Hardware and software for single-molecule fluorescence analysis

Ben Gamari <bgamari@physics.umass.edu>[1], Laura Dietz <dietz@cs.umass.edu>[2], and Lori Goldner <lgoldner@physics.umass.edu>[1]

[1]Department of Physics, University of Massachusetts, Amherst
[2]Department of Computer Science, University of Massachusetts, Amherst

August 1, 2013

Over the last three decades, fluorescence spectroscopy has developed into a unique tool for probing biologically relevant systems. In particular, fluorescence correlation spectroscopy (FCS) is now a routine measurement enabling biologists access to reaction kinetics (Wu et al. 2006; Heyduk and Niedziela-Majka 2002; Allen and Thompson 2006) and sizes of biomolecular assemblies (Pitschke et al. 1998). Likewise, single molecule Förster resonance energy transfer (smFRET) has proved to be useful for a variety of biologically relevant measurements ranging from DNA/RNA mechanics (Iqbal et al. 2008; Vafabakhsh and Ha 2012; Laurence et al. 2005; Wo´zniak et al. 2008) to protein conformation (Wo´zniak et al. 2005) and protein function (Gopich et al. 2009; Sevenich et al. 1998). When used in tandem these two techniques can provide unique perspective onto dynamics on molecular length-scales and over a wide range of time-scales (Nettels, Hoffmann, and Schuler 2008).

These techniques share a need for hardware and software for photon timing and associated analysis tools. A typical experiment may require timestamping on two to four detection channels, with nanosecond resolution. The resulting data can reveal details of the nature of molecular interactions, flexibility, and conformation through intensity and temporal correlations.

Commercial hardware for recording photon arrivals is available but expensive. For instance, time-correlated single photon counting modules are available for $40,000 or less capable hardware correlator devices for $10,000, not including the cost of software. These devices generally provide only few analysis tools, have little potential for extension, and are not flexible enough for use with recent experimental extensions such as alternating laser excitation (Lee et al. 2005; Kalinin et al. 2008).

With open tools and high-resolution timing hardware we support novel statistical methods for single-molecule fluorescence analysis. Our toolchain provides tools ranging from hardware and low-level acquisition utilities, including data aggregation and filtering, up to high-level probabilistic inference methods. The off-the-shelf board around which the hardware is built is available for less than $200 and can be assembled into a working instrument with little knowledge of hardware or software. The entire toolchain is open-source and well-documented, enabling customization and extension. This extensibility allows the instrument to be integrated into applications ranging from fluorescence imaging to fixed confocal studies.

Background: Fluorescence spectroscopy

A dye is a fluorescent molecule; that is, one which will emit a photon after having been excited by photon absorption or other excitation. While dyes have a long history of use in probing chemical systems (Fernandez and Berlin 1976), steady improvement in dye brightness and detector electronics over the last decades have enabled the use of fluorescence spectroscopy for probing a wide variety of biological systems. These techniques have been used to observe single-molecule protein folding within cells (Choi et al. 2008; Chung, Louis, and Eaton 2009), characterize enzyme interactions (Liu and Lu 2002), and measure *in vivo* protein expression

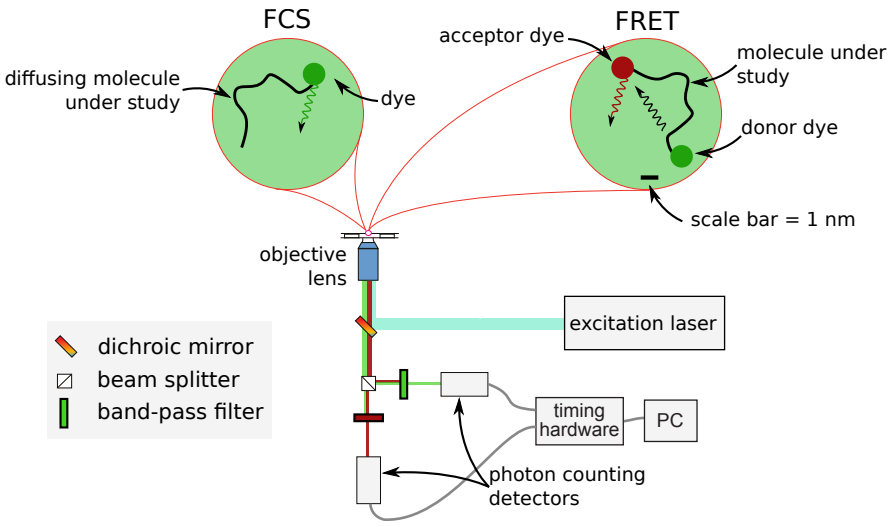

Figure 1: A schematic representation of a typical fluorescence spectroscopy experiment. The bubble on the left shows a diffusing singly-labelled molecule as might be used in an FCS experiment. The right bubble shows a pair of dyes undergoing energy transfer, as would be seen in a FRET study.

(Goedhart et al. 2011). By using dyes with distinct emission spectra multiple probes can be placed on a single molecule or assembly, enabling access to rich conformational details.

Figure 1 shows a typical fluorescence spectroscopy apparatus built around a confocal microscope. A laser illuminates an observation volume. As a molecule passes through the observation volume, its dyes are excited and emissions collected by photon-counting detectors. These detectors may collect different colors or polarizations of light, depending upon the experiment. Timestamps encoding the photon arrival events are then recorded by the timing hardware and sent to a computer for storage and later analysis.

A variety of experimental artifacts are present in a typical fluorescence spectroscopy experiment. Even high-quality photon counting detectors sporadically report events even with no fluorescence signal. These background events add noise to the fluorescence signal. Furthermore, optical filters do not perfectly separate wavelengths, giving rise to an artifact known as spectral crosstalk. Finally, despite improvements in dye brightness and stability, photon emission rates often limit the timescales accessible to fluorescence experiments. This is especially true in experiments where the molecule under study is freely diffusing in solution yielding residence times below one millisecond. In this case it is common to collect less than 50 photons per residency.

We will focus on two fluorescence methods in particular: Förster resonance energy transfer and fluorescence correlation spectroscopy.

Förster Resonance Energy Transfer

Förster resonance energy transfer (FRET) spectroscopy is a fluorescence technique that exploits the fact that two dyes in close proximity to one another can undergo energy transfer. That is, after exciting one of the dyes, there is a non-zero probability that an emission would be observed from the other dye and shifted in wavelength.

The efficiency of this transfer process is modulated by the distance and relative orientation of the dyes. Therefore, it is possible to infer intra-molecular distances given the FRET efficiency which in turn can be estimated from relative dye intensities. This is often used to observe gross structural changes in enzymes and other biological molecules. Furthermore, with careful characterization and modelling, quantitative FRET measurements are also possible, allowing measurement of intra-molecular distances and conformation.

In recent years, FRET has enabled studies of macromolecular conformation (Iqbal et al. 2008), mechanical flexibility (Vafabakhsh and Ha 2012), and enzyme kinetics (Kalinin et al. 2010), revealing structure and dynamics inaccessible by other means.

While FRET offers a unique perspective on molecular systems, accurately interpreting the resulting experimental data requires consideration of a number of confounding factors: detector background, imperfect spectral separation, and differences in dye fluorescence efficiency all require characterization and correction for accurate quantitative interpretation. Additionally, studies of molecules in free solution—a very common class of experiment—poses even greater challenges as we'll see in the example below.

Fluorescence Correlation Spectroscopy

Fluorescence Correlation Spectroscopy (FCS) (Haustein and Schwille 2007) is an experimental tool widely used to characterize reaction kinetics (Allen and Thompson 2006) and molecular size. By examining temporal correlations in fluorescence intensity fluctuations within an observation volume, one can infer characteristics of a molecular species' diffusion.

Using FCS in conjunction with FRET can provide further insights into dynamics on molecular length-scales (Nettels, Hoffmann, and Schuler 2008).

Contribution

While fluorescence techniques provide useful insight, they pose both experimental and theoretical challenges that must be addressed. We present a suite of end-to-end tools for the collection and analysis of fluorescence spectroscopy data. Our package includes hardware and software for experimental data collection, as well as a diverse set of tools for manipulating the resulting data. Finally, we introduce a set of tools for robust analysis of FCS and FRET data.

Hardware

We developed a photon arrival timing instrument (B. D. Gamari et al. 2013) based around a readily available FPGA development board[1]. Assembly instructions[2], pre-compiled firmware[3] and extensively documented[4] sources[5] are provided. Further, several introductory experiments using the hardware and tools are described in our paper (B. D. Gamari et al. 2013).

The hardware is capable of recording event times for up to eight input channels with 8ns accuracy. Furthermore, the hardware includes support for driving output channels with programmable binary waveforms, for use in experimental techniques (Lee et al. 2005) which require fast modulation of experimental parameters.

Software for data acquisition

The instrument is complemented with a set of software tools[6] for low-level manipulation of collected data along with a user-friendly graphical interface for setting up and monitoring an experiment. These include real-time monitors of useful experimental quantities including a running FRET efficiency histogram, correlation function, photon counting histogram, and bin count timeseries. This is in stark contrast to existing commercial software which provides little—if any—support to the experimentalist during data collection.

[1] http://www.knjn.com/FPGA-FX2.html
[2] http://goldnerlab.physics.umass.edu/wiki/FpgaTimeTagger?action=AttachFile&do=view&target=construction.pdf
[3] http://goldnerlab.physics.umass.edu/~bgamari/timetag/
[4] http://github.com/bgamari/timetag-fpga/tree/master/docs
[5] http://github.com/bgamari/timetag-fpga
[6] http://github.com/bgamari/timetag-tools

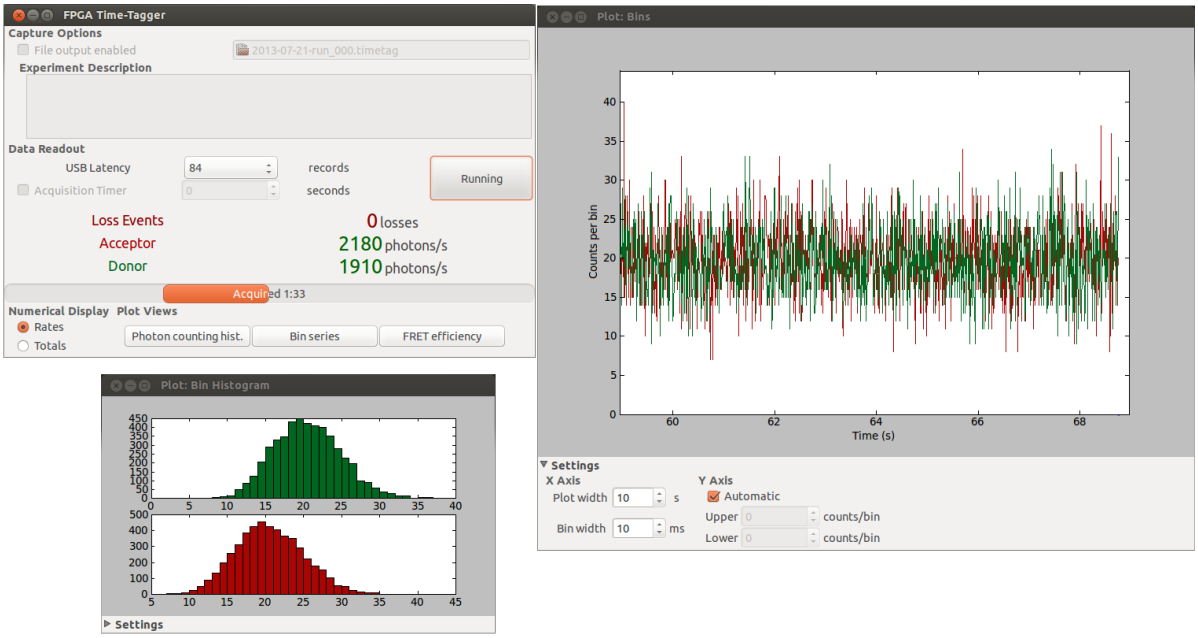

Figure 2: The `timetag_ui` interface during data acquisition

This interface makes it straightforward to setup and acquire fluorescence data. While the tools should run in any POSIX-compliant environment, the usage tutorial[7] provides step-by-step instructions for installation and operation on Ubuntu Linux.

The package also provides a set of low-level command-line utilities for slicing, filtering, and converting data produced by the instrument to a variety of common formats (including plain text, binary, and Matlab `.mat` format).

Tools for data analysis

The `photon-tools` package provides a set of Python utilities and underlying libraries for manipulating and analyzing photon data. The utilities support a variety of file formats including formats of several commercial instruments. This package includes an extensive tutorial[8] as well as comprehensive documentation for the provided Python interfaces.

The `fcs-corr` utility provides a convenient way to compute and plot correlation functions of timestamp data. The `fcs-fit` utility takes one or more correlation functions produced by `fcs-corr` and allows the user to fit these data collectively to a variety of physical models. The tool supports parameter-tying over multiple datasets, and produces a variety of diagnostic statistics and goodness-of-fit metrics.

The `hphoton` package provides the `fret-analysis` and `alex-analysis` tools for analysis of both FRET data and FRET with alternating laser excitation (ALEX) (Lee et al. 2005). These programs support both automatic and manual correction for detector background, cross-talk, and the detection efficiency factor which is necessary for inferring quantitatively accurate FRET efficiencies. Moreover, the tools produce both human-readable HTML providing a succinct summary of each data set, as well as machine-readable results.

[7] http://goldnerlab.physics.umass.edu/wiki/FpgaTimeTagger?action=AttachFile&do=view&target=construction.pdf
[8] https://github.com/bgamari/photon-tools/blob/master/readme.mkd

Interoperability

Due to the modular nature of the packages, the provided tools are highly interoperable with external packages. All output is provided as either tab-delimited text or a well-documented binary format. This makes it trivial to import data and analysis results into tools such as R, OpenBUGS, Matlab, and others for further processing.

The data manipulation tools provided in the `timetag-tools` package hold to the UNIX philosophy, making it easy to chain together individual utilities into more sophisticated pipelines.

Furthermore, the libraries on which the tools are built are well documented and can be easily used on their own. The modules provided by `photon-tools` expose easy-to-use interfaces which are amenable to use in an interactive environment such as ipython notebook[9].

Example application: FRET in solution

In the submission package we provide two datasets[10] from a single-molecule fluorescence experiment taken with our time-stamping hardware, from our recent publication (Milas et al. 2013).

The "RNA 16-mer" dataset was taken on an RNA system labelled terminally with Cy3 and Cy5 dyes. The goal of this experiment was an experimental validation of findings from a molecular dynamics simulation.

A common difficulty in fluorescence spectroscopy is measuring the species of interest for sufficiently long durations. While chemical linkers allow many molecules to be immobilized on a surface, the interaction between the molecule, the linker, and the nearby surface will often perturb the dynamics under study. For this reason, many biological experiments are conducted with the molecule of interest in dilute aqueous solution. In this class of experiment, single molecules occasionally diffuse through the observation volume, giving rise to short bursts of emissions of tens of microseconds in length during which between ten and several hundred photons are observed. Due to presence detector background, one challenge of FRET in solution is distinguishing the short bursts from background noise.

Typically, background is distinguished from signal by a bin-threshold method wherein photon arrivals are partitioned into homogeneous temporal bins. We consider those bins with counts below some threshold to contain no fluorescence signal.

After the background bins have been identified, we examine their count statistics to determine the background count rates of the detectors. These rates can then be used to correct for the background contribution to those bins containing signal.

Another complication is the presence of unwanted donor-only signals. Due to imperfection in sample preparation and the fragility of the dyes, many FRET experiments will have a measurable population of molecules without a functional acceptor dye. Being able to distinguish this donor-only population from the molecule of interest is crucial for quantitatively accurate FRET measurements. Moreover, the ability to reliably isolate the donor-only population provides statistics for correcting spectral crosstalk and differences in dye quantum yields.

To accomplish this, the present tools use approximate probabilistic inference (Markov chain Monte Carlo sampling) to infer a mixture model of Beta distributions over the empirical FRET efficiency distribution, allowing accurate separation of multiple species within a dataset. In this two-stage inference, a component assignment is first sampled for each bin. After assignments have been drawn for each bin, the parameters of each mixture component are reestimated from their assigned samples in a maximum likelihood manner.

The FRET analysis tool provided by the `hphoton` package can carry out these corrections in an automated manner. For instance, to estimate a corrected FRET efficiency distribution from the supplied `fret.timetag` dataset, one might run,

[9] http://ipython.org/notebook.html

[10] experimental data taken by Peker Milas of the Goldner group

```
$ fret-analysis --fit-comps=2 --burst-size=20 --nbins=20
--crosstalk=auto --gamma=auto --bin-width=5e-3 fret.timetag
```

Figure 3 shows the FRET histogram and fits estimated by this tool. On the left of the histogram we see the remnants of the corrected donor-only peak, most of which has been shifted below $E = 0$. Centered around $E = 0.6$ we see the FRET peak from the 16-mer. The blue line shows a Beta fit to this component while the purple line shows the fit that would be expected in the case of a Poisson-limited FRET process (Nir et al. 2006).

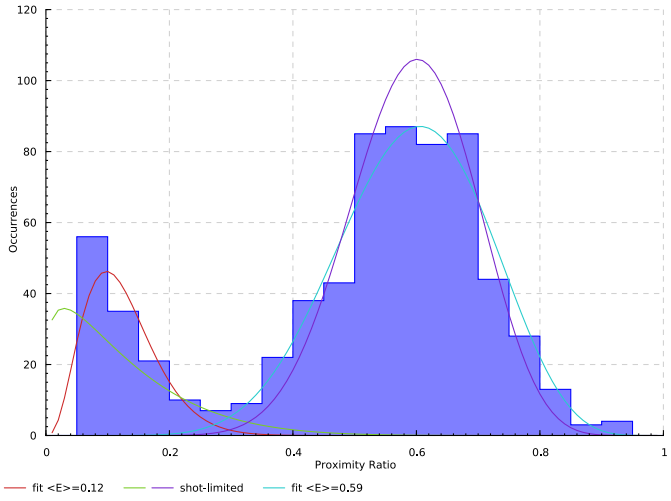

Figure 3: The FRET efficiency histogram inferred from the given RNA 16-mer dataset.

FRET in solution is only one scenario which shows how our tools integrate with each other and benefit from probabilistic inference. Further applications can be found in the Walkthough[11] document accompanying this submission.

Conclusion

We present a package for fluorescence spectroscopic data acquisition, processing, and analysis. Built on off-the-shelf hardware and open-source software, the tools provide a uniquely flexible platform for experimental data acquisition and analysis. By combining high-resolution photon time-stamping with active experimental control, our solution enables multi-channel fluorescence experiments with active excitation modulation. Our toolchain both lowers the barrier to entry for these important experimental tools and enables even more sophisticated methods for higher fidelity fluorescence measurements.

Acknowledgments

This work was supported by NSF-MCB grant #0920139.

References

Allen, Noah W., and Nancy L. Thompson. 2006. “Ligand Binding by Estrogen Receptor Beta Attached to Nanospheres Measured by Fluorescence Correlation Spectroscopy.” *Cytometry* 532 (May): 524–532. doi:10.1002/cyto.a.

[11] file://./walkthrough.pdf

Choi, Paul J., Long Cai, Kirsten Frieda, and X. Sunney Xie. 2008. "A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell." *Science* 322 (5900) (oct): 442–446. doi:10.1126/science.1161427.

Chung, Hoi Sung, John M. Louis, and William A. Eaton. 2009. "Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories." *Proc. Nat. Acad. Sci.* 106 (29): 11837–11844. doi:10.1073/pnas-0901178106.

Fernandez, S. M., and R. D. Berlin. 1976. "Cell surface distribution of lectin receptors determined by resonance energy transfer." *Nature* 264: 411–415.

Gamari, Benjamin D., Dianwen Zhang, Richard E. Buckman, Peker Milas, John S. Denker, Hui Chen, Hongmin Li, and Lori S. Goldner. 2013. "Inexpensive hardware and software for photon statistics and correlation spectroscopy." *American Journal of Physics* (in print).

Goedhart, Joachim, Laura van Weeren, Merel J. W. Adjobo-Hermans, Ies Elzenaar, Mark a Hink, and Theodorus W. J. Gadella. 2011. "Quantitative co-expression of proteins at the single cell level–application to a multimeric FRET sensor." *PloS one* 6 (11) (jan): 27321. doi:10.1371/journal.pone.0027321.

Gopich, Irina V., Daniel Nettels, Benjamin Schuler, and Attila Szabo. 2009. "Protein dynamics from single-molecule fluorescence intensity correlation functions." *J Chem Phys* 131 (9) (sep): 95102. doi:10.1063/1.3212597.

Haustein, Elke, and Petra Schwille. 2007. "Fluorescence Correlation Spectroscopy: Novel Variations of an Established Technique." *Annual Review of Biophysics and Biomolecular Structure* 36 (jan): 151–169. doi:10.1146/annurev.biophys.36.040306.132612.

Heyduk, T., and A. Niedziela-Majka. 2002. "Fluorescence resonance energy transfer analysis of escherichia coli RNA polymerase and polymerase-DNA complexes." *Biopolymers* 61: 201–213.

Iqbal, Asif, Sinan Arslan, Burak Okumus, Timothy J. Wilson, Gerard Giraud, David G. Norman, Taekjip Ha, and David M. J. Lilley. 2008. "Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids." *Proceedings of the National Academy of Sciences* 105 (32) (aug): 11176. doi:10.1073/pnas.0801707105.

Kalinin, Stanislav, Suren Felekyan, Alessandro Valeri, and Claus a. M. Seidel. 2008. "Characterizing multiple molecular States in single-molecule multiparameter fluorescence detection by probability distribution analysis." *The journal of physical chemistry. B* 112 (28) (jul): 8361–74. doi:10.1021/jp711942q.

Kalinin, Stanislav, Alessandro Valeri, Matthew Antonik, Suren Felekyan, and Claus a. M. Seidel. 2010. "Detection of Structural Dynamics by FRET: A Photon Distribution and Fluorescence Lifetime Analysis of Systems with Multiple States." *The journal of physical chemistry. B* 114 (23) (jun): 7983–95. doi:10.1021/jp102156t.

Laurence, Ted A., Xiangxu Kong, Marcus Jäger, and Shimon Weiss. 2005. "Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins." *Proceedings of the National Academy of Sciences of the United States of America* 102 (48) (nov): 17348–53. doi:10.1073/pnas.0508584102.

Lee, Nam Ki, Achillefs N. Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhyay, Richard H. Ebright, and Shimon Weiss. 2005. "Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation." *Biophysical journal* 88 (4) (apr): 2939–53. doi:10.1529/biophysj.104.054114.

Liu, Juewen, and Yi Lu. 2002. "FRET Study of a Trifluorophore-Labeled DNAzyme." *Journal of the American Chemical Society* 124 (51) (nov): 15208–15216. doi:10.1021/ja027647z.

Milas, Peker, Ben D. Gamari, Louis Parrot, Sheema Brent P. Krueger, Rahmanseresht, James Moore, and Lori S. Goldner. 2013. "Indocyanine Dyes Approach Free Rotation at the 3 Terminus A-RNA: A Comparison with the 5 Terminus and Consequences for FRET." *Journal of Physical Chemistry B2* in print.

Nettels, Daniel, Armin Hoffmann, and Benjamin Schuler. 2008. "Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds." *The journal of physical chemistry. B* 112 (19) (may): 6137–46. doi:10.1021/jp076971j.

Nir, Eyal, Xavier Michalet, Kambiz M. Hamadani, Ted A. Laurence, Daniel Neuhauser, Yevgeniy Kovchegov, and Shimon Weiss. 2006. "Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments." *J. Phys. Chem. B* 110 (44) (nov): 22103–24. doi:10.1021/jp063483n.

Pitschke, M., R. Prior, M. Haupt, and D. Riesner. 1998. "Detection of single amyloid [beta]-protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy." *Nat Med* 4 (7) (jul): 832–834.

Sevenich, F. W., J. Langowski, V. Weiss, and K. Rippe. 1998. "DNA binding and oligomerization of NtrC studied by fluorescence anisotropy and fluorescence correlation spectroscopy." *Nucleic Acids Research* 26 (6) (mar): 1373–1381.

Vafabakhsh, Reza, and Taekjip Ha. 2012. "Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization." *Science (New York, N.Y.)* 337 (6098) (aug): 1097–101. doi:10.1126/science.1224139.

Wo´zniak, Anna K., Stephanie Nottrott, E. V. A. Kühn-Hölsken, Gunnar F. Schröder, Helmut Grubmüller, Reinhard Lührmann, Claus A. M. Seidel, and Filipp Oesterhelt. 2005. "Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements." *RNA* 11 (10) (oct): 1545–1554. doi:10.1261/rna.2950605.

Wo´zniak, Anna K., Gunnar F. Schröder, Helmut Grubmüller, Claus A. M. Seidel, and Filipp Oesterhelt. 2008. "Single-molecule FRET measures bends and kinks in DNA." *Proceedings of the National Academy of Sciences of the United States of America* 105 (47) (nov): 18337–42. doi:10.1073/pnas.0800977105.

Wu, Yunxia, Da Xing, Shiming Luo, Yonghong Tang, and Qun Chen. 2006. "Detection of caspase-3 activation in single cells by fluorescence resonance energy transfer during photodynamic therapy induced apoptosis." *Cancer Letters* 235: 239–247.