JED: A Java Essential Dynamics Program for

Comparative Analysis of Protein Trajectories

Charles C. David

University of North Carolina at Charlotte

Department of Bioinformatics and Computational Biology

Donald J. Jacobs

University of North Carolina at Charlotte

Department of Physics and Optical Science

NOTE: April 1 is the deadline. I would like to see a first draft much sooner than near the deadline. Something like March 8 would be good. The paper length needs to be 4 pages, which is less than 30 hours of work for a 1st draft. Note that originally we were thinking this would be a 3-page paper. Should be possible.

Background

Essential Dynamics (ED) is a common application of principal component analysis (PCA) to extract biologically relevant motions from atomic trajectories of proteins. A covariance (Q) or correlation (R) based form of PCA for a selected set of atoms is applied to yield PCA modes (eigenvectors) and their eigenvalues. The displacement vectors (DV) relative to a reference structure are projected onto the top set of modes to produce a set of principal components, which are used to cluster conformations. The subspace defined by a relatively small set of PCA modes with largest eigenvalues quantifies conformational space explored by a protein. In bioinformatics studies, it is necessary to made subspace comparisons over a set of different protein trajectories for similarity assessment, but software to facilitate this comparative-analysis is needed.

Results

We developed the Java Essential Dynamics (JED) package to perform complete ED for multiple protein trajectories. JED reads the trajectory data from sets of PDB files or from a matrix of atomic coordinates. The PDB files may be single or multi-chained. A set of residues can be chosen that need not be contiguous. The analysis is based on alpha carbon atoms from each of the selected residues. Cartesian-based coordinates (cPCA) or internal distance coordinates (dPCA) can be applied with Q and R based PCA. JED outputs results as text files for further processing using generic graphing software. Key data include the transformed coordinates, conformation and residue RMSDs, PCA modes with their eigenvalues and DV projections onto the top principal components. A set of PDB files is generated along with a Pymol™ script to visualize each of the top modes derived from Q and R cPCA as either a static picture or as a movie. For comparative studies, JED compares subspaces defined by the top eigenvectors using several metrics to quantify the similarity/overlap of high dimensional vector spaces.

Conclusions

We present JED, a new Java package that encourages best-practices ED, allowing one to perform comparative studies over multiple protein trajectories to extract essential motions for the entire protein, user-defined sub-regions of interest, and to interrogate the significance of the results.

**Keywords**

Essential Dynamics, PCA, distance PCA, vector space, subspace, RMSIP, principal angle

**Background**

The Background section should be written in a way that is accessible to researchers without specialist knowledge in that area and must clearly state - and, if helpful, illustrate - the background to the research and its aims. It should clearly described the relevant context and the specific issue which the software described is intended to address.

**Implementation**

This should include a description of the overall architecture of the software implementation, along with details of any critical issues and how they were addressed.

**Results and Discussion**

The Results and Discussion may be combined into a single section or presented separately. They may also be broken into subsections with short, informative headings. In any case what should be described is the functionality of the software together with data on how its performance and functionality compare with and improve on functionally similar existing software. There should then be a discussion of the intended use of the software, and the benefits that are envisioned together, if possible, with an outline for the planned future development of new features.

**Conclusions**

This should state clearly the main conclusions of the article and give a clear explanation of the importance and relevance of the software.

**Availability and requirements**

* **Project name:** Java Essential Dynamics: JED
* **Project home page:** http://sourceforge.net/projects/JED
* **Operating system(s):** Platform independent
* **Programming language:** Java
* **Other requirements:** Java JDK 1.5 or higher, an amount RAM appropriate to the size of Q or R: JED performs a full eigenvalue decomposition.
* **License:** GNU GPL.
* **No restrictions to use:** For reproduction and development, cite thelicense