JED: A Java Essential Dynamics Program for

Comparative Analysis of Protein Trajectories

Charles C. David1 and Donald J. Jacobs2,3\*

1Computing Resource Center {cdavid.uncc.edu}, 2Department of Physics and Optical Science and 3Center for Biomedical Engineering and Science {djacobs1.uncc.edu} University of North Carolina at Charlotte, USA

**Abstract.** Principal component analysis (PCA) is commonly used to extract the essential dynamics of a protein described by eigenvectors (modes) of a covariance or correlation matrix constructed from atomic trajectories. The subspace defined by a small set of PCA modes with largest eigenvalues quantifies conformational space explored by a protein. In bioinformatics studies there is a need to compare subspaces among different proteins for similarity assessment, but software to facilitate this comparative-analysis is lacking. We developed the Java Essential Dynamics (JED) package to perform many variants of PCA applied to user-defined regions within a protein. Operationally, JED compares subspaces defined by the top eigenvectors using several metrics to quantify the similarity/overlap of high dimensional vector spaces with features that help to quantify statistical significance.

**Keywords:** Essential Dynamics, PCA, distance PCA, vector space, subspace, RMSIP, principal angle

**1. Introduction**

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 make subspace comparisons over a set of different protein trajectories for similarity assessment, but software to facilitate this comparative-analysis is lacking. We describe the key features in a Java Essential Dynamics (JED) software package open to the public that fills this need.

**2. Methods**

For comparative studies, JED compares subspaces defined by the top eigenvectors using several metrics to quantify the similarity/overlap of high dimensional vector spaces. The metrics are defined as: Describe the key metrics used to compare subspaces, as this is new, and outline some of the features.

**3. Results**

JED reads 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 to be processed using graphing software. The output includes 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.

Show at least 2 figures but no more than 4 that highlight the metrics described above. Take from the write up that you already have on the new to be web pages, and the book chapter for the equations. If you can send something by 10:30pm or 11:00pm, we probably can finish it tonight. Also, add references where appropriate. I left much more room than you will probably need for references, which can come from your thesis and/or the book chapter. I can edit what you give me, but the ball is in your court now. Be aware that you need to generate new figures (can be old data) and, preferably use data we never showed in a publication. In other words, I would much rather you show results on things in your thesis that have not been published rather than recycling the book chapter, much of which was recycled form the first paper. Also, think about something unique about JED that is not available in other statistical packages, and something that is eye catchy. Remove this cyan stuff (instructions). Also, make sure you modify this WORD document into the correct format using the link they gave and that I sent to you via email last night. Do this first, because I hate to have to worry about reformatting something later. Thanks.

**4. Conclusions**

JED encourages best practices for essential dynamics analysis that includes comparative studies over multiple protein trajectories over the entire protein or user-defined regions of interest, and to interrogate the significance of the results.

**5. References**