# Experimental data and program files

for the paper

**"A possible universal intermediate identified during the folding of a metamorphic protein" by C Correa et al. (in preparation)**

## Overview

The program files cover three main tasks:

1. Postprocessing of experiment files: Extract information about rips and zips to Matlab tables.
2. Fit mathematical models to the extracted data, resulting in estimated model parameters and their confidence intervals.
3. Create figures and tables shown in the paper and supplementary documents.

Matlab code for the three tasks is organised in three folders:

Task 1: Top7paper/Experiment\_postprocessing

Task2: Top7paper

Task 3: Top7paper/Figures\_and\_tables

All programs are written in Matlab and were run under release R2024a. In addition to basic Matlab the following toolboxes are used:

|  |  |
| --- | --- |
| Signal Processing toolbox | Postprocessing |
| Optimization toolbox  Statistics and Machine Learning toolbox | Fitting models |

## Experiment files

The Optical Tweezers instrument records a set of data at regular time intervals. The recording frequency may vary between 10 and 4000 records per second but is normally 200 or 400 records per second. The record files have one column per variable recorded. The relevant variables for this study are

CycleCount: Number of controller cycles since the experiment started.   
The cycle frequency is 4000 cycles per second, so the recording time in seconds is CycleCount/4000

Y\_force: Pulling force (pN)

A\_dist-Y, B\_dist-Y: Position of the trapped pulling bead (trap position, nm).   
We use the mean of the two values.

Status: Used to code the heating power in the chamber. Can be translated to temperature difference between chamber and bath.

The uncompressed experiment files take up about 3GB of storage. We include only a few, in folder Experiment\_postprocessing/SampleExperiments, to enable testing of the postprocessing functions. The full set is available from the authors.

## Postprocessing

### Main functions

function [Trip,Tzip] = analyse\_experiment(file)

Input is a file generated by the optical tweezers instrument. Output is two Matlab tables with 16 columns and one row for each identified rip or zip.

|  |  |
| --- | --- |
| **Column** | **Content** |
| Filename | Name of experiment file |
| Time | Time of rip (seconds since experiment start) |
| Deltax | Δx = Change in trap position at force = Force (nm) |
| Force | Pulling force at start of rip/zip (pN) |
| Temperature | At protein (may be higher than in bath) °C |
| Forceshift | Shift in pulling force at rip/zip (pN) |
| Trapx | Position of trapped bead (nm) |
| Fdot | Rate of change of pulling force (pN/s) |
| Slope\_b | d(pulling force=/d(Trapx) before rip/zip |
| Slope\_a | d(pulling force=/d(Trapx) after rip/zip |
| Pullingspeed | Speed of trappd bead (nm/s) |
| Topforce | Maximum force in trace pN) |
| Noise | Measure of rapid change in pulling force (pN/s) |
| Cycleno | Index of pull/relax trace |
| Work | Crooks work |
| Timestep | Length of recording time step (s) |

[TRIP,TZIP] = analyse\_many(files);

This function runs analyse\_experiment for all files in the string array “files” and collects the output Trip tables into TRIP and the Tzip tables into TZIP

For the paper, we have quantified nearly 22000 rips and zips in 239 experiment files. This takes about four minutes on a standard laptop computer. The resulting tables form the basis for most tables and figures in the paper. They are found in the file “Tables.mat”.

### Input parameters

One challenge is to strike the best balance between maximising the rip/zip candidates found while minimising the number of spurious rips/zips. The parameter file params.m contains several tuning parameters that were found by trial and error. Here the user can also specify if he/she wants to look for more than one rip per trace, and whether to include rips in the relaxing trace (late rips). For this paper, the search is limited to a single rip in the pulling trace.

## Fitting models

### Grouping experiments

Experiments have been performed under a range of temperatures and at varying speeds of the trapped bead (pulling speed). We have defined four temperature ranges and three speed ranges. Experiments were performed at eight combinations and separate fits were made for two or three of the clusters shown in Figure S3. The number of rips found for each combination is shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Temperature** | **Speed (m/s)** | **Zips** | **Cluster** | **Rips** |
|  |  |  |  |  |
| 3-7°C | <50 |  | 1 | 214 |
|  | <50 | 454 | 2 | 24 |
|  | <50 |  | 3 | 43 |
|  | 50-250 |  | 1 | 3008 |
|  | 50-250 | 6633 | 2 | 1127 |
|  | 50-250 |  | 3 | 1030 |
|  | >250 |  | 1 | 54 |
|  | >250 | 248 | 2 | 56 |
|  | >250 |  | 3 | 42 |
| 7-14°C | 50-250 |  | 1 | 387 |
|  | 50-250 | 663 | 2 | 51 |
|  | 50-250 |  | 3 | 76 |
| 14-20°C | 50-250 |  | 1 | 842 |
|  | 50-250 | 884 | 2 | 83 |
|  | 50-250 |  | 3 | 99 |
| 20-30°C | <50 |  | 1 | 29 |
|  | <50 | 85 | 2 | 9 |
|  | <50 |  | 3 | 3 |
|  | 50-250 |  | 1 | 1102 |
|  | 50-250 | 1255 | 2 | 91 |
|  | 50-250 |  | 3 | 65 |
|  | >250 |  | 1 | 70 |
|  | >250 | 504 | 2+3 | 130 |

At high temperature and slow pulling speed, the number of rips in clusters 2 and 3 is too low to fit any models. At high temperature and fast pulling speed, clusters 2 and 3 are lumped together as there are only two peaks in the observed probability density.

The experiments were thus grouped into 14 unfolding and 8 refolding groups based on temperature and pulling speed. For every combination (8 for zips and 21 for rips), the experiment probability density (pdf) for the rip and zip forces was calculated as the number of events with force within 1pN bins, divided by the total number of rips or zips for the given combination.

### Model fitting: *run\_fit*

The function

Tout = run\_fit(TRIP,TZIP)

This function calculates parameters for the Bell and Dudko unfolding models for all the 21 unfolding combinations as well as the Bell refolding model for the eight temperature and pulling speed combinations. Using those parameters, it calculates the kinetic and the Crooks value for the energy barrier. The input can be any pair of rip and zip tables, such as the set for a given solute in the bath. (See the paragraph Tables.mat, below)

The Bell and Dudko models calculate force pdfs. For each group and cluster, the models were fitted to the experiment pdfs using Matlab’s *lsqcurvefit* function. The fitting process gives values for the model parameters that give the best fit between model and experiment pdfs. For the Bell model the parameters are the distance from the potential minimum to the energy barrier, and the unfolding rate at zero force. The Dudko model has the height of the energy barrier as an additional parameter.

To assess the accuracy of the parameter values we calculate confidence intervals for the parameters. This is done by bootstrapping using the Matlab function *bootci*. In the paper we show confidence intervals calculated using 200 bootstrap samples. The results stored in Tout in the repository were calculated using 1000 samples and thus should be more accurate. The number of samples is specified in function *run\_fit*.

### Tables.mat

This file contains rips and zips tables from *analyse\_many*.

* TRIP, TZIP: All experiments used in the paper
* T(RIP/ZIP)Top7 : Experiments without solutes in the bath
* T(RIP/ZIP)Top7BSA: With BSA in the bath
* T(RIP/ZIP)Top7FOXP1: With FOXP1 in the bath
* T(RIP/ZIP)Top7Top7: With Top7 in the bath

In addition, it contains Tout: The output from run\_fit(TRIP,TZIP).

## How to reproduce figures and tables

First run *setup* to add all three folders to your Matlab path.

To reproduce individual figures run the corresponding m-files in the Top7paper/ Experiment\_postprocessing folder. *create\_figures* creates all the figures.

The *Table* functions create output to the Command Window. This can be copied and pasted directly into a MS Word table:

Create an empty table as specified in the help text (top comment) for the Table function. Copy the output and paste into the empty table. Set the font to Times New Roman, font size 9. The values for contain the multiplier 10xx. Mark xx and set it to superscript.

## Acknowledgement

The function *movingslope*, used for identifying rips and zips, was created by John D’Errico and was copied from the Matlab Central File Exchange.

Ref.: John D'Errico (2024). Movingslope (https://www.mathworks.com/matlabcentral/fileexchange/16997-movingslope), MATLAB Central File Exchange. Retrieved April 19, 2021.

## Dependency graphs

### Analyse\_many

A screenshot of a computer

AI-generated content may be incorrect.

### run\_fit

A screenshot of a computer

AI-generated content may be incorrect.

### Figures

A screenshot of a computer

AI-generated content may be incorrect.

### Tables

****

**A diagram of a table

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A diagram of a science and table

AI-generated content may be incorrect.**

**A diagram of a computer

AI-generated content may be incorrect.**

**A diagram of a company

AI-generated content may be incorrect.**