A Comprehensive Single Molecule Analysis Software Package

**Abstract** (3 sentences, 70 words)

The analysis of single molecule studies is often performed through a combination of unpublished custom scripts. To address this we have created a comprehensive software package to analyze raw data from single molecule co-localization, flow-stretching and fluorescence resonance energy transfer experiments. Here we present the methods used to automatically analyze these videos and the user oriented graphical interface to allow all researchers to easily perform these analyses.

**Brief Communication** (1000-1500 words, 2 figures, no headings, 20 references)

Single molecule studies using co-localization, flow stretching, and fluorescence resonance energy transfer (FRET) have become a vital tool for many researchers studying the dynamics and structure of biomolecules. With its growth several labs have developed more streamlined techniques and protocols to assist researchers as they enter the field. However, the software for analysis of the raw data generated from these studies remains largely unpublished, varies from between labs, and is often broken into many uncoordinated steps.

Of the excellent programs that have been designed to analyze single molecule data they are limited to one component of the analysis for specific types of experiments. This makes it hard for new users to work with and lowers the analysis throughput. Here we hope to build on these analyses to make a more comprehensive program. We have particularly paid attention to the analysis of co-localization experiments for which there are no published methods for analysis. These vital experiments are uniquely suited to answer questions of biomolecular dynamics. As they have grown in popularity it has become more important to develop tools for their analysis.

Here we present an open source analysis package that is accessible to researchers with and without programming experience. With this software we intend to streamline the analysis of raw data from co-localization, fluorescence resonance energy transfer and flow stretching experiments. To this end we have created a clean intuitive graphical user interface with robust analysis methods. The data generated from the software can easily be analyzed in the program or exported in familiar formats (HaMMY, Excel, ect) for further experiment specific analysis. All of the code is made using Matlab so that it is accessible to as many researchers as possible and is freely available as an open source program at mmr.osu.edu. Furthermore, the program can be downloaded as a pre-compiled version that does not required the installation of Matlab.

For all the analyses the program relies on three core algorithms: a particle detector, channel register, and a particle tracker. The particle detector identifies particles and determines their position. This is handled by first convolving the image with a gaussian kernel to smooth the image and then selecting bright local maximums. The algorithm then fits a gaussian to each of these spots to determine the sub-pixel position.

The channel register is used for experiments where several channels are recorded from different wavelength excitations. The channel register determines the transformation to go from the coordinate space of one channel to each of the other channels. This is typically performed by recording multi-fluorescents beads in each channel. From this video the algorithm can determine the transformation and apply it to videos of multi-channel experiments. To do this the algorithm first uses the particle detector to determine the location of all the beads. The beads in each region are then randomly iterated through in sets of three to find an affine transformation that approximately matches the regions. The approximately matched positions in each channel are then matched together and used to determine the thin-plate spline that transforms from one channel to the next. This transformation can then be applied to analyzed movies to accurately overlap them.

The particle tracker uses a similar Tri-Track algorithm as used in DiaTrack. This algorithm works by optimizing for the most tracks that would result from the most continuous movement based on the speed and direction of the particles over three frames. The algorithm takes in a set of particle positions in each from of a movie and then returns the set of continuous tracks that occur during the movie.

Using these three core algorithms, the program can analyze then analyze the three types of experiments. The user simply provides the program with the original video from the experiment to get started. If there are multiple channels in the video user will also provide a mapping movie of multi-florescent beads which the channel register can then process. Next, the user can then select which type of analysis to perform.

For co-localization experiments kymographs are generated. These are typically generated based on movies of stained DNA. However, there are sometimes not possible to recorded such as when the DNA is broken down by a nuclease. In these cases, the program can use the tracking algorithm to detect molecules that slide on the DNA during the experiment and use those to estimate the DNA location.

Additionally, there are settings to automatically correct for x-y drift of the field of view if needed. At all stages there are automatic settings that adjust but all settings can be finely tuned by the user to fit the experiment.