α} atoms, and (b) the mean square fluctuations of each residue for LAC (pdb: 1EFA). The cross correlation of the C_{α} atom motion is calculated from Equation 11 which defines how similar the motion direction is, 1 is identical motion, 0 is completely different motion, while -1 is exactly inverse motion. The mean square fluctuations of each residue is calculated from Equation 12 and represents how much the distance between each residue varies during the natural protein motion.

of residues that are part of separate molecules into one RTB residue, different chain identifiers should be used. Alternate amino-acid conformations should be removed (if present) and hydrogen atoms should be erased, as their presence will have only a minor influence on the results but a large effect on the solution time (using the $-ca$ flag will automatically ignore any atoms that are not C_{α} atoms).

After solving the interaction matrix or covariance matrix from a simulation, the eigenvalues and eigenvectors will be output into a single file. These can be analysed with any of the tools mentioned and the normal modes can be conveniently viewed with the NMWIZ plugin for VMD [43]. The NMWIZWT tool will convert the calculated values into the relevant input file for the NMWIZ plugin.

Table 1 $\Delta\Delta$ PT tools

Tool	Description
GENENMM	Generates interaction matrix for an ENM
DIAGSTD	Diagonalises an interaction matrix
DIAGRTB	Diagonalises an interaction matrix with the RTB approximation
DOMAINS	Rewrites pdb into custom domain order for domain RTB
GNMPROD	Calculates the normal modes for a GNM
FREQ/EN	Calculates the frequencies and energies for a set of normal modes
RMS/COL	Calculates the motion, B-factors and collectivity for a set of normal modes
CROSCOR	Calculates the cross correlation for a set of normal modes
OVERLAP	Calculates the overlap of a set of normal modes
GROAMED	Converts Gromacs and Amber output for use with the toolbox
COVAR	PCA decomposition for Gromacs, Amber, or DL_POLY trajectories
FULL2CA	Reduces outputs for full protein into C_{α} only outputs
TRAJPDB	Converts trajectories from Gromacs, Amber, or DL_POLY into a pbc fixed pdb file
NMWIZWT	Produces input for the nmwiz plugin for VMD
PROJECT	Produces a set of pdb files for each normal mode perturbed from the input structure
EGNPROJ	Plots the trajectory frames onto the eigenvector space
PDBDIFF	Produces the vector between two pdb structures
SPACING	Calculates the inter-atom distances for plotting
MOVEING	Calculates the change in atom positions due to the normal modes
PLOTPDB	Writes a file of residue based values onto a pdb for plotting in VMD

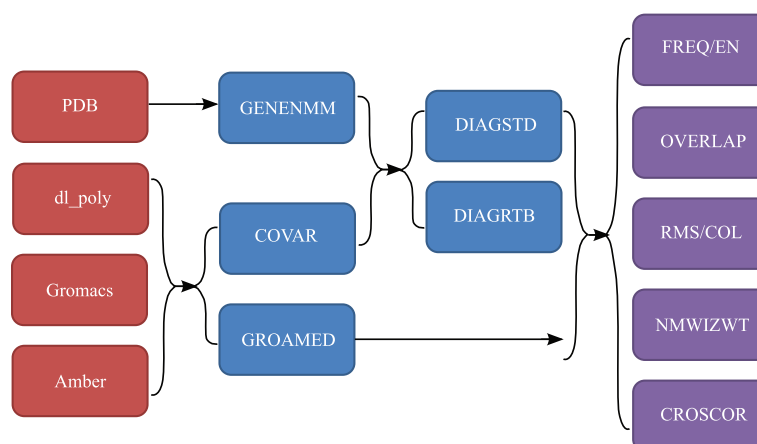


Figure 4 Minimal flow sheet for $\Delta\Delta PT$. Minimal flow sheet for $\Delta\Delta PT$. Red boxes are the types of input files which can be used with $\Delta\Delta PT$, blue boxes are the main processing programs, while red boxes are the subsequent analysis programs provided by $\Delta\Delta PT$.

Table 1 contains a list of, and a brief description of, the programs included in the $\Delta\Delta PT$ toolbox; Figure 4 shows a minimal flow sheet for $\Delta\Delta PT$.

Conclusions

NMA is a powerful tool for the study of protein movements, conformational changes, and protein entropy. It compliments experimental techniques such as X-ray crystallography and NMR, has been used extensively in identifying different structural biology domains, and provides new insights into entropy changes on binding.

This toolbox has increased functionality over those of the currently available web servers, e.g. EL-Nemo [15] and ANM web server [40]. Its main advantages are its abilities to provide the user with tools for analysing elastic network models and molecular dynamics simulations, and for users to add their own extra modules and functions if needed.

This toolbox makes the respective tools available to a wide community of potential NMA users, and allows them unrivalled ability to analyse normal modes using a variety of techniques and current software. With consistent file types, information can be easily exchanged and compared between methods. The availability of a comprehensive and easy-to-use dedicated NMA downloadable software will therefore facilitate further research into this interesting technique.

Availability and requirements

Project name: $\Delta\Delta PT$

Project home page: <https://sourceforge.net/projects/durham-ddpt/>

Operating system(s): Platform independent

Programming language: fortran90

Other requirements: gfortran 4.4.1 or higher, or ifort 11.1 or higher

License: GNU GPL

Any restrictions to use by non-academics: none

Nomenclature

Roman

B	B-factor	—
C	Cross correlation	—
D	Mass weighted Hessian matrix	$\text{J mol}^{-1} \text{m}^{-2} \text{Da}^{-1}$
e	Eigenvector	m Da^{-1}
F	Mass weighted covariance matrix	Da m^2
G	Free energy	J mol^{-1}
$\hbar$	Reduced Plank constant	$1.05457148 \times 10^{-34} \text{m}^2 \text{kg s}^{-1}$
I	Overlap	—
i	Atom number	—
i	Atom number	—
i	Atom number	—
k	Boltzmann constant	$8.314 \text{J mol}^{-1} \text{K}^{-1}$
k_{ij}	Hookean spring constant	$\text{J mol}^{-1} \text{m}^{-2}$
m	Mass	Da
N	Number of atoms	—
Q	Eigenvalue matrix	—
R	Equilibrium distance between atoms	m
r	Distance between atoms	m
S	Entropy	$\text{J mol}^{-1} \text{K}^{-1}$
T	Temperature	K
u	Difference from equilibrium distance between atoms	m
V	Potential Energy	J mol^{-1}
v	Normal mode number	—
x	Atomic position matrix	m

Greek

α	Collectivity constant	—
α	Direction	—

β Direction —
 κ Collectivity —
 ω Eigenfrequency s^{-1}

Endnotes

^aThe Gaussian network model is explicitly represented as $V = \frac{\kappa}{2} \sum_{ij}^N (\mathbf{r}_j \Gamma_{ij} \mathbf{r}_i) [11]$, where $\mathbf{r}$ is the distance between sites and Γ is the Kirchhoff matrix.

^bNote that for the ENM, there are always six frequencies that are several orders of magnitude lower than the others (the eigenvalues of these are essentially zero), these correspond to six solid block rotational and translational modes. If more than six very low frequency normal modes are obtained, this means that a group of atoms is at a distance larger than the cut-off radius from the other atoms.

^cThe B-factors calculated by Equation 10 give only the contribution of the thermal fluctuations while the experimental B-factors also contain contributions from factors [59].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TLR wrote the paper. TLR and DB wrote the toolbox. All authors contributed to the design of the toolbox and substantially edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by EPSRC grant EP/H051759/1.

Author details

¹Biophysical Sciences Institute, Durham University, Durham, UK. ²Department of Chemistry, Durham University, Durham, UK. ³School of Biological and Biomedical Sciences, Durham University, Durham, UK. ⁴Department of Physics, Durham University, Durham, UK.

Received: 15 November 2012 Accepted: 24 May 2013

Published: 7 June 2013

References

- Tama F: **Normal mode analysis with simplified models to investigate the global dynamics of biological systems.** *Protein Pept Lett* 2003, **10**:119–132.
- Go N, Noguti T, Nishikawa T: **Dynamics of a small globular protein in terms of low-frequency vibrational modes.** *Proc Natl Acad Sci U S A* 1983, **80**:3696–3700.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.** *J Comput Chem* 1983, **4**:187–217.
- Brooks B, Karplus M: **Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme.** *Proc Natl Acad Sci U S A* 1985, **82**:4995–4999.
- Mouawad L, Perahia D: **Diagonalization in a mixed basis: a method to compute low-frequency normal modes for large macromolecules.** *Biopolymers* 1993, **33**:599–611.
- Durand P, Trinquier G, Sanejouand YH: **A new approach for determining low-frequency normal modes in macromolecules.** *Biopolymers* 1994, **34**:759–771.
- Marques O, Sanejouand YH: **Hinge-bending motion in citrate synthase arising from normal mode calculations.** *Proteins: Struct Funct Genet* 1995, **23**:557–560.
- Tama F, Gadea FX, Marques O, Sanejouand YH: **Building-block approach for determining low-frequency normal modes of macromolecules.** *Proteins: Struct Funct Genet* 2000, **41**:1–7.
- Li G, Cui Q: **A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+} -ATPase.** *Biophys J* 2002, **83**:2457–2474.
- Tirion KM: **Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.** *Phys Rev Lett* 1996, **77**:1905–1908.
- Bahar I, Atilgan AR, Erman B: **Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.** *Fold Des* 1997, **2**:173–181.
- Hinsen K: **Analysis of domain motions by approximate normal mode calculations.** *Proteins: Struct Funct Genet* 1998, **33**:417–429.
- Hinsen K, Thomas A, Field MJ: **Analysis of domain motion in large proteins.** *Proteins: Struct Funct Genet* 1999, **34**:369–382.
- Tama F: **Conformational change of proteins arising from normal mode calculations.** *Protein Eng* 2001, **14**:1–6.
- Delarue M, Sanejouand YH: **Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model.** *J Mol Biol* 2002, **320**:1011–1024.
- Tama F, Valle M, Frank J, Brooks III CL: **Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy.** *Proc Natl Acad Sci U S A* 2003, **100**:9319–9323.
- Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M: **Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic.** *Proteins: Struct Funct Genet* 2002, **48**:682–695.
- Valadié H, Lacapère JJ, Sanejouand YH, Etchebest C: **Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis.** *J Mol Biol* 2003, **332**:657–674.
- Kim MK, Jernigan RL, Chirikjian GS: **An elastic network model of HK97 capsid maturation.** *J Struct Biol* 2003, **143**:107–117.
- Reuter N, Hinsen K, Lacapère JJ: **Transconformations of the SERCA1 Ca-ATPase: a normal mode study.** *Biophys J* 2003, **85**:2186–2197.
- Thomas A, Hinsen K, Field MJ, Perahia D: **Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study.** *Proteins: Struct Funct Genet* 1999, **34**:96–112.
- Tirion M, ben Avraham D, Lorenz M, Holmes K: **Normal modes as refinement parameters for the F-actin model.** *Biophys J* 1995, **68**:5–12.
- Gofman Y, Shats S, Attali B, Haliloglu T, Ben-Tal N: **How does KCNE1 regulate the Kv7.1 potassium channel? Model-structure, mutations, and dynamics of the Kv7.1-KCNE1 complex.** *Structure* 2012, **20**:1343–1352.
- Liu M, Wang S, Sun T, Su J, Zhang Y, Yue J, Sun Z: **Insight into the structure, dynamics and the unfolding property of amylosucrases: implications of rational engineering on thermostability.** *PLoS ONE* 2012, **7**:e40441.
- Kidera A, Gö N: **Normal mode refinement: crystallographic refinement of protein dynamic structure I. Theory and test by simulated diffraction data.** *J Mol Biol* 1992, **225**:457–475.
- Diamond R: **On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor.** *Acta Crystallogr A* 1990, **46**:425–435.
- Miller WW, Agard DA: **Enzyme specificity under dynamic control: a normal mode analysis of α -lytic protease.** *J Mol Biol* 1999, **286**:267–278.
- Faure P, Micu A, Pérahia D, Doucet J, Smith J, Benoit J: **Correlated intramolecular motions and diffuse X-ray scattering in lysozyme.** *Nat Struct Mol Biol* 1994, **1**:124–128.
- Wang Z, Schröder GF: **Real-space refinement with DireX: from global fitting to side-chain improvements.** *Biopolymers* 2012, **97**:687–697.
- Gaillard T, Martin E, San Sebastian E, Cossio FP, Lopez X, Dejaegere A, Stote RH: **Comparative normal mode analysis of LFA-1 Integrin I-domains.** *J Mol Biol* 2007, **374**:231–249.
- Nojima H, Takeda-Shitaka M, Kanou K, Kamiya K, Umeyama H: **Dynamic interaction among the platform domain and two MembraneProximal immunoglobulin-like domains of class I major histocompatibility complex: normal mode analysis.** *Chem Pharm Bull* 2008, **56**:635–641.
- Kurihara Y, Watanabe T, Nojima H, Takeda-Shitaka M, Sumikawa H, Kamiya K, Umeyama H: **Dynamic character of human growth**

- hormone and its receptor: normal mode analysis.** *Chem Pharm Bull* 2003, **51**:754–758.
33. Rodgers TL, Burnell D, Wilson MR, Pohl E, Cann M, Townsend PD, McLeish TCB, Toncrova H: **Modelling allosteric signalling in protein homodimers.** *Eur Biophys J Biophys Lett* 2011, **40**:121.
 34. Mouawad L, Perahia D: **Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, Quaternary R-like intermediate structures.** *J Mol Biol* 1996, **258**:393–410.
 35. Tirion MM, ben Avraham D: **Normal mode analysis of G-actin.** *J Mol Biol* 1993, **230**:186–195.
 36. Jackson JE: *A User's Guide to Principal Components*. New York: John Wiley & Sons, Inc.; 1991.
 37. Hayward S, Kitao A, Gö N: **Harmonic and Anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis.** *Protein Sci* 1994, **3**:936–943.
 38. Ramanathan A, Agarwal PK: **Computational identification of slow conformational fluctuations in proteins.** *J Phys Chem B* 2009, **113**:16669–16680.
 39. Gorfe AA, Grant BJ, McCammon JA: **Mapping the nucleotide and Isoform-dependent structural and dynamical features of Ras proteins.** *Structure* 2008, **16**:885–896.
 40. Eyal E, Yang LW, Bahar I: **Anisotropic network model: systematic evaluation and a new web interface.** *Bioinformatics* 2006, **22**:2619–2627.
 41. Camps J, Carrillo O, Emperador A, Orellana L, Hospital A, Rueda M, Cicin-Sain D, D'Abramo M, Gelpi JL, Orozco M: **FlexServ: an integrated tool for the analysis of protein flexibility.** *Bioinformatics* 2009, **25**:1709–1710.
 42. Hinsen K: **The molecular modeling toolkit: a new approach to molecular simulations.** *J Comput Chem* 2000, **21**:79–85.
 43. Bakan A, Meireles LM, Bahar I: **ProDy: protein dynamics inferred from theory and experiments.** *Bioinformatics* 2011, **27**:1575–1577.
 44. Hess B, Kutzner C, van der Spoel D, Lindahl E: **GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation.** *J Chem Theory Comput* 2008, **4**:435–447.
 45. Case D, Darden T, Cheatham III T, Simmerling C, Wang J, Duke R, Luo R, Walker R, Zhang W, Merz K, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolossvai I, Wong K, Paesani F, Vanicek J, Liu J, Wu X, Brozell S, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh MJ, Cui G, Roe D, et al: **AMBER 11.** San Francisco, US: University of California; 2010.
 46. Todorov I, Smith W, Trachenko K, Dove M: **DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism.** *J Mater Chem* 2006, **16**:1911–1918.
 47. Smith W, Todorov IT: **A short description of DL POLY.** *Mol Simul* 2006, **32**:935–943.
 48. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG: **Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation.** *Nature* 2011, **474**:521–525.
 49. Frishman D, Argos P: **Knowledge-based secondary structure assignment.** *Proteins: Struct Funct Genet* 1995, **23**:566–579.
 50. Doruker P, Jernigan RL, Bahar I: **Dynamics of large proteins through hierarchical levels of coarse-grained structures.** *J Comput Chem* 2002, **23**:119–127.
 51. Brink J, Ludtke SJ, Kong Y, Wakil SJ, Ma J, Chiu W: **Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis.** *Structure* 2004, **12**:185–191.
 52. Dykeman EC, Sankey OF: **Normal mode analysis and applications in biological physics.** *J Phys Condensed Matter* 2010, **22**:423202.
 53. Schlitter J: **Estimation of absolute and relative entropies of macromolecules using the covariance matrix.** *Chem Phys Lett* 1993, **215**:617–621.
 54. Brüschweiler R: **Collective protein dynamics and nuclear spin relaxation.** *J Chem Phys* 1995, **102**:3396–3403.
 55. Pohl E, Holmes RK, Hol WG: **(Motion of the DNA-binding domain with respect to the core of the diphtheria toxin repressor (DtxR) revealed in the crystal structures of apo- and holo-DtxR.** *J Biol Chem* 1998, **273**:22420–22427.
 56. Russo S, Schweitzer JE, Polen T, Bott M, Pohl E: **Crystal structure of the caseinolytic protease gene regulator, a transcriptional activator in actinomycetes.** *J Biol Chem* 2009, **284**:5208–5216.
 57. Ichiye T, Karplus M: **Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations.** *Proteins* 1991, **11**:205–217.
 58. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The protein data bank.** *Nucleic Acids Res* 2000, **28**:235–242.
 59. Hinsen K: **Structural flexibility in proteins: impact of the crystal environment.** *Bioinformatics* 2008, **24**:521–528.

doi:10.1186/1471-2105-14-183

Cite this article as: Rodgers et al.: $\Delta\Delta$ PT: a comprehensive toolbox for the analysis of protein motion. *BMC Bioinformatics* 2013 **14**:183.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

