Nucleic Acids Research, 2006, Vol. 34, Web Server issue ***W57–W62***
doi:10.1093/nar/gkl039

UMMS: constrained harmonic and anharmonic analyses of macromolecules based on elastic network models

Yunho Jang, Jay I. Jeong[1] and Moon K. Kim*

Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA 01003, USA and [1]School of Mechanical and Automotive Engineering, Kookmin University, Seoul, South Korea

Received October 28, 2005; Revised December 22, 2005; Accepted January 13, 2006

ABSTRACT

UMass Morph Server (UMMS) has been developed for the broad impact on the study of molecular dynamics (MD). The elastic network model (ENM) of a given macromolecule has been proven as a useful tool for analyzing thermal behaviors locally and predicting folding pathways globally. UMMS utilizes coarse-grained ENMs at various levels. These simplifications remarkably save computation time compared with all-atom MD simulations so that one can bring down massive computational problems from a supercomputer to a PC. To improve computational efficiency and physical reality of ENMs, the symmetry-constrained, rigid-cluster, hybrid and chemical-bond ENMs have been developed and implemented at UMMS. One can request both harmonic normal mode analysis of a single macromolecule and anharmonic pathway generation between two conformations of a same molecule using elastic network interpolation at http://biomechanics.ecs.umass.edu/umms.html.

INTRODUCTION

As a number of protein and nucleic acid structures have been obtained experimentally and deposited in the Protein Data Bank (1), biological research areas such as computational biology, bioinformatics and protein dynamics have been growing rapidly. Among the many ways to analyze the dynamic characteristics of macromolecules, the conventional engineering disciplines such as kinematics and mechanical vibrations have been proved as a powerful tool to reduce computational cost and the generated results have shown a good agreement with the experimentally observed dynamics of macromolecules. From this fact, it is convincing that they can play an important role in developing much more computationally efficient methods than traditional molecular dynamics (MD) simulations, as well as establishing theoretical foundation for linking the structural information of macromolecules to their biological functions.

MD is one of the most common tools we have widely used for the simulation of macromolecular dynamics. By computing the interactions of all forces between atoms involved (2,3), MD provides time-dependent behaviors of a molecular system including the effects of surrounding solvent at atomic detail. However, empirically-derived force fields yield problems with local minima traps and the current computer power limits the investigation of the relevant dynamics to only its early stage. Moreover, even if one can simulate millions of MD time steps, it is very difficult to interpret the output data in order to get the long time-scale collective motions because the obtained time-evolving conformations fluctuate rapidly near local minima like Brownian motion.

Alternatively, normal mode analysis (NMA), a traditional harmonic analysis tool in mechanical vibrations, has been used to calculate thermal fluctuations of a macromolecule around its energy equilibrium conformation (4–6). In NMA, an all-atom empirical potential function is approximated as a harmonic function. One can obtain vibrational frequencies and directions of corresponding motions of the system by solving the generalized eigenvalue problem of the stiffness (Hessian) matrix. To reduce computational burden of all-atom NMA in the case of large macromolecules, coarse-grained elastic network models (ENMs) have been developed (7–9). For example, a protein structure is modeled as a spring network among representative alpha-carbon atoms.

In these days, many web-based services for calculating normal modes are available. ProMode is a database of NMA of proteins in which the global stiffness matrix is derived from an empirical potential function with respect to only dihedral angles (10). In contrast with ProMode, ElNe'mo uses a much simplified Hookean potential with respect to Cartesian coordinates of atoms involved (11).

*To whom correspondence should be addressed. Tel: +1 413 545 2357; Fax: +1 413 545 1027; Email: mkkim@ecs.umass.edu

The rotation-translation block (RTB) approximation is uniquely implemented to solve the limitation of the size of system. Three other online servers MoViEs (12), WEBnm (13) and *i*GNM (14) are launched recently to perform various types of structural analyses including NMA and B-factor calculation. NMA has been also used to refine docked protein–ligand and protein–DNA structures (15).

However, NMA is not able to predict anharmonic and large pathways because the ensuing prediction of motions only occurs near the equilibrium state. For conformational transitions between meta-stable conformations, several interpolation strategies such as linear interpolation, internal variable interpolation and elastic network interpolation (ENI) have been used. ENI is a purely geometry-based technique developed by the author (16,17). The essence of ENI is to uniformly interpolate the distances in two different conformations within the context of the coarse-grained ENM.

Currently the most widely used morph server is the 'Database of Macromolecular Movements' run by Gerstein at Yale University (http://www.molmovdb.org/) (18). This server has used the adiabatic mapping in which predetermined hinge or shear regions in macromolecules are linearly interpolated and each iteration step is followed by energy minimization of the computed intermediate conformations to produce chemically reasonable morphs.

Although the conventional coarse-graining techniques substantially reduce the degrees-of-freedom of the system, harmonic and anharmonic analyses for large macromolecules such as GroEL-GroES complex, virus capsid and ribosome are still impracticable in a PC (19). Motivated by this situation, we introduce further simplified models such as symmetry-constrained, rigid-cluster, hybrid and chemical-bond based ENMs. These new features can bring down massive computational problems, which have been traditionally solved in a supercomputer, to a PC level. For more details, see the following section of applications.

As the demand on computational (theoretical) approaches to MD has increased, we have developed UMass Morph Server (UMMS) to provide one with both harmonic and anharmonic analysis tools online. UMMS cannot only cross-validate the results of other morph servers based on NMA (10–14) and other interpolation techniques (18), but also provide several unique features described above to improve computational efficiency and physical reality of the models.

UMASS MORPH SERVER

Figure 1 illustrates the flow chart of UMMS proposed here. Once the server receives the request of NMA and/or ENI for a particular macromolecule, the first step is to standardize the set of coordinates of representative atoms. In ENI, a sequence alignment is necessary to make sure that the given two sets of coordinates have the same number of residues for the purpose of interpolation. Next, we can create the linking matrix based on the input cutoff values. Using this geometry-based elastic network, NMA and ENI will be performed. Finally, the mode shapes calculated from NMA and the intermediate conformations morphed by ENI will be visualized with the 3D molecular graphics program Rasmol (20).

One of the important features of UMMS is that it has more user-oriented interface than other servers do. For example, UMMS allows one to override the default setting for both the coarse-graining approaches (e.g. all-atom, backbone and alpha-carbon only) and the spring connection (cutoff) rules.

In the case of NMA, specific cutoff distances are provided for each coarse-grained method in order to generate an elastic network mimicking both chemically bonded and non-bonded interactions within a macromolecule. For instance, the range from 8 to 16 Å has been chosen empirically for an alpha-carbon coarse-grained ENM to get the robust outputs in NMA, which mathematically means that there are no more than six zero eigenvalues. These cutoff values are somewhat unrealistic (overestimated) compared with the actual van der Waals interaction range. On the other hand, a relatively short cutoff range from 5 to 8 Å can be selected for a chemical-bond ENM because there are additional backbone connections modeled to stabilize the network (see Chemical-bond ENM section).

Figure 2 illustrates the UMMS homepage, the graphic user interface for query building, and the result windows from the standard NMA and ENI, respectively. One can first browse the NMA or ENI database of macromolecules posted on UMMS in order to find one of interest. Then one can also submit the query specifying its PDB name and simulation options. All of requests are to be processed offline and the simulation results will be uploaded to UMMS. The animated GIF movies and the computed data files are downloadable at the server.

APPLICATIONS

As aforementioned, NMA has been used to predict harmonic fluctuations (i.e. flexibilities) of a single macromolecule at an equilibrium conformation, whereas ENI has been used to generate anharmonic transition pathways between two different conformations without any steric clashes. However, harmonic and anharmonic analyses for large macromolecules may be still impracticable in a PC because of memory limitation. In addition, the current ENI method takes into account only geometric information with linear spring connections replacing chemically oriented interactions between atoms. It means that the computed pathways cannot be considered as time-evolving motions like MD results, but be interpreted as plausible conformational changes between two given conformers. The following applications aim to not only simplify the current ENMs but also get more realistic simulation results.

Symmetry-constrained ENM

The assembly of repeated units is one of the ways that nature adopts to produce large macromolecular structures. For example, most viruses adopt this strategy to reproduce their protein shells called capsids. Only a short sequence of nucleic acids is needed to encode a small repeated unit of a capsid. Therefore, one can significantly reduce computational cost for NMA or ENI of a symmetric system by using its symmetry feature induced by the manner of assembly. This symmetry-constrained NMA and ENI have been applied for the study of maturation process of the HK97 virus capsid (21). The degree-of-freedom of this symmetric system is reduced by the factor of 60. The resulting symmetric normal modes explain the

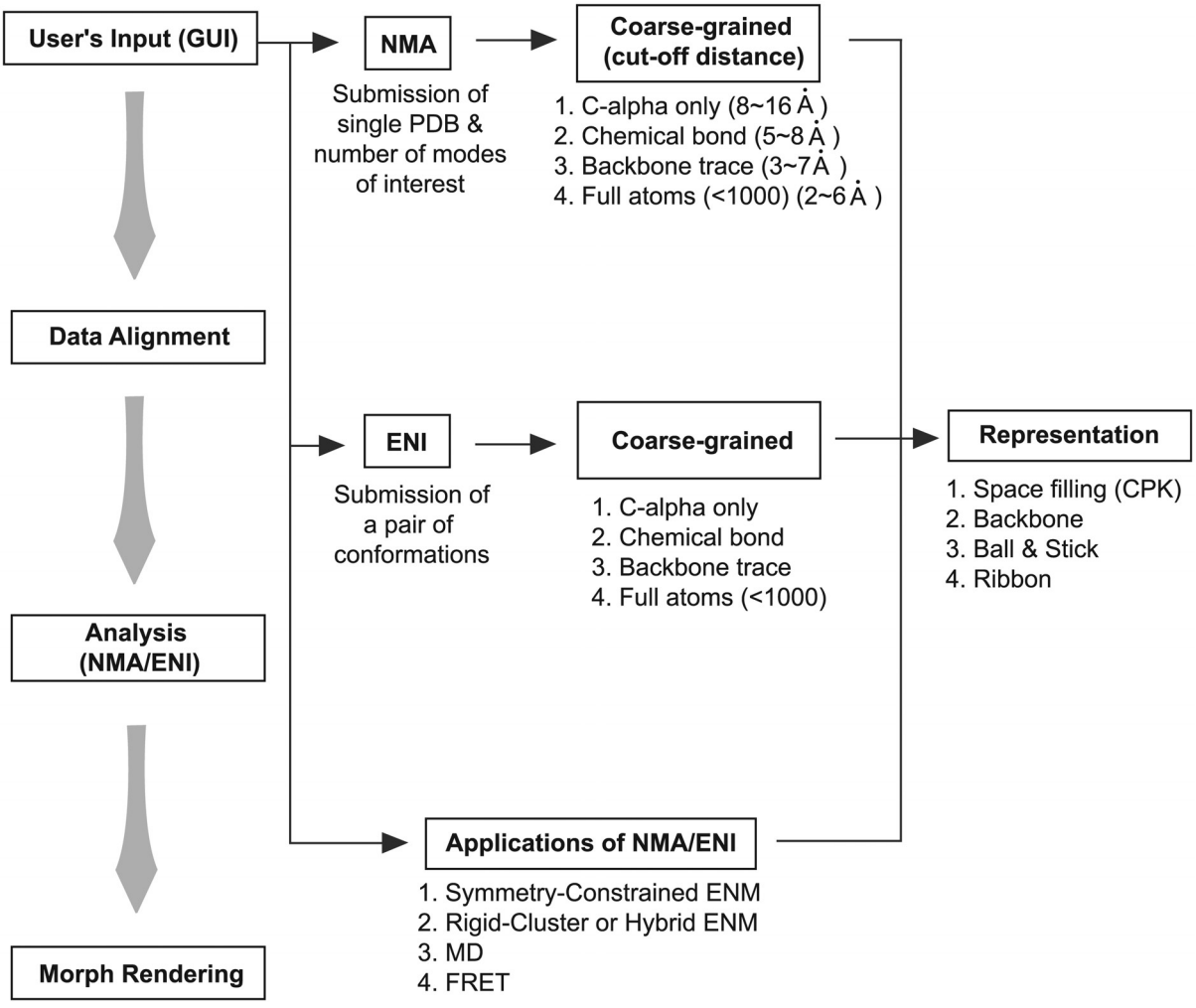

Figure 1. Schematic representation of the working procedure of UMMS. Harmonic NMA and anharmonic pathway generation can be performed based on user's requests. The default setting of coarse-graining method for protein is to take only alpha-carbon atoms per residue. Its resolution is probably enough to investigate the global dynamics of proteins in most of cases. For small proteins (approximately <1000 atoms), backbone-trace ENMs including only heavy atoms along the protein backbone and full-atom ENMs are also practicable in a PC. Regardless of the level of coarse-graining of ENMs, a single-parameter Hookean potential is defined in Cartesian space and then differentiated resulting in the stiffness matrix for NMA. MD data interpretation and incorporation with experimental data such as FRET and NMR are also available at UMMS. For more details, see the section of applications.

maturation mechanism of HK97 well, and show a good agreement with those obtained by expensive computation for the whole capsid (22,23). UMMS can efficiently simulate the dynamics of many other symmetric macromolecules, currently impracticable in a PC, using this symmetry-constrained method.

Rigid-cluster or hybrid ENM

Most of large conformational changes in macromolecules can be resolved into collective motions such as hinge and shear motions (18). For instance, the major conformational change of HK97 is the capsid expansion induced by the shear motion between skewed trimers in each asymmetric unit (24). Hence, the rigid-cluster ENM can be applied to macromolecules in which such collective motions occur during the transition (25). The key idea is to only consider inter-connections among rigid clusters defined from the standard ENM because intra-connections within a cluster are preserved under rigid-body motions. The number of clusters and the size of each cluster will be adjusted by the user's preference between resolution and efficiency.

In the case of a complex structure which contains both rigid domains and flexible loop regions, the rigid-cluster ENM may destroy the generality of the flexible parts of the system too much. One can accommodate this problem by a hierarchical modeling in which the flexible regions are represented with higher resolution than other rigid regions (26).

The mathematical models for both rigid-cluster and hybrid ENMs have been developed and also extended to symmetry-constrained systems (25). This computational advantage enables UMMS to perform both harmonic and anharmonic analyses of large macromolecules within reasonable time (less than a day) in a PC.

MD data interpretation with NMA/ENI

One of the unique features provided by UMMS is MD data interpretation. As mentioned in the introduction part, interpretation of a massive amount of MD data which just resemble

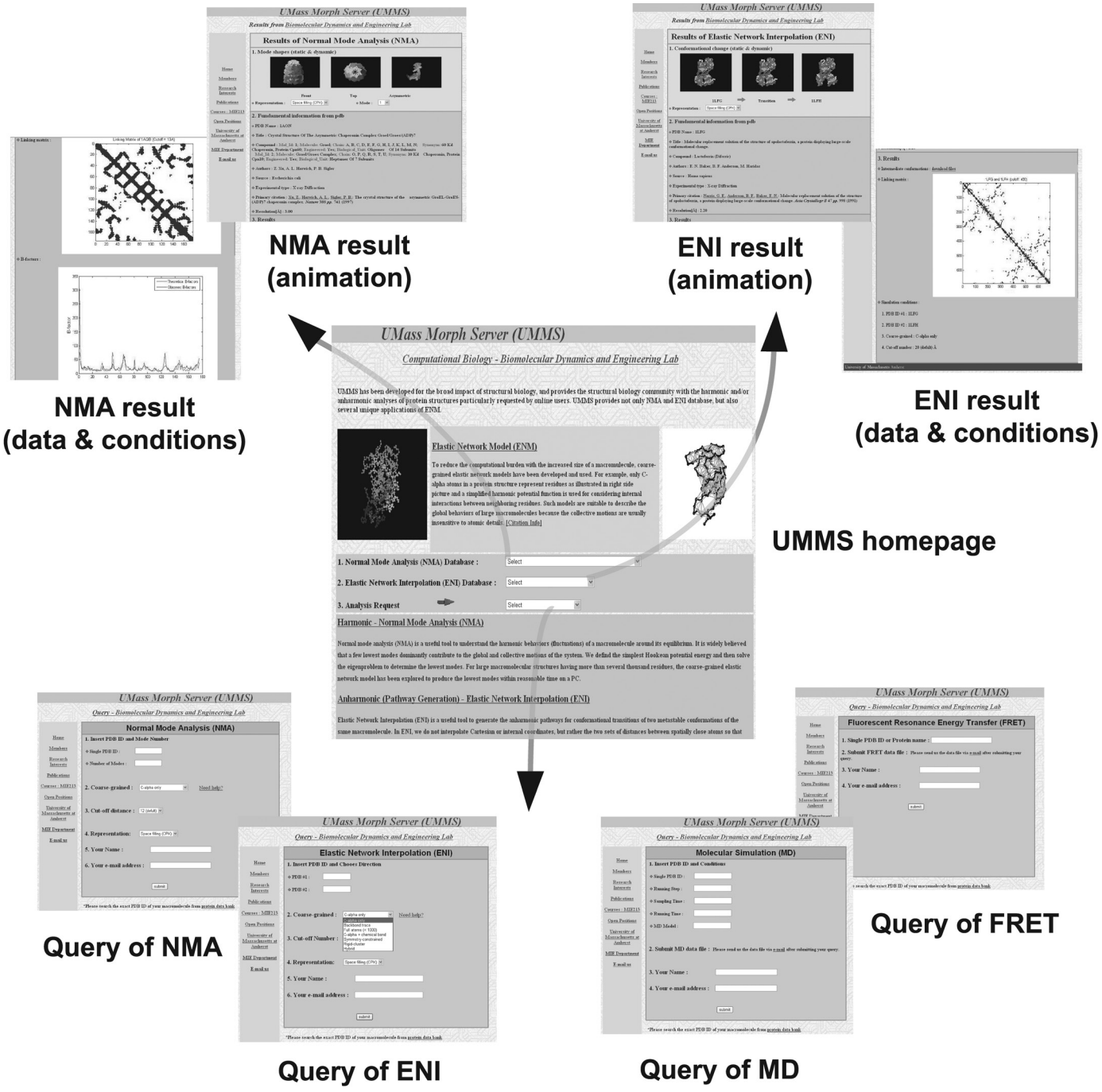

Figure 2. Overview of UMMS. The main page (middle panel) consists of NMA database, ENI database and query windows. One can not only browse NMA or ENI database to access the existing result of a specific macromolecule of interest, but also request the analysis via the query windows. A typical NMA window (upper left panels) shows animated mode shapes with various representations, the connectivity map of ENM called linking matrix, and the B-factor comparison plot. An ENI window (upper right panels) presents the computed ENI pathway between the two end conformations. One can download both animations and data files (PDB and text format) from those windows. Four different query windows (lower panels) are used for submitting the structural information of macromolecules (PDB ID, Cartesian coordinate file, MD trajectory data, FRET distance data) and choosing the simulation parameters (coarse-graining method, cutoff rule, output representation type).

a series of Brownian motions is another painstaking task after computationally expensive MD simulation. NMA and ENI can act as a supplementary tool to interpret MD results with ease (27). When MD trajectory data computed by Amber or Charmm are provided through UMMS, we first test if the given MD trajectory includes large and collective motions by measuring root mean square deviation with respect to the initial conformation. ENI can be applied to generate a highly directed pathway between the two extreme MD conformers (i.e. one is the initial conformer and the other is the farthest structure from the initial one). This pathway is considered as an ensemble of MD trajectory without random

fluctuations along the way. One can also utilize NMA to interpret MD data. If there is no significant conformational change, a reference structure is determined by finding the closest conformer to all other MD data on average. The displacement errors of each MD conformer with respect to the reference structure can be considered as thermal fluctuations. The predicted lowest normal modes collaboratively capture those fluctuations well. Consequently, NMA and ENI may serve as a paradigm for reduced-DOF dynamic simulations of large macromolecules as well as a method for the reduced-parameter interpretation of MD data.

ENI incorporated with partial conformational data

ENI can also be extended to incorporate partial structural information into computer simulation of macromolecular motions. When the initial conformation and only partial distance information for the target conformation are available, a modified ENI method can serve as a prediction tool for global motions that are consistent with conformational data experimentally measured by fluorescent resonance energy transfer (FRET) or NMR (16). In this method, any inter-residue distance that is not specified in the target conformation is allowed to relax unless it results in steric clashes during the transition. The related mathematical description can be found at our previous paper (16). Once partial information about conformational change like FRET distance data is provided through UMMS, those data are to be superimposed onto the known crystal structure to examine the structural implications of measured inter-residue distances for the prediction of the second conformation.

Chemical-bond ENM

The distance-cutoff method, a traditional connection rule for ENMs, sometimes fails to generate physically reliable results due to the uncertainty of system stiffness and connectivity. This model has limitation on representing the nature of chemical interactions in a real protein structure because a cutoff distance is empirically chosen longer than the maximum range to which atomic interactions can be felt. Therefore, the resulting elastic network over-represents the global stiffness of the system. To solve this discrepancy between the actual interaction range of a real macromolecule and the virtual spring connection range in an ENM, a chemical-bond ENM has been proposed (28). First, four consecutive alpha-carbon atoms along the protein backbone (ith through $i + 3$th representative atoms in a coarse-grained model) are linked to one another. These constraints in Cartesian space replace $3N - 6$ internal coordinates including $N - 1$ bond lengths, $N - 2$ bond angles and $N - 3$ torsion angles for an N-particle system. These backbone connections stabilize the elastic network so that the derived stiffness matrix cannot have more than six zero eigenvalues which correspond to rigid-body motions. It enables to reduce the cutoff distance down to the range of van der Waals interactions in the chemical-bond ENM. Then, the chemical-bond information such as disulfide bonds, hydrogen bonds and ionic bonds is utilized to add more spring connections for the realism of output results. These chemical bonds play an important role in characterizing the dynamics of macromolecules. By testing several example proteins, it is observed that the chemical-bond ENM is computationally more efficient as well as more robust than the other conventional models. NMA and ENI based on the chemical-bond ENM are also available at UMMS.

CONCLUSIONS

We have introduced UMMS which can analyze harmonic (NMA) and anharmonic (ENI) motions of macromolecules based on various levels of coarse-grained ENMs. Since we have exploited a lot of mathematical expressions (not displayed here) to implement those sophisticated procedures, it is probably difficult for some biologists who are not mathematically oriented to directly adopt the methodologies presented here for their research topics. Therefore, we recently launched a new online morph server called UMMS. Once one requests the dynamic analysis for a specific macromolecule through the server, we then perform it offline and post the result on the server visually and numerically to share the benefits of this work with the general public.

Our focus is also placed on the theoretical modification of the standard coarse-grained NMA and ENI methods in order to improve computational efficiency and physical realism with aid of engineering disciplines such as robot kinematics, group theory and mechanical vibrations.

Symmetry-constrained, rigid-cluster and hybrid models can together bring down the high-dimensional problems to a PC level. The chemical-bond ENM can generate more robust and reliable network models. UMMS also provides two unique services such as MD data interpretation and time-resolved ENI. This online service will enhance scientific understanding of the relationship between molecular structure and dynamics (i.e. function) in a cost-effective fashion.

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges for this article was provided by the University of Massachusetts Amherst.

Conflict of interest statement. None declared.

REFERENCES

1. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res*., **28**, 235–242.
2. Levitt,M. (1983) Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol*., **168**, 595–617.
3. Levitt,M. (1983) Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol*., **168**, 621–657.
4. Brooks,B. and Karplus,M. (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic typsin inhibitor. *Proc. Natl Acad. Sci. USA*, **80**, 6571–6575.
5. Scott,W.R.P., Hunenberger,P.H., Tironi,I.G., Mark,A.E., Billeter,S.R., Fennen,J., Torda,A.E., Huber,T., Kruger,P. and van Gunsteren,W.F. (1999) The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, **103**, 3596–3607.
6. Tirion,M.M. and Ben-Avraham,D. (1998) Normal modes analyses of macromolecules. *Physica A*, **249**, 415–423.
7. Tirion,M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett*., **77**, 1905–1908.
8. Atilgan,A.R., Durell,S.R., Jernigan,R.L., Demirel,M.C., Keskin,O. and Bahar,I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J*., **80**, 505–515.

9. Bahar,I., Atilgan,A.R. and Erman,B. (1997) Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Fold. Des.*, **2**, 173–181.
10. Wako,H., Kato,M. and Endo,S. (2004) ProMode: A database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, **20**, 2035–2043.
11. Suhre,K. and Sanejouand,Y.H. (2004) ElNe'mo: a normal mode web server for protein movement analysis and the generation of templates for moelular replacement. *Nucleic Acids Res.*, **32**, W610–W614.
12. Cao,Z.W., Xue,Y., Han,L.Y., Xie,B., Zhou,H., Zheng,C.J., Lin,H.H. and Chen,Y.Z. (2004) MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids. *Nucleic Acids Res.*, **32**, W679–W685.
13. Hollup,S.M., Salensminde,G. and Reuter,N. (2005) WEBnm@: a web application for normal mode analyses of proteins. *BMC Bioinformatics*, **6**, 52.
14. 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. *Bioinformatics*, **21**, 2978–2987.
15. Lindahl,E. and Delarue,M. (2005) Refinement of docked protein-ligand and protein–DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Res.*, **33**, 4496–4506.
16. 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.
17. Kim,M.K., Chirikjian,G.S. and Jernigan,R.L. (2002) Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.*, **21**, 151–160.
18. Gerstein,M. and Krebs,W. (1998) A database of macromolecular motions. *Nucleic Acids Res.*, **26**, 4280–4290.
19. Wang,Y., Rader,A.J., Bahar,I. and Jernigan,R.L. (2004) Global ribosome motions revealed with elastic network model. *J. Struct. Biol.*, **147**, 302–314.
20. Sayle,R.A. and Milnerwhite,E.J. (1995) RasMol: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374–376.
21. Kim,M.K., Jernigan,R.L. and Chirikjian,G.S. (2003) An elastic network model of HK97 capsid maturation. *J. Struct. Biol.*, **143**, 107–117.
22. Rader,A.J., Vlad,D.H. and Bahar,I. (2005) Maturation dynamics of bacteriophage HK97 capsid. *Structure*, **13**, 413–421.
23. Tama,F. and Brooks,C.L.,III (2005) Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.*, **345**, 299–314.
24. Conway,J.F., Wikoff,W.R., Cheng,N., DudaR.L., Hendrix,R.W., Johnson,J.E. and Steven,A.C. (2001) Virus maturation involving large subunit rotations and local refolding. *Science*, **292**, 744–748.
25. Kim,M.K., Jernigan,R.L. and Chirikjian,G.S. (2005) Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys. J.*, **89**, 43–55.
26. Doruker,P., Jernigan,R.L. and Bahar,I. (2002) Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.*, **23**, 119–127.
27. Kim,M.K., Li,W., Shapiro,B.A. and Chirikjian,G.S. (2003) A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA. *J. Biomol. Struct. Dyn.*, **21**, 395–405.
28. Jeong,J.I., Jang,Y. and Kim,M.K. (2006) A connection rule for α-carbon coarse-grained elastic network model using chemical bond information. *J. Mol. Graph. Model.*, **24**, 296–306.