vbFRET User Guide

March 22, 2010

Welcome to vbFRET! This program is an open source[1] MATLAB package, and graphics user interface (GUI), for performing statistical inference on FRET data. It models individual FRET traces using hidden Markov modeling (HMM) and finds the most probable fit of the data (i.e. the idealized trace) using evidence maximization and the variational Bayesian expectation maximization algorithm (VBEM). This work is described in our paper "Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data" [2]. Please cite this paper if you use vbFRET.

Although the primary purpose of this program is to fit FRET data with idealized trajectories, the program is also intended to be a user friendly environment for visualizing your data. Consequently, there are many options and parameters that can be customized to display your data in the most useful way possible. All of these options are described in this manual.

If you are not interested in fancy display options and are looking for a simple program that will effectively fit your data while applying model selection to individual traces, then vbFRET is still the program for you. The tutorial on p. 19 will show you how to fit your data with a few simple point and click steps.

Any questions, comments or bugs you notice should be directed to Jonathan Bronson (jeb2126@columbia.edu).

[1] You are welcome to download, modify, and/or redistribute this software to your friends and family free of charge. While we tried hard to make a bug-free, efficient and effective data fitting program, *we make no guarantees about the quality or performance of this program.* Should this program freeze, crash your computer, delete your data or do anything else to diminish the quality of your data fitting life, we will be very sad but take no responsibility.

[2] Bronson J.E., Fei J., Hofman J.M., Gonzalez R.L., Wiggins C.H. "Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and single-molecule FRET Data". *Biophysical Journal* 97(12)3196-3205, 2009. Also available on the arXiv: 0907.3156 [q-bio.QM], http://arxiv.org/abs/0907.3156 (2009)

Contents

1 What does vbFRET do?

vbFRET takes data files as inputs and fits idealized traces to them. It also returns information about the FRET states in the data: the number of states, their most probable means and standard deviations, and transition frequencies between states.

2 Getting started

2.1 Installation

As a MATLAB software package, VBFRET does not need to be installed but must be run from within the MATLAB environment. Before running VBFRET, make sure the vbFRET folder is present somewhere in your MATLAB directory. To run VBFRET: (1) open MATLAB; (2) change the MATLAB current directory to the vbFRET folder[3]; and (3) type 'vbFRET' in the command window.

2.2 Analyzing data with the default settings

To analyze data using all of VBFRET's default settings, simply load the data (Sec. 3) and click the large 'Analyze Data' button in the center of the main GUI. Saving data is discussed in Sec. 6.

2.3 Fixing the GUI's appearance

It has been our experience that MATLAB will sometimes resize the VBFRET GUI windows when VBFRET is run with a different computer, version of MATLAB or screen resolution. Fortunately, it is easy to perform cosmetic touch-ups to GUI displays in MATLAB using the GUI Design Environment (GUIDE). To resize any of the GUI windows, type 'guide' in the Command Window. Select 'Open Existing Gui'. Open 'vbFRET.fig'. All of the textboxes, pulldown menus, etc. in the vbFRET GUI can be resized in the window that opens. If you want to touch-up the sub-GUIs, they can be found in the folder 'vbFRET/src/'.

[3]If the vbFRET folder is added to the matlabpath (see 'path' or 'addpath' in the MATLAB help file) then typing 'vbFRET' in the MATLAB Command Window will open VBFRET from any directory.

3 Loading data

To load data click 'File' → 'Load' or press Ctrl+L. This will open the Load Data sub-GUI. Click the 'Add Files' button. Select the data file(s) you want to load (multiple files can be added at once). The name(s) of the file(s) will appear in the display window of the Load Data sub-GUI.

File names accidentally added can be removed by highlighting the files and clicking 'Delete Files'. To clear all file names in the display window, click 'Clear List'.

Once all the desired files names have been added, click the 'Load Data' button. The screen will flicker for a moment, the Load Data sub-GUI will close and the first of the loaded traces will appear in the display window of the Main GUI.

If the 'Sort Traces' checkbox is checked, then traces will be loaded in alphanumeric order (otherwise, vbFRET will load the traces in the order it encounters them). If desired, the traces can be relabeled from 1 to [number of traces] by checking the 'Relabel Traces' checkbox before loading the data. Traces will be relabeled before being sorted, so the 'Sort Traces' checkbox does not do anything if the 'Relabel Traces' checkbox is checked.

After 'Load Data' is pressed, the Load Data sub-GUI will automatically close, but the file names will be saved in the display window of the main GUI and will appear if the Load Data sub-GUI is reopened.

If you try to load traces into vbFRET and there already are traces loaded, you will be given the option to overwrite the existing traces or append the new traces to the end of the existing ones.

For a list of allowed input data formats, see section 15 on page 17.

4 Removing photobleaching

Photobleached segments at the end of FRET traces can be removed by vbFRET. To do so click 'Traces' → 'Remove Photobleaching'. The Remove Photobleaching sub-GUI will open. Set the criteria you wish to use to determine when a trace begins to show a photobleached signal. Once the criteria are set, click the 'Remove Photobleaching!' button to truncate the traces.

Backup copies of the traces will be saved when you remove photobleaching from traces. If you are unhappy with where the traces have been truncated, you can simply change the truncation criteria, click 'Remove Photobleaching!' again and the saved copies will be used to retruncate the data.

Clicking 'Remove Photobleaching!' or 'Save Settings' will store the Remove Photobleaching parameter settings for the session so they will appear in the Remove Photobleaching sub-GUI the next time it is opened. Clicking the 'Restore Default Settings' button restores the default settings and discards the user selected settings.

WARNING: several processes, such as deleting traces, appending new traces to the current data set and saving to certain file types, will cause all backup copies of traces to be deleted. In addition, clicking 'Remove Photobleaching!' will cause any data analysis already performed to be cleared.

There are several options and parameters in the Remove Photobleaching sub-GUI that may be set. They are (from top to bottom):

Identification method

This pulldown menu allows you to chose from three methods of identifying the point at which photobleaching begins in each trace. They are:

1. 1D FRET: Examine the 1D FRET transformation of the raw data. Truncate the data when the signal becomes less than 0 or greater than 1 by more than a value specified in the 'Truncation threshold' textbox.

2. Single Channel: Examine each channel of the raw data separately. Truncate the data when either channel of the raw data falls below a value specified in the 'Truncation threshold' textbox.

3. Summed Channel: Sum the two raw data channels for each point in the FRET trace. Truncate the data when the sum of the channels falls below a value specified in the 'Truncation threshold' textbox.

Truncation threshold

The truncation threshold should be set in this text box, located immediately below the 'Identification method' pulldown menu. Note the actual label of this text box changes with the identification method selected to describe the truncation criterion. Input must be an integer.

Smooth traces

To avoid having a single, wild data point (such as a blinking event) cause premature truncation of the trace, the data can be smoothed (i.e. averaged) over several time steps before the trace is examined for photobleaching. To do so, check the 'Smooth traces?' checkbox and indicate the number of time steps which the traces should be smoothed over in the 'Smoothing width' text box. Input must be a non-negative integer.

The smoothed version of the traces will only be used for photobleaching removal. The original, non-smoothed data will be used for all other analysis.

Truncate extra

When set to 0, each trace is truncated starting with the first data point to meet the photobleaching criteria. Setting 'Truncate extra' to a positive integer t causes the trace to be truncated t steps earlier. Setting 'Truncate extra' to a negative integer $-t$ causes it to be truncated t steps later.

Minimum trace length

If, after photobleaching has been removed, any traces are shorter than the length entered in this textbox they are discarded and not used in further analysis. Input must be a non-negative integer.

5 Analyzing data

Once traces have been loaded, push the 'Analyze Data!' button to fit the data. The text of the 'Analyze Data!' button will change to say 'Analyzing Data...' and vbFRET will begin fitting the traces, one at a time, in sequential order. Each trace will be fit with as many states and as many restarts as specified in 'Analysis Settings'. Only the fit of the trace with the highest evidence is saved, however.

The plot window will display the trace currently being fit and the fit of that trace with the highest evidence. The status box (p. 12) will update with information on the number of states being fit and the fitting attempt of the trace. These display options can be changed in 'Advanced analysis settings'.

Analysis can be paused at any point by clicking the 'Pause Analysis' button. To resume analysis, click the 'Analyze Data!' button (the text of this button will now say 'Resume Analysis'). While data analysis is paused you may load, view and save traces. Photobleaching removal, trace deletion and analysis settings cannot be changed while analysis is paused, however. If you want to change any of these settings you must clear the analysis first and restart the analysis from the beginning.

Note: Once all the traces are fit, pushing the 'Analyze Data!' button will cause the data to be analyzed again from the beginning and all previous analysis will be erased.

6 Saving and exporting data, plots and analysis summaries

Saving and exporting data files, plots and analysis summaries are all considered saving data in vbFRET. To save data click 'File' → 'Save' or press Ctrl+S. This will open the Save Data sub-GUI. This sub-GUI has 3 pulldown menus. The top one specifies the saved data format, the middle one specifies the file extension and the bottom one specifies the number of traces per figure if plots are saved. Directly below these pulldown menus is a text box where the name of the file to be saved should be entered. If no file name is given the file will be called 'Save Name'.

Unless otherwise specified, saved files will be saved in the matlab Current Directory. To save data to a different directory click 'Browse', select a new directory in the 'Browse for Folder' window, click 'OK' and type the file name at the end of the folder's full name, which will now appear in the 'Save Name' text box. If traces are saved as individual files, the name of the trace will be appended to the file name. Checking the 'Date Stamp' and 'Time Stamp' checkboxes will append the date and time to the file name, respectively. The date will be written as 'MMDDYY' and is immediately preceded by the letter 'D'. The time will be written as 'HHMM' in military time and is immediately preceeded by the letter 'T'.

After the name and type of file to be saved have been selected, click the 'Save' button. If the file name exists you will receive a warning asking if you wish to overwrite the existing file. A message box will appear to notify you that the file(s) have been successfully saved. The Save Data sub-GUI will remain open after the file is saved. Unlike most of the other sub-GUIs, any settings for the Save Data sub-GUI are lost when it is closed. vbFRET can save 5 different types of files.

Save session

Using 'Saving session' will create a .mat file with all the information from your current vbFRET session (traces, data fits and parameter settings). Saving data as a 'Save session' will allow you to close vbFRET, reopen it and resume exactly where you left off. It is recommended that you save all data using 'Save session' and then use that file to generate other forms of saved data (such as plots or path files).

Save analysis summary using 'Save analysis summary' creates a file containing the following information for each trace:

-Trace label
-Log(evidence) (a.k.a. $p(D|K)$)
-Number of FRET states in the trace
-Each FRET state's mean and standard deviation
-A transition matrix (the number in the i^{th} row and j^{th} column is the probability of transitioning from state i to state j)

Analysis summaries can be saved as text files (.dat) or .mat files. If the analysis summary is saved as a .mat file then this information will be saved in a cell array called 'vbFRET-summary'. Each cell the contains a structure with the trace summary information. To see the summary of the n^{th} trace, type 'vbFRETsummary{n}' in the command window (make sure the summary file is loaded). To just view one field of the summary, type 'vbFRETsummary{n}.field_name' (for example typing 'vbFRETsummary{n}.transition_mtx' will show the transition matrix for the n^{th} trace).

Analysis summaries saved as .mat files will also save all the information saved when 'Save traces and best fits (path data)' is selected.

Note: saving an analysis summary is the only way to view means and standard deviations of FRET states and transition matrices for FRET traces.

Save traces and best fits (path data)

Saving data this way will only save the traces and their best fits (sometimes referred to as the 'path data'). Path data may be saved as a .mat file, individual text files (.dat) or one large, concatenated text file (.dat). If the data is saved using either of the latter two formats, the characters '_PATH' will be appended to the file name.

Data saved as a .mat file will contain 5 variables: FRET (the 1D FRET transformed data); data (the raw 2D FRET data); labels (the names of each trace); path (the 1D fit of the FRET data); and path2D (the extrapolated 2D fit of the data (see plotting section)). Each variable will be a 1xN cell array, where N is the number of traces.

When the data is saved as individual text files, each will contain 5 columns of data (from left to right): the time step of the data; the donor intensity; the acceptor intensity; the 1D FRET transform; and the fit (idealized FRET) of the data point.

Data saved as one large, concatenated text file will have two columns of text in the .dat file. The first column will contain the trace number and the second column will contain the idealized FRET trajectory for the data. The trace labels are not included in this output file. This format is only recommended for calculating data-set wide properties such as 1D FRET histograms and TDPs.

Save traces only

This save option will save only the traces. Traces may be saved as a .mat file, individual text files (.dat) or one large text file (.dat).

Data saved as a .mat file will contain 3 variables: FRET (the 1D FRET transformed data); data (the raw 2D FRET data); and labels (the names of each trace). Each variable will be a 1xN cell array, where N is the number of traces.

When the data is saved as individual text files, each will contain 3 columns of data (from left to right): the time step of the data; the donor intensity; and the acceptor intensity.

When the data are saved as one large text file the traces are stored as adjacent T_n x 2 arrays, where T_n is the length of the n^{th} trace. The first column of each trace is the donor intensity and the second column of each trace is the acceptor intensity. Trace labels will be printed in the first row of the file. Both the donor and acceptor columns are labeled. Any traces shorter than the longest trace saved will have NaN's in the blank space at the end of the trace.

Save plots (using settings from main display)

This option will save plots of the data. No other information will be saved. The current plot settings on vbFRET will be used to generate the saved plots (so the saved plots will look the same way as they appear in the display window).

Plots can be saved in any of the following file formats: Adobe Illustrator (.ai), Bitmap (.bmp), EPS (.eps), Matlab figure (.fig), JPEG (.jpg), PDF (.pdf), Portable Network Graphic (.png) and compressed TIFF image (.tif).

Multiple traces can be plotted in each save plot file. Any of the following number of traces per file is allowed: 1, 2, 4, 6 or 9.

7 Autosave

When autosave is enabled, vbFRET will save a 'saved session' .mat file (see the above section on saving data) whenever data analysis is paused or completed[4]. A small window will open when you click 'Analyze Data', asking you to name the autosave file. If you 'Cancel', no autosave file will be created (pausing and resuming analysis will cause VBFRET to ask you again if you would like to create an autosave file, however). Autosave can be disabled by clicking 'File' → 'Disable Autosave'.

8 Deleting traces

Unwanted traces can be deleted both before and after traces have been analyzed (but not during analysis). To delete data click 'Traces' → 'Delete Traces' or press Ctrl+D. This will open the Delete Traces sub-GUI.

The sub-GUI contains two large listboxes with an 'Add -->' button above and a '<-- Remove' button below. The listbox on the left contains the names of the traces that will not be deleted. The listbox on the right contains the names of the traces that will be deleted. Initially, all the trace names will be on the left side.

[4]The autosave frequency can be changed in the 'Advanced Analysis Settings' sub-GUI (sec. 11, p. 11).

Tag traces for deletion by highlighting them and clicking 'Add -->' (multiple files can be added at once by holding down Shift, Ctrl or the left mouse button). This will move the trace names from the left listbox to the right listbox. Traces accidentally tagged for deletion can be untagged by highlighting them and clicking the '<-- Remove' button. Once all the traces you want to delete have been moved to the right listbox, click 'Delete Traces'. The screen will flicker, the Delete Traces sub-GUI will close and the traces will be deleted.

If desired, the remaining traces can be relabeled from 1 to [number of traces] by checking the 'Relabel Traces' checkbox before clicking 'Delete Traces'. Even if no traces are tagged for deletion, checking the 'Relabel Traces' checkbox and clicking 'Delete Traces' will relabel all the traces loaded into vbFRET.

If traces have had photobleaching removed (sec. 4, p. 4), clicking 'Delete Traces' will also delete the photobleached backup copies of all the traces.

9 Closing/restarting

To close vbFRET click 'File' → 'Exit'. Any unsaved changes to your data will be lost.

There is also a 'Reset' button located in the File menu. Clicking the 'Reset' will clear all analysis, delete all traces and restore all settings to the default. It is the same as closing vbFRET and reloading it.

10 Analysis settings

The most commonly adjusted analysis options can be set in the 'Analysis Settings' box in the bottom left corner of the Main GUI. More advanced options can be set in the Advanced Analysis Settings sub-GUI (see sec. 11, p. 11). The 'Analysis Settings' options are:

Number of FRET states

vbFRET will attempt to fit each trace with every integer number of FRET states between 'Min' and 'Max'. For example, setting 'Min' to 1 and 'Max' to 4 will cause vbFRET to fit each trace with 1, 2, 3 and 4 FRET states. Only the fit of the trace with the highest evidence will be saved.

Fitting attempts per trace

The fitting algorithm used by vbFRET requires starting guesses for the locations of each FRET state as well as the transition matrix. If these guesses are substantially different than the true parameters of the system then the algorithm might converge to a sub-optimal

solution (often called a local maximum)[5]. This can be avoided by fitting each trace multiple times using different starting guesses. Only the fit of the trace with the highest evidence score will be saved.

Fitting each trace multiple times will improve the fit of the data but requires additional time to perform the calculation. It is recommended that you fit the data as many times as time/CPU power allows. On our test data sets, we have found a noticeable improvement when fitting attempts are increased from 1–10, slight improvement when they are increased from 10–25 and no noticeable improvement above 50 fitting attempts per trace.

The first time a trace is fit, FRET states are initialized to be evenly distributed between 0 and 1. For all subsequent fitting attempts they are initialized to be randomly distributed between 0 and 1. Providing guesses for FRET states overrides these settings. The transition matrix is initialized randomly for each fitting attempt.

Use guesses for FRET states

Unless this option is checked, the initial guesses for the locations of FRET states will be set at random (see the 'Fitting attempts per trace' section above). Initial guesses for FRET states can be provided by the user by checking this checkbox and inputting guesses in the textbox immediately below. Guesses can be separated by spaces or commas. Starting guesses can only be input when the 'Use guesses for FRET states' checkbox is checked.

Note: If fewer states are input than are fit to the data, the additional states will be initialized at random. If more states are input than are fit to the data, than random subsets of the guesses are used.

11 Advanced analysis settings

The FRET analysis settings which are less commonly customized by the user can be found in the Advanced Analysis Settings sub-GUI, located in the 'Settings' menu. Once new settings have been selected they can applied by clicking 'OK' or 'Apply' (clicking 'OK' will close the sub-GUI as well). The default advanced analysis settings can be restored by clicking 'Default'. Clicking 'Cancel' will close the sub-GUI and discard any unsaved changes. The are three different groups of advanced analysis settings.

11.1 VBEM Options

These options pertain to vbFRET while the VBEM data fitting algorithm is running.

[5]This feature is not unique to the vbFRET inference algorithm. Programs such as HaMMy and QUB require starting guesses for FRET states and can converge to local maxima as well.

Max iterations per VBEM

vbFRET uses an iterative algorithm to fit the data. The maximum number of iterations allowed per fitting of a trace is set here.

Convergence threshold

The vbFRET algorithm will iteratively fit the data until two consecutive iterations of data fitting result in an insignificant increase in the fit's evidence (or the maximum number of allowed iterations is exceeded). The algorithm will cease iterations once the increase in evidence from the last iteration is less than the value specified here.

Plot results during analysis

When this box is checked, vbFRET will display the trace currently being fit and the best fit of the trace found so far.

Display analysis progress

When this box is checked, the white status box located below the 'Analyze Data!' button will display: (1) the trace currently being analyzed; (2) The current number of states being fit to the trace and the fitting attempt number; (3) how long the last fitting attempt took (in seconds); and (4) how many iterations of the fitting program were necessary.

11.2 Clean Blurred Data Options

Clean camera blurred data

FRET data collected by CCD camera shows the binned FRET signal intensity over each time step in the trace. When a change in FRET intensity occurs during a time step, the observed FRET signal is the *average* of the two intensities, weighted by the fraction of time spent at each intensity during the time step. The observed signal at these time steps is essentially an artifact of camera blurring, but is sometimes fit by vbFRET as a unique intermediate state[6].

If the data is fit while the 'Clean camera blurred data' checkbox is checked, vbFRET will attempt to correct these camera blur artifacts using the following algorithm: First, the data is fit normally. Next, The data are examined for single data point intermediates. These are data points where the datum immediately preceding is fit to a lower state and the datum immediately following is fit to a higher state (or vice versa). Single data point intermediates are then moved to the mean value of the closest state to the datum. Finally, traces which

[6]This is not a shortcoming of vbFRET – camera blur intermediate states are real, just not biologically relevant. These states can be differentiated from real, biological, short-lived intermediates by re-recording your data with different exposure times. Camera blur artifact states should not change with exposure time, but the lifetime (in terms of time steps) of real biological intermediates will change.

had single data point intermediates are then reanalyzed. The data is reanalyzed because the presence of blur artifacts can sometimes alter the evidence for the existence of an intermediate state and cause data points that are the noisy extrema of a state to be fit as part of a second spurious intermediate state.

Of course if your data has biologically meaningful data that meets this criteria for being single data point intermediates, then you should not use this option.

Both the original fit of the data and the camera blur cleaned fit are saved. You can toggle between the two fits using the 'Plot cleaned results' checkbox.

The 'Clean camera blurred data' option can be activated at any point in the analysis, even after analysis has been completed (i.e. checking 'Clean camera blurred data' and then pushing 'Analyze Data' again caused vbFRET to continue on to analyzing camera blur cleaned data, rather than starting its analysis over from the beginning).

Plot cleaned results

If data has been analyzed with the 'Clean camera blurred data' option, then checking this box toggles between the original data (and original idealized traces) and the camera blur fixed data. To let you know that you are viewing blur fixed data, the plot title will change from 'Best fit of trace X' to 'Best fit of cleaned trace X'. Traces that did not show signs of camera blurring will have a '*' after the word 'cleaned' to denote that the trace has not been altered during camera blur cleaning. If data has not been analyzed with the 'Clean camera blurred data' option, then checking this checkbox has no effect.

Important: This checkbox also toggles between the type of data saved (except for VBFRET*saved sessions, where everything is saved). If you save results, plots, idealized traces or raw data and 'Clean camera blurred data' is checked, then the traces and analysis results for the cleaned data will be saved. If 'Clean camera blurred data' is unchecked then traces and analysis results for the non-camera blur cleaned data will be saved.*

11.3 Hyperparameter Priors

Note: This section is intended to document the specific hyperparameters for the priors used in vbFRET and only provides a cursory discussion of hyperparameters and priors. Hyperparameters are an advanced topics in Bayesian statistics. Unless you are familiar with this topic and have a good understanding of the Bayesian model of the HMM with Gaussian observables, we strongly discourage changing the hyperparameter priors from their default settings.

vbFRET fits traces by using variational Bayes to calculate the evidence of the fit of the data. Like most Bayesian statistical methods, vbFRET treats all unknown parameters as random variables and assigns probability distributions to them. So, for example, rather

than say that the mean of a FRET state is 0.7, vbFRET says that the mean of the state is most probably 0.7, but could also take on any other value with a probability given by a Gaussian distribution (with a specified standard deviation, σ). The probability distributions over the parameters have parameters as well, which are known as hyperparameters. When vbFRET fits data, it is calculating the hyperparameters for the probability distributions over the unknown parameters, not the unknown parameters themselves.[7]

Another component of Bayesian statistics is that each random variable must be assigned a *prior* – a probability distribution reflecting your belief about the unknown parameter's value prior to seeing the data. The prior is used in conjunction with the observed data to calculate the *posterior* probability distributions for the hyperparameters, which are then used to find the most probable parameters for the system and fit the data.

If the hyperparameter priors are set to be weak (which they are in vbFRET), then the priors have only a minor impact on the fit of the data. When the data is reasonably well resolved (or very poorly resolved for that matter), the values of the priors do not impact the idealized fit of the trace. They can, however, have some impact on the fit of the data for traces that are right on the boarder of being resolvable. The effects of hyperparameter priors on FRET data fitting is analyzed in the supporting material of our paper by Bronson et. al (reference on p. 1).

upi

The probability that the system starts in a given FRET state is given by a K dimensional multinomial distribution: $\vec{\pi} = (\pi_1, ..., \pi_K) : \sum_{k=1}^{K} \pi_k = 1$, where π_k is the probability of starting in the k^{th} state. The prior $p(\vec{\pi})$ is given by a Dirichlet distribution[8]: $p(\vec{\pi}) = \frac{\Gamma(K*upi)}{\prod_{k=1}^{K}\Gamma(upi)} \prod_{k=1}^{K} \pi_k^{upi-1}$. Although this prior could take up to K hyperparameters, a single value of upi is used for all of them since, prior to seeing the data, the trace is equally likely to start in any of the K states.

mu, beta, W, v

The probability of an observed FRET signal, y_t, at time step t, when the system is in the k^{th} FRET state is given by a Gaussian distribution[9] : $p(y_t) = \sqrt{\frac{\lambda_k}{(2\pi)}} e^{-0.5*\lambda_k(y_t-\mu_k)^2}$. The prior probabilities for μ_k and λ_k are given by a Gaussian-Gamma distribution[10]:

[7] These probability distributions turn out to be sharply peaked around their maxima, however, so the most probable values for the parameters may be used to calculate idealized trace trajectories.

[8] In addition, the Dirichlet distribution has the following constraints:
(1) $0 \leq \pi_k \leq 1 \quad \forall\, \pi_k$
(2) $\sum_{k=1}^{K} u_k = 1$

[9] The precision, λ, is used instead of the variance, σ^2, because it simplifies the math vbFRET uses to fit the data. The precision is the inverse of the variance, $\lambda = \frac{1}{\sigma^2}$.

[10] The slightly unusual representation of the Gamma distribution here simplifies the internal math of vbFRET as well.

$$p(\mu_k, \lambda_k) = p(\mu_k|\lambda_k)p(\lambda_k) = \sqrt{\frac{beta*\lambda_k}{2\pi}}e^{-0.5*beta*\lambda_k(\mu_k-mu)^2}\frac{1}{\Gamma(v/2)}(2W)^{-v/2}\lambda^{(v/2)-1}e^{-\frac{\lambda}{2W}}$$

Prior to seeing the data, all FRET states are interchangeable, so mu, beta, W and v are used as the hyperparameter priors for all K FRET states.

ua, uad

The probability of transitioning between FRET states is stored in a transition matrix, A, where the matrix entry a_{jk} holds the probability that the system is in the k^{th} FRET state at time t, given that it was in the j^{th} FRET state during the previous time step (i.e. the rows of A are multinomial distributions). The rows of A must be normalized and the prior probability of the k^{th} row is given by a Dirichlet distribution: $p(a_{j1}, ..., a_{jK}) = \frac{\Gamma(K*ua+uad)}{\prod_{k=1}^{K}\Gamma(ua_*)}\prod_{k=1}^{K}a_{jk}^{ua_*-1}$, where $ua_* = ua$ when $k \neq j$ and $ua_* = (ua + uad)$ when $k = j$. Although this prior could take up to K^2 hyperparameters only two, ua and uad, are used because prior to seeing data all states are interchangeable. The only important quantities are the probability of switching to new states (governed by ua) and the probability of staying in the current state (governed by the sum of ua and uad).

11.4 Autosave Options

Autosave frequency

When set to < 1, vbFRETwill autosave your data whenever analysis is paused or completed. If autosave frequency is set to $n \geq 1$ then vbFRETwill autosaved your data every n traces and whenever analysis is paused or completed.

12 Plot settings

The plot settings panel, located in the bottom right quadrant of the vbFRET GUI, controls the plot display of vbFRET. The textbox in the top left corner of the panel displays the name of the trace currently displayed in the plot display. The text box in the top right of the panel contains the current trace's number as well as the total number of traces loaded (i.e. trace n of N). The scrollbar directly below the trace name textbox incrementally changes the trace currently displayed. In addition, it is possible to jump to a trace by name or number by typing the trace name into the trace name textbox or the trace number into the trace number textbox.

Clicking the 'Enlarge Plot' button will open up a new MATLAB figure window with the current plot in it. In addition to being larger, the plot can be manipulated with all of MATLAB's plotting tools (zoom, Property Editor, etc.).

If the current trace has already been fit, then the 'Plot' sub-panel toggles between displaying the raw (unfit) and analyzed (fit) data. If a trace has not been analyzed yet, then the raw trace will be displayed regardless of the plot setting.

The 'Plot in:' sub-panel toggles between displaying the FRET transformation of the data (1-D) and the donor and acceptor signals separately (2-D). If fit data is displayed in 2-D, vbFRET will use the fit of the data in 1-D to generate a 2-D fit of the data[11]. The donor is plotted in green and the acceptor in red. The idealized trace is shown as dashed lines as a reminder that the idealization is an extrapolation from the 1-D fit of the trace rather than a fit of the donor and acceptor signals.

The 'Show:' sub-panel has two checkboxes, 'Data Points' and 'Fit Points'. Checking these boxes will cause Xs to appear over the observed data and idealized trace, respectively.

13 Advanced plot settings

The plot display can be customized using the Advanced Plot Settings sub-GUI, located in the 'Settings' menu. The background color of the plot can be changed using the 'Background Color' pulldown menu. Grid lines can be added/removed with the 'Enable Grid Lines' checkbox. When 'Fix Y-axis' is checked, the Y-axis of the 1-D plot display will range from -0.1 to 1.1. When this button is unchecked, the Y-axis will range from $MIN - 0.05 * MAX$ to $1.05 * MAX$, where MIN and MAX are the smallest and largest valued data points, respectively, being plotted. This box should be unchecked when the data do not take on values between 0 and 1 (e.g. if the program is being used to analyze a non-smFRET time series).

The line color, style and thickness and marker type, size and color (for plotting individual data points) for each type of plot line can be customized in this sub-GUI as well. To customize a type of plot line, select it from the 'Plot Line' pull down menu. The customization options below the 'Plot Line' pulldown menu will pertain to the type of line selected. The plot line options are: FRET Data (1D), FRET Fit (1D), Donor Data (2D), Acceptor Data (2D), Donor Fit (2D) and Acceptor Fit (2D).

14 Customize default settings

There are many analysis and plotting settings in vbFRET which the user can customize. Since reconfiguring all these settings every time you open vbFRET would get tedious,

[11] vbFRET calculates the idealized donor and acceptor values for each FRET state by taking the mean donor and acceptor values for all the data assigned to that FRET state in the 1-D fit of the trace.

you can change the default setting which load when you open vbFRET, click 'Reset' or click 'Default' on any of the sub-GUI windows. To reconfigure your settings to be the default, simply click 'Make Current Settings Default' under the 'Settings' menu. The current VBFRET settings will become the default settings. If you want to restore the default settings that came VBFRET, click 'Restore Original Settings' under the 'Settings' menu.

15 Input file formats

Terminology: In this section N will denote the total number of traces in a file, T_n the length of the n^{th} trace, I_D the observed intensity of the FRET donor and I_A the observed intensity of the FRET acceptor.

vbFRET can read several different types of data files. Multiple data files can be loaded at once, even if they are different types of file formats. The only exception is 'Saved Session' .mat files. When one of these is loaded, all other files being loaded will be ignored and any traces already loaded will be deleted. Additional traces can still be loaded after loading a 'Saved Session' though. For all types of imported data, the 2-D FRET trace must be loaded[12]. Any of the following types of files can be read by vbFRET[13].

15.1 vbFRET saved session

vbFRET saved sessions are .mat files created by vbFRET. They can be loaded without any modifications (and probably should not be modified by the user anyway).

15.2 .mat files

Traces and idealized traces may be saved and loaded as .mat files. The .mat file must contain a cell array named 'data' and can also contain cell arrays named 'labels' and 'path'. All other saved variables will be ignored.

- **data:** This variable must be a 1xN cell array. Each cell of must be a T_nx2 array. The first column of the array must contain the time series for the donor signal and the second column must contain the time series for the acceptor signal.

[12] If your data is already 1D FRET transformed, set I_A equal to your 1D FRET signal and I_D equal to 1-I_A.

[13] These allowed formats are compatible with input files prepared for HaMMy, as well as path output files generated by HaMMy.

- **labels:** If this variable is present it must be a 1xN cell array. Each cell must contain a text string. The n^{th} cell will be used as the label for the trace in the n^{th} cell of the 'data' variable. If this variable is absent, traces will be labeled 1 to N.
- **path:** If this variable is present it must be a 1xN cell array. Each cell must be a T_nx1 vector containing the idealized 1D FRET trajectory for the n^{th} trace.

15.3 .dat (.txt) files

vbFRET can load several types of .dat and .txt files (it makes no distinction between the two extensions). Any lines at the beginning of the file that begin with '@', '#', '$', '%' or '&' are treated as comments and are ignored. If only one trace is present in the data file, the name of the data file is used as the trace's label. If multiple traces are present, the first row of the traces is used for the trace labels. The first row of the traces can be strings of text in this case.

The first column of data is assumed to be I_D for trace 1. The second column is assumed to be I_A for trace 1. If there is a third column it is assumed to be I_A for trace 2, and so on. Columns must be separated by tabs or white space (multiple tabs/spaces are treated as a single one).

The first column may also denote the time steps of the trace (i.e. be sequential integers). If this is the case, the first column of the data file is discarded. The second column is treated as I_D for trace 1, the third column is treated as I_A for trace 1 and so on.

15.4 Path files

Path files for individual traces can be loaded as well. vbFRET will assume any .dat or .txt file name with the phrase 'path' (case insensitive) is a path file. A path file contains 5 columns of data (from left to right): the time step of the data; I_D; I_A; the 1D FRET transform; and the fit (idealized FRET) of the data point.

Both the trace and its idealized fit will be loaded into vbFRET. The name of the file will be used as the label for the trace.

vbFRET will assume it is a .dat file unless it is a path file or a .mat file (so extensions like .txt will be treated the same way as a .dat file).

16 Tutorial

This is a short tutorial to guide you through the basics of loading data, removing photobleaching, analyzing data and viewing the results. The tutorial assumes you have already opened MATLAB and changed the directory to 'vbFRET'.

1. Type 'vbFRET' to load the vbFRET GUI.
2. Click 'File' → 'Load Data'.
3. Click 'Add Files'. Go to the 'example_files' folder and double click 'tutorial_data.dat'.
4. Click 'Load Data' in the Load Data sub-GUI. There should now be three traces loaded into the vbFRET display window of the main GUI. Use the buttons in the Plot Settings panel of the main GUI to scroll through the plots and view the traces in both 1-D and 2-D. The wild looking signal at the end of each 1-D trace is a result of photobleaching (which is easy to see when the traces are viewed in 2-D). Before analyzing the data, we'll want to get rid of some of the photobleaching.
5. Click 'Traces' → 'Remove Photobleaching'. In the Remove Photobleaching sub-GUI, change 'Photobleach identification method' to 'Summed Channel' and click 'Remove Photobleaching!'. This will not remove all of the photobleaching from the traces, but it will get rid of most of it. The traces are now cleaned up enough to analyze.
6. Two of the traces in this data set have blur states (see sec. 11.2 for a discussion of blur states). We'll want to get rid of these blur states during analysis, so before analyzing the data click 'Settings' → 'Advanced Analysis Settings'. Check the 'Clean camera blurred data' checkbox in the Advanced Analysis Settings sub-GUI and click 'OK'. Now you're all set to analyze your data!
7. Push the large 'Analyze Data!' button on the main GUI. A window will open asking you for an autosave name. Enter a name for the autosave name and hit 'OK' or just push 'Cancel' to continue without autosaving. The data will now begin analyzing. The vbFRET display will update with the best fit of the current trace. After all the fitting, traces 1 and 2 should be fit with 3 states (2 real and 1 photobleached) and trace 3 should be fit with 2 states.
8. To see a summary of the results, click 'File' → 'Save Data'. Change the top pulldown menu from 'Save Session'[14] to 'Save analysis summary'. You can save the results of your analysis as either a text file (.dat) or MATLAB file (.mat). That's it!

[14] Remember, you must save your data using 'Save Session' if you wish to be able to reload your data later and then save an analysis summary.