bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mars, a molecule archive suite for reproducible analysis and reporting of single molecule properties from bioimages

Nadia M Huisjes[1,*], Thomas M Retzer[1,2,*], Matthias J Scherr[1], Rohit Agarwal[1,2], Barbara Safaric[1], Anita Minnen[1], and Karl E Duderstadt[1,2,✉]

[1]Structure and Dynamics of Molecular Machines, Max Planck Institute of Biochemistry, Martinsried, Germany
[2]Physik Department, Technische Universität München, Garching, Germany
[*]these authors contributed equally to this work.

The rapid development of new imaging approaches is generating larger and more complex datasets revealing the time evolution of individual cells and biomolecules. Single-molecule techniques, in particular, provide access to rare intermediates in complex, multistage molecular pathways, but few standards exist for processing these information-rich datasets, posing challenges for wider dissemination. Here, we present Mars, an open-source platform for storage and processing of image-derived properties of biomolecules. Mars provides Fiji/ImageJ2 commands written in Java for common single-molecule analysis tasks using a Molecule Archive architecture that is easily adapted to complex, multistep analysis workflows. Three diverse workflows involving molecule tracking, multichannel fluorescence imaging, and force spectroscopy, demonstrate the range of analysis applications. A comprehensive graphical user interface written in JavaFX enhances biomolecule feature exploration by providing charting, tagging, region highlighting, scriptable dashboards, and interactive image views. The interoperability of ImageJ2 ensures Molecule Archives can easily be opened in multiple environments, including those written in Python using PyImageJ, for interactive scripting and visualization. Mars provides a flexible solution for reproducible analysis of image-derived properties facilitating the discovery and quantitative classification of new biological phenomena with an open data format accessible to everyone.

Single-Molecule Microscopy | Reproducibility | Bioimaging | Software | Image Analysis | FRET | Single-Particle Tracking | Flow Magnetic Tweezers

Correspondence: ***duderstadt@biochem.mpg.de***

Introduction

Reproducible analysis of bioimaging data is a major challenge slowing scientific progress. New imaging techniques generate datasets with increasing complexity that must be efficiently analyzed, classified, and shared. This challenge has gained wide recognition (Carpenter *et al.*, 2012; Eliceiri *et al.*, 2012; Meijering *et al.*, 2016; Ouyang and Zimmer, 2017), which has led to the development of software frameworks for curating images (Allan *et al.*, 2012; Kvilekval *et al.*, 2010) and analysis workflows (Rubens *et al.*, 2020), metadata reporting standards (Goldberg *et al.*, 2005; Lerner *et al.*, 2021) and the creation of public archives to increase data availability (Williams *et al.*, 2017). However, few standards exist for the reproducible analysis and re-use of image-derived properties and those that have been developed for single molecule fluorescence (Greenfeld *et al.*, 2015; Ingargiola *et al.*, 2016) offer limited options for adaption to other experimental configurations. There is an increasingly powerful toolkit to quantitatively follow the time evolution of the position, shape, composition, and conformation of individual cells and complexes, but these precious information-rich observations are generated in heterogenous formats that do not provide easy, transparent access to key features. As a consequence, the promise of new technologies is often unrealized, and new biological phenomena remain undiscovered in existing datasets, due to the lack of tools that enable robust classification and interactive exploration.

Single-molecule techniques provide access to rare intermediates in complex, multistage molecular pathways. Multicolor fluorescence imaging has revealed the conformational dynamics of membrane transport (Akyuz *et al.*, 2013; Erkens *et al.*, 2013), molecular states underlying assembly and transcription by RNA polymerase (Baek *et al.*, 2021; Duchi *et al.*, 2016), and DNA replication dynamics (Duderstadt *et al.*, 2016; Scherr *et al.*, 2018; Ticau *et al.*, 2015). These approaches have been combined with spatial tracking on biological structures to clarify how exchange events and conformational changes modulate the function of motor proteins, as well as replication, transcription and DNA repair machineries (Crickard *et al.*, 2020; Lewis *et al.*, 2020; Niekamp *et al.*, 2021; Scherr *et al.*, 2021). Biological macromolecules are routinely attached to microspheres which allow for controlled studies of the forces and torques involved in basic biological reactions (Agarwal and Duderstadt, 2020; Dulin *et al.*, 2013; Neuman and Nagy, 2008; Revyakin *et al.*, 2006). And finally, improvements in the sensitivity of camera sensors, fluorophore brightness, and new illumination strategies (Gao *et al.*, 2012; Tokunaga *et al.*, 2008) have enabled time-resolved studies of biological processes in live cells. These developments have revealed frequent exchange of factors during normal operation of replisomes (Beattie *et al.*, 2017; Kapadia *et al.*, 2020) and the dynamics of the transcription-factor target site search (Chen *et al.*, 2014).

The discovery of new biological phenomena from these multidimensional observations, depends on long, multistage image analysis workflows, followed by careful grouping and feature classification. Images are typically pre-processed, to

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

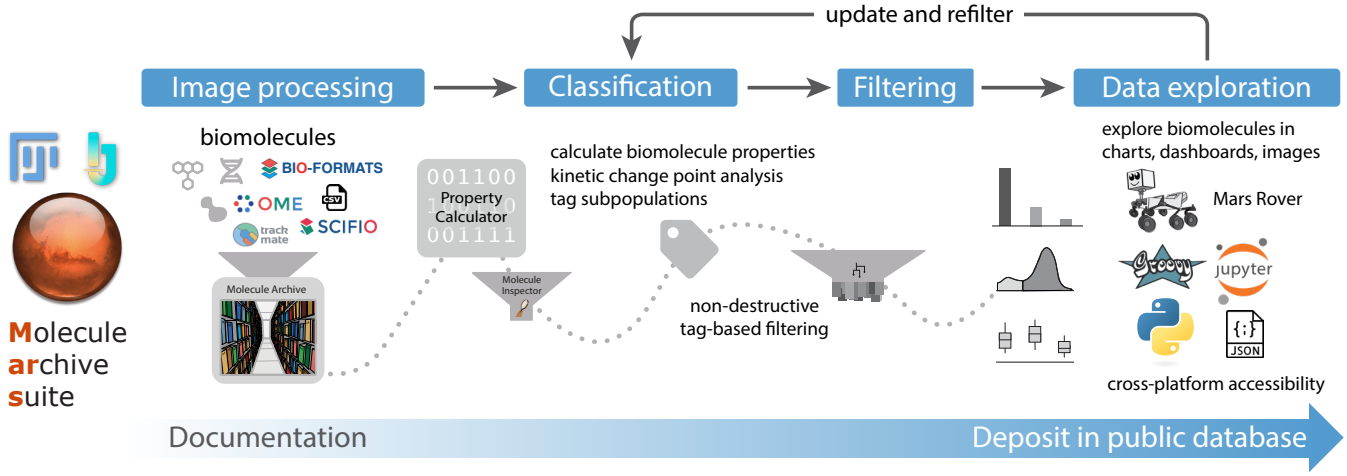


Fig. 1. The process of reproducible data analysis with Mars starting from image processing to iterative rounds of classification and filtering to the final stage of data exploration and deposition into a public database.

correct for non-uniform beam profiles (Peng *et al.*, 2017) and filtered (Marr and Hildreth, 1980) to enhance detection of individual molecules and structures over background. Additional tools are then used to follow the properties of individual biomolecules through time. These often provide some form of tabular data containing aggregated results, which are either manually evaluated or filtered using a feature unique to the desired group of observations. Typically, at this point, results can no longer be re-evaluated in the context of the original images due to unidirectional transformations or lack of interactive image evaluation software. The final data are frequently migrated to another platform ideally suited for the generation of publication-quality figures. This necessitates further data restructuring during transfers between software platforms. Large datasets are downsized to cope with limited storage by removal of rejected observations in ways that irreversibly alter datasets. As rejection criteria change and additional data must be incorporated, precious time and reproducibility are lost. This process is often repeated for each experimental application and research group, leading to substantial duplication of efforts.

Recognizing these issues in common single-molecule image processing workflows, we developed Mars, which provides a collection of ImageJ2 commands and integrations with well-established Fiji workflows for processing images and image-derived properties based on a versatile Molecule Archive architecture (**Fig. 1**). This architecture allows for seamless virtual storage, merging, and multithreaded processing of very large datasets. A simple, yet powerful interface allows for fast and easy access to subsets of records based on any biomolecule property. This framework allows for the same data structure to be used from initial image processing all the way to the generation of final figures. All image metadata and biomolecule records are assigned universally unique ids (uids) upon creation, which together with comprehensive logging, ensures the history of each record remains traceable through long and complex analysis workflows involving numerous data merging steps. Non-destructive tag-based filtering ensures no observations are lost and updates to rejection criteria only require retagging of processed records. These design principles facilitate the reproduction of analysis workflows and ensure Molecule Archives provide a comprehensive, transparent format for deposition of final datasets.

The modular design of commands and minimal definitions of image metadata and biomolecule records ensure flexibility that facilitates the development of varied Mars workflows. To illustrate the range of image processing and kinetic analysis tasks that can be accomplished, we present the results of three applications. We demonstrate the versatility of Mars record types in representing both single molecules as well as large macromolecular structures by tracking single RNA polymerases on long DNAs. We benchmark a multichannel fluorescence integration workflow using a well-established single-molecule FRET framework. And finally, we show how high-throughput imaging of DNA-tethered microspheres reveals the results of complex topological transformations induced by controlled forces and torques. In each example, we highlight image processing commands and methods for accessing and manipulating Molecule Archives using scripts. Each workflow provides a basic framework that can be further adapted to custom applications by introduction of additional processing steps and modification of the scripts provided. The graphical user interface provided by Mars Rover facilitates workflow refinement through interactive data exploration and manual classification with multichannel image views. Taken together, these features make Mars a powerful platform for the development of imaging-processing workflows for the discovery and quantitative characterization of new biological phenomena in a format that is open to everyone.

Results

Molecule Archive architecture. Molecule Archives provide a flexible standard for storing and processing image-derived properties adaptable to a broad range of experimental configurations. Molecule Archives contain three record types: *Properties*, *Metadata*, and *Molecule* (**Fig. 2**). Each

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

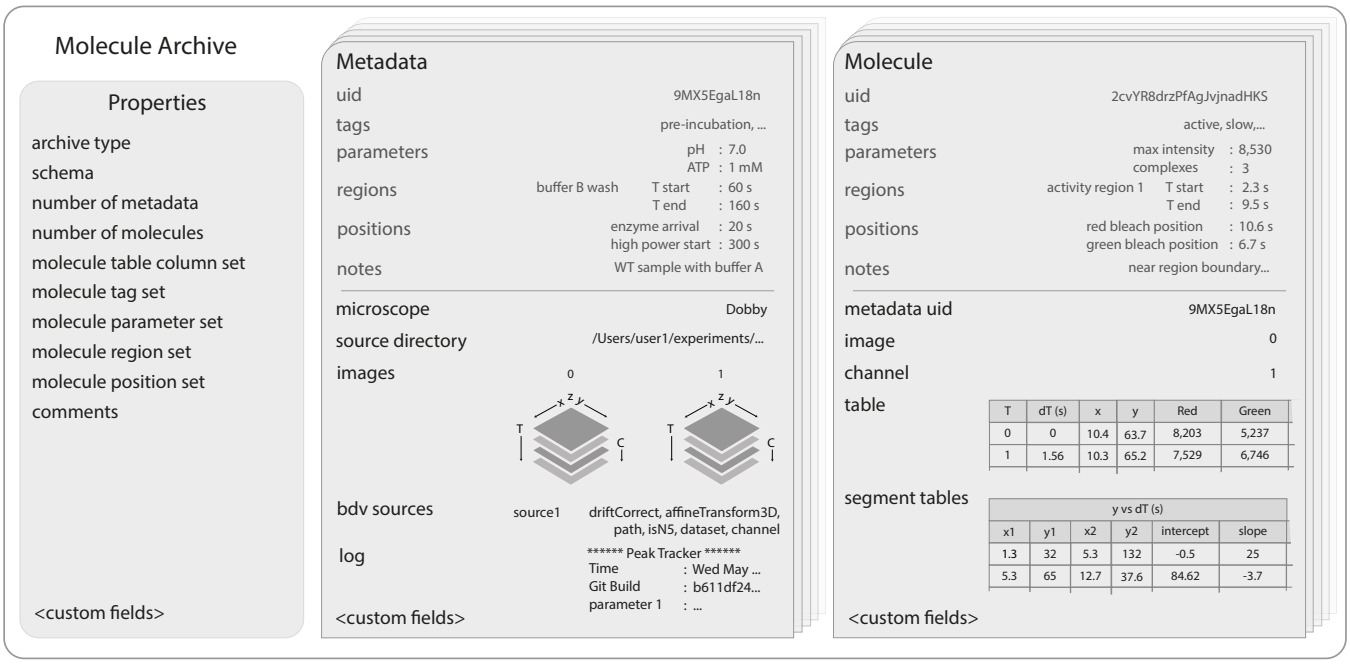


Fig. 2. Schematic representation of the structure of Molecule Archives consisting of three types of records: *Properties*, *Metadata*, and *Molecule*. The single *Properties* record contains global information about the Molecule Archive contents, the *Metadata* records store information about the images used for biomolecule analysis (e.g. image dimensions, the analysis log), and the *Molecule* records store molecule-specific information (e.g. position over time, intensity).

type is defined in an interface, independent of implementation details. This abstraction ensures Molecule Archives support a variety of biomolecule, metadata, and property implementations. Moreover, this allows for the creation of new implementations that seamlessly work with the existing code base, algorithms, and user interface. To simplify the mechanics of record retrieval and the process of dataset merging, all *Molecule* and *Metadata* records are assigned human-readable, base58-encoded universally unique ids (uids). Storing records using uids reduces indexing requirements, facilitates scalable processing using uid-to-record maps that support multithreaded operations, and ensures the traceability of records through analysis workflows.

Molecule Archives contain a single *Properties* record with the type of the Molecule Archive and global information about the Molecule Archive contents. This includes the number of *Metadata* and *Molecule* records and unique names used for tags, parameters, regions, positions, and table columns. The *Properties* record stores global Molecule Archive comments typically containing important information about the analysis strategy and naming scheme for tags, parameters, regions, positions, and other fields to orient a new researcher that did not perform the original analysis. To improve the organization and readability of comments, the Mars Rover provides a convenient Markdown editor.

Metadata records contain experimental information about the images used for biomolecule analysis. To ensure compatibility with the broadest set of image formats and maximum reusability, image dimensionality, timepoints, channels, filters, camera settings, and other microscope information are stored in OME format (Goldberg *et al.*, 2005). Molecule Archives do not store the raw images, only the image-derived properties and the source directory containing the project. Nevertheless, interactive image views linked to molecule records are supported through Big Data Viewer (BDV) integration (Pietzsch *et al.*, 2015) with HD5 and N5 formatted images. The image file locations, coordinate transforms, and other settings are stored in *Metadata* records. Global events in time influencing all biomolecules are documented using Metadata regions and positions that are available for kinetic analysis and scripting. Global experimental conditions are stored as parameters in the form of key-value pairs (e.g. buffer composition and temperature). *Metadata* records are categorized using tags to filter biomolecules by whole experiments. *Metadata* records contain a log where all commands and settings used for processing are recorded, thus maintaining the entire history of data processing throughout the analysis.

Molecule records contain fields for convenient storage of common image-derived properties of biomolecules. This includes a table that typically contains the position and intensity over time, the uid of the *Metadata* record containing the primary experimental information, and the index of the image containing the biomolecule. Events of interest in time can be marked with regions and positions (e.g. activity bursts, dye bleaching), which allow for event-specific calculations and kinetic analyses. Calculated global biomolecule properties are stored as parameters in the form of key-value pairs (e.g. mean intensity, distance traveled, mean position). Mars comes with kinetic change point (KCP) commands for unbiased identification of distinct linear regimes and steps (e.g. polymerase synthesis rate, FRET states, dye bleaching) (Hill *et al.*, 2018), which are stored in segment tables. And finally, *Molecule* records are categorized for later analysis using custom tags and notes for manual assessments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Molecule Archives can be created, saved, and reloaded in a variety of formats in multiple environments both Desktop-based and without a user interface for parallel processing on high-performance clusters. Molecule Archives are saved in JSON format using the field schema outlined in **Fig. 2**. By default, Molecule Archives are written to single files with a yama extension using smile encoding to reduce file size, but can also be saved and reopened in plaintext JSON. Mars supports processing of very large datasets, that typically do not fit in physical memory, using a virtual storage mode in which records are retrieved only on-demand supported by a simple filesystem-backed record hierarchy. This architecture provides a powerful and flexible framework for multistep analysis workflows involving very large datasets.

Mars Rover – Interactive molecule feature exploration and image views. The discovery of new biological phenomena using single-molecule techniques often relies on manual exploration of individual biomolecules in primary images together with image-derived measurements. To simplify this process, we developed a user interface in JavaFX, called the Mars Rover, that provides access to all information stored in Molecule Archives. The Mars Rover is integrated into Fiji with windows available for all open Molecule Archives. Open Molecule Archives are available as inputs and outputs in ImageJ2 commands and in supported scripting languages.

In the Mars Rover, Molecule Archive windows contain tabs and subpanels that provide complete access to all fields of *Molecule* and *Metadata* records as well as global properties, comments, and interface settings (**Supplementary Fig. 1**). A customizable global dashboard provides information about the Molecule Archive contents and scriptable chart widgets that can be adapted to specific workflows. The Molecule tab features an interactive chart panel that reveals the time evolution of biomolecule properties and provides region and position highlighting tools. To ensure image-derived measurements of biomolecule properties are evaluated together with primary observations, interactive image views linked to molecule records are supported through Big Data Viewer integration (Pietzsch *et al.*, 2015).

The Mars Rover was written to provide extensive possibilities for customization. All tabs and panels are defined in interfaces, independent of implementation details. This facilitates extension of the Mars Rover to support custom icons and display elements based on Molecule Archive type. This will enable further refinement and the development of workflow-specific displays by extending the core architecture in the future.

Commands for image processing and biomolecule analysis. Mars comes with a collection of several dozen ImageJ2 commands for common image processing and analysis operations (**Supplementary Table 1** – Command list). This includes commands to find, fit, integrate, and track intensity peaks in images. Together with a standard beam profile correction and region of interest transformation commands, all basic operations in common single-molecule analysis workflows can be performed. KCP commands provide options for more refined kinetic analysis and identification of unique features through classification of rates, pauses, and steps (Hill *et al.*, 2018). The UI-agnostic format of ImageJ2 commands ensures they can be used in many different contexts with or without a graphical user interface available. To facilitate these applications, methods have been added for all required settings and example scripts for all commands. Commands can be easily combined into larger scripts that support multistep analysis workflows runnable on high-performance computing clusters. This facilitates a smooth transition from a dialog-based workflow development phase to high-performance parallel processing of many experiments in environments lacking a user interface.

TrackMate interoperability. Fiji provides a comprehensive open-source platform for scientific image analysis containing well-established software for common image processing tasks. These technologies are integrated in a modular fashion as plugins that can be combined in limitless combinations. Mars commands and data structures are fully integrated into Fiji, simplifying interoperability with these technologies. The applications presented below provide examples of how Fiji plugins can be combined with Mars commands. To illustrate how this interoperability can be further extended, we developed an Action to export TrackMate results to Molecule Archive format (Source code available at https://github.com/duderstadt-lab/mars-trackmate). TrackMate is a Fiji plugin for single-particle tracking that offers several tracking algorithms with a powerful user interface with many spot filtering and track editing tools (Tinevez *et al.*, 2017). The Action we developed adds an export option in the final TrackMate panel called 'Go to Mars' that opens a Molecule Archive with the converted results. This feature is installed with Mars and requires no additional configuration. The Mars *Peak Tracker* is ideal for everyday single-molecule tracking problems with a few simple options in a single dialog but does not offer all the capabilities of TrackMate. This extension gives users more possibilities for complex problems such as tracking the shape and position of objects using machine learning algorithms (Ershov *et al.*, 2021) and exporting the results from TrackMate to a Mars ObjectArchive.

Applications. To demonstrate the range of analysis tasks that can be performed with Mars, we have developed three workflows based on real-world applications. In the first workflow, we determine the rate of transcription of single RNA polymerases by tracking their position as a function of time on long DNAs based on (Scherr *et al.*, 2021). In the second workflow, we demonstrate how to accurately analyze images containing observations of single-molecule FRET based on (Hellenkamp *et al.*, 2018). And finally, we highlight the virtual storage capability of Mars working with very large datasets from high-throughput imaging of DNA-tethered microspheres manipulated with forces and torques based on

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

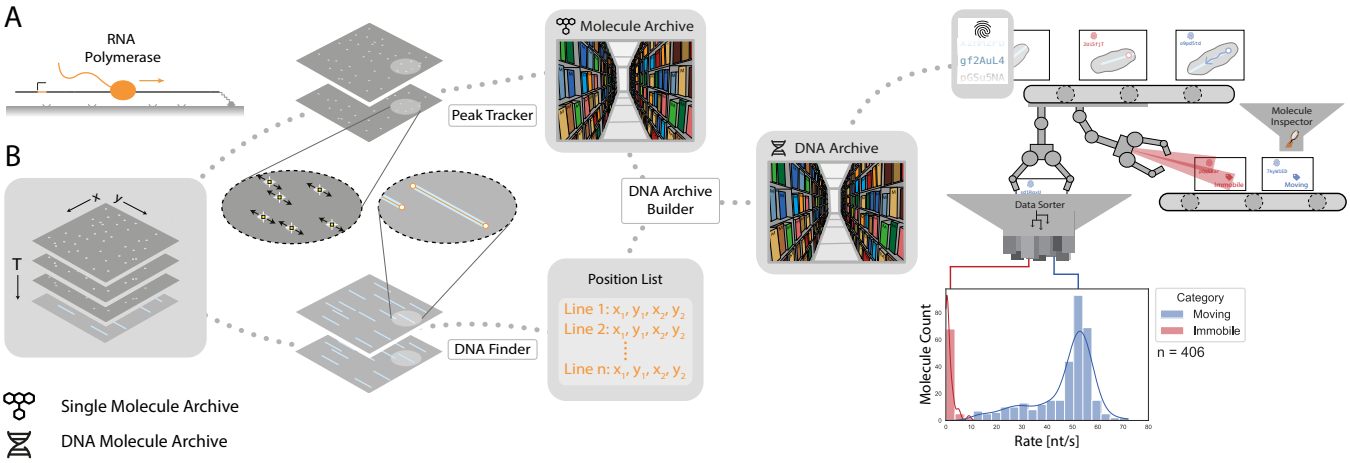


Fig. 3. A. Schematic of the RNA polymerase assay. Promoter-containing surface-immobilized 21 kb DNA was incubated with fluorescently-labeled RNA polymerase after which transcription was tracked over time. B. Representation of the analysis pathway showing the analysis steps starting from the raw image stack on the left to a final plot on the right. First, the *Peak Tracker* extracted position vs. time information from each fluorescent RNA polymerase creating a Single Molecule Archive. In parallel, the *DNA Finder* located the long, line-shaped, DNA molecules and generates a list of start and end positions. The information yielded from both tools was merged into a final DNA Molecule Archive. A classification and sorting process was applied resulting in a final plot showing the abundance of tracked molecules at various transcription rates (nt/s). A Gaussian fit to the population with rates > 40 nt/s revealed a population average rate of 53 ± 3.6 nt/s.

(Agarwal and Duderstadt, 2020). Detailed step-by-step instructions for each workflow can be found on the Mars documentation website (https://github.com/duderstadt-lab/mars-docs) and the raw data used in each workflow are freely available on either GitHub or Zenodo.

Workflow 1 – Tracking RNA polymerase position during transcription. Single-molecule total internal reflection fluorescence (smTIRF) microscopy has become an indispensable tool to study bio-macromolecular structure and functionality allowing the observation of, for example, molecule position and dynamics. Examples of such studies include kinetic studies on DNA replication (Ha *et al.*, 2002), studies of the polymerization of structural elements like actin (Amann and Pollard, 2001), the direct observation of flagellar motor rotation (Sowa *et al.*, 2005), as well as tracking of processes in vivo (Vizcay-Barrena *et al.*, 2011). To illustrate the use of Mars for the analysis of such datasets, this example shows a typical Mars analysis workflow for the interpretation of a smTIRF dataset studying the kinetics of a fluorescently-labeled RNA polymerase transcribing an immobilized, promoter-containing, 21 kb DNA molecule (Scherr *et al.*, 2021) (**Fig. 3A**). In the presence of all four nucleotides, RNA polymerase could initiate transcription from the promoter and progress on the DNA which could be temporally and spatially visualized by measuring fluorescent emission upon excitation. After transcription was completed, DNA was post-stained with SYTOX Orange to reveal the position of the DNA molecules in the last frames of the video. By correlating the RNA polymerase movement with the position of the DNA molecule, information about the polymerase processivity and progression rates were obtained.

To analyze the data quantitatively, first, a beam profile correction was applied to remove the non-uniform laser excitation in the field of view due to the Gaussian beam profile of frequently employed light sources in TIRF microscopy. The *Peak Tracker* (**Fig. 3B**) then determined the location of each fluorescent spot throughout the progressing frames and stored this information in a Single Molecule Archive. The identified positions were subsequently corrected for sample drift, that occurred over the course of the measurement, with the *Drift Corrector*. In parallel, the last frames of the video, showing the DNA molecules, were fit with the *DNA Finder* to generate a coordinate list with the positions of all DNA molecules. This information was correlated with the positional information of the polymerases and a DNA Molecule Archive was created. The obtained DNA Molecule Archive contains the positional information of all polymerase molecules found to be on the DNA and serves as the basis for further kinetic studies. Distinguishing between sub-populations is possible by sorting the molecule records either by means of parameter values and/or assigned molecule tags. Concluding the analysis, data exploration revealed a population-specific rate distribution (**Fig. 3B**, right) showing an observed transcription rate of 53 ± 3.6 nt/s which is well in line with previous studies reporting transcription rates between 40 to 80 nt/s (Thomen *et al.*, 2008).

Workflow 2 – Measuring intramolecular distances with smFRET. Single-molecule Förster Resonance Energy Transfer (smFRET) microscopy is used extensively to study protein dynamics (Lerner *et al.*, 2018; Mazal and Haran, 2019; Michalet *et al.*, 2006), RNA (Seidel and Dekker, 2007; Shaw *et al.*, 2014; Xiaowei, 2005) and DNA (Seidel and Dekker, 2007) interactions, to elucidate enzyme mechanisms (Smiley and Hammes, 2006) as well as to study protein structure (Dimura *et al.*, 2016; Schuler and Eaton, 2008) and molecular machines (Hildebrandt *et al.*, 2014; Stein *et al.*, 2011). Mars comes with multi-color fluorescence integration commands well-suited for the analysis of such datasets. In this example, a typical Mars analysis workflow for dual-color smFRET data is presented by analyzing a smFRET TIRF dataset (Hellenkamp *et al.*, 2018) collected using alternating laser exci-

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

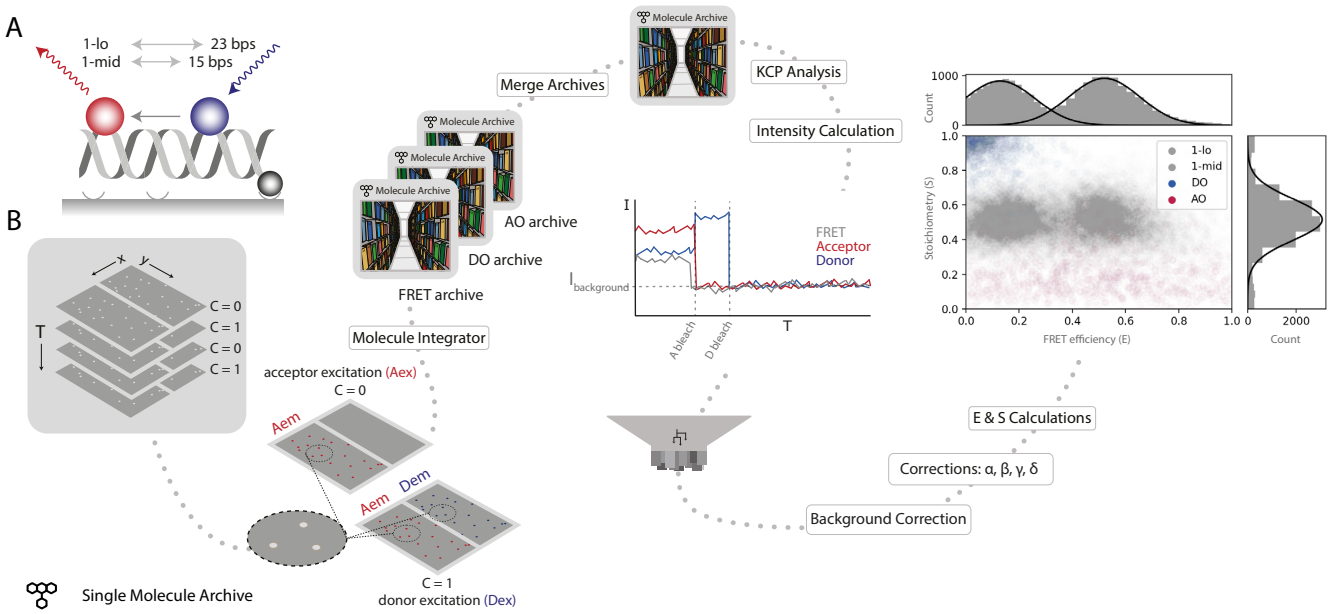


Fig. 4. A. Schematic of the FRET assay. The FRET efficiency between two covalently coupled dyes (donor (Atto550, shown in blue) and acceptor (Atto647N, shown in red)) on a short, immobilized, dsDNA oligo was measured. The inter-fluorophore distance was probed for two constructs: 23 bps (1-lo) or 15 bps (1-mid). B. Representation of the analysis pathway starting from the raw image stack on the left to a stoichiometry vs. FRET efficiency plot on the far right. First, the *Molecule Integrator* is used to extract the different colored intensity vs. time traces at each peak position (Aex: acceptor excitation, Aem: acceptor emission, Dem: donor emission, Dex: donor excitation) resulting in the corresponding three Single Molecule Archives (FRET archive, donor only (DO) archive, and acceptor only (AO) archive). After merging, the data are corrected for background and other photo-physical effects and classified according to the observed molecular features. Finally, the single-molecule data is represented in a common 2D histogram showing a scatterplot with the Stoichiometry (S) and FRET efficiency (E) information for both FRET samples (1-lo and 1-mid) as well as the AO and DO populations. The accompanying histograms plot the data from the 1-lo and 1-mid populations in gray bars and corresponding population-specific Gaussian fits as a solid black line.

tation. In this workflow, we show how the FRET efficiency (E) and stoichiometry (S) values for the measured samples were obtained using Mars. A direct comparison between the analysis outcomes with Mars and the extensive benchmark approach from Hellenkamp *et al.* (Hellenkamp *et al.*, 2018) shows that Mars reports E and S values in complete agreement with their study.

The TIRF dataset discussed (Hellenkamp *et al.*, 2018) measures the FRET efficiency (E) and stoichiometry (S) of two short dsDNA-based samples covalently labeled with both a donor (Atto550) and acceptor (Atto647N) fluorophore at two different inter-fluorophore distances (15 bps (1-lo) and 23 bps (1-mid), **Fig. 4A**). Data were gathered by recording the emission from the surface-attached DNA molecules upon alternating donor and acceptor excitation (ALEX) to ensure accurate FRET measurements. At first, the *Peak Finder* and *Molecule Integrator* identified the location of these DNA molecules and subsequently integrated their intensities to find the intensity vs. time (I vs. T) traces and stored those in molecule records in the generated Molecule Archive (**Fig. 4B**, Molecule Archives). In this particular analysis, three Single Molecule Archives were generated: (i) a FRET archive including all DNA molecules that were found to fluoresce in both emission channels, (ii) an acceptor only (AO) archive including all DNA molecules that fluoresce upon acceptor excitation only, and (iii) a donor only (DO) archive containing all molecules fluorescing only upon donor excitation. Separating these three species at the start of analysis facilitated easier downstream processing and data correction calculations. All three Molecule Archives were tagged and merged to a single master Molecule Archive containing the information of all three DNA molecule population types before further data corrections were applied.

First, a position-specific excitation correction was applied normalizing donor and acceptor intensities in relation to their specific position in the field of view. A subsequent kinetic change point analysis procedure (Hill *et al.*, 2018) identified all intensity transitions associated with a bleaching event. A manual selection of relevant traces was carried out to exclude all molecules: (i) with more or less than exactly one donor and acceptor per molecule, (ii) displaying large intensity fluctuations not caused by the bleaching event, as well as, (iii) molecules with a too low signal to noise ratio. To all remaining traces, a trace-wise background correction was applied by subtracting the mean background intensity after bleaching from the measured intensity at each time point of the respective molecule. Further corrections accounting for leakage, intensity, quantum yield normalization, and direct excitation of the acceptor were applied to obtain accurate FRET parameters (see Methods for details). Finally, the fully corrected traces allowed for the calculation of E and S values for each molecule. Value averaging of the E values calculated for molecules in each of the two populations revealed the fully corrected FRET parameters with accompanying standard error of the mean: $E_{1\text{-lo}} = 0.14 \pm 0.13$ and $E_{1\text{-mid}} = 0.51 \pm 0.09$. These values are in full agreement with the benchmarked values as reported by Hellenkamp *et al.* ($E_{1\text{-lo}} = 0.15 \pm 0.02$ and $E_{1\text{-mid}} = 0.56 \pm 0.03$) and hence underline the usability of Mars for the interpretation, correction, and analysis of dual-color smFRET datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

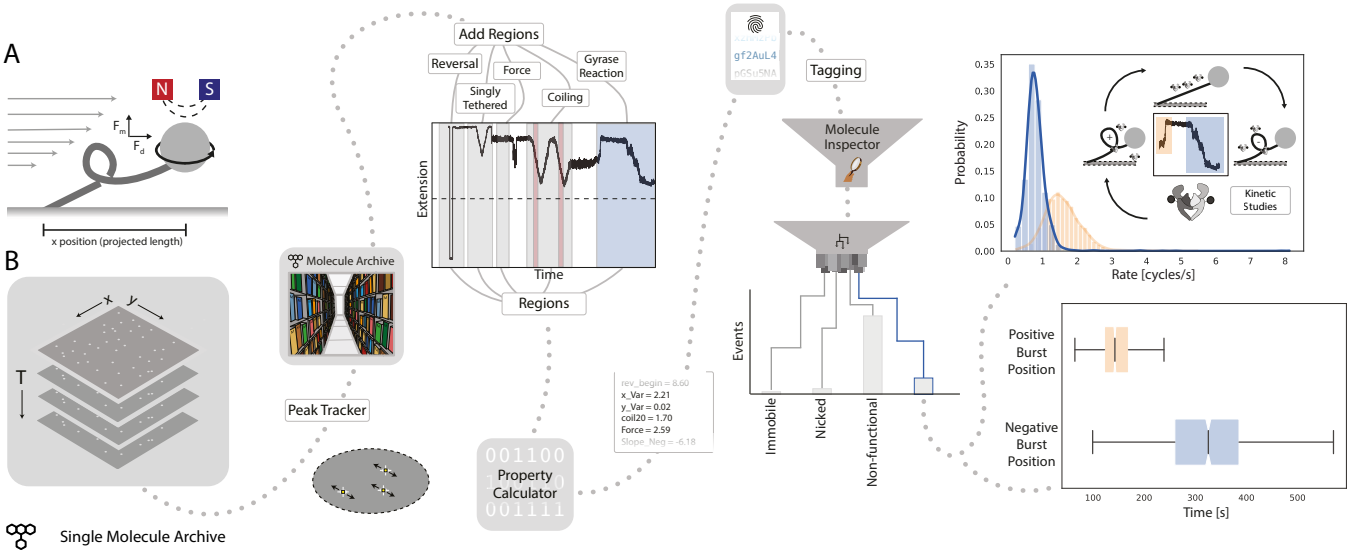


Fig. 5. A. Schematic of Flow Magnetic Tweezers. The projected length of the surface-immobilized DNA molecule attached to a magnetic bead was measured under different flow as well as magnet height and rotation conditions to study changes in DNA topology. B. Representation of the analysis workflow starting from the raw image stack on the left to the fully analyzed plots on the right. First, positional information is extracted by the *Peak Tracker* to yield a Single Molecule Archive. Regions are assigned to specific parts of the experiment highlighted in the example trace ('reversal', 'singly tethered', 'force', 'coiling', and 'gyrase reaction') and are used to calculate different DNA-related properties and parameters. Subsequent classification and tagging allows for easy exploration of subpopulations. The top graph shows the rate distribution (enzymatic cycles/s) found for gyrase activity resolving positive supercoils (orange) and introducing negative supercoils (blue), respectively. The lower graph shows the time past between the introduction of the enzyme to the system (T=0) and the observed enzymatic activity.

Workflow 3 – Characterizing the kinetics of DNA topology transformations. Force spectroscopy methods, utilizing the high precision tracking of beads attached to biomolecules, have led to great insights into biological processes. Magnetic tweezers, in particular, are routinely used to study the physical behavior of DNA (Nomidis *et al.*, 2017; Strick *et al.*, 1998) as well as nucleic acid transformations by essential types of cellular machinery, such as structural maintenance of chromosomes complexes (Eeftens *et al.*, 2017), the DNA replication machinery (Burnham *et al.*, 2019; Hodeib *et al.*, 2016; Manosas *et al.*, 2012; Seol *et al.*, 2016) and topoisomerases (Charvin *et al.*, 2005; Gore *et al.*, 2006; Nollmann *et al.*, 2007; Strick *et al.*, 2000). We have recently improved the throughput of this approach by combining magnetic tweezers with DNA flow stretching to create an instrument called Flow Magnetic Tweezers (FMT) (Agarwal and Duderstadt, 2020). The very large datasets generated by this instrument pose a challenge for analysis and initially led to the development of Mars. To illustrate the advantages offered by Mars in the analysis of these types of data, we present a workflow studying the behavior of DNA gyrase, a topoisomerase from *E.coli* (Nollmann *et al.*, 2007). This workflow illustrates the analysis steps required from raw tracking data to a well-structured single-molecule dataset with a traceable processing history utilizing the virtual storage infrastructure of Mars.

In the experimental set-up of the FMT, surface-immobilized DNA molecules are attached to paramagnetic polystyrene beads in a flow cell (**Fig. 5**). Block magnets placed above the flow cell create a magnetic field that orients the beads and applies a vertical force. A constant flow through the flow cell results in a drag force on the bead and the DNA. Similar to a conventional magnetic tweezers set-up, rotation of the magnets will change the topology of the DNA according to the direction of rotation. Furthermore, by inverting the flow direction the DNA tether will flip and reveal the location of surface attachment. A low magnification telecentric lens makes it possible to image a massive field of view providing high throughput observations. At low applied forces, changes in DNA topology induced through magnet rotation result in the formation of DNA supercoils and DNA compaction. This results in a decrease in the projected length of the DNA molecule observable as bead motion in the image. Despite the low magnification of the lens required for high throughput, subpixel fitting allows for high resolution tracking of DNA length over time. The tracking results are rich in information that allows for detailed molecule classification and quantification of enzyme kinetics.

In the experiment presented in **Fig. 5B**, first, a series of predetermined flow and magnet transformations were executed to check tether quality. Then gyrase was introduced, resolving positive supercoils and subsequently introducing negative supercoils. Regions were assigned to the trace in which parameters were calculated and tags were allocated to the molecules based on these parameter values. In this particular example, four final categories discriminated molecules that were either retained for analysis, immobile, nicked, or rejected for various other reasons (e.g. attachment of multiple DNA molecules to one magnetic bead). The molecules retained for analysis were background corrected and the rate of positive relaxation and negative introduction was calculated and plotted (**Fig. 5B**, right). A positive burst rate of 1.59 cycles/s, as well as a negative burst rate of 0.80 cycles/s, were found in this experiment. These values are in agreement with those reported previously (Nollmann *et al.*, 2007).

This workflow illustrates how Mars can be used to analyze

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

large datasets virtually with on-the-fly data retrieval. Careful documentation is even more critical for large datasets where many stages of analysis must be entirely automated. The documentation framework provided in Mars ensures that the analysis is done reproducibly by entering each step in the log of the metadata record. This ensures a fully reversible workflow in combination with keeping all observations in the Molecule Archive including those rejected from final analysis. Mars can be of value for other major force spectroscopy methods even though many do rely on z-directional movement instead of x-y tracking. Several options for external initial image processing are available from which results could be imported to Mars either in tabular form or using a scripting environment. Furthermore, this example proves that Mars is very flexible and can be used for any camera-based data where single-molecules are localized or tracked. Subsequently, Mars provides a platform with improved classification options, documentation, and data reusability for downstream analysis. The expandability of Mars will allow diversification of the workflow to integrate a broader range of force spectroscopy input data, for example, from optical and traditional magnetic tweezers.

Discussion

Rapid improvements in bioimaging technologies have led to powerful new approaches to follow the time evolution of complex biological systems with unprecedented spatial and temporal resolution generating vast information-rich datasets. These observations must be efficiently and reproducibly analyzed, classified, and shared to realize the full potential of recent technological advances and ensure new biological phenomena are discovered and faithfully quantified. Single-molecule imaging approaches, in particular, have moved from obscurity in specialized physics laboratories to the forefront of molecular biology research. Nevertheless, surprisingly few reporting standards or common formats exist. While significant progress has been made in establishing common file formats for single-molecule FRET datasets that can store raw photon information from point detectors (Greenfeld *et al.*, 2015) and time-binned trajectories from images (Ingargiola *et al.*, 2016), they offer limited options for adaption to other experimental modalities. Mars provides a solution to bridge this gap in the form of a common set of commands for single-molecule image processing, a graphical user interface for molecule exploration, and a Molecule Archive file format for flexible storage and reuse of image-derived datasets adaptable to a broad range of experiment types. Mars provides a solution to these issues in the form of a common set of commands for single-molecule image processing, a graphical user interface for molecule exploration, and a Molecule Archive file format for flexible storage and reuse of image-derived datasets. To ensure Mars is accessible to a large community, Mars is developed open source and freely available as a collection of SciJava commands distributed through a Fiji update site. The Mars project is a member of the Scientific Community Image Forum which provides a vibrant platform for new users to get help and for advanced users to find solutions to difficult image analysis problems.

The workflows described illustrate how the basic collection of modular commands and Molecule Archive transformations provided can be reused to analyze data from very different experimental configurations. Nevertheless, we recognize the commands are ultimately limited and will not address all problems easily. Therefore, we provide interoperability between Mars and other platforms available in Fiji. In particular, to provide access to a broader range of particle tracking options, results can be exported from TrackMate to Molecule Archive format. We plan to expand interoperability to include other common formats generated from single-molecule imaging experiments as they emerge. Additionally, Mars uses the SCIFIO framework (Hiner et al., 2016) to convert different image formats into OME format and includes a specialized image reader for more comprehensive support of images recorded using MicroManager (Edelstein *et al.*, 2010) frequently used for single-molecule imaging. This will allow Mars to process new image formats as they are developed and SCIFIO or Bio-Formats readers are written (Linkert *et al.*, 2010). These are only a few examples of the many workflow options that integrate with core Fiji technologies. In case these integrations do not provide a solution, Mars has several built-in extension mechanisms. New Molecule Archive types and custom user interface elements can be added separately and discovered by Mars at runtime using the SciJava discovery mechanism. Mars was written with script and command development in mind to allow for the analysis of observations resulting from new approaches beyond single-molecule applications by extending the existing framework. Finally, Molecule Archives can easily be opened in Python environments by directly loading them as JSON files or using PyImageJ (Rueden *et al.*, 2021), which provides access to many other data manipulation and visualization libraries.

Single-molecule imaging approaches have gained widespread usage and become an indispensable tool for the discovery of new biological mechanisms. Unfortunately, data reporting standards have lagged far behind. Raw images together with a record of the processing history for reproduction are rarely provided. The development of new formats like the Molecule Archive format presented here will make it easier for researchers to faithfully report their results and ensure reproducibility. This will increase the level of confidence and quantitative accuracy of findings and allow for broader reuse of existing information-rich datasets. However, Mars only provides a framework to aid reproducibility and does not ensure it. Individual researchers are ultimately responsible for maintaining reporting standards sufficient for reproduction and following best practice recommendations for single molecule imaging experiments (Lerner *et al.*, 2021). For example, scripts developed with Mars should be version controlled, made publicly accessible, and report all essential parameters to the processing log. Moreover, the version numbers and settings of all software integrated into workflows must be documented and reported to ensure

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reproducibility.

The discovery of new biological phenomena from single-molecule observations often depends on time-consuming manual classification of individual molecules and behaviors. Machine learning algorithms are now offering the possibility to automate these tasks (Kapadia *et al.*, 2021; Thomsen *et al.*, 2020), but their accuracy depends on robust training datasets. The powerful record tagging tools provided with Mars provide the ideal platform for the creation of large training datasets for machine learning based classifiers. Future work will focus on further development of interoperability of Mars with other platforms and machine learning workflows.

Methods

Workflow 1 - Tracking RNA polymerase position during transcription. Specific details about protein purification and labeling, the microscope set-up, sample preparation, and the imaging procedure can be found in the publication by Scherr *et al.* (Scherr *et al.*, 2021). The raw data accompanying this example workflow is freely available through the Mars tutorials GitHub repository (https://github.com/duderstadt-lab/mars-tutorials/tree/master/Example_workflows/TrackonDNA).

Extensive background information about all described Mars commands, specific settings used, and screenshots of example Molecule Archives and algorithm outcomes can be found on the Mars documentation pages (https://duderstadt-lab.github.io/mars-docs/docs/). Scripts and example Molecule Archives accompanying this analysis can be found in the Mars tutorials GitHub repository (https://github.com/duderstadt-lab/mars-tutorials/tree/master/Example_workflows/TrackonDNA).

Workflow 2 – Measuring intramolecular distances with smFRET. In this workflow, parameter nomenclature, as well as data correction and calculation procedures, were performed as described by (Hellenkamp *et al.*, 2018) to facilitate a clear comparison between their benchmark study and our work. We acknowledge major efforts in the field of single-molecule FRET towards further improvements to these procedures that could be incorporated into an application-tailored workflow by future users of our software utilizing its flexible and modular structure. Specifics on sample design and preparation and the data acquisition procedure for the dataset used in this analysis can be found in the publication of (Hellenkamp *et al.*, 2018). The raw image data is freely available through their published Zenodo link (https://doi.org/10.5281/zenodo.1249497).

The analysis starts with the identification of the fluorescent DNA molecule positions and subsequent intensity vs. time trace extractions. To make the data compatible with our software, first, a script was applied that converts the original video to a video with channel information corresponding to acceptor excitation (C=0) and donor excitation (C=1). Next, the peaks were identified with the *Peak Finder* and their locations are exported to the ImageJ ROI manager. To account for the split view data recording, in which data was collected simultaneously at acceptor emission and donor emission wavelengths, the coordinates of the peaks were transformed to the other half of the split view using an Affine2D matrix. This matrix was calculated from the available calibration data using the 'Descriptor-based registration (2d/3d) (Preibisch *et al.*, 2010) plugin in Fiji (Rueden *et al.*, 2017; Schindelin *et al.*, 2012; Schneider *et al.*, 2012). Next, the *Molecule Integrator* was used to extract the intensity vs. time traces of all molecules at all specified emission and excitation colors. This generated a Single Molecule Archive containing molecule records with intensity traces for each identified molecule. For easier downstream analysis, the described procedure was repeated, thereby generating three different Molecule Archives from each video: (i) FRET archive – a Molecule Archive containing molecules that have both donor and acceptor emission; (ii) AO archive – a Molecule Archive containing molecules with acceptor emission after acceptor excitation only; and (iii) DO archive – a Molecule Archive containing molecules with donor emission after donor excitation only. These Molecule Archives were tagged accordingly before merging them into a master Molecule Archive using the *Merge Archives* command.

In the master Molecule Archive, first, a position-specific excitation correction was applied that normalizes the donor and acceptor intensities dependent on their location in the field of view. Next, for each intensity trace, the *Single Change Point Finder* was used to automatically detect large intensity shifts indicating bleaching events. Subsequent manual tagging of molecule traces allowed for the exclusion of certain molecules in further analysis: (i) molecules with either more than or fewer than exactly one donor and acceptor fluorophore per molecule respectively, (ii) molecules showing large intensity changes other than the bleaching event, and (iii) molecules with a too low signal to noise ratio. This yielded a Molecule Archive with tagged molecules to be included in the forthcoming calculation and data correction steps while the excluded molecules remain unaltered.

To yield reliable intensity values to be used in the calculation of the FRET parameters (E and S) five subsequent corrections were applied:

Step 1: The background correction step subtracts the mean background intensity after bleaching as measured from the traces in the respective FRET Molecule Archive from all pre-bleaching intensities in a trace-wise fashion. This yielded the background-corrected intensity values ($^{ii}I_{\text{emission colorlexcitation color}}$). These corrected intensity values were used to calculate $^{ii}E_{app}$ and $^{ii}S_{app}$ (**Eq. 1 & 2**).

$$^{ii}S_{app} = \frac{^{ii}I_{Aem|Dex} + ^{ii}I_{Dem|Dex}}{^{ii}I_{Aem|Dex} + ^{ii}I_{Dem|Dex} + ^{ii}I_{Aem|Aex}} \quad \textbf{(1)}$$

$$^{ii}E_{app} = \frac{^{ii}I_{Aem|Dex}}{^{ii}I_{Aem|Dex} + ^{ii}I_{Dem|Dex}} \quad \textbf{(2)}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Step 2 & 3: Next, the leakage of donor fluorescence into the acceptor channel (α) and direct acceptor excitation by the donor excitation laser (δ) were calculated from the AO and DO molecules respectively according to equation 3 (**Eq. 3**). FA|D stores the fully corrected intensity value for each molecule at each time point before the first photobleaching event (**Eq. 4**).

$$\alpha = \frac{\langle {}^{ii}E_{app}^{(DO)} \rangle}{1 - \langle {}^{ii}E_{app}^{(DO)} \rangle} \quad \text{and} \quad \delta = \frac{\langle {}^{ii}S_{app}^{(AO)} \rangle}{1 - \langle {}^{ii}S_{app}^{(AO)} \rangle} \tag{3}$$

$$F_{A|D} = {}^{ii}I_{Aem|Dex} - \alpha {}^{ii}I_{Dem|Dex} - \delta {}^{ii}I_{Aem|Aex} \tag{4}$$

Step 4 & 5: To then account for the normalization of excitation intensities and cross-sections of the acceptor and donor (β) and the normalization of effective fluorescence quantum yields and detection efficiencies of the acceptor and donor (γ) the respective correction factors were determined based on the relationship between the 1-lo and 1-mid population. To do so, ${}^{ii}E_{app}$ and ${}^{ii}S_{app}$ values were averaged in a molecule-wise fashion and linear regression against the values of the entire population yielded correction factors β & γ (**Eq. 5**) that were applied to calculate $F_{A|A}$ and $F_{D|D}$ (**Eq. 6**). These subsequently yielded the fully corrected E and S parameters for each molecule (**Eq. 7**). A population-specific molecule averaging revealed the respective population average values.

$$\frac{1}{\langle {}^{iii}S_{app}^{(FRET)} \rangle} = b * \langle {}^{iii}E_{app}^{(FRET)} \rangle + a$$
$$\text{where}: \quad \beta = a + b - 1 \quad \text{and} \quad \gamma = \frac{a-1}{a+b-1} \tag{5}$$

$$F_{D|D} = \gamma * {}^{ii}I_{Dem|Dex} \quad \text{and} \quad F_{A|A} = \frac{1}{\beta} * {}^{ii}I_{Aem|Aex} \tag{6}$$

$$E = \frac{F_{A|D}}{F_{D|D} + F_{A|D}} \quad \text{and} \quad S = \frac{F_{A|D} + F_{D|D}}{F_{D|D} + F_{A|D} + F_{A|A}} \tag{7}$$

More information regarding the derivation of the discussed formula as well as information about the applied corrections can be found in the publication by (Hellenkamp *et al.*, 2018). Extensive background information about all described Mars tools, specific settings used in the built-in tools, and screenshots of expected outcomes can be found on the Mars documentation pages (https://duderstadt-lab.github.io/mars-docs/). Scripts, Jupyter notebooks, and Molecule Archives accompanying this analysis can be found in the Mars tutorials GitHub repository (https://github.com/duderstadt-lab/mars-tutorials/tree/master/Example_workflows/FRET

Workflow 3 – Characterizing the kinetics of DNA topology transformations. The raw video is available through Zenodo (https://zenodo.org/record/3786442#.YTns2C2B1R0). This video is a reduced dataset from one of the FMT experiments investigating the topological changes gyrase performs on the DNA. The Groovy scripts for analyzing the dataset can be found on GitHub (https://github.com/duderstadt-lab/fmt-scripts). The first three scripts have been used to create a CSV file and the data was then plotted with a Python script. The data analysis procedure using Mars has been described in (Agarwal and Duderstadt, 2020).

After tracking the molecules in the dataset and the generation of the Single Molecule Archive, the traces were sorted and classified. The experiment was designed in such a way that steps in the process of the assay could used as indicators. Based on these indicators, a discrimination was made between the molecules classifying them either as immobile (not mobile, not coilable), nicked (mobile, not coilable) or fit for analysis (mobile, coilable). A fourth category contains every rejected molecule due to various reasons like multiple DNA attachments on a single bead or getting stuck during the experiment before the enzymatic activity was detected. The tethers can appear to be mobile and coilable but certain thresholds are not passed like being coilable at high force. Specifically, the main indicators for this dataset are derived from: (i) the observation of a positional change of the DNA-bead after flow direction reversal (stuck or not stuck), (ii) a test for DNA molecule coilability by rotating the magnets at high force (single or multiple DNA molecules attached to a single bead), (iii) a molecular force measurement investigating the force on the DNA, and (iv) a second rotating step at lower force to introduce positive twist which is resolved by gyrase. The positive and negative introduction of supercoiling is used to calibrate the extension change of each DNA molecule to the change of twist during the analysis. The response of the molecule to these indicator tests is determined by parameter calculation scripts that only consider a certain assigned region of interest in the trace. When a certain threshold for a parameter is met, a tag is added to the molecule record. The final set of tags present in the molecule record determines in which group the molecule is classified. A sliding window approach is applied to all relevant molecules revealing kinetic information such as coiling and compaction rates as well as influences on these parameters after the introduction of gyrase. The size of the window depends on whether gyrase is relaxing positive twist or introducing negative twist. For the positive relaxation a window of 12.5 s and for the negative introduction a window of 25 s is used. In this case, one cycle means the linking number of the DNA is changed by -2. Since the DNA has to be relaxed to get negatively supercoiled, the burst position for positive relaxation comes earlier than the negative introduction. The time points have been corrected such that gyrase introduction coincides with T=0. The calculated information was either exported to the CSV format or was directly interpreted with Python. This enables data visualization on various platforms.

The entire analysis, including more background information, can be found on our website (https://duderstadt-lab.github.io/mars-docs/examples/flow-Magnetic-Tweezers/). Furthermore, the Mars tools are explained in great detail on our documentation site (https://duderstadt-lab.github.io/mars-docs/docs/).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements. We are grateful to Giovanni Cardone (MPIB imaging facility) and Evelyn Plötz for their helpful feedback on the project. We would like to thank Prof. Thorsten Hugel for providing the Igor source code used for the analysis of smFRET data (Hellenkamp *et al.*, 2018). We are grateful for the recent development efforts in the ImageJ2 and Fiji communities that have greatly enhanced the applications across platforms. We would like to thank Curtis Rueden, Jan Eglinger, Tobias Pietzsch, and Jean-Yves Tinevez for taking the time to answer the many questions that have come up during development and considering our pull requests. We thank Curtis Rueden for helping to configure our repositories for deployment to SciJava maven. We would like to thank the Henriques lab for making this pre-print format available. This work was funded by the Deutsche Forshungsgemeinshaft (DFG, German Research Foundation)–SFB863–11166240, a starting grant from the European Research Council (Project Number: 804098, REPLISOMEBYPASS), and the Max Planck Society.

Author Contributions. Conceptualization, N.M.H., T.M.R., M.J.S., R.A. and K.E.D.; Software, N.M.H., T.M.R., M.J.S., R.A. and K.E.D.; Workflow Development, N.M.H., T.M.R., M.J.S., R.A. and K.E.D.; Writing – Original Draft, N.M.H., T.M.R., and K.E.D.; Writing – Review & Editing, N.M.H., T.M.R., M.J.S., R.A., B.S., A.M. and K.E.D.; Visualization, N.M.H., T.M.R., and K.E.D.; Resources, N.M.H., T.M.R., M.J.S., R.A., B.S., A.M. and K.E.D.; Software Beta Testing, N.M.H., T.M.R., M.J.S., R.A., B.S., A.M. and K.E.D.; Supervision, K.E.D.; Funding Acquisition, K.E.D.

Declaration of Interest. The authors have no competing interests.

Data Availability. The analysis software described is publicly available in several repositories on GitHub at https://github.com/duderstadt-lab. The core library used for the analysis and storage of data is contained in the mars-core repository. The graphical user interface is contained in the mars-fx repository. The videos used in all workflows have been made available in public databases. Links can be found in the methods sections for each workflow. Extensive documentation and links to many additional resources and scripts used in all workflows can be found at https://duderstadt-lab.github.io/mars-docs/.

References

Agarwal, R., and Duderstadt, K.E. (2020). Multiplex flow magnetic tweezers reveal rare enzymatic events with single molecule precision. Nat Commun 11, 4714.

Akyuz, N., Altman, R.B., Blanchard, S.C., and Boudker, O. (2013). Transport dynamics in a glutamate transporter homologue. Nature 502, 114-118.

Allan, C., Burel, J.M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W.J., Neves, C., Patterson, A., *et al.* (2012). OMERO: flexible, model-driven data management for experimental biology. Nat Methods 9, 245-253.

Amann, K.J., and Pollard, T.D. (2001). Direct Real-Time Observation of Actin Filament Branching Mediated by Arp2/3 Complex Using Total Internal Reflection Fluorescence Microscopy. Proc Natl Acad Sci USA 98, 15009-15013.

Baek, I., Friedman, L.J., Gelles, J., and Buratowski, S. (2021). Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. Mol Cell 81, 3576-3588 e3576.

Beattie, T.R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A.J., Leake, M.C., and Reyes-Lamothe, R. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. Elife 6.

Burnham, D.R., Kose, H.B., Hoyle, R.B., and Yardimci, H. (2019). The mechanism of DNA unwinding by the eukaryotic replicative helicase. Nat Commun 10, 2159.

Carpenter, A.E., Kamentsky, L., and Eliceiri, K.W. (2012). A call for bioimaging software usability. Nat Methods 9, 666-670.

Charvin, G., Strick, T.R., Bensimon, D., and Croquette, V. (2005). Tracking topoisomerase activity at the single-molecule level. Annu Rev Biophys Biomol Struct 34, 201-219.

Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., *et al.* (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 156, 1274-1285.

Crickard, J.B., Moevus, C.J., Kwon, Y., Sung, P., and Greene, E.C. (2020). Rad54 Drives ATP Hydrolysis-Dependent DNA Sequence Alignment during Homologous Recombination. Cell 181, 1380-1394 e1318.

Dimura, M., Peulen, T.O., Hanke, C.A., Prakash, A., Gohlke, H., and Seidel, C.A.M. (2016). Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. Current Opinion in Structural Biology 40, 163-185.

Duchi, D., Bauer, D.L., Fernandez, L., Evans, G., Robb, N., Hwang, L.C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., *et al.* (2016). RNA Polymerase Pausing during Initial Transcription. Mol Cell 63, 939-950.

Duderstadt, K.E., Geertsema, H.J., Stratmann, S.A., Punter, C.M., Kulczyk, A.W., Richardson, C.C., and van Oijen, A.M. (2016). Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes. Mol Cell 64, 1035-1047.

Dulin, D., Lipfert, J., Moolman, M.C., and Dekker, N.H. (2013). Studying genomic processes at the single-molecule level: introducing the tools and applications. Nat Rev Genet 14, 9-22.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer control of microscopes using microManager. Curr Protoc Mol Biol Chapter 14, Unit14 20.

Eeftens, J.M., Bisht, S., Kerssemakers, J., Kschonsak, M., Haering, C.H., and Dekker, C. (2017). Real-time detection of condensin-driven DNA compaction reveals a multistep binding mechanism. EMBO J 36, 3448-3457.

Eliceiri, K.W., Berthold, M.R., Goldberg, I.G., Ibanez, L., Manjunath, B.S., Martone, M.E., Murphy, R.F., Peng, H., Plant, A.L., Roysam, B., *et al.* (2012). Biological imaging software tools. Nat Methods 9, 697-710.

Erkens, G.B., Hanelt, I., Goudsmits, J.M., Slotboom, D.J., and van Oijen, A.M. (2013). Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. Nature 502, 119-123.

Ershov, D., Phan, M.-S., Pylvänäinen, J.W., Rigaud, S.U., Le Blanc, L., Charles-Orszag, A., Conway, J.R.W., Laine, R.F., Roy, N.H., Bonazzi, D., *et al.* (2021). Bringing TrackMate into the era of machine-learning and deep-learning. biorxiv.

Gao, L., Shao, L., Higgins, C.D., Poulton, J.S., Peifer, M., Davidson, M.W., Wu, X., Goldstein, B., and Betzig, E. (2012). Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. Cell 151, 1370-1385.

Goldberg, I.G., Allan, C., Burel, J.M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P.K., and Swedlow, J.R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biol 6, R47.

Gore, J., Bryant, Z., Stone, M.D., Nollmann, M., Cozzarelli, N.R., and Bustamante, C. (2006). Mechanochemical analysis of DNA gyrase using rotor bead tracking. Nature 439, 100-104.

Greenfeld, M., van de Meent, J.W., Pavlichin, D.S., Mabuchi, H., Wiggins, C.H., Gonzalez, R.L., Jr., and Herschlag, D. (2015). Single-molecule dataset (SMD): a generalized storage format for raw and processed single-molecule data. BMC Bioinformatics 16, 3.

Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M., and Chu, S. (2002). Initiation and Re-Initiation of DNA Unwinding by the Escherichia Coli Rep Helicase. Nature 419, 638-641.

Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Adariani, S.R., Ambrose, B., Aznauryan, M., Barth, A., Birkedal, V., *et al.* (2018). Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. Nature Methods 15, 669-676.

Hildebrandt, L.L., Preus, S., Zhang, Z., Voigt, N.V., Gothelf, K.V., and Birkedal, V. (2014). Single Molecule FRET Analysis of the 11 Discrete Steps of a DNA Actuator. Journal of the American Chemical Society 136, 8957-8962.

Hill, F.R., van Oijen, A.M., and Duderstadt, K.E. (2018). Detection of kinetic change points in piece-wise linear single molecule motion. J Chem Phys 148, 123317.

Hiner, M.C., Rueden, C.T., and Eliceiri, K.W. (2016). SCIFIO: an extensible framework to support scientific image formats. BMC Bioinformatics 17, 521.

Hodeib, S., Raj, S., Manosas, M., Zhang, W., Bagchi, D., Ducos, B., Allemand, J.F., Bensimon, D., and Croquette, V. (2016). Single molecule studies of helicases with magnetic tweezers. Methods 105, 3-15.

Ingargiola, A., Laurence, T., Boutelle, R., Weiss, S., and Michalet, X. (2016). Photon-HDF5: Open Data Format and Computational Tools for Timestamp-based Single-Molecule Experiments. Proc SPIE Int Soc Opt Eng 9714.

Kapadia, N., El-Hajj, Z.W., and Reyes-Lamothe, R. (2021). Bound2Learn: a machine learning approach for classification of DNA-bound proteins from single-molecule tracking experiments. Nucleic Acids Res 49, e79.

Kapadia, N., El-Hajj, Z.W., Zheng, H., Beattie, T.R., Yu, A., and Reyes-Lamothe, R. (2020). Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. Mol Cell 80, 114-126 e118.

Kvilekval, K., Fedorov, D., Obara, B., Singh, A., and Manjunath, B.S. (2010). Bisque: a platform for bioimage analysis and management. Bioinformatics 26, 544-552.

Lerner, E., Barth, A., Hendrix, J., Ambrose, B., Birkedal, V., Blanchard, S.C., Borner, R., Sung Chung, H., Cordes, T., Craggs, T.D., et al. (2021). FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices.

Lerner, E., Cordes, T., Ingargiola, A., Alhadid, Y., Chung, S., Michalet, X., and Weiss, S. (2018). Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. Science 359.

Lewis, J.S., Spenkelink, L.M., Schauer, G.D., Yurieva, O., Mueller, S.H., Natarajan, V., Kaur, G., Maher, C., Kay, C., O'Donnell, M.E., *et al.* (2020). Tunability of DNA Polymerase Stability during Eukaryotic DNA Replication. Mol Cell 77, 17-25 e15.

Linkert, M., Rueden, C.T., Allan, C., Burel, J.M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., Macdonald, D., et al. (2010). Metadata matters: access to image data in the real world. J Cell Biol 189, 777-782.

Manosas, M., Perumal, S.K., Croquette, V., and Benkovic, S.J. (2012). Direct observation of stalled fork restart via fork regression in the T4 replication system. Science 338, 1217-1220.

Marr, D., and Hildreth, E. (1980). Theory of edge detection. Proc R Soc Lond B Biol Sci 207, 187-217.

Mazal, H., and Haran, G. (2019). Single-molecule FRET

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

methods to study the dynamics of proteins at work. Current Opinion in Biomedical Engineering 12, 8-17

Meijering, E., Carpenter, A.E., Peng, H., Hamprecht, F.A., and Olivo-Marin, J.C. (2016). Imagining the future of bioimage analysis. Nat Biotechnol 34, 1250-1255.

Michalet, X., Weiss, S., and Jäger, M. (2006). Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics. Chem Rev 106, 1785-1813

Neuman, K.C., and Nagy, A. (2008). Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods 5, 491-505.

Niekamp, S., Stuurman, N., Zhang, N., and Vale, R.D. (2021). Three-color single-molecule imaging reveals conformational dynamics of dynein undergoing motility. Proc Natl Acad Sci U S A 118.

Nollmann, M., Stone, M.D., Bryant, Z., Gore, J., Crisona, N.J., Hong, S.C., Mitelheiser, S., Maxwell, A., Bustamante, C., and Cozzarelli, N.R. (2007). Multiple modes of Escherichia coli DNA gyrase activity revealed by force and torque. Nat Struct Mol Biol 14, 264-271.

Nomidis, S.K., Kriegel, F., Vanderlinden, W., Lipfert, J., and Carlon, E. (2017). Twist-Bend Coupling and the Torsional Response of Double-Stranded DNA. Phys Rev Lett 118, 217801.

Ouyang, W., and Zimmer, C. (2017). The imaging tsunami: Computational opportunities and challenges. Current Opinion in Systems Biology 4, 105-113.

Peng, T., Thorn, K., Schroeder, T., Wang, L., Theis, F.J., Marr, C., and Navab, N. (2017). A BaSiC tool for background and shading correction of optical microscopy images. Nat Commun 8, 14836.

Pietzsch, T., Saalfeld, S., Preibisch, S., and Tomancak, P. (2015). BigDataViewer: visualization and processing for large image data sets. Nat Methods 12, 481-483.

Preibisch, S., Saalfeld, S., Schindelin, J., and Tomancak, P. (2010). Software for bead-based registration of selective plane illumination microscopy data. Nature Methods 7, 418-419.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science 314, 1139-1143.

Rubens, U., Mormont, R., Paavolainen, L., Backer, V., Pavie, B., Scholz, L.A., Michiels, G., Maska, M., Unay, D., Ball, G., *et al.* (2020). BIAFLOWS: A Collaborative Framework to Reproducibly Deploy and Benchmark Bioimage Analysis Workflows. Patterns (N Y) 1, 100040.

Rueden, C.T., Yang, L., Pinkert, M., Liu, Y., Hiner, M., Eglinger, J., Macarse, M.H., Hereñú, D., Hanslovsky, P., Ouyang, W. (2021). imagej/pyimagej: v1.0.2 (1.0.2). In Zenodo.

Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., and Eliceiri, K.W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics 18, 529.

Scherr, M., Wahab, S.A., Remus, D., and Duderstadt, K.E. (2021). Born to Slide: Mobile Origin Licensing Factors Confer Resistance to Transcription Conflicts. SSRN Electronic Journal.

Scherr, M.J., Safaric, B., and Duderstadt, K.E. (2018). Noise in the Machine: Alternative Pathway Sampling is the Rule During DNA Replication. Bioessays 40.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676-682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671-675

Schuler, B., and Eaton, W.A. (2008). Protein folding studied by single-molecule FRET. Current Opinion in Structural Biology 18, 16-26

Seidel, R., and Dekker, C. (2007). Single-molecule studies of nucleic acid motors. Current Opinion in Structural Biology 17, 80-86

Seol, Y., Strub, M.P., and Neuman, K.C. (2016). Single molecule measurements of DNA helicase activity with magnetic tweezers and t-test based step-finding analysis. Methods 105, 119-127.

Shaw, E., St-Pierre, P., McCluskey, K., Lafontaine, D.A., and Penedo, J.C. (2014). Using sm-FRET and Denaturants to Reveal Folding Landscapes. Methods in Enzymology 549, 313-341.

Smiley, R.D., and Hammes, G.G. (2006). Single Molecule Studies of Enzyme Mechanisms. Chem Rev 106, 3080-3094.

Sowa, Y., Rowe, A.D., Leake, M.C., Yakushi, T., Homma, M., Ishijima, A., and Berry, R.M. (2005). Direct Observation of Steps in Rotation of the Bacterial Flagellar Motor. Nature 437, 916-919.

Stein, I.H., Steinhauer, C., and Tinnefeld, P. (2011). Single-Molecule Four-Color FRET Visualizes Energy-Transfer Paths on DNA Origami. Journal of the American Chemical Society 133, 4193-4195.

Strick, T.R., Allemand, J.F., Bensimon, D., and Croquette, V. (1998). Behavior of Supercoiled DNA. Biophysical Journal 74, 2016-2028.

Strick, T.R., Croquette, V., and Bensimon, D. (2000). Single-molecule analysis of DNA uncoiling by a type II topoisomerase. Nature 404, 901-904.

Thomen, P., Lopez, P.J., Bockelmann, U., Guillerez, J., Dreyfus, M., and Heslot, F. (2008). T7 RNA Polymerase

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Studied by Force Measurements Varying Cofactor Concentration. Biophys J 95, 2423-2433.

Thomsen, J., Sletfjerding, M.B., Jensen, S.B., Stella, S., Paul, B., Malle, M.G., Montoya, G., Petersen, T.C., and Hatzakis, N.S. (2020). DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. Elife 9.

Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J., and Bell, S.P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. Cell 161, 513-525.Tinevez, J.Y., Perry, N., Schindelin, J.,

Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: An Open and Extensible Platform for Single-Particle Tracking. Methods 115, 80-90.

Tokunaga, M., Imamoto, N., and Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5, 159-161.

Vizcay-Barrena, G., Webb, S.E.D., Martin-Fernandez, M.L., and Wilson, Z.A. (2011). Subcellular and Single-Molecule Imaging of Plant Fluorescent Proteins Using Total Internal Reflection Fluorescence Microscopy (TIRFM). J Exp Bot 62, 5419-5428.

Williams, E., Moore, J., Li, S.W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R.K., Sarkans, U., *et al.* (2017). The Image Data Resource: A Bioimage Data Integration and Publication Platform. Nat Methods 14, 775-781.

Xiaowei, Z. (2005). Single-Molecule RNA Science. Annual Review of Biophysics and Biomolecular Structure 34, 399-414.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Figures and Tables

Image

Peak Finder	Finds high-intensity pixel clusters (peaks) in an image. Additionally, the sub-pixel position can be determined utilizing a 2D Gaussian fit.
DNA Finder	Finds vertically aligned DNA molecules in an image. Additionally, the sub-pixel position of both ends of the molecule can be determined utilizing a 2D Gaussian fit.
Peak Tracker	Finds, fits, and tracks peaks in images.
Object Tracker	Identifies unspecified objects in images utilizing classification by segmentation and tracks their center of mass.
Molecule Integrator	Integrates the intensity of a peak over all frames.
Molecule Integrator (dual view)	Integrates the intensity of a peak over all frames in an image stack with dual view images.
Beam Profile Corrector	Corrects for the beam profile-generated image intensity deviations.
Gradient Calculator	Calculates the gradient of consecutive pixels from top to bottom or from left to right to identify long linear objects such as DNA molecules.
Overlay Channels	Combines several individual videos into one creating a single video with the information stored along the 'Channel (C)' dimension.

Molecule

Open Archive	Opens a Molecule Archive.
Open Virtual Store	Opens a virtual Molecule Archive.
Build Archive from Table	Converts an opened data table with a "molecule" column into a Molecule Archive.
Build DNA Archive	Builds a DNA Molecule Archive from a Single Molecule Archive and a list of DNA ROIs in the ROI manager. It uses the location of the DNA molecules to search for molecules in the Single Molecule Archive that overlap with (parts of) this location.
Merge Archives	Merges multiple Molecule Archives (placed in a single folder) into one.
Merge Virtual Stores	Merges multiple virtual Molecule Archives (placed in a single folder) into one.
Add Time	Adds a column to the molecule tables to convert frames (T) to real time values as specified in the metadata or by a user-defined time increment.
Drift Corrector	Calculates and corrects for the sample drift given a Molecule Archive and a tag corresponding to all immobile molecules in the dataset. Generates new columns for each molecule table.
Region Difference Calculator	Calculates the difference between the regions specified for all molecules in the Molecule Archive and adds the outcome as a molecule parameter.
Variance Calculator	Calculates the variance on a specified molecule table column and adds the outcome as a molecule parameter.

Table

Open Table	Imports a comma or tab-delimited table to the MarsTable format.
Sort	Sorts a MarsTable based on values in a specified column.
Filter	Filters the rows of a MarsTable based on the specified criterium.
Import IJ1 Table	Imports any ImageJ1 table to the MarsTable format.
Import TableDisplay	Imports any SciJava table to the MarsTable format.

KCP

Change Point Finder	Detects linear regions in single-molecule traces where a specific parameter linearly changes with respect to another parameter. This command generates molecule segments tables listing the line fits for these linear regions as well as their limits and fit errors.
Single Change Point Finder	Detects a single change point in the progression of each single-molecule trace. A segments table is generated containing the coordinates of the identified change point, the calculated slopes, intercepts, and corresponding fit error values.
Sigma Calculator	Calculates the error value in a specific region of interest in all single-molecule traces that can be used as input for the change point calculation commands.

ROI

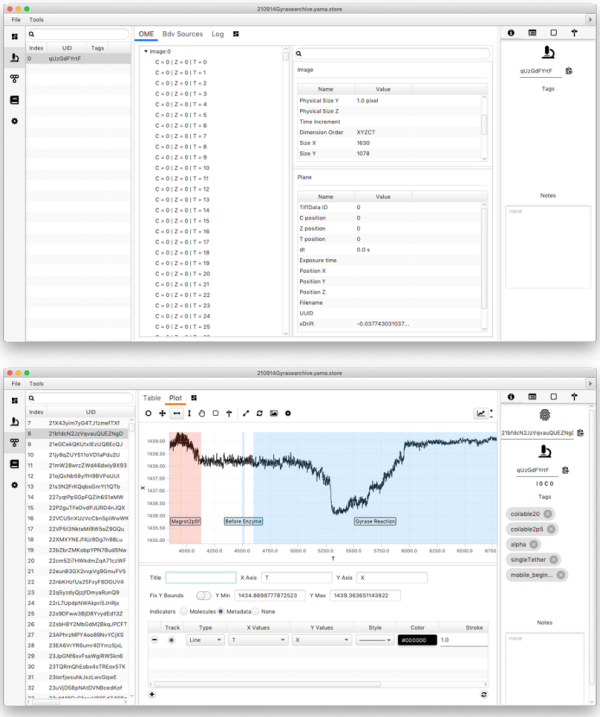
Transform ROIs	Transforms detected peaks from one channel of a dual view image to the other.

Supplementary Table. 1: Description of all ImageJ2 Mars Commands supporting the analysis of image-derived biomolecule data in Mars. Detailed documentation can be found on the Mars documentation website (https://duderstadt-lab.github.io/mars-docs/).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.26.470105; this version posted December 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

B



Supplementary Fig. 1. A. Mars Rover metadata tab showing OME image metadata information for each analyzed image as well as the metadata-specific tagging and other classification options features. B. Mars Rover molecules tab highlighting the user-friendly data exploration features such as the trace plotter displaying region highlighting as well as molecule-specific tags, notes, uids, and metadata uids.