BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 9 2001
Pages 851–852

# VISTRAJ: exploring protein conformational space

John J. Salama[1], Howard J. Feldman[2, 3] and Christopher W. V. Hogue[2, 3,*]

[1]MDS Proteomics Inc., 480 University Ave, Toronto, Ontario, Canada, [2]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8 and [3]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Ave, Toronto, Ontario, Canada M5G 1X5

Received on March 19, 2001; revised on April 27, 2001; accepted on May 30, 2001

## ABSTRACT

**Summary:** VISTRAJ is an application which allows 3D visualization, manipulation and editing of protein conformational space using probabilistic maps of this space called 'trajectory distributions'. Trajectory distributions serve as input to FOLDTRAJ which samples protein structures based on the represented conformational space. VISTRAJ also allows FOLDTRAJ to be used as a tool for homology model creation, and structures may be generated containing post-translationally modified amino acids.

**Availability:** Binaries are freely available for non-profit use as part of the FOLDTRAJ package at ftp://ftp.mshri.on.ca/pub/TraDES/foldtraj/.

**Contact:** hogue@mshri.on.ca

**Supplementary Information:** The Trajectory Directed Ensemble Sampling (TraDES) web site can be visited at http://bioinfo.mshri.on.ca/trades/. The web-based version of INITTRAJ/FOLDTRAJ may be accessed at http://foldtraj.mshri.on.ca/.

## INTRODUCTION

Recent development of FOLDTRAJ (Feldman and Hogue, 2000) has made available an application that rapidly generates plausible probabilistic protein conformers. FOLDTRAJ utilizes 2D probability distributions in conformational space, called 'trajectory distributions', to place each amino acid backbone. The dimensions of this space may either be the backbone dihedral angles at each residue (Ramachandran and Sasisekharan, 1968) space or the $\alpha$-carbon trajectories at each residue. $\alpha$-carbon space is similar to that employed by Levitt (1976); Gregoret and Cohen (1991); de la Cruz *et al.* (1997) where we denote by $\phi$ the angle between the three previous $\alpha$-carbons and by $\theta$ the torsional angle between the previous four. This space was found to be optimally discretized into $400 \times 400$ 'bins' (Feldman and Hogue, 2000). The height at any such bin in the distribution is proportional to the probability of that conformation being chosen, and normally corresponds to the frequency of occurrence of the conformation, based on residue type, in a non-redundant set of PDB structures. FOLDTRAJ takes as input a sequence of trajectory distributions containing these probabilistic plots, one for each amino acid. These were originally produced by a tool called INITTRAJ (Feldman and Hogue, 2000). Also contained within the trajectory distribution file are all global and local parameters that are used by FOLDTRAJ to construct the probabilistic protein structures. By editing or manipulating the trajectory distributions, or any of the local or global parameters, one can bias the resulting conformers generated by FOLDTRAJ. Until the development of VISTRAJ the editing of the trajectory distribution matrix and its visualization would have been extremely tedious if not impossible.

## OVERVIEW

VISTRAJ is a multi-platform application, which has been cross-compiled on Windows 98/ME/NT/2000, Linux, HP-UX, Compaq Tru64, IRIX, Solaris and PowerPC-Linux operating systems and is based on the Vibrant software system in the NCBI toolkit. The application loads an entire trajectory distribution sequence corresponding to a protein, and allows for the 3D visualization of each amino acid's trajectory distribution in both the spherical map representation ($\alpha$-carbon space) and an 'unrolled' map representation ($\alpha$-carbon or Ramachandran space). (See ftp://ftp.mshri.on.ca/pub/TraDES/figs/vistraj/.) In both types of representations, the entire $400 \times 400$ matrix, is plotted in order to preserve visual integrity and resolution. The plot of up to 160 000 polygons can present a performance problem depending on the graphical API used. With this in mind, VISTRAJ was programmed using OpenGL for plotting, rotation, translation and scaling of trajectory distributions, using the Vibrant/OpenGL interface from the structure viewer tool Cn3D (Hogue, 1997).

*To whom correspondence should be addressed.

## TRAJECTORY DISTRIBUTION CREATION

An interactive graphical 'wizard' to create new trajectory distributions has been incorporated into VISTRAJ to simplify the previous command-line driven interface of INITTRAJ. There are three different methods of creating a trajectory distribution file: (1) from an NCBI-MMDB (Hogue *et al.*, 1996; Wang *et al.*, 2000) structure file; (2) from a FASTA input file containing the amino acid sequence; or (3) by directly entering the amino acid sequence within VISTRAJ. In the case of a structure file, the actual conformation from the structure at each residue is placed as a sharp peak or Gaussian in trajectory space. In this way it guides FOLDTRAJ in the reconstruction of a protein with a similar backbone trajectory. Global settings that control both the way the trajectory distributions are created and the structure generation process, such as which space to use (Ramachandran or $\alpha$-carbon), backbone error tolerance, atom hardness, and number of tries to place residues and sidechains before backtracking, can be easily changed to the user's preference. The user may also specify a distance constraint file to be used during the FOLDTRAJ run, for example, to include information about disulfides or NOEs. The user may also provide a secondary structure prediction.

## TRAJECTORY DISTRIBUTION EDITING

VISTRAJ contains two main functions for editing trajectory distributions: filtering and adding or replacing noise. The filtering function gives the user the option of using a smooth, Gaussian or low pass filter to alter the distributions, as well as to help correct for the slight biases caused by the discretization of space into $400 \times 400$ bins.

The add/replace noise function allows the user to explicitly manipulate conformational space by adding, or replacing the current distribution with, either single peak probability values, or adding to the helical, sheet or coil regions of space using precomputed 'dictionaries' for each of the 20 amino acids, based on observed secondary structure conformations in a non-redundant set of the PDB. One can specify percentages of helix, sheet and coil, for example corresponding to a three-state secondary structure prediction. The ability to directly manipulate the trajectory distributions allows the user to bias any amino acid into a specific region of conformational space when subsequently sampling the space. This bias is particularly useful for proteins of known fold or secondary structure.

## AMINO ACID SEQUENCE EDITING

The amino acid sequence loaded in with the trajectory distribution database may be edited by removing, inserting or mutating any residue. An application for this type of editing is in homology modeling. This is accomplished by storing the backbone conformation of a known structure in a trajectory distribution file as described above, and then the sequence could be modified to correspond to the target. Thus at insertions, all possible residue conformations would still be sampled, while for the aligned portions the backbone would be reconstructed from the template but with the target's sidechains. Torsional angles of sidechains can be selectively preserved as well.

When inserting or mutating an amino acid, the user may pick from a list comprised of the 20 standard L-amino acids and various post-translationally modified amino acids. A complete list of these 60 modified amino acids may be found in the VISTRAJ help file.

## RESIDUE AND GLOBAL PROPERTIES EDITING

Properties relating to a residue's trajectory distribution and the trajectory distribution database can be changed in VISTRAJ as well. Residue properties that may be changed include the probability of a bond being cis (in the case of a residue preceding proline), the mean and standard deviation to use for choosing the peptide dihedral angle between two residues and the number of tries before backtracking.

A detailed description of all global settings used by FOLDTRAJ can be found in the VISTRAJ help file.

## ACKNOWLEDGEMENTS

This research was supported by grants to CWV Hogue by the Natural Sciences and Engineering Research Council of Canada. H.J.Feldman is supported in part by an Ontario Graduate Scholarship.

## REFERENCES

de la Cruz,X.F., Mahoney,M.W. and Lee,B. (1997) Discrete representations of the protein C$\alpha$ chain. *Fold. Des.*, **2**, 223–234.

Feldman,H.J. and Hogue,C.W.V. (2000) A fast method to sample real protein conformational space. *Proteins*, **39**, 112–131.

Gregoret,L.M. and Cohen,F.E. (1991) Protein folding. Effect of packing density on chain conformation. *J. Mol. Biol.*, **219**, 109–122.

Hogue,C.W. (1997) Cn3D: a new generation of three-dimensional molecular structure viewer. *Trends Biochem. Sci.*, **22**, 314–316.

Hogue,C.W., Ohkawa,H. and Bryant,S.H. (1996) A dynamic look at structures: www-Entrez and the molecular modeling database. *Trends Biochem. Sci.*, **21**, 226–229.

Levitt,M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.*, **104**, 59–107.

Ramachandran,G.N. and Sasisekharan,V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283–438.

Wang,Y., Addess,K.J., Geer,L., Madej,T., Marchler-Bauer,A., Zimmerman,D. and Bryant,S.H. (2000) MMDB: 3D structure data in Entrez. *Nucleic Acids Res.*, **28**, 243–245.