BIOINFORMATICS APPLICATIONS NOTE Vol. 25 no. 13 2009, pages 1709–1710
doi:10.1093/bioinformatics/btp304

Structural bioinformatics

FlexServ: an integrated tool for the analysis of protein flexibility

Jordi Camps[1,2], Oliver Carrillo[1], Agustí Emperador[1], Laura Orellana[1], Adam Hospital[1,3], Manuel Rueda[1], Damjan Cicin-Sain[1,3], Marco D'Abramo[1,4], Josep Lluís Gelpí[1,2,4,*] and Modesto Orozco[1,3,4,*]

[1]Joint IRB-BSC Program on Computational Biology, Institute of Research in Biomedicine, Barcelona Science Park, Josep Samitier 1-5, Barcelona 08028 and Barcelona Supercomputing Centre, [2]National Institute of Bioinformatics, Computational Bioinformatics, Jordi Girona 31, Barcelona 08034, Spain, [3]National Institute of Bioinformatics, Structural Bioinformatics, Josep Samitier 1-5 and [4]Department of Biochemistry and Molecular Biology, University of Barcelona, Av Diagonal 645, Barcelona 08028, Spain

Received on December 28, 2008; revised on April 23, 2009; accepted on May 4, 2009

Advance Access publication May 7, 2009

Associate Editor: Thomas Lengauer

ABSTRACT

Summary: FlexServ is a web-based tool for the analysis of protein flexibility. The server incorporates powerful protocols for the coarse-grained determination of protein dynamics using different versions of *Normal Mode Analysis* (NMA), *Brownian dynamics* (BD) and *Discrete Dynamics* (DMD). It can also analyze user provided trajectories. The server allows a complete analysis of flexibility using a large variety of metrics, including basic geometrical analysis, B-factors, essential dynamics, stiffness analysis, collectivity measures, Lindemann's indexes, residue correlation, chain-correlations, dynamic domain determination, hinge point detections, etc. Data is presented through a web interface as plain text, 2D and 3D graphics.

Availability: http://mmb.pcb.ub.es/FlexServ; http://www.inab.org

Contact: modesto@mmb.pcb.ub.es or gelpi@mmb.pcb.ub.es

Supplementary information: Additional information and methodology details can be found at http://mmb.pcb.ub.es/FlexServ/help.

1 INTRODUCTION

Proteins have evolved to display not only optimal structures, but also functionally optimal deformability properties (Henzler-Wildman *et al*., 2007). Many evidences show that evolution carefully conserved deformation patterns as a mechanism to maintain function (Falke, 2002).

Despite recent advances in experimental techniques, the study of flexibility is mostly a task for theoretical methods. The most powerful of them is atomistic molecular dynamics (MD) (McCammon *et al*., 1977), a rigorous method which provides accurate representations of protein flexibility under physiological-like environments. Unfortunately, MD is complex, computationally expensive and their use requires a certain degree of expertise (Rueda *et al*., 2007b). Coarse-grained methods coupled to simple potentials (see Dokholyan *et al*., 1998; Tirion, 1996) are a cheap alternative to MD. By using them we lose atomic detail to gain formal and computational simplicity in the representation of near-native state flexibility of proteins (Emperador *et al*., 2008; Ma, 2005). Unfortunately, despite its power, the practical use of coarse-grained methods is still limited, due mostly to the lack of standardized protocols for analysis and the existence of a myriad of different algorithms distributed in different websites.

We present here FlexServ, an integrated tool that links most important coarse-grained based methods, structural databases, and atomistic models with a powerful analysis platform. FlexServ incorporates three coarse-grained algorithms for the representation of protein flexibility: (i) discrete dynamics (DMD), (ii) normal mode analysis (NMA) and (iii) Brownian dynamics (BD). DMD assumes that residue-residue interactions are controlled by infinite square wells centered at equilibrium distance with width fitted to reproduce atomistic flexibility (Emperador *et al*., 2008). Within this approach a particle is either moving at constant velocity or colliding with a wall, which allows the derivation of trajectories using simple ballistic equations (for discussion, see Emperador *et al*., 2008). On the other hand, NMA and BD assume that inter-residue interactions are controlled by a harmonic-like potential energy expressed as:

$$E_{ij} = \Gamma_{ij} K_{ij}(r_{ij}^0)(r_{ij} - r_{ij}^0)^2 \quad (1)$$

where r_{ij} is the distance between residues i and j, and r_{ij}^0 the equilibrium value, Γ_{ij} the Kirchhoff connectivity matrix, and $K_{ij}(r_{ij}^0)$ the stiffness force constant. In a pure harmonic model $K_{ij}(r_{ij}^0)$ takes a single value K_{ij} and $\Gamma_{ij} = -1$ for atoms within a given cutoff and 0 otherwise. In Kovacs' pseudo-harmonic model (Kovacs *et al*., 2004), $K_{ij}(r_{ij}^0)$ has a $(1/r_{ij}^0)^6$ dependence and no cutoff is applied. FlexServ incorporates both Hamiltonian definitions as well as a new hybrid one (Orellana *et al*., unpublished results) which treats differently covalently bonded and non-bonded residues and that is fitted to reproduce atomistic MD results.

NMA [within the Anisotropic Network Model formalism (Altigan *et al.*, 2001)] uses the energy function to define the Hessian whose diagonalization provides the eigenvectors and eigenvalues characterizing the harmonic deformability of the protein. NMA based pseudo-trajectories are generated by activating movements along the eigenvectors (Rueda, *et al*., 2007a). Within the BD approach the energy functional is used to compute forces, from

*To whom correspondence should be addressed.

which trajectories are derived by using a Brownian algorithm (Emperador *et al*., 2008).

FlexServ incorporates a large variety of methods to characterize flexibility. Thus, the server performs basic analysis like structural oscillation using either standard (RMSd) or Gaussian (gRMSd) root mean square deviation. gRMSd (Damm and Carlson, 2006) fits the molecules reinforcing the alignment of rigid moieties, localizing more clearly the movements into flexible regions. Essential dynamics routines (Amadei *et al*., 1993) are used to characterize the most important deformation modes, which are obtained by diagonalization of the trajectory covariance matrix. Essential deformation movements (after either fitting) are ranked by importance, and can be visualized and processed to obtain information (Meyer *et al*., 2006). Analysis include B-Factor profiles, the 'collectivity' index (a measure of the collective nature of protein motions), the variance profile, the dimensionality (the number of movements defining a percentage of variance), or the size of the essential space (i.e. the number of relevant modes). Lindemann's indexes are computed to evaluate the liquid/solid nature of the entire or partial regions of the protein (Rueda *et al*., 2007a).

Advanced capabilities of FlexServ include calculation of the apparent stiffness between interacting residues (obtained by inverting the inter-residue covariance (Rueda *et al*., 2007a) and the determination of residue to residue correlations. Determination of dynamic domains and hinge points is implemented using a variety of techniques: (i) exploration of the B-factor landscape after fitting with the gRMSd method, (ii) analysis of the force-constant profile (Sacquin-Mora and Lavery, 2006) and (iii) clustering by inter-residue correlation (Navizet *et al*., 2004). Calculations are performed using different sliding windows to reduce noise and false positives.

2 IMPLEMENTATION

FlexServ is written in PHP and implemented as a web-based interface, acting as a front-end of a series of simulation and analysis modules (see the full workflow diagram at http://mmb.pcb.ub.es/FlexServ/img/diagram.png). Results can be obtained through the web or retrieved later using a unique key. The modules are also available as web services in the Spanish National Institute of Bioinformatics platform (http://www.inab.org).

2.1 Input data

The user can upload a coordinate file, retrieve the structure using just the PDB code (Berman *et al.*, 2000), or upload his/her trajectory as a PCZ compressed file [PCAZIP (Meyer *et al*., 2006); the software is available at the server]. Users can also upload the protein sequence, either in FASTA format or using a UniprotKB code (The Uniprot Consortium, 2008). In this case, analysis is performed on the closest homologue available in the PDB, identified by a standard BLAST analysis (Altschul *et al.,* 1990). Additionally, the program allows the user to analyze already available atomistic MD trajectories included in the MoDEL database (Rueda *et al*., 2007b).

2.2 Workflow

Once the reference PDB structure is loaded, the program generates a C_α-model, computes several protein descriptors and provides basic visualization. FlexServ offers three engines to generate protein trajectories: (i) NMA, (ii) BD and (iii) DMD. In an elastic NMA calculation, the Hessian is computed and diagonalized. The obtained eigenvectors and eigenvalues are then used to derive a Monte-Carlo pseudo-trajectory in the NMA space. As for BD or DMD, trajectories are obtained directly. Trajectories are then processed with PCAZIP: RMS fits are applied to the data, and trajectory is finally compressed. Uploaded and MoDEL trajectories are processed directly. At this point, the compressed format contains the eigenvectors/eigenvalues data which can be used to animate the movement of the protein along essential dynamics modes as well as to derive the analysis offered by FlexServ (see above).

2.3 Visualization

Data is presented as plain text or 2D plots. When appropriated, 3D data is presented using a JMol applet (http://www.jmol.org). The produced trajectories and raw analysis data can be downloaded from the server for further off-line processing. Besides, a full snapshot of the results can also be downloaded to repeat the analysis at any later time.

Funding: Spanish Ministry of Education and Science [grant numbers: BIO2006-01602, CTQ2005-09365, CONSOLIDER Project in Supercomputation]; the Spanish Ministry of Health [COMBIOMED network]; the Fundación Marcelino Botín; and the National Institute of Bioinformatics.

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al*. (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
Amadei,A. *et al*. (1993) Essential dynamics of proteins. *Proteins*, **17**, 412–425.
Atilgan,A.R. *et al*. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J*., **80**, 505–515.
Berman,H.M. *et al*. (2000) The protein data bank. *Nucleic Acid Res*., **28**, 235–242.
Damm,K.L. and Carlson,H.A (2006). Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. *Biophys. J.*, **90**, 45558–45573.
Dokholyan,N.V. *et al*. (1998) Discrete molecular dynamics studies of the folding of a protein-like model, *Folding Des.*, **3**, 577–587.
Emperador,A. *et al*. (2008) Exploring the suitability of coarse-grained techniques for the representation of protein dynamics, *Biophys. J.*, **95**, 2127–2138.
Falke,J.J. (2002) Enzymology—a moving story. *Science*, **295**, 1480–1481.
Henzler-Wildman,K.A. *et al*. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450**, 913–916.
Kovacs,J.A. *et al*. (2004) Predictions of protein flexibility: first-order measures. *Proteins*, **56**, 661–668.
Ma,J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, **13**, 373–380.
McCammon,J.A. *et al*. (1977) Dynamics of folded proteins. *Nature*, **267**, 585–590.
Meyer,T. *et al*. (2006) Essential dynamics: a tool for efficient trajectory compression and management. *J. Chem. Theory Comput.*, **2**, 251–258.
Navizet,I. *et al*. (2004) Probing protein mechanics: residue-level properties and their use in defining domains. *Biophys. J.*, **87**, 1426–1435.
Rueda,M. *et al*. (2007a) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure*, **15**, 565–575.
Rueda,M. *et al*. (2007b) A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. USA*, **104**, 796–801.
Sacquin-Mora,S. and Lavery,R. (2006) Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.*, **90**, 2706–2717.
The Uniprot Consortium (2008) *Nucleic .Acid Res.*, **36**, D190–D195.
Tirion,M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, **77**, 1905–1908.