



HAL
open science

NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis

Erik Lindahl, Cyril Azuara, Patrice Koehl, Marc Delarue

► To cite this version:

Erik Lindahl, Cyril Azuara, Patrice Koehl, Marc Delarue. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Research*, 2006, 34 (Issue suppl_2), pp.W52-W56. 10.1093/nar/gkl082 . pasteur-02175154

HAL Id: pasteur-02175154

<https://pasteur.hal.science/pasteur-02175154v1>

Submitted on 5 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis

Erik Lindahl^{1,2}, Cyril Azuara¹, Patrice Koehl³ and Marc Delarue^{1,*}

¹Unite de Dynamique Structurale des Macromolécules, URA 2185 du C.N.R.S., Institut Pasteur, 75015 Paris, France, ²Stockholm Bioinformatics Center, Stockholm University, 106 91 Stockholm, Sweden and ³Computer Science Department and Genome Center, University of California, Davis, CA 95616, USA

Received February 14, 2006; Revised and Accepted March 6, 2006

ABSTRACT

Normal mode analysis (NMA) is an efficient way to study collective motions in biomolecules that bypasses the computational costs and many limitations associated with full dynamics simulations. The NOMAD-Ref web server presented here provides tools for online calculation of the normal modes of large molecules (up to 100 000 atoms) maintaining a full all-atom representation of their structures, as well as access to a number of programs that utilize these collective motions for deformation and refinement of biomolecular structures. Applications include the generation of sets of decoys with correct stereochemistry but arbitrary large amplitude movements, the quantification of the overlap between alternative conformations of a molecule, refinement of structures against experimental data, such as X-ray diffraction structure factors or Cryo-EM maps and optimization of docked complexes by modeling receptor/ligand flexibility through normal mode motions. The server can be accessed at the URL <http://lorentz.immstr.pasteur.fr/nomad-ref.php>.

INTRODUCTION

Structural flexibility is an important property of most biological macromolecules, and often crucial for substrate or drug binding or protein–protein interactions (1). Collective normal mode motions provide a unique way to tackle this flexibility problem, and can therefore be very efficient in principle to describe structural changes between homologous proteins or in solving crystal structures through molecular replacement techniques.

Normal modes are straightforward to calculate, particularly in the simplified framework of elastic network models (ENMs) (2–4), and provide a basis set of orthogonal vectors to drive a conformational transition with as few degrees of freedom as possible; emphasizing the large amplitude and collective movements if one focuses on low-frequency modes. While the underlying model is a coarse-grained one (no solvent, frequency scale is arbitrary) it turns out that the low-frequency motions are remarkably conserved using different models of increasing complexity (4).

Gerstein and coworkers (5) showed that it is useful to explain known structural transitions as documented in their database of proteins whose structure has been solved in at least two different conformations. Indeed, an average of only 2 modes is involved in known structural transitions, generally identified among the first (slowest) 10–15 lowest frequency ones. This result has been used to build databases of protein movements, based both on experimental structures and normal mode analysis (NMA) (6–8). Amplitudes are generally adjusted to match a chosen cRMS, after applying thermal averaging.

NMA has proved useful for structural refinement against experimental data (9,10). The addition of a small number of collective degrees of freedom is sufficient to capture most of the intrinsic flexibility of the macromolecule, while retaining local connectivity and stereochemical properties. In contrast to using rigid bodies, NMA is almost model-free, and the level of detail can be adjusted freely by changing the number of modes used. In some sense, normal modes can be regarded as completely arbitrary collective displacements. The fact that they provide such an efficient refinement space suggests however that they actually capture the most important biological motions, with obvious applications to docking methods and drug design in the presence of induced fit (11–13).

Here we describe NOMAD-Ref, a web server that provides access to a number of online tools that calculate and use normal modes for visualization and refinement problems. A flow chart of the different options is given in Figure 1.

*To whom correspondence should be addressed. Tel: +33 1 45 68 86 05; Fax: +33 1 45 68 86 04; Email: delarue@pasteur.fr

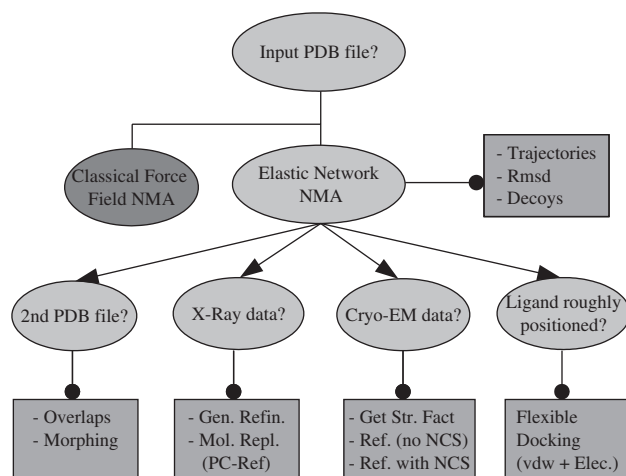


Figure 1. Flow chart of the NOMAD-Ref server.

The next section describes the underlying formalism. The result section clarifies the use of the web server through test applications. We conclude with a description of future work centered on NOMAD-Ref.

MATERIALS AND METHODS

NMA and visualization

Normal modes are simply the eigenvectors of the Hessian matrix obtained from an approximation of an energy landscape around a local minimum. This is theoretically straightforward to calculate for classical force fields provided all atoms are present in the structure and that a local minimum has been located. To get the molecule to a local minimum requires however a CPU intensive minimization that frequently leads to major distortion, not to mention the prohibitive memory and CPU requirements during the normal mode calculation.

Paradoxically, the properties of the low-frequency modes are almost entirely insensitive to force field details—they only seem to be affected by the overall molecular connectivity. Tirion (2) was the first to note this and introduced what became later the ENM where any molecular system is plainly represented by a set of harmonic potentials between all atoms within a given cutoff—usually in the order of 10 Å. A simplified version using only C α coordinates and a $N \times N$ Kirchhoff matrix (3), the so-called Gaussian Network Model, yielded cRMS residue fluctuations. Subsequently, a $3N \times 3N$ Hessian matrix was used (4), whose eigenvectors gave the directions of each mode for each C α . The striking simplicity of this method has made it quite popular (14,15). Computation of elastic normal modes does not require any prior energy minimization since the starting state is designed to be the global minimum; there are virtually no size limitation for the molecules and missing side chains or even backbone segments can be handled transparently. The cutoff length and the interaction weight are the only adjustable parameters (see below). Elastic normal modes are ideally suited to study global collective motions since interatomic distances tend to be preserved, and the low computational cost makes them perfect for online usage. Some of the currently available web servers that implement elastic normal mode calculations include Elnemo (16)

(<http://igs-server.cnrs-mrs.fr/elnetmo/>), Webnm@ (17) (<http://www.bioinfo.no/tools/normalmodes>), and AD-ENM (18) (<http://enm.lobos.nih.gov/>).

In most implementations only C α coordinates are retained in the actual Elastic Network (4), but a rotation-translation block (19) (RTB) approach with rigid residues can be used if all coordinates are needed. For the NOMAD-Ref server, we have additionally implemented sparse matrix data storage and sparse diagonalization using the ARPACK library (20) for the Hessian matrices, which makes it possible to retain all atomic coordinates and degrees of freedom in the calculation to obtain true eigenvectors to the full system, even for structures with over 100 000 atoms.

The strength of real pairwise potentials decays with distance, so to reduce the effect of the fairly arbitrary cutoff the server supports an exponential interaction weight parameter as first proposed by Hinsen (4). The user can further select the type of diagonalization algorithm to use, as well as the output mode amplitudes. Results for each mode are presented in terms of raw eigenmode vectors, relative frequency, mode collectivity measures, cRMS displacement as a function of residue index, and finally as PDB format output trajectories that can be played in visualization programs such as PyMol (21) (<http://www.pymol.org>) and rendered into movies.

For comparison and reference the NOMAD-Ref server also supports structure minimization and normal mode calculation with true force fields, using the GROMACS package (22). For server performance reasons, this is currently limited to structures with less than 3000 atoms.

Normal mode-based deformation

The most obvious application of normal modes is the analysis of structural flexibility, for instance how well they describe transformations between a pair of structures, such as the open and closed conformations of a receptor. The server provides functionality to calculate ‘overlap coefficients’ (scalar products) between the coordinate difference vectors of two superimposed structures and the eigenvectors of the 100 lowest frequency normal modes (23). It is also possible to generate plausible continuous PDB trajectories between two given forms of a macromolecule (currently restricted to C α representation) for visualization in PyMol, using the algorithm of Kim, Jernigan and Chirikjian (24). Finally, the normal modes provides an excellent way to generate artificial ‘decoy’ structures with mostly correct stereochemistry around the initial state, e.g. to test database-derived potentials or benchmark refinement algorithms. Decoy generation is accomplished by randomly assigning amplitudes to low-frequency modes, with subsequent scaling to reproduce a user-specified average cRMS value (e.g. 3.5 Å) for the produced set of structures.

Structure refinement

NOMAD-Ref provides access to two different options for refinement of flexible models against low or medium resolution experimental data using normal mode amplitudes as degrees of freedom. The data can be either X-ray diffraction data or cryo-EM data; the actual refinement is carried out with a conjugate gradient algorithm in reciprocal space. A complete description of the method can be found in Ref. (9). The user can choose the resolution of the data and the number of modes,

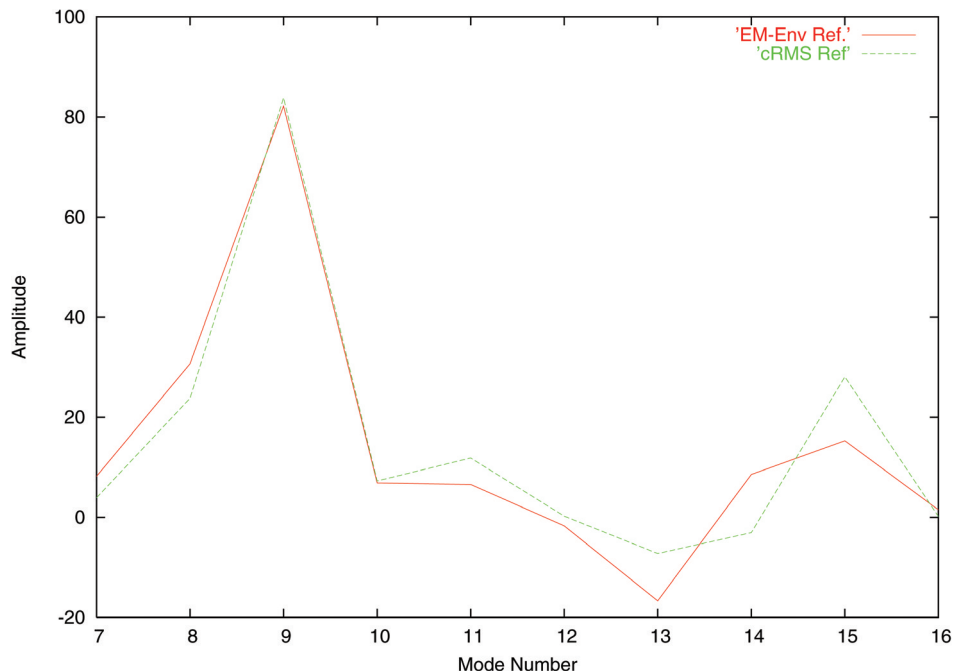


Figure 2. Refinement of the first 10 lowest frequency amplitudes of citrate synthase open form (5CSC) based on simulated Cryo-EM data at 15 Å resolution for the closed form (6SCS). In green, the control experiment is made by conducting the refinement against a cRMS score between the two forms.

including the first six modes, thereby allowing for correcting for any initial slight mispositioning of the model. A new feature has been added that allows pre-scanning the amplitudes of each mode in the range specified by the user; the minimizer then starts from the amplitudes found by these 1D scans. Another option intended to help in difficult molecular replacement cases has also been implemented, whereby the user can submit a list of rotation (Eulerian) angles and translations. The program will then try to optimize each of these oriented and positioned models one by one, using the desired number of modes. If used in P1 instead of the true space group, this is equivalent to PC-refinement (25) and can therefore be used just after the rotation function to identify the correct solution.

For cryo-EM data, a phased correlation coefficient involving $\mathbf{F}\mathbf{F}^*$ products (with \mathbf{F} the complex structure factor and \mathbf{F}^* its complex conjugate) are used instead of the X-ray correlation coefficient, which uses only structure factor amplitudes. The web site includes an option to get the complex structure factors of the cryo-EM map, using CCP4 tools (26). Two versions of the cryo-EM refinement program have been implemented: the first one accepts classical crystallographic space groups, and the second one assumes there exist some internal symmetry in the object to be fitted in the cryo-EM map, which is supposed to be in P1 spacegroup; the corresponding symmetry operators are read from a file containing their rotation matrices and translations. In this way, very large models (e.g. entire viral particles) can be studied with only limited memory requirements, as only the monomer coordinates are needed. Examples are provided for each option.

Finally, the website further provides services for direct-space refinement of docked complex structures; primarily modeling of receptor structural change upon binding a small rigid ligand or vice versa (frequently the case for DNA). This optimization is done entirely without experimental

data to guide, and instead is based on the nonbonded interactions energy between the two molecules. The receptor distortion can simultaneously be controlled by adding restraints on the mode amplitudes. The entire normal mode docking optimization has recently been described in a related paper (11).

RESULTS

We focus on the application of NMA for Cryo-EM refinement, and refer the reader to the original papers for a description of NMA for the refinement of X-ray data (9), and for the application of NMA in docking experiments (11).

In the original article (9) lattice points filling the closed form envelope of Citrate synthase, a dimer of about 850 residues (6CSC), were used to calculate normal modes and deform them into the open form calculated envelope (5CSC). Here we show the result of the direct refinement of the α -coordinates taken from the open form into the electron density for the closed form calculated at low resolution (15 or even 25 Å). The refined amplitudes for the first 10 lowest frequency normal modes are very close to the ones obtained when one just minimizes the cRMS between the two forms (see Figure 2). The radius of convergence of this refinement is therefore much larger than conventional refinement methods.

The next test involves the experimental map used by Hinsen and colleagues [kindly provided by the authors, see also Ref. (27)]: the phased correlation coefficient of the initial model increased from 0.383 to 0.568 using 21 (15 + 6) modes and data between 100 and 10 Å resolution. As a control, the final model of Hinsen and coworkers, obtained in a completely independent way, had a phased correlation coefficient of 0.547. The program also outputs a list of violations of α - α distances in the refined model. If there are too many such

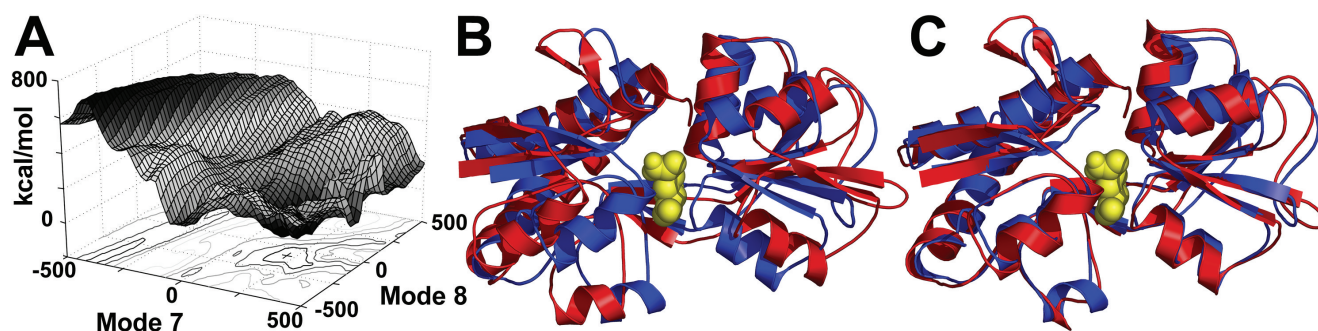


Figure 3. Receptor-ligand docking refinement for glutamine-binding protein (1GGG/1WDN). (A) Intermolecular soft-core nonbonded energy landscape as a function of the two first non-rigid-body normal mode amplitudes. (B) Ligand docked with initial (open) receptor conformation (red) superimposed on closed state (blue)—5.33 Å cRMS. (C) Receptor structure refined using the five lowest normal modes compared to the target closed state—2.16 Å cRMS. The five degrees of freedom reduce the cRMS by a factor two.

violations, the model should be considered as doubtful and the procedure should be repeated with a different number of modes.

To test non-crystallographic symmetry (NCS), we used the experimental cryo-EM map of the ATP-bound form of GroEL (28). This map was Fourier transformed to get structure factors between 25 and 150 Å resolution (1225 reflections, out of which 10% were left out of the refinement process to calculate a CC-free agreement factor). Normal modes were calculated on the monomer and the program made full use of the 7-fold NCS. Normal Modes 1–21 were allowed to adjust their amplitudes, resulting in an increase of the phased correlation coefficient from 0.55 (CC-free = 0.64) to 0.83 (CC-free = 0.91). The first six normal modes are rigid-body overall modes (three rotations and three translations), allowing to adjust potential initial mispositioning of the model in the map.

Figure 3 provides an illustration example of docking optimization for the Glutamine binding protein. The cRMS difference between the open (1GGG) and closed (1WDN) forms is as large as 5.33 Å. By using a small number of normal mode degrees of freedom (1–5) and soft-core interactions it was possible to refine the open receptor structure with a docked ligand down to 2.16 Å. In contrast, full Cartesian energy minimization just deteriorates the complex structure (11).

WEB SERVER IMPLEMENTATION

Usage

The general organization of the NOMAD-Ref web server is as follows: applications are listed in a menu, with links to a short description of the underlying algorithm, to examples for the method, and to a form for submitting a job. Once a job is submitted, it is put in a batch queue currently running on a dual Opteron server. No registration is required, and the user is immediately forwarded to a status page that is automatically refreshed until the run is finished—the job queue can also be tracked in real-time online. All results are stored in a non-published location on the server for two weeks, so only the user can retrieve the results. For convenience it is also possible to provide an email address for automatic notification with a web link if a job is large or the server queue heavily loaded.

Input/output formats

All input structures are submitted as standard PDB files. Both ATOM and HETATM records are included, but alternate-residue-flagged lines discarded. A handful of algorithms are limited to C α -only PDB files but this is always clearly stated on the submission page. Symmetry data for X-ray refinement is represented in AMoRe format (29) with several examples on the server. Reflection data are described in a free-format text file with one reflection per line: the three Miller indices (**h,k,l**), the amplitude and error (or the phase, in the case of EM maps refinement).

All output structures are written in PDB format, including normal mode trajectories consisting of several models that can be viewed as movies e.g. in PyMol or VMD. Raw normal mode data can be downloaded as compressed text files, and other results are presented interactively online.

Performance

Normal mode calculations are usually fast, except for very large structures. As a typical example, calculating the lowest 50 modes of a 3000-atom protein with full atomic detail (i.e. 9000 degrees of freedom) takes less than 60 s with the ENM, and about 20 min including minimization if the GROMACS force field-based method is used. Note that memory and execution time requirements do increase as more modes are required. As of January 2006, the site has been active for about 18 months and served almost 2000 submissions.

FUTURE WORK

New methods are continuously being added to the web server, and algorithms improved. Projects currently under development and/or testing include more efficient optimization procedures based on simulated annealing instead of simple minimization, and options to gradually increment the number of degrees of freedom as the optimization is advancing in order to enable both low- and high-resolution refinement in a single submission. For receptor-ligand docking we are working on enabling simultaneous normal mode flexibility in both structures, and also better prediction of the active site surface complementarity to increase the radius of convergence. Recent work by Tobi shows that protein interactions are well

described by normal modes, at least in some cases, which suggests another highly interesting refinement/prediction application.

ACKNOWLEDGEMENTS

The authors thank K. Hinsén and J. Lacapère for providing them with the experimental ATCase cryo-EM map. The authors acknowledge financial support from C.N.R.S. (ACI IMPB-405 and GDR 2417). Funding to pay the Open Access publication charges for this article was provided by Swedish Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Wodak, S.J. and Mendez, R. (2004) Predictions of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.*, **14**, 242–249.
2. Tirion, M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, **77**, 1905–1908.
3. Bahar, I., Atligan, A.R. and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.*, **2**, 173–181.
4. Hinsén, K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins: Struct. Funct. and Genet.*, **33**, 417–429.
5. Krebs, W.G., Alexandrov, V., Wilson, C.A., Echols, N., Yu, H. and Gerstein, M. (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins: Struct. Funct. and Genet.*, **48**, 682–695.
6. Alexandrov, V., Lehnert, U., Echols, N., Milburn, D., Engelman, D. and Gerstein, M. (2005) Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. *Prot. Sci.*, **14**, 633–643.
7. Yang, L.W., Liu, X., Jursa, C.J., Holliman, M., Rader, A.J., Karimi, H.A. and Bahar, I. (2005) iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics*, **21**, 2978–2987.
8. Wako, H., Kato, M. and Endo, S. (2005) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, **20**, 2035–2043.
9. Delarue, M. and Dumas, P. (2004) On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci.*, **101**, 6957–6962.
10. Suhre, K. and Sanejouand, Y.-H. (2004) On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallogr.*, **D60**, 796–799.
11. Lindahl, E. and Delarue, M. (2005) Docking refinement using low frequency normal mode amplitude optimization. *Nucleic Acids Res.*, **33**, 4496–4506.
12. Cavasotto, C.N., Kovacs, J.A. and Abagyan, R.A. (2005) Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.*, **127**, 9632–9640.
13. May, A. and Zacharias, M. (2005) Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochem. Biophys. Acta*, **1754**, 225–231.
14. Delarue, M. and Sanejouand, Y.-H. (2002) Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.*, **320**, 1011–1024.
15. Bahar, I. and Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology. *Curr. Op. Struct. Biol.*, **15**, 586–592.
16. Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, **32**, W610–W614.
17. Hollup, S.M., Salensminde, G. and Reuter, N. (2005) WEBnm@: a web application for normal mode analysis of proteins. *BMC Bioinformatics*, **6**, 52.
18. Zheng, W. and Brooks, B.R. (2005) Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. *Biophys. J.*, **89**, 167–178.
19. Tama, F., Gadea, F.X., Marques, O. and Sanejouand, Y.-H. (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. and Genet.*, **31**, 1–7.
20. Lehoucq, R.B., Sorensen, D.C. and Yang, C. (1998) *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia. ISBN: 0-89871-407-9.
21. DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA.
22. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. (2005) GROMACS: fast, flexible and free. *J. Comp. Chem.*, **26**, 1701–1719.
23. Tama, F. and Sanejouand, Y.-H. (2001) Conformational change of proteins arising from normal mode calculations. *Prot. Eng.*, **14**, 1–6.
24. Kim, M.K., Jernigan, R.L. and Chirikjian, G.S. (2002) Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.*, **83**, 1620–1630.
25. DeLano, W. and Brunger, A.T. (1995) The direct rotation function: patterson correlation search applied to molecular replacement. *Acta Crystallogr.*, **D51**, 740–748.
26. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
27. Hinsén, K., Reuter, N., Navaza, J., Stokes, D.L. and Lacapère, J.J. (2005) Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.*, **88**, 818–827.
28. Saibil, H.R., Horwich, A.L. and Fenton, W.A. (2001) ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, **107**, 869–879.
29. Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst.*, **A50**, 157–163.
30. Tobin, D. and Bahar, I. (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci.*, **102**, 18908–18913.