



FRET restrained high-precision structural modeling of biomolecules

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Simon Sindbert

aus Nürnberg

Düsseldorf, Mai 2012

aus dem Institut für Physikalische Chemie II
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Claus A. M. Seidel
Koreferent: Prof. Dr. Dieter Willbold

Tag der mündlichen Prüfung:

02.07.2012

ABSTRACT

Förster Resonance Energy Transfer (FRET) is a highly distance dependent mechanism particularly useful for the quantification of distances between fluorescent dyes in the range of 20-100 Å. It can be used for measurements on the single molecule level (smFRET). Thus, structural heterogeneities and dynamic transitions in biomolecules are directly measurable under physiological conditions with high time resolution which is a large advantage over most other techniques in structural biology.

However, the main limitation when using quantitative FRET measurements for structural modeling is the lack of accuracy primarily originating from the long linkers which are used to couple fluorophores to the molecules of interest. Even though interdye distances can be measured with high accuracy, translating them into structurally relevant distances between linker attachment atoms yields uncertainties in the order of magnitude of the distances accessible to FRET. Another limitation for the accuracy of FRET distances originates from the uncertainties of the orientations of the dyes which have a significant influence on the efficiency of FRET and are difficult to estimate unless fast rotational diffusion can be assumed. So far, there has been no comprehensive approach to use smFRET measurements for structural modeling which also takes linker effects into account in a meaningful way.

The goal of this thesis was to establish a methodology for the generation of accurate smFRET based structural models of biomolecules.

In a first step, to overcome the above mentioned problem of uncertainty due to linker effects and orientational uncertainties, it was vital to realistically describe linker and dye dynamics (translational and rotational diffusion, respectively) and model dye position distributions. To achieve this, fluorescence labeled dsDNAs and dsRNAs were chosen as model systems and the local environment and mobility of dyes was characterized through fluorescence intensity and anisotropy decay measurements for different linker types. Furthermore, it was shown that by proper consideration of accessible volumes (AVs) of fluorophores, modeled by a simple geometric algorithm for different donor and acceptor positions along a dsRNA, it is possible to accurately predict distances measured by FRET (RMSD = 1.3 Å). Also a rigorous procedure is introduced to estimate and minimize dye orientation uncertainties (κ^2 -problem) which significantly contribute to the errors of measured distances. Considering the case of undefined dye environments we introduce short linkers which significantly reduce position uncertainties. A detailed

analysis of possible orientation effects indicated that, for short linkers and unknown local environments, additional κ^2 -related uncertainties are clearly outweighed by better defined dye positions.

In a second step, a complete set of tools is introduced for smFRET based structural modeling. It includes the proper consideration of FRET distance restraints and measurement uncertainties and, furthermore, structure determination via docking of rigid bodies or screening of structural ensembles. A dramatic improvement in the precision is achieved through explicitly considering spatial distributions of dye positions, which greatly reduces uncertainties due to flexible dye linkers. Furthermore, possible solutions are evaluated and their precision estimated via bootstrapping. A DNA/DNA 19/35 primer/template (dp/dt) to HIV-1 reverse transcriptase (RT) complex was chosen as a validation system. The structure of the double stranded part of dp/dt in complex with the protein was recovered by docking rigid bodies. The model agrees with the known X-ray structure with an RMSD of 0.5 Å. Furthermore, a large structural ensemble of the flexible single stranded template overhang was created by molecular dynamics simulations and filtered with respect to agreement with FRET data yielding a preferential conformation bound to the fingers domain of RT. This might have important implications concerning proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis

Finally, the method was successfully tested on an RNA four-way junction (RNA4WJ) based on the hairpin ribozyme, a dynamic and heterogeneous molecule of unknown structure. 51 independent smFRET measurements were performed, the presence of one major and two minor coexisting conformers was proven and for each dataset three distances and corresponding errors were extracted and successfully assigned to the three conformers of the RNA4WJ using their distinct Mg^{2+} -affinities. A rigid body model for each conformer was obtained by docking four dsRNA helices. The three rigid body models were refined by MD simulations and coarse grained RNA folding using FRET-restrains resulting in meaningful all-atom structural models for all conformers. A cluster analysis gives confidence levels for the proposed ensemble of rigid body models (> 99.99 %, 84 % and 97 %) and their quality was assessed via rigorous error estimation. The achieved precisions are significantly better than the uncertainty of dye position with respect to macromolecule.

ZUSAMMENFASSUNG

Förster Resonanz Energie Transfer (FRET) ist ein stark abstandsabhängiger Prozess, der besonders zur Quantifizierung von Abständen zwischen Fluoreszenzfarbstoffen im Bereich zwischen 20-100 Å geeignet ist. Da FRET auch auf Einzelmolekülebene gemessen werden kann (smFRET), sind strukturelle Heterogenität und dynamische Übergänge in Biomolekülen unter physiologischen Bedingungen direkt zugänglich und können mit einer hohen zeitlichen Auflösung gemessen werden. Dies ist ein großer Vorteil gegenüber den meisten anderen strukturb biologischen Methoden.

Der größte Nachteil bei der Nutzung von FRET-Messungen zur Strukturmodellierung ist jedoch die geringe Genauigkeit, die sich aus der großen Länge der Linker zur Kopplung von Fluoreszenzfarbstoffen an Biomolekülen ergibt. Obwohl Abstände zwischen Farbstoffen sehr genau gemessen werden können, ist die strukturell relevante Distanz zwischen den Atomen, an die die Farbstofflinker gekoppelt sind mit einem Fehler behaftet, der in der Größenordnung der für FRET zugänglichen Abstände liegt. Eine zweite Limitierung der Genauigkeit bei der Abstandsbestimmung ergibt sich aus den schwer zu bestimmenden Orientierungen der Farbstoffe, die einen großen Einfluss auf die FRET-Effizienz haben können. Bislang gibt es keinen umfassenden Ansatz zur FRET-basierten Strukturmodellierung, dem eine sinnvolle Berücksichtigung von Farbstofflinkereffekten zu Grunde liegt.

Ziel dieser Arbeit war die Entwicklung einer Methode zur Generierung von genauen FRET-basierten Strukturmodellen von Biomolekülen.

Zunächst war es wichtig das Problem der Abstandsungenauigkeiten zu lösen, die durch die Farbstofflinker und die Unsicherheit der Farbstofforientierungen verursacht werden. Hierzu ist es notwendig, die Linker- und Farbstoffdynamik (Translations- und Rotationsdiffusion) realistisch zu beschreiben und Farbstoffpositionsverteilungen möglichst genau zu modellieren. Für verschiedene Linkertypen wurden dazu Fluoreszenzintensitäts- und Anisotropiezerfälle an fluoreszenzmarkierter doppelsträngiger (ds) DNA und RNA gemessen, um die Umgebung und Mobilität der Farbstoffe zu charakterisieren. Durch die Berücksichtigung von Farbstoffpositionsverteilungen (accessible volumes, AVs), die durch einen simplen geometrischen Algorithmus modelliert wurden, konnten außerdem präzise Vorhersagen über experimentell gemessene FRET-Abstände an einer dsRNA gemacht werden. Des Weiteren wurde eine rigorose Prozedur eingeführt um die Ungenauigkeiten von Farbstofforientierungen (κ^2 -Problem) abzuschätzen und zu minimieren. Die hier erstmals verwendeten kurzen Farbstofflinker eignen sich besonders für unbekannte Farbstoffumgebungen, wo AV-

Modellierung nicht möglich ist. Für sie ergeben sich kleine zusätzliche Ungenauigkeiten aufgrund Orientierungsunsicherheiten. Diese werden jedoch durch die besser bestimmte Position mehr als aufgewogen.

Als nächster Schritt wurde ein vollständiges Toolset zur Einzelmolekül-FRET-basierten Strukturmodellierung eingeführt. Es beinhaltet die sinnvolle Einbeziehung von gemessenen FRET-Abständen und deren Messfehlern. Des Weiteren die Strukturmodellierung durch das Docken rigider Strukturelemente oder durch „Screening“ von Strukturensambles. Durch die explizite Berücksichtigung von Farbstoffpositionsverteilungen werden Ungenauigkeiten durch Linkereffekte stark reduziert und so die Präzision der Strukturmodelle dramatisch erhöht. Die Methode beinhaltet außerdem die Evaluation möglicher Lösungen und die rigorose Bestimmung ihrer Präzision. Als Validierungssystem wurde ein Komplex aus einer DNA/DNA 19/35 primer/template (dp/dt) und der Reversen Transkriptase (RT) des HIV-1 benutzt. Die Struktur des rigiden doppelsträngigen Teils der DNA im Komplex mit dem Protein wurde durch Docken der beiden Elemente bestimmt. Die mittlere quadratische Abweichung (RMSD) zwischen der gedockten und der Röntgenkristallstruktur betrug 0.5 Å. Mit Hilfe von MD-Simulationen wurde ein großes Strukturensamble für den flexiblen einzelsträngigen Teil der DNA generiert. Anschließend wurde das Ensemble gefiltert, indem Farbstoffverteilungen modelliert wurden und die sich ergebenden FRET-Abstände mit den gemessenen verglichen wurden. Eine bevorzugte Position des Template-Einzelstrangs in der Fingerregion des Proteins konnte so bestimmt werden. Dieses Ergebnis könnte wichtige Konsequenzen für die Ausrichtung des 5'-Endes der DNA im aktiven Zentrum des Enzyms haben und sich auf die DNA-Synthesegenauigkeit auswirken.

Zuletzt wurde die Methode erfolgreich an einer RNA „four-way junction“ (RNA4WJ) getestet, die auf dem Hairpinribozym basiert. Diese RNA4WJ ist ein Molekül, das dynamische Heterogenität aufweist und dessen Struktur unbekannt ist. 51 unabhängige Einzelmolekül-FRET-Messungen wurden durchgeführt und die Existenz von einem Hauptkonformer und zwei gering bevölkerten Konformeren nachgewiesen. Aus jedem Datensatz konnten drei Abstände und die dazugehörigen Fehler extrahiert und den jeweils zugehörigen Konformeren aufgrund ihrer unterschiedlichen Mg^{2+} -Affinitäten zugeordnet werden. Für jedes Konformer konnten Strukturmodelle erstellt werden, indem vier als rigide angenommene dsRNA-Helices gedockt wurden. Diese wurden unter Berücksichtigung der FRET-Abstandsbeschränkungen mit Hilfe von MD- und „coarse grained“-Simulationen verfeinert. So ergaben sich sinnvolle atomare Modelle. Für die sich ergebenden Strukturen konnten Konfidenzniveaus durch Clusteranalyse bestimmt

werden. Die Präzision der jeweiligen Modelle ergab sich aus einem Bootstrappingverfahren und war jeweils um eine Größenordnung kleiner als die Ungenauigkeit die sich aus den Linkereffekten ergibt.

LIST OF PAPER and MANUSCRIPTS

This thesis is based on the following paper and two manuscripts, which and will be referred to in the following way:

Table 0-1.

	Title	Status	impact factor of Journal	contribution by the author	referred to as
Appx I	Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Müller, S.; Seidel, C. A. M.: "Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity" <i>J Am Chem Soc</i> 2011 , 133, 2463	published in "Journal of the American Chemical Society"	9.0	~ 30 %	linker paper
Appx II	Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M.: "FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang"	manuscript submitted to "Nature Methods"	20.7	~ 15 %	RT manuscript
Appx III	Sindbert, S.; Kalinin, S.; Vardanyan, H.; Hanke, C.; Soltysinski, T.; Lach, G.; Müller, S.; Bujnicki, J.; Gohlke, H.; Seidel, C. A. M.: "Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling"	manuscript in preparation, to be submitted to "Science"	31.4	~ 40 %	RNA4WJ manuscript

They are included in the appendices **I** (linker paper), **II** (RT manuscript) and **III** (RNA4WJ manuscript).

Contributions by the author in detail:

“linker paper”:

- buffer preparation
- sample preparation for single-molecule measurements
- Measurement of fluorescence quantum yields
- Alignment and maintenance of the optical single-molecule setup
- All ensemble and most single-molecule fluorescence experiments
- Data analysis of all fluorescence measurements and interpretation
 - e.g. fitting fluorescence and anisotropy decay curves
 - multi-parameter fluorescence analysis
- Writing approx. half of the paper manuscript
- Preparation of most of the figures in the paper manuscript

“RT manuscript”:

- Analysis of approx. half of the 44 single-molecule fluorescence data sets
 - fluorescence burst cutting
 - data quality assessment
 - extraction of FRET distances and corresponding uncertainties
- Structural model generation via rigid body docking
- Estimation quality and precision of rigid body models
- Writing parts of the manuscript

“RNA4WJ manuscript”

- buffer preparation
- sample preparation for single-molecule measurements
- Experiment design: e.g. choosing optimal fluorescence labeling positions
- Optimization of buffer conditions
- Alignment and maintenance of the optical single-molecule setup
- All ensemble and most single-molecule fluorescence experiments
 - 51 single-molecule measurements for all donor-acceptor combinations
 - approx. 80 single-molecule measurements for Mg^{2+} -titrations
 - 18 ensemble measurements of single labeled samples
- Analysis of all ensemble and single-molecule fluorescence data
 - determination of fluorescence quantum yields
 - fluorescence burst cutting
 - data quality assessment
 - extraction of FRET distances and corresponding uncertainties via photon distribution analysis
 - minor conformer assignment

- Structural model generation via rigid body docking
- Estimation of model quality and precision
- FRET-guided “screening” of MD simulation data
- Writing most of the manuscript

Table of Contents

1	Introduction	1
2	Theory of FRET	7
3	FRET restrained high-precision structural modeling.....	11
3.1	Step 1.1: Starting model	12
3.2	Step 1.2: Dye position distributions.....	13
3.3	Step 2: Labeling scheme.....	14
3.4	Step 3.1: Quantitative FRET measurements via smMFD	15
3.5	Step 3.2: Extraction of distances and errors	17
3.6	Step 4: Generation of structural models via rigid body docking	18
3.6.1	Search run	19
3.6.2	Refinement run	19
3.7	Step 5: Model discrimination via cluster analysis	20
3.8	Step 6: Error estimation via bootstrapping.....	21
3.9	Refinement via molecular dynamics (MD) simulations	21
4	Results	23
4.1	Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity	23
4.1.1	Interactions between dyes and nucleic acids	23
4.1.2	Mobility of the dyes	24
4.1.3	Translational movements of the dyes.....	25
4.1.4	AV accurately predicts dye position distributions for nucleic acids	26
4.1.5	Benchmark study for dsRNA	27
4.1.6	Estimation of errors due to orientational uncertainties (κ^2)	27
4.1.7	Short dye linkers reduce position uncertainties	28
4.1.8	Dye diffusion due to linker movement: additional unpublished data	28
4.2	FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang	32
4.2.1	Recovering the position of the rigid dsDNA through FRET restrained rigid body docking	33
4.2.2	Structure determination of the single-stranded template overhang through FRET guided screening of a structural ensemble.....	34
4.3	Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling	37

4.3.1	Assignment of the minor FRET states to the corresponding conformers.....	38
4.3.2	Rigid body docking and model validation	40
4.3.3	Model refinement	42
5	Methods and Materials.....	45
5.1	Materials	45
5.1.1	Materials for “linker paper”	45
5.1.2	Materials for “RT manuscript”	47
5.1.3	Materials for “RNA4WJ manuscript”	49
5.2	Experimental conditions	50
5.2.1	Buffer conditions for RT: dp/dt	50
5.2.2	Buffer conditions for nucleic acids.....	51
5.3	Methods	52
5.3.1	The AV approach.....	52
5.3.2	Definitions of differently averaged DA distances	53
5.3.3	Time-resolved polarized fluorescence experiments and data analysis	54
5.3.4	Single-molecule multi-parameter fluorescence detection (smMFD) measurements	57
5.3.5	Distance determination via photon distribution analysis (PDA).....	58
5.3.6	Static FRET line and distribution of possible σ_{DA} -values	60
5.3.7	Determination of detection efficiency ratio $g_{\text{G}}/g_{\text{R}}$	62
5.3.8	Confidence intervals for fit parameters in PDA	63
5.3.9	Estimation of the κ^2 related distance uncertainties.....	63
5.3.10	Fitting of correlation curves.....	65
5.3.11	Rigid body docking: implementation details	66
5.3.12	R_{mp} to $\langle R_{\text{DA}} \rangle_{\text{E}}$ conversion function.....	69
5.3.13	Model discrimination via cluster analysis	70
6	Conclusions and Outlook	73
7	Acknowledgements.....	77
8	References	79

1 Introduction

Förster Resonance Energy Transfer (FRET) is a highly distance dependent mechanism describing energy transfer between two chromophores (Förster 1948). Acting as a “spectroscopic ruler” (Stryer 1978), it is particularly useful for the quantification of distances between fluorescent dyes in the range of 20-100 Å. Upon excitation of a fluorophore (donor, D) energy is transferred via dipole-dipole coupling to another fluorophore (acceptor, A) provided they are in close proximity.

Apart from its high distance sensitivity ($1/R^6$ - dependence), FRET offers the following advantages: (1) Fluorescent dyes can be site-specifically covalently bound to nucleic acids, proteins and lipids. (2) FRET labeled molecules can be measured while immobilized on surfaces as well as freely diffusing in solution. (3) Due to the availability of fluorophores with high quantum yield and photostability (Gonçalves 2009) and highly sensitive detection systems (Moerner and Kador 1989; Orrit and Bernard 1990; Shera, Seitzinger et al. 1990) with high time resolution, measurements on the single-molecule level are possible (Ha, Enderle et al. 1996; Deniz, Dahan et al. 1999) (smFRET). Thus, structural heterogeneities and dynamic transitions in biomolecules are directly accessible under physiological conditions (*in vitro* (Ha, Enderle et al. 1996; Weiss 1999; Margittai, Widengren et al. 2003; Borgia, Borgia et al. 2011) and even *in vivo* (Sakon and Weninger 2010)) with high time resolution determined by fluorescence lifetime of the dyes in the order of a few nanoseconds. Furthermore, there are practically no limitations to the size of the molecule of interest. In summary, since most biomolecules are dynamic and undergo intrinsic motions (Henzler-Wildman and Kern 2007; Tokuriki and Tawfik 2009), FRET has key advantages over X-ray crystallography (no heterogeneities and dynamics, only molecules that crystallize), NMR spectroscopy (strong limitations in size) and cryo-EM (no dynamics, not under physiological conditions), the three most well established techniques in structural biology.

There are, however, two limitations on quantitative FRET measurements which originate (1) from uncertainties of the dye orientations and (2) from uncertainties of the dye position due to linker effects: (1) The FRET efficiency, E , depends not only on the distance between D and A but is also strongly dependent on the mutual orientation of their transition dipole moments, expressed by the orientation factor κ^2 (see Eq. 2.5). κ^2 is difficult to estimate unless D and A can be assumed to rotate relatively freely. To ensure free rotational diffusion but also for better labeling efficiency, fluorophores are usually coupled to the molecule of interest via long and flexible linkers with overall lengths from the attachment

point of the linker to the center of the chromophore of up to $\sim 20 \text{ \AA}$ (see Figure 1.1). (2) The use of these linkers, however, is the reason for the second and more significant limitation of FRET. It results in significant displacements between mean position of the dye and attachment atom and in large distance distributions between the dyes. Thus, translating accurate interdye distances into structurally relevant distances between attachment atoms yields uncertainties in the order of magnitude of the distances accessible to FRET making those distances impossible to use for accurate structural modeling. To illustrate the problem of flexibly linked dyes, we consider Alexa488 and Cy5 as donor and acceptor dyes, respectively, attached to dsDNA using standard C6 (hexamethylene) linkers (Figure 1.1). There is ample theoretical evidence for the existence of donor-acceptor distance distributions due to flexible dye linkers (Olofsson, Kalinin et al. 2006; Best, Merchant et al. 2007; Dolgikh, Roitberg et al. 2007; VanBeek, Zwier et al. 2007; Dolgikh, Ortiz et al. 2009; Hoefling, Lima et al. 2011). In Figure 1.1, the green and red surfaces show the space accessible to D and A fluorophores, respectively, as determined by a geometric accessible volume (AV) algorithm (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008) (see Section 5.3.1). It is obvious that (1) the mean dye positions (colored spheres) are far from dye attachments points (black crosses) and that (2) AVs are large, requiring an averaging of FRET observables over a distribution of donor-acceptor distances. Thus, distances measured by FRET must not be (mis)interpreted as distances between the mean positions of the dyes or, even worse, as distances between the dyes' attachment points.

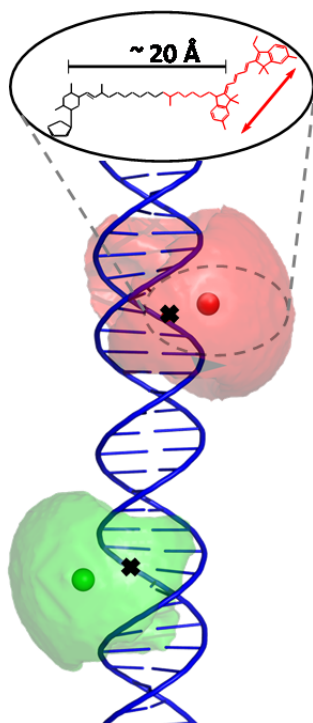


Figure 1.1. Effect of dye position distributions on FRET. (A) Accessible volumes of Alexa488 (green) and Cy5 (red) attached to a dsRNA via a C6 linker (ellipse). The mean positions of the dyes are depicted as spheres (green for Alexa488, red for Cy5) and the attachment atoms of the linkers as black crosses. The distance between the dyes' attachment points (C5 atoms) is 43.6 Å, whereas the distance between the dyes' mean positions (R_{mp}) is 52.6 Å. The structure was rendered via PyMOL (DeLano 2002).

The goal of this thesis is to develop a method for the generation of accurate smFRET based structural models of biomolecules and, thus, to establish FRET as a quantitative tool in the field of structural biology. During my thesis I worked on three main projects which are described in detail in Section 4. The full manuscripts are given in the appendix.

In Section 4.1 approaches are presented to overcome the two main limitations of FRET: The uncertainties in distance measurements due to large dye position distributions and due to orientational uncertainties. The key to increase the spatial resolution of FRET with flexibly linked dyes is the explicit modeling of dye behavior (Best, Merchant et al. 2007; Woźniak, Schröder et al. 2008; Hoefling, Lima et al. 2011) by considering the structure of the biomolecule and calculating the distribution of dye positions. We apply the aforementioned AV approach and verify its assumptions experimentally for nucleic acids and dye linkers of different length and rigidity. To check for possible interactions between dyes and macromolecules, the local environment and the mobility of dyes are characterized through fluorescence intensity and anisotropy decay measurements. Furthermore, in a benchmark study several FRET distances along a dsRNA were predicted through proper averaging over modeled AVs and

compared to measurement results. We also introduce a rigorous procedure to estimate and minimize uncertainties in the orientation factor κ^2 . For cases when the local environment of the dye is unknown and, thus, AV modeling cannot be applied, we propose an alternative approach: We introduce new short dye linkers and test their influence on dye position uncertainties and on orientational dye diffusion. ds B-DNAs and ds A-RNAs were chosen as model systems for the following reasons: (1) Their structure is well known. (2) Fluorescence labeled DNA and RNA samples are commercially available. (3) They are easy to handle.

In Section 4.2 we introduce a complete set of tools for smFRET based structural modeling and demonstrate the accuracy of this approach by docking a DNA/DNA 19/35 primer/template (dp/dt) to HIV-1 reverse transcriptase (RT). Several approaches to derive FRET-restrained structures of biomolecules and of their complexes have been published (Mekler, Kortkhonjia et al. 2002; Margittai, Widengren et al. 2003; Andrecka, Lewis et al. 2008; Muschielok, Andrecka et al. 2008; Choi, Strop et al. 2010; Balci, Arslan et al. 2011; Brunger, Strop et al. 2011; McCann, Zheng et al. 2011; Sabir, Schroder et al. 2011). However, many questions remain unsolved yet. These include the uncertainty of dye positions due to flexible dye linkers, averaging of the FRET efficiency over distributions of donor-acceptor distances, and potential effects of the spatial arrangement of structural units on dye distributions. Another issue is the missing information on the quality of resulting structural models and how it is influenced by uncertainties of “input” FRET data. There is also little evidence as to the accuracy of FRET-restrained 3D modeling, which could be gained by comparison to known structures. Finally, a productive combination of FRET with state-of-the-art *in silico* modeling approaches for the generation of candidate model structures is needed. The method introduced in Section 4.2 includes the following steps: (1) Use of prior knowledge to choose an adequate starting model (e.g. assume protein and dsDNA, both from a crystal structure, as rigid bodies). (2) Selection of meaningful labeling positions. (3.1) Perform quantitative smFRET measurements via single-molecule multi-parameter fluorescence detection (smMFD). (3.2) Extract distance constraints and their corresponding errors from measured data. (4) Generation of structural models. For the double-stranded part of dp/dt in complex with RT by rigid body docking and for the ssDNA template overhang by screening a large structural ensemble created by molecular dynamics with respect to agreement with FRET data. In both cases the dye position distributions are explicitly taken into account. (5) Model discrimination by checking for ambiguous solutions. (6) Once a unique model is found its precision is determined, e.g. by a bootstrapping procedure. The RT:(dp/dt) complex was chosen for following reasons: (1) For the double-stranded part

of dp/dt in complex with RT, there is a number of known crystal structures (Kohlstaedt, Wang et al. 1992; Jacobo-Molina, Ding et al. 1993; Esnouf, Ren et al. 1995; Peletskaya, Kogon et al. 2004) which makes it possible to validate the approach. (2) The ssDNA template overhang has not been resolved by X-ray crystallography, there is, however, biochemical evidence that the properly bound template overhang plays an important role in translocation of nucleic acid during processive DNA synthesis (Goette, Rausch et al. 2010) (3) The complex is easy to label as only one dye has to be attached to the protein (single mutant) and the DNA, respectively. (4) A large data set was available in the Seidel group.

In Section 4.3, the method was applied to an RNA four-way junction (RNA4WJ) derived from the hairpin ribozyme. It has been chosen for the following reasons: (1) The molecule has been proven to exhibit at least three distinct coexisting structural conformations (Hohng, Wilson et al. 2004), none of which have been resolved, and undergoes dynamic transitions between them. Therefore, it is highly suitable for structural determination via FRET. Moreover, transient minor populations have never been structurally characterized before. (2) Helical junctions are highly biologically relevant motifs as they are very common in natural RNA species (e.g. Hairpin (Murchie, Thomson et al. 1998; Fedor 1999; Walter, Burke et al. 1999; Zhao, Wilson et al. 2000; Tan, Wilson et al. 2003), hammerhead (Pley, Flaherty et al. 1994; Scott, Finch et al. 1995), VS (Lafontaine, Norman et al. 2002) ribozymes, riboswitches (Chowdhury, Maris et al. 2006) and ribosomal RNA (Shen and Hagerman 1994; Orr, Hagerman et al. 1998)). (3) So far, there are no structural models of any RNA4WJs unperturbed by interactions with proteins or between stems, e.g. via bases of internal loops. The effects caused by sole staking interactions are unknown. 51 smMFD measurements with different DA-pairs were performed yielding the presence of one major and two minor conformers. For each dataset three distances and corresponding errors were extracted and successfully assigned to the three conformers of the RNA4WJ using their distinct Mg^{2+} -affinities. Rigid body models for each conformer were obtained by docking ds A-RNA helices. The confidence levels of the models are determined via cluster analysis and bootstrapping is applied to provide precision for each model. The three docked structural models are further refined by MD simulations and coarse grained RNA folding using the measured FRET-restraints yielding meaningful all atom structures.

2 Theory of FRET

In this section, the theory behind Förster Resonance Energy Transfer (FRET) and how it can be used for distance determination is explained. FRET is a strongly distance dependent non-radiative energy transfer and occurs between the coupled excited states of a donor (D) and an acceptor (A) dye. For the derivation of the equations shown in this section see refs. (Förster 1948; Andrews 1989; van der Meer, Cooker et al. 1994; Andrews and Bradshaw 2004; Braslavsky, Fron et al. 2008). Throughout this work Alexa488 and Cy5 were used as D and A, respectively (see e.g. Figure 5.1). The rate of D de-excitation via FRET to A, k_{FRET} , is given by Eq. 2.1 (Sisamakias, Valeri et al. 2010):

$$k_{FRET} = \frac{1}{\tau_{D(0)}} \left(\frac{R_0}{R_{DA}} \right)^6 \quad \text{Eq. 2.1}$$

Here $\tau_{D(0)}$ the donor lifetimes in the absence of FRET, R_{DA} is the distance between the donor and acceptor dye and R_0 is the Förster radius which, among other parameters, depends on the properties of the dye pair involved in FRET. R_0 depends on fluorescence quantum yield of the donor dye in the absence of FRET, $\Phi_{FD(0)}$, the refractive index n of the medium between the dyes, the dye orientation factor κ^2 (see Eq. 2.5) and the spectral overlap integral $J(\lambda)$ between fluorescence of D and absorbance of A (see Figure 2.1A). R_0 is given by Eq. 2.2 (Sisamakias, Valeri et al. 2010):

$$R_0 = \left[\frac{9(\ln 10)}{128\pi^5 \cdot N_A} \cdot \frac{J(\lambda) \cdot \kappa^2 \cdot \Phi_{FD(0)}}{n^4} \right]^{\frac{1}{6}} \quad \text{Eq. 2.2}$$

where N_A is the Avogadro constant.

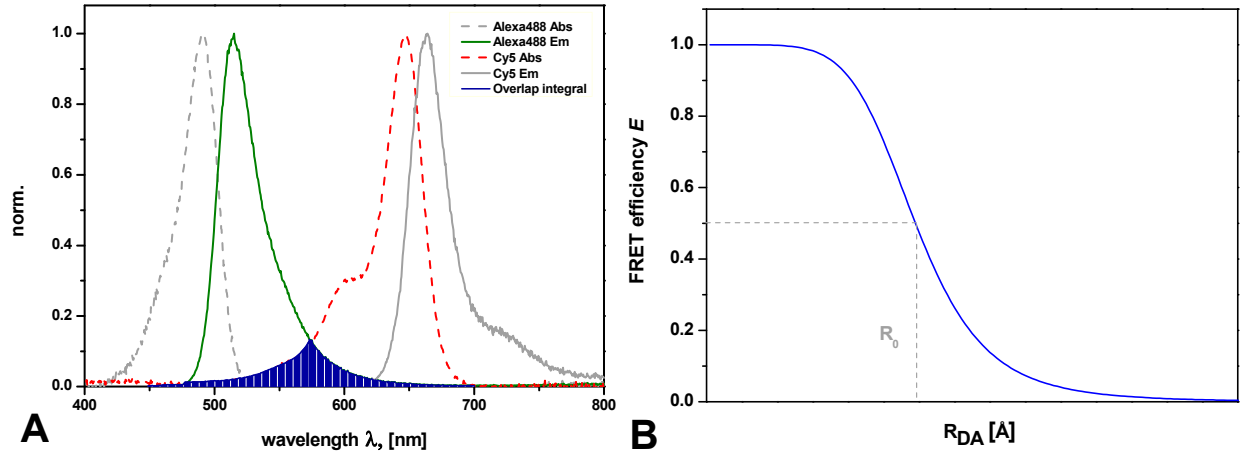


Figure 2.1. (A) Overlap integral $J(\lambda)$ (blue) between normalized fluorescence and absorbance spectra of Alexa488 (green) and Cy5 (red dashed), respectively. The absorption of Alexa488 and the emission of Cy5 are shown as gray dashed and gray solid lines, respectively. (B) The FRET efficiency E is strongly dependent on R_{DA} (see Eq. 2.4). E is most sensible to changes in R_{DA} in the region of R_0 ($E(R_0) = 0.5$).

The FRET efficiency (E) can be determined by measuring the fluorescence of the donor (F_D) and the acceptor (F_A) but also by measuring the donor lifetimes in the presence and in the absence of FRET ($\tau_{D(A)}$ and $\tau_{D(0)}$, respectively) (Sisamakris, Valeri et al. 2010):

$$E = \frac{F_A}{\gamma' F_D + F_A} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad \text{with } \gamma' = \frac{\Phi_{FA}}{\Phi_{FD(0)}} \quad \text{Eq. 2.3}$$

where Φ_{FA} are the fluorescence quantum yield of the acceptor dye. F_D and F_A are the signals in the green and red channels corrected for background and different detection efficiencies (see Eq. 5.16 and Eq. 5.17). Distances can be calculated according to Eq. 2.4 (van der Meer, Cooker et al. 1994; Lakowicz 2006) (see also Figure 2.1B):

$$E = 1 / (1 + R_{DA}^6 / R_0^6) \quad \text{Eq. 2.4}$$

Throughout this work E is mostly determined from measured fluorescence intensities F_D and F_A . If the dyes move slowly through their accessible volumes with respect to the fluorescence lifetimes (static averaging of distances) Eq. 2.4 yields the *FRET-averaged distance* $\langle R_{DA} \rangle_E$ (see Section 5.3.2.2 for the definition of $\langle R_{DA} \rangle_E$ and see Section 5.3.5 for further details on distance determination). This is due to the fact that E and R_{DA} average differently (see Eq. 5.2 and Eq. 5.3). $\langle R_{DA} \rangle_E$ must not be mistaken as the mean distance between the dyes $\langle R_{DA} \rangle = \langle | \mathbf{R}_D - \mathbf{R}_A | \rangle$ (\mathbf{R}_D and \mathbf{R}_A are the position vectors of D and A, respectively, see Section 5.3.2.1) which is determined by measuring the fluorescence decay of the donor, e.g. via time-correlated single-photon-counting (TCSPC, see Eq. 5.8 in Section 5.3.3). In Eq. 2.4

Eq. 2.2 states that R_0 and, therefore, also R_{DA} (see Eq. 2.4) depend on κ^2 which is a measure for the mutual orientation of the transition dipole moments of D and A, μ_D and μ_A (see Figure 2.2) and can be calculated by Eq. 2.5 (Sisamakris, Valeri et al. 2010):

$$\begin{aligned}\kappa^2 &= [\hat{\mu}_A \cdot \hat{\mu}_D - 3(\hat{\mu}_A \cdot \hat{R}_{DA})(\hat{\mu}_D \cdot \hat{R}_{DA})]^2 \\ &= (\sin \theta_D \sin \theta_A \cos \varphi - 2 \cos \theta_D \cos \theta_A)^2\end{aligned}\quad \text{Eq. 2.5}$$

In Eq. 2.5 $\hat{\mu}_D$ and $\hat{\mu}_A$ are the unit vectors corresponding to μ_D and μ_A , respectively, while \hat{R}_{DA} is their connecting unit distance vector. According to Eq. 2.5, κ^2 can range from 0 to 4, however, if rotational diffusion of the dye is fast with respect to k_{FRET} , a mean orientation factor can be used ($\langle \kappa^2 \rangle$). If, additionally, rotational diffusion is unrestricted (isotropic dynamic averaging regime), $\langle \kappa^2 \rangle = 2/3$. In this case $R_0 = 52 \text{ \AA}$ for Alexa488 and Cy5. For most of the linker types for the coupling of dyes to the molecules of interest used in this work (see e.g. Figure 5.1), fast and only weakly restricted rotational diffusion and, therefore, $\langle \kappa^2 \rangle \sim 2/3$ can be assumed (see Sections 3.1.2 and 3.5 in “linker paper” and Section 4.1).

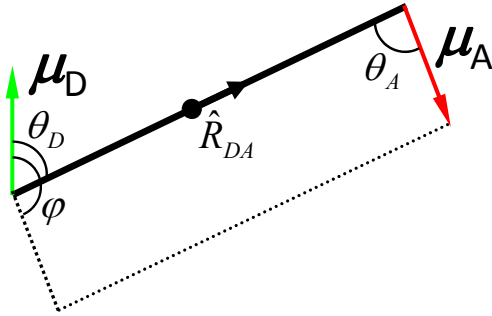


Figure 2.2. Sketch showing angles which define the orientation factor κ^2 in Eq. 2.5: θ_D and θ_A are the angles between the transition dipole moments of the donor and acceptor dyes (μ_D and μ_A) and the distance unit vector \hat{R}_{DA} , respectively, and φ is the angle between μ_D and μ_A .

3 FRET restrained high-precision structural modeling

In this section the procedure of FRET restrained high-precision structural modeling is described according to the flow diagram in Figure 3.1 (see also “RT manuscript”). Following its workflow we end up with a refined structural model. Each step is explained for the major conformer of the RNA four-way junction (RNA4WJ) as described in “RNA4WJ manuscript”.

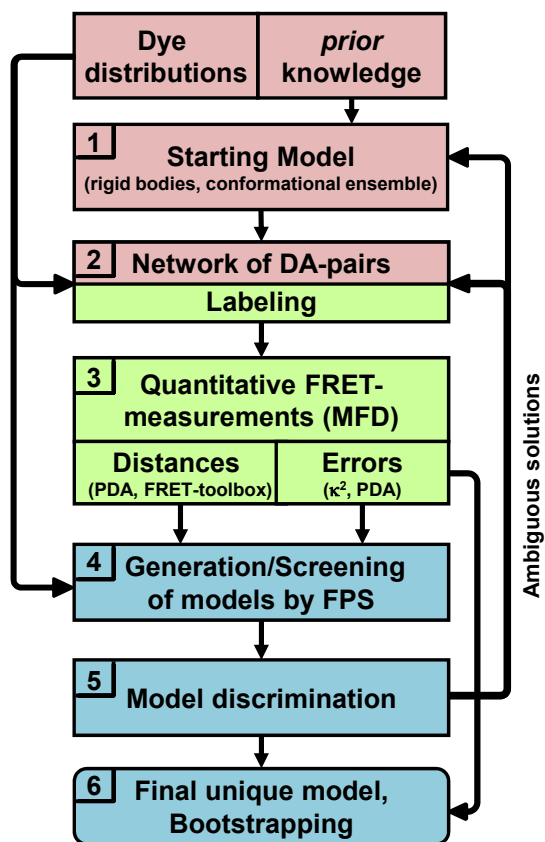


Figure 3.1. Flow diagram of FRET restrained high-precision structural modeling. It comprises three main parts: design of the experiment (shaded in red), measurement and analysis of a set of samples (green), and generation and validation of structural models (blue).

As depicted in Figure 3.1, overall, six steps are needed to generate a FRET-restrained structural model: (1) A starting model is generated utilizing prior knowledge from known structures, homology modeling or *ab initio* modeling (see Section 3.1). (2) Taking the positional distributions of the coupled dyes into account, we use the starting model to design a network of dye positions (see Section 3.3) most useful for FRET Positioning and Screening (FPS, step 4). (3) FRET is quantitatively measured by single-molecule

multiparameter fluorescence detection (smMFD, see Sections 3.4 and 5.3.4). We perform a rigorous data analysis and error estimation of FRET-derived donor-acceptor distances by analyzing photon distributions and time-resolved anisotropies of the dyes (see Sections 3.5, 5.3.5, 5.3.8 and 5.3.9). (4) Possible structural models are searched for and evaluated with respect to their agreement with the FRET data by FPS (see Sections 3.6 and 5.3.11). (5) The possible models are ranked according to their violation of FRET restraints and are assigned to clusters of related structure organization to judge the uniqueness of the structure models (see Sections 3.7 and 5.3.13). (6) In the final step, the precision of the structure models, expressed by the root mean square deviation (RMSD, see Eq. 3.4), is determined by bootstrapping (see Section 3.8 and also Section “Step 6” in “RT manuscript”).

3.1 Step 1.1: Starting model

As a first step, prior knowledge is used to choose an adequate starting model. In the case of the RNA4WJ the sequence and the secondary structure are known (see Figure 3.2). Since all base pairs are Watson-Crick base pairs, it is reasonable to assume the four arms of the RNA4WJ to be rigid ds A-RNA helices. Significant deviations from A-RNA structure are expected only in close proximity to the junction. An alternative approach, e.g. for dynamic molecules or domains, is to use prior information (e.g. sequence) to generate a conformational ensemble using bioinformatical methods, e.g. homology modeling or MD simulations, which can then be used as a starting model (see Section 4.2.2 and see Section “Extension to flexible parts of the complex: ...” in “RT manuscript”).

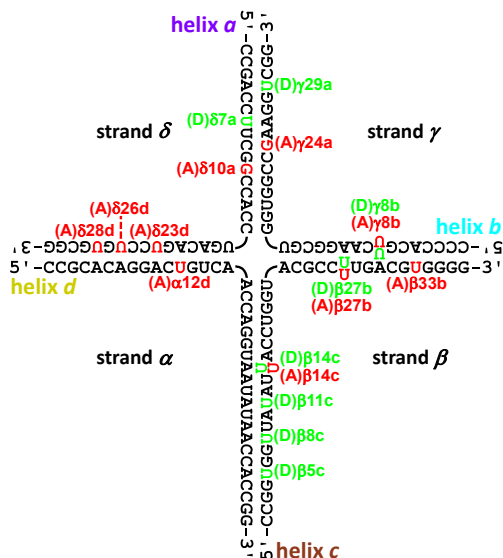


Figure 3.2. Secondary structure of the RNA4WJ with labeling positions for donor (green) and acceptor (red) dyes.

3.2 Step 1.2: Dye position distributions

The key to high accuracy FRET-based structure determination is the proper consideration of position distributions of fluorophores and linker dynamics. In this work we use the accessible volume (AV) approach (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008) for the modeling of dye position distributions (see Section 5.3.1). These are needed for the design of a network of dye positions (see “Step 2” Section 3.3) and, more importantly, for the determination of expected distances and the comparison to measured ones during structural modeling (FPS, see “Step 4” Section 3.6). In AV modeling a simple geometric algorithm is used to find all positions within linker length from the attachment atom of the linker which do not cause sterical clashes with the macromolecule (see Figure 3.3A). As input parameters the algorithm needs the distance between attachment point and center of mass of the dye, L_{link} , width of the linker, w_{link} , and, to account for the three quite different dimensions of a fluorophore, the three radii of the dye, $R_{\text{dye}(i)}$ (see Figure 3.3B). Also the structure in close proximity of the dye needs to be known (see Section 5.3.1 for further details and all values used for L_{link} , w_{link} and $R_{\text{dye}(i)}$ throughout this work). We assume an equally distributed probability throughout the AV. The applicability of the simple AV approach for nucleic acids and long C6 dye linkers (as used throughout “RNA4WJ manuscript”) has been experimentally proven in “linker paper”: (1) There are no significant specific interactions between dyes and the macromolecules (see Section 4.1.1). (2) The movements of the dyes

through their sterically accessible volumes are unrestricted (see Section 4.1.2). (3) By properly averaging over AVs, we were able to predict distances measured by FRET with high accuracy (see Section 4.1.5). The equally distributed probability throughout the AVs for nucleic acids can be explained by two counterbalancing effects: As the dye linker behaves like a Gaussian chain, the fully extended linker is entropically unfavorable. On the other hand, there is repulsion between the phosphate backbone and the negatively charged dyes (SO_3^- groups, see e.g. Figure 5.1). For proteins (as in “RT manuscript”, see Section 4.2) the assumptions of the AV approach are not necessarily given. However, its use is still justified as long as oversampling with sufficient FRET distance restraints compensates for systematic errors.

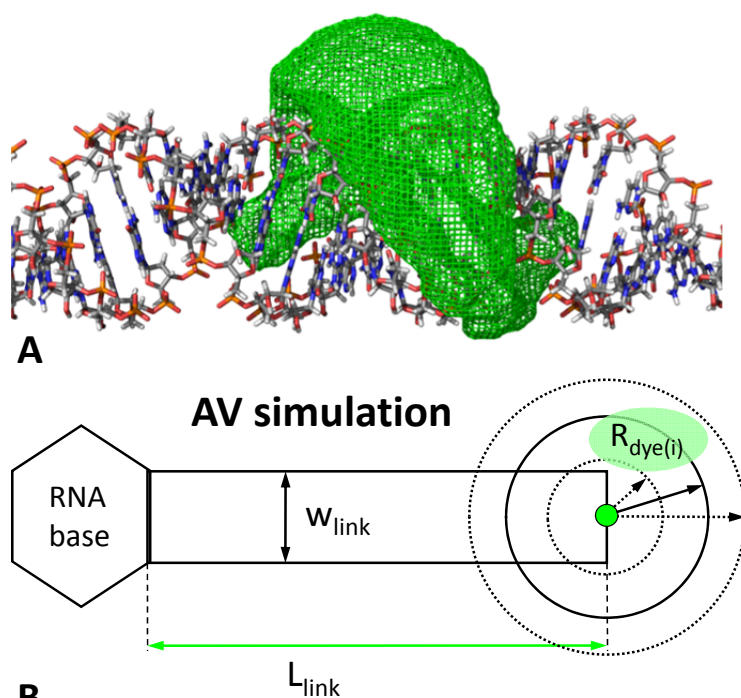


Figure 3.3. (A) AV simulation of possible positions of Alexa488 attached to a ds A-RNA via a C6 linker. The structure was rendered via PyMOL (DeLano 2002). **(B)** For AV simulations for a C6-linker in “RNA4WJ manuscript” we used $w_{\text{link}} = 4.5 \text{ \AA}$, and $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$, and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for the three dimensions of Alexa488 and $R_{\text{dye}(1)} = 11 \text{ \AA}$, $R_{\text{dye}(2)} = 3 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for Cy5. We used $L_{\text{link}} = 20 \text{ \AA}$ for Alexa488 and $L_{\text{link}} = 22 \text{ \AA}$ for Cy5.

3.3 Step 2: Labeling scheme

The next step is to choose meaningful labeling positions for D and A yielding sufficient FRET restraints. For the case of the RNA4WJ we chose 8 D and 10 A positions on the different helices of the RNA4WJ (see Figure 3.2) yielding 51 measurable FRET-pairs. It is important that there are sufficient restraints between

all possible helix pairs so that the positions and orientations of the helices in space are well defined. For labeled uracils the dye linker points into the major groove of dsRNA (see Figure 5.4A), thus, the mean position of the dye is close to the helix axis (see Figure 3.4A) and is, therefore, invariant to mirror image transformation of the structure. For labeled guanines, dye linkers point into the minor groove resulting in AVs which are significantly displaced from the helical axis (see Figure 3.4B and Figure 5.4B). Labeling positions (A) γ 24a and (A) δ 10a (see Figure 3.2) were specifically chosen to distinguish between the two quasi-mirrored solutions for the major conformer of the RNA4WJ (see Figure 4.3 in the SI of “RNA4WJ manuscript”) resulting in well separable solutions according to agreement with FRET distances (see Figure 3.7B). Without FRET distances involving (A) γ 24a and (A) δ 10a positions the agreement is very similar.

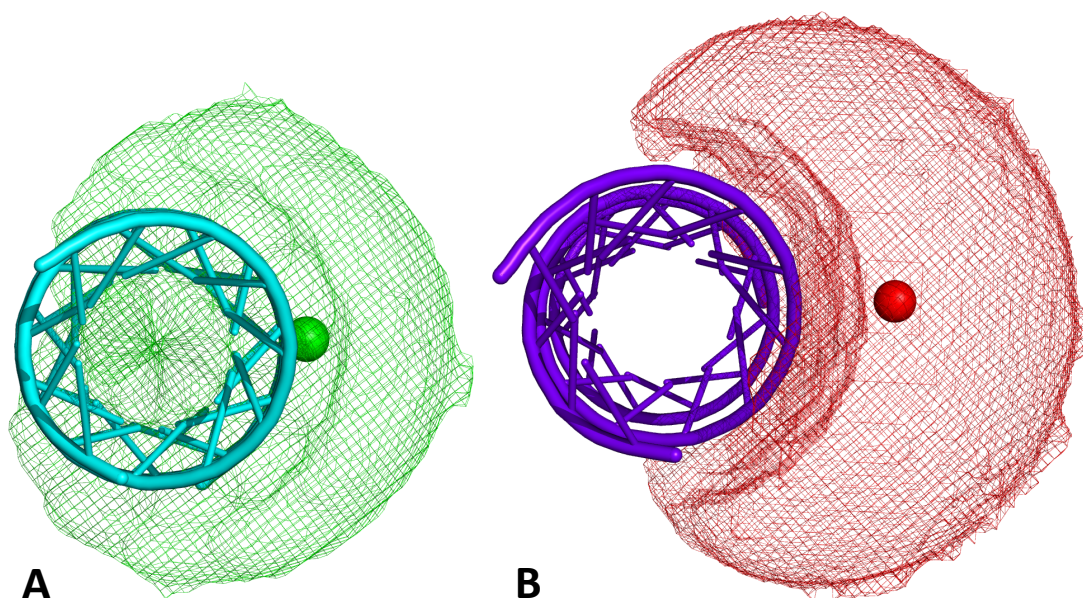


Figure 3.4. Accessible volumes of Alexa488 (panel A, green) and Cy5 (panel B, red) attached at positions (D) β 8c (A, modified U) and (A) δ 10a (B, modified G), respectively. The mean dye positions are displayed as green and red spheres. Their distances from the helical axis are ~ 10 Å (A) and ~ 17 Å (B), respectively. The following parameters were used for the AV simulation: Alexa488: $L_{\text{linker}} = 20$ Å; $w_{\text{linker}} = 4.5$ Å; $R_{\text{dye}(1)} = 5$ Å; $R_{\text{dye}(2)} = 4.5$ Å; $R_{\text{dye}(3)} = 1.5$ Å; Cy5: $L_{\text{linker}} = 22$ Å; $w_{\text{linker}} = 4.5$ Å; $R_{\text{dye}(1)} = 11$ Å; $R_{\text{dye}(2)} = 3$ Å; $R_{\text{dye}(3)} = 1.5$ Å. The structures were rendered via PyMOL (DeLano 2002).

3.4 Step 3.1: Quantitative FRET measurements via smMFD

All different DA-pairs are then measured via single-molecule multi-parameter fluorescence detection (smMFD, see Figure 3.5A for experimental setup and Section 5.3.4). In smMFD an inverted confocal microscope setup is used while the sample is in solution and diluted sufficiently high so that there is

never more than one molecule diffusing through the focus at the same time. After excitation by linearly polarized pulsed laser light the fluorescence light is split according to spectral range (green and red) and polarization and detected by avalanche photo diodes with ps-time resolution (see Section 5.3.4 for further details). Hence, smMFD simultaneously acquires all fluorescence parameters (Sisamakos, Valeri et al. 2010) e.g. fluorescence intensities, fluorescence lifetimes and anisotropies of the donor and the acceptor dye, respectively. For the analysis of smFRET experiments 2D frequency histograms of the fluorescence intensity ratio F_D/F_A (or equivalently any other FRET indicator) versus the donor fluorescence lifetime $\tau_{D(A)}$ are the most important plots. According to Eq. 2.3, all changes in F_D/F_A must correlate with observed changes in $\tau_{D(A)}$. The theoretically expected relation between the two observables for static FRET states can be plotted (see Eq. 3.1 and Figure 3.5B). Deviations from this “static FRET line” indicate fluorophore quenching or bleaching or dynamic transitions between FRET states. Furthermore, in 2D histograms of F_D/F_A versus $\tau_{D(A)}$ species due to incomplete labeling and, most importantly, multiple FRET states are well resolved (Margittai, Widengren et al. 2003; Rothwell, Berger et al. 2003; Sisamakos, Valeri et al. 2010). Figure 3.5B shows the dataset for the RNA4WJ sample with labeling positions $(D)\beta 11c$ and $(A)\delta 23d$ (see Figure 3.2) plotted in 2D frequency histograms of F_D/F_A and the donor anisotropy r_D versus $\tau_{D(A)}$. The D-only species and the three FRET states (indicated by horizontal lines) follow the dependencies between $\tau_{D(A)}$ and F_D/F_A (see Eq. 3.1) and r_D (Perrin equation).

$$F_D/F_A = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left(\frac{\langle \tau_{D(0)} \rangle_x}{c_3 \langle \tau_{D(A)} \rangle_f^3 + c_2 \langle \tau_{D(A)} \rangle_f^2 + c_1 \langle \tau_{D(A)} \rangle_f + c_0} - 1 \right) \right. \quad \text{Eq. 3.1}$$

In Eq. 3.1 $\langle \tau_i \rangle_x$ and $\langle \tau_i \rangle_f$ are species and fluorescence averaged mean lifetimes, respectively, and a polynomial approximation with coefficients c_i is used for the $F_D/F_A(\tau_{D(A)})$ dependence to correct for non-exponential fluorescence decay of the donor dye in absence of FRET (see Section 4.1 in the SI of “RNA4WJ manuscript” and Table S6 in the SI “RT manuscript”) and distributions of donor-acceptor distances due to flexible dye linkers (see Section 5.3.6 for further details).

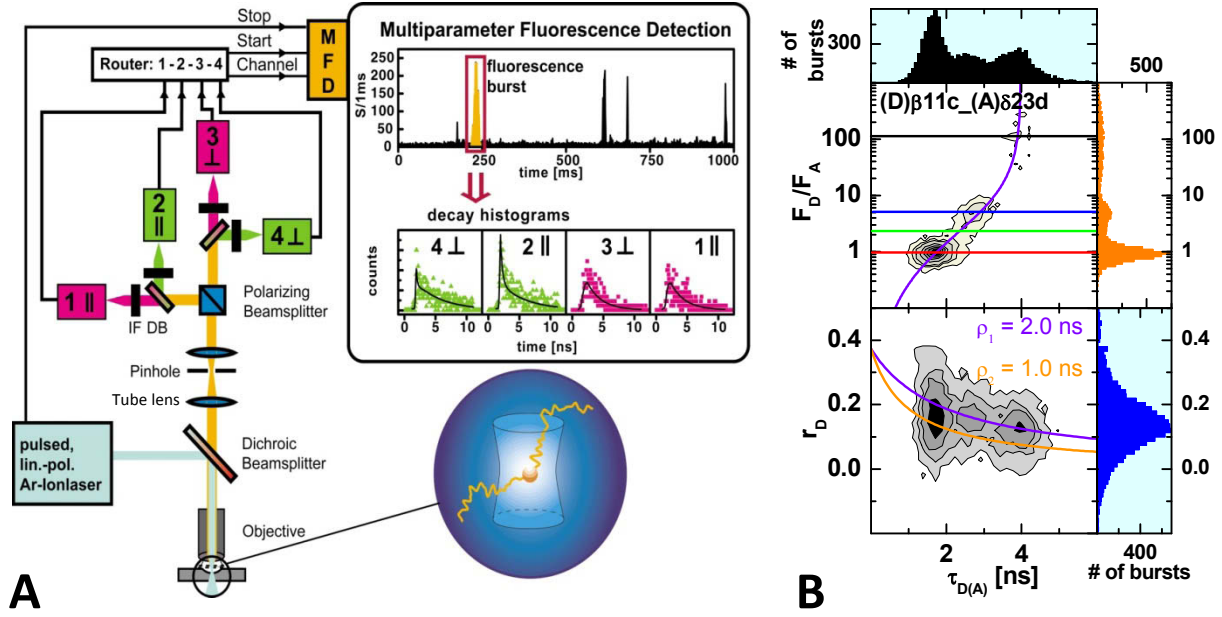


Figure 3.5. (A) Experimental setup for smMFD measurements. For further details see Section 5.3.4 and (Kühnemuth and Seidel 2001). **(B)** 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for sample $(D)\beta11c_{(A)}\delta23d$. The number of molecules (fluorescence bursts) in each bin is gray scaled, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; purple, see Eq. 3.1 and Section 5.3.6 for details) and horizontal lines indicating the D only species (black) and three FRET states (red, green and blue) are overlaid. The solid purple and orange lines in the r_D - $\tau_{D(A)}$ diagram are given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$, with rotation correlation times $\rho_1 = 1.0$ ns (orange) and $\rho_2 = 2.0$ ns (purple), $r_0 = 0.374$.

3.5 Step 3.2: Extraction of distances and errors

In order to extract values of $\langle R_{DA} \rangle_E$ from noisy single molecule data we use photon distribution analysis (PDA) (Antonik, Felekyan et al. 2006; Kalinin, Felekyan et al. 2008) (see Figure 3.6A and Section 5.3.5 for further details) as it takes into account shot noise and provides a meaningful reduced chi-squared value (χ_r^2) directly derived from photon statistics. Thus, PDA can resolve even minor FRET populations states with high accuracy (see Figure 3.6A). Furthermore, PDA allows us to estimate the errors of the fitted parameters due to photon statistics, in particular for $\langle R_{DA} \rangle_E \{ \Delta R_{DA}(E) \}$ (see Section 5.3.8 for more details). Another contribution to ΔR_{DA} originates from the uncertainty of κ^2 (see Eq. 2.3 and Eq. 2.2). $\Delta R_{DA}(\kappa^2)$ can be derived from the distribution of possible values for κ^2 (see Figure 3.6B) which can be determined from the offsets of the anisotropy decays of D-only and A-only molecules and the FRET-sensitized acceptor ($r_{\infty,D}$, $r_{\infty,A}$ and $r_{\infty,A(D)}$, respectively, see Section 5.3.9 for more details). The overall error ΔR_{DA} is then calculated according to error propagation rules by Eq. 3.2:

$$\Delta R_{DA}^2 = \Delta R_{DA}^2(E) + \Delta R_{DA}^2(\kappa^2) \quad \text{Eq. 3.2}$$

We finally end up with a complete set of 51 distance restraints $\langle R_{DA} \rangle_E$ and their corresponding uncertainties ΔR_{DA} for the major conformer of the RNA4WJ.

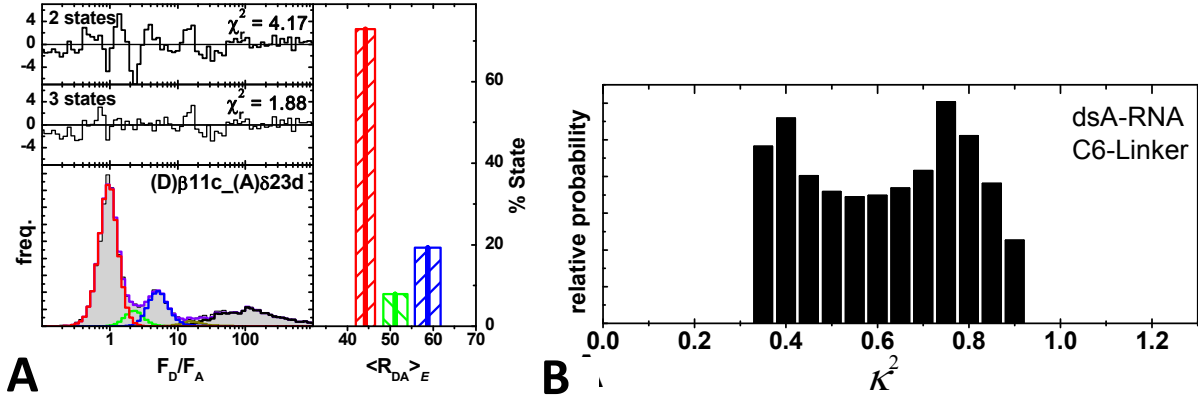


Figure 3.6. (A) PDA for sample (D)β11c_(A)δ23d (selected bursts). F_D/F_A histogram of experimental data (gray area) is fitted (purple solid line) using the following parameters: 45.1 % of $\langle R_{DA} \rangle_{E(1)} = 44.2 \text{ Å}$ (red); 4.9 % of $\langle R_{DA} \rangle_{E(2)} = 51.2 \text{ Å}$ (green); 11.9 % of $\langle R_{DA} \rangle_{E(3)} = 58.8 \text{ Å}$ (blue); $\sigma_{app} = 3.8 \text{ %}$ of $\langle R_{DA} \rangle_E$, respectively; 36.2 % of D-only; 1.9 % of impurities with apparent $R_{DA} = 71.5 \text{ Å}$ (also present in D-only samples); $\chi_r^2 = 1.88$. Weighted residuals are shown in the middle plot. (see Section 4.4 in the SI of “RNA4WJ manuscript” for all PDA parameters). The upper plot shows the weighted residuals for PDA with only two FRET states ($\chi_r^2 = 4.17$) (see Section 4.2 in the SI of “RNA4WJ manuscript”). The right panel shows the distances and relative amplitudes (solid lines) and the overall confidence intervals for the distances (striped boxes) of the three FRET states ($\Delta R_{DA(1)} = 5.0 \text{ %}$, green, $\Delta R_{DA(2)} = 5.6 \text{ %}$, red and $\Delta R_{DA(3)} = 5.1 \text{ %}$, blue). (B) Calculated probability distribution for possible values of κ^2 in dsA-RNA and for C6 linkers determined by the offsets from the anisotropy decays of the D-only and A-only molecules ($r_{\infty,D} = 0.05$ and $r_{\infty,A} = 0.13$, respectively) and the FRET-sensitized acceptor ($r_{\infty,A(D)} = 0.012$). The distribution results in a value for $\Delta R_{DA}(\kappa^2)$ of 5.1 %. (see Section 3.5 in “linker paper”)

3.6 Step 4: Generation of structural models via rigid body docking

To find those structures which agree best with the measured distances we apply the rigid body dynamics approach (for details see Section 5.3.11) while explicitly modeling the AVs of fluorophores (see Section 3.2). This makes it possible to estimate mean dye positions with respect to the dsRNA helices and average over distributions of DA distances.

To avoid repeated calculations of AVs during the iterative structure optimization we convert $\langle R_{DA} \rangle_E$ (see Eq. 5.3) into distances between the mean positions of the dyes $R_{mp} = |\langle R_D \rangle - \langle R_A \rangle|$ (see Eq. 5.4) using an $\langle R_{DA} \rangle_E$ to R_{mp} conversion function (see Section 5.3.12 for details).

3.6.1 Search run

In a first “search” run the weighted data-model deviation for a set of $n = 51$ distances (see Eq. 3.3), here defined for values of R_{mp} , is minimized by docking the four dsRNAs. This is done sufficiently often (1000 iterations) yielding all local minima.

$$\chi_E^2 = \sum_{i=1}^n \frac{(R_{\text{DA}(i)} - R_{\text{model}(i)})^2}{\Delta R_{\text{DA}(i)}^2} \quad \text{Eq. 3.3}$$

3.6.2 Refinement run

In a second “refinement” run the AVs are re-modeled for all solutions found in the first run. This takes into account possible sterical clashes of fluorophores with RNA4WJ arms they are *not* attached to. All structures are then optimized using the new AVs. The resulting structure for the major RNA4WJ conformer with the lowest χ_r^2 (Eq. 5.29) after the refinement run is shown in Figure 3.7A.

For the case on non-rigid domains (as mentioned in Section 3.1) the conformational ensemble can be filtered by first, modeling dye distributions and calculating expected distances and, second, screening with respect to agreement with FRET data as shown in Section “Extension to flexible parts of the complex: ...” in “RT manuscript”.

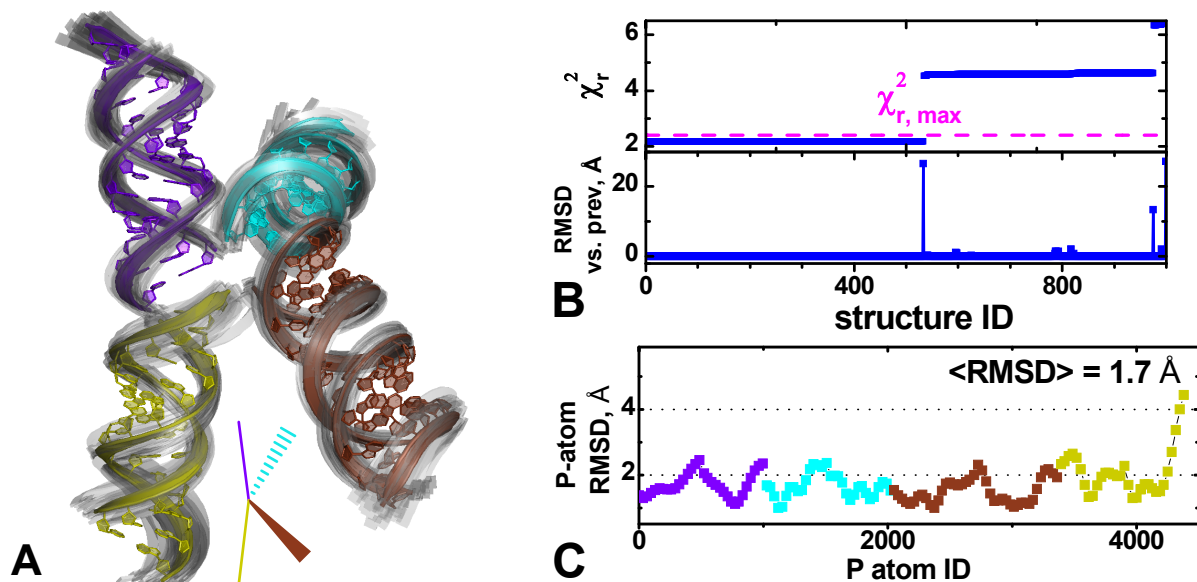


Figure 3.7. (A) Resulting structure (cartoon representation) with the lowest χ_r^2 after refinement run of rigid body docking (colored) for the major RNA4WJ conformer overlaid with 100 structures (grey transparent) indicating the precision (uncertainties) of the helix positions and orientations resulting from the bootstrapping procedure. Bottom: Sketch depicting the mutual orientation of helices *a* (purple), *b* (cyan), *c* (brown) and *d* (dark yellow). The structures were rendered via PyMOL (DeLano 2002). (B) χ_r^2 (upper panel) and RMSD vs. the previous structure (lower panel) plotted against the (number of the) structure ID found after the refinement run for all 1000 solutions in ascending order with respect to χ_r^2 . The dashed magenta line represents a 84 % confidence level: $\chi_{r,\max}^2 = 2.46$ (see Eq. 5.36) (C) uncertainty of phosphate atom positions of helices *a* (purple), *b* (cyan), *c* (brown) and *d* (dark yellow), respectively, calculated for each P atom for the solution with the lowest χ_r^2 after the refinement procedure for the major RNA4WJ conformer. The average RMSD value for all P atoms is 1.7 Å.

3.7 Step 5: Model discrimination via cluster analysis

For the purpose of model discrimination the χ_r^2 and the RMSD vs. the previous structure for all structures from the refinement run are plotted to recognize clusters of similar solutions (see Figure 3.7B). Steps in the χ_r^2 plot and corresponding peaks in the RMSD plot separate groups of very similar solutions (clusters). For the major conformer of the RNA4WJ only one group of solutions with the same $\chi_r^2 = 2.2$ and zero RMSD within the cluster is found below the 84 % confidence threshold ($\chi_{r,\max}^2$, see Eq. 5.36). The cluster of solutions with the second lowest χ_r^2 deviates strongly from the first one (RMSD = 26.6 Å) and can be excluded with a confidence level of > 99.99 % ($\chi_r^2 = 4.5$). Hence, the first solution can be considered to be unique. See Section 5.3.13 for further details. For the case of ambiguous

solutions, one can either go back to “step 1.1” to rethink the starting model, or go back “step 2” and add additional labeling positions to distinguish between the possible solutions as described in Section 3.3.

3.8 Step 6: Error estimation via bootstrapping

To estimate the uncertainties of the best solution we apply a procedure comparable to bootstrapping (Efron 1986) (see Section “Step 6” in “RT manuscript” for details). All model distances found for the optimized structure are simultaneously perturbed by normally distributed random numbers with a mean of zero and the standard deviation given by the experimental errors $\{\Delta R_{DA}\}$. Afterwards the structures are re-optimized using the perturbed distances. This procedure is repeated 100 times yielding a set of structures representing the distribution of possible positions of the helices. In Figure 3.7A the best solution after rigid body docking for the major RNA4WJ conformer is overlaid with a set of perturbed structures and in Figure 3.7C the uncertainties of all P-atoms are plotted which were calculated according to Eq. 3.4:

$$\text{RMSD} = \left\langle \left| \vec{R}_{\text{best model}} - \vec{R}_{\text{perturbed model}} \right|^2 \right\rangle^{1/2} \quad \text{Eq. 3.4}$$

The average RMSD value for P atoms can be regarded as the precision of the structural model (1.7 Å).

3.9 Refinement via molecular dynamics (MD) simulations

In this Section the refinement of the rigid body model through molecular dynamics (MD) simulations is described. Alternatively, it can also be refined via coarse-grained simulations as described in Section “SimRNA” in “RNA4WJ manuscript”. The final rigid body model is used as starting structure for further refinement through all-atom MD simulations in explicit solvent (Case, Cheatham et al. 2005) (see Section 3.12 in the SI of “RNA4WJ manuscript”). Throughout the simulation the positions of all P-atoms except for those within 6 base pairs distance from the junction were restrained by harmonic potentials such that their fluctuations are equal to positional uncertainties (RMSD values) obtained by bootstrapping for the rigid body models (see Figure 3.7C). The resulting trajectories were then filtered with respect to agreement with FRET data. For this, AVs were modeled to the MD-derived structures, resulting DA distances were calculated and compared to measured ones. Figure 3.8 shows the resulting

structure with the lowest value of χ_r^2 , i.e., best agreement with FRET distance restraints. The model of the RNA4WJ from FRET-filtered molecular dynamics simulations has both the global geometry consistent with the FRET restraints and the local stereochemistry encoded in the MD force field (see Section 3.12 in the SI of “RNA4WJ manuscript”). In particular, all bases at the junction are properly stacked after MD-refinement. Compared to the best docked model, the RMSD over all P atoms, excluding those within 6 bps distance from the junction, is 3.17 Å.

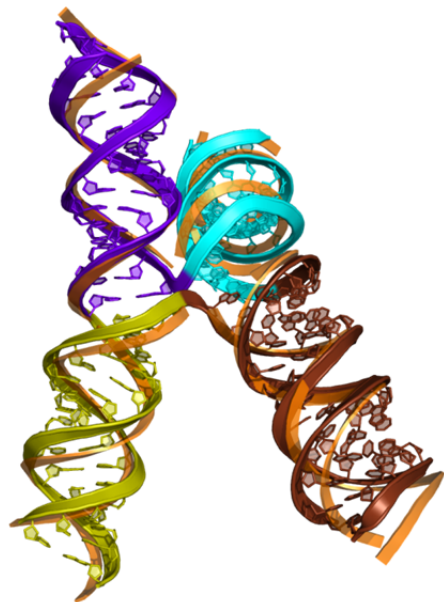


Figure 3.8. Structure (cartoon representation) with the lowest χ_r^2 after refinement of the rigid body model (orange) by MD simulations (helices *a* (purple), *b* (cyan), *c* (brown) and *d* (dark yellow)) for the major RNA4WJ conformer. RMSD over all P atoms, excluding those within 6 bps distance from the junction, is 3.17 Å. The structure was rendered via PyMOL (DeLano 2002).

4 Results

In this section the most important results of the paper and manuscripts this thesis is based on (see Table 4-1 and Table 0-1) are summarized.

Table 4-1. List of paper and manuscripts this thesis is based on

Title	Status	Section
Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity	published	4.1
FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang	submitted	4.2
Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling	manuscript in preparation	4.3

4.1 Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity

In this work approaches are presented to overcome the two main limitations of FRET: The uncertainties in distance measurements due to large dye position distributions and due to orientational uncertainties (κ^2 -problem). As model systems dsDNA and dsRNA internally labeled with Alexa488 (donor, D) and/or Cy5 (acceptor, A) and different dye linker types (long, L, intermediate, I, short with an alkenyl unit, S_d and short with a propargyl unit, S_t , see Figure 5.1) were used. See “linker paper” for further details.

4.1.1 Interactions between dyes and nucleic acids

The local environment of dyes for the different dye linker types was characterized by analyzing fluorescence decays of single labeled dsDNA and ds RNA via eTCSPC measurements (see Section 5.3.3). It was shown that no strong quenching due to interactions with nucleobases occurs and, except for linker S_t , quenching is purely dynamic (see Sections 3.1.1 and 3.4 in “linker paper”). In summary, except

for linker S_t , there are no significant interactions between the dyes and the nucleic acids other than sterical ones.

4.1.2 Mobility of the dyes

The mobility of dyes for the different linker types was characterized by analyzing fluorescence anisotropy decays of single labeled dsDNA and ds RNA via eTCSPC measurements (see Section 5.3.3). It was shown that for all dyes and linker types the local orientation fluctuations with the rate constant of $k_R=1/\rho_1$ (ρ_1 is the fastest rotation correlation time of the anisotropy decay) are fast (see Table 2 in “linker paper”). However, the case isotropic dynamic averaging is not given for any linker type as dye movement is never completely unrestricted. The opening half-angles describing cones in which the linkers wobble (θ_D and θ_A for D and A, respectively, see Section 3.1.2 and Figure 3C and D in “linker paper” and Figure 4.1A) are largest for Alexa488 with L linkers and for Cy5 with all linker types (Cy5 has its own 6 atom linker, see e.g. Figure 5.1) for which cases they correspond to the opening angles of the major grooves of DNA and RNA (45° and 32° , respectively, see Table 2 in “linker paper”). In case of Alexa488 and short linkers θ_D is 0° . However, θ_D and θ_A are never 90° , as it would be the case for unrestricted movement. In summary, due to fast rotation, a mean effective orientation factor κ^2 can be used for the determination of R_0 , however, due to the restriction of dye motions, it is not necessarily equal to $2/3$ (see Section 2). Furthermore, for L linkers, movement is only restricted by the groove of the nucleic acid.

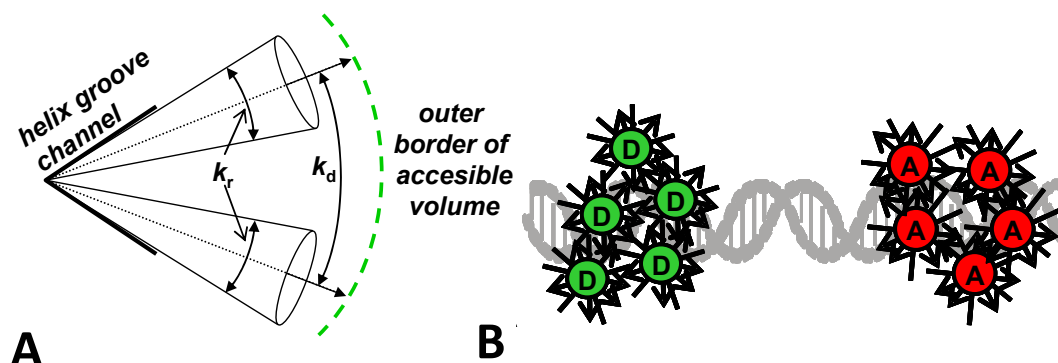


Figure 4.1. (A) The diffusion of the dyes in the sterically allowed volume (helical groove) with the characteristic diffusion rate constant k_d and the orientation fluctuations (wobbling within a cone) with the rate k_R ($k_R=1/\rho_1$). (B) Sketch illustrating the case of fast rotational and slow translational diffusion ($k_R \gg k_{FRET} \gg k_d$).

4.1.3 Translational movements of the dyes

smMFD measurements (see Section 5.3.4) of FRET labeled nucleic acids show that for all linker types except for linker S_t , the translational motions of the dyes average out in the ms time scale (see Section 3.4 of “linker paper” and (Kalinin, Sisamakos et al. 2010)). However, ensemble time-correlated single-photon-counting (eTCSPC) measurements (ns resolution, see Section 5.3.3) show that they are slow on the timescale of FRET (few ns). The fluorescence intensity decay of the donor in the presence of FRET cannot be fitted with a single exponential decay. Instead, a Gaussian distribution of distances (see Eq. 5.2 and Eq. 5.3) needs to be assumed (see Figure 4.2). For all linker types except S_t , the widths of the resulting distance distributions are typically $\sigma_{DA} \sim 7 \text{ \AA}$ (see Table 4 in “linker paper”). In summary, the diffusion of the dyes in the sterically allowed volume with the characteristic diffusion rate constant k_d is much slower than the FRET rate constant k_{FRET} (see Eq. 2.1), i.e. the distribution of individual DA distances $R_{DA, i}$ is quasi-static on the FRET time scale. This finding is further supported by the data presented in Section 4.1.8. In Figure 4.1 the case for $k_R \gg k_{FRET} \gg k_d$ is depicted.

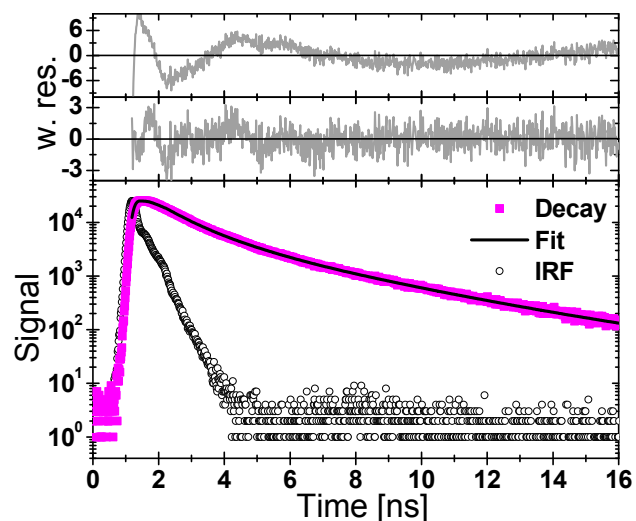


Figure 4.2. eTCSPC measurement of the fluorescence decay of Alexa488 obtained for a ds B-DNA FRET sample with C6 linker. Experimental data (magenta filled squares), instrument response function (IRF, black open circles) and the fit assuming a Gaussian distribution of distances (see Section 5.3.3.1; bottom panel, solid black line) are shown. Weighted residuals are presented in the upper plots; (top): formal biexponential fluorescence decay with a single FRET state corresponding to $\tau_{D(A)} = 1.0\text{ns}$ (87.5%), and a Donor-only decay with $\tau_{D(0)} = 4.1\text{ns}$ (12.5%); $\chi_r^2 = 10.3$; (middle): Gaussian distribution of distances ($\langle R_{DA} \rangle = 39.5 \text{ \AA}$, $\sigma_{DA} = 7.5 \text{ \AA}$, 93.3 %) and Donor-only decay (see Eq. 5.8) with $\tau_{D(0)} = 4.1\text{ns}$ (6.7%); $\chi_r^2 = 1.35$. The fit ranges from the maximum of the IRF to the first time channel with less than 100 detected photons.

4.1.4 AV accurately predicts dye position distributions for nucleic acids

We demonstrate that, for the cases when the local structure in the proximity of the dye is known, an easily applicable accessible volume (AV) simulation method (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008) provides realistic dye position distributions. The use of AV modeling, which does not take into account interactions between the dye and the macromolecule other than sterical ones, is justified for L linkers as consequence of Section 4.1.1 (no significant quenching) and Section 4.1.2 (unrestricted movement through the sterically accessible volume). Furthermore, the resulting mean dye positions and position distribution widths are consistent with MD simulation data (see Table 3 in “linker paper” and Figure 4.3A) and the resulting distance distribution widths are similar to those found for the eTCSPC data (see Section 4.1.3). See Section 5.3.1 for further details on AV modeling.

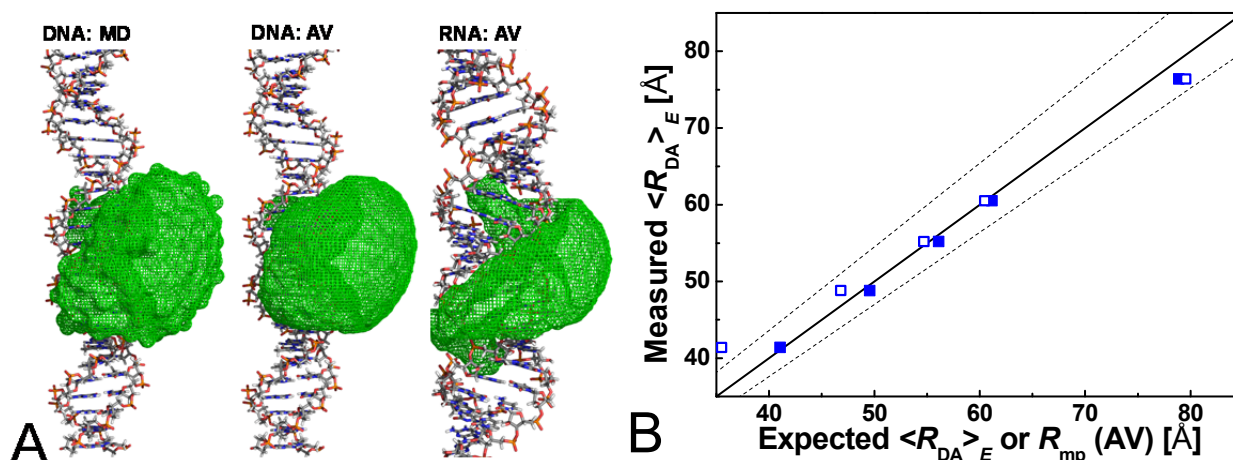


Figure 4.3. (A) Molecular dynamics (left) and AV (middle, right) simulations of possible positions of Alexa488 attached to a ds B-DNA (left, middle) or ds A-RNA (right) via a C6 linker (L). The structures were rendered via PyMOL (DeLano 2002). (B) DA-distances $\langle R_{DA} \rangle_E$ measured by smFRET (compiled also in Table S2 in the SI of “linker paper”) plotted versus expected (simulated using AV) distances $\langle R_{DA} \rangle_E$ (filled squares) and R_{mp} (open squares). The following polynomial approximation of a $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ -conversion function was used: $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$ (see Section 5.3.12 for further details). The solid line represents equal expected and measured distances. The statistical experimental errors are smaller than the symbol size. The dashed lines represent the expected uncertainties due to the distributions of possible values for κ^2 (Section 3.5 in “linker paper” and Section 5.3.9) given by the typical precision of 5.1% in Table 5 in “linker paper”.

4.1.5 Benchmark study for dsRNA

In a benchmark study five distances along a dsRNA were measured via smFRET. Additionally, AV modeling was used to predict those distances. It became clear that experimentally measured distances, especially for L linkers, cannot be directly compared with the structurally relevant distance between the mean positions of the dyes R_{mp} (see Figure 4.3B) or the distance between the attachment points of the dye linkers (see Figure 1.1). Because of broad distributions of DA-distances there is a large discrepancy between R_{mp} and the FRET-averaged mean distance $\langle R_{DA} \rangle_E$ (measured by smFRET) or mean distance $\langle R_{DA} \rangle$ (measured by eTCSPC (see Section 5.3.2 for the definitions of R_{mp} , $\langle R_{DA} \rangle_E$ and $\langle R_{DA} \rangle$)), which must be always taken into account. When doing so, high precision distance predictions are possible. Comparison of the five measured and expected distances $\langle R_{DA} \rangle_E$ yields an RMSD value of 1.3 Å (see Figure 4.3B). The correction for differently averaged distances ($R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$) and, therefore, modeling of dye positions, is less important for linkers I, S_d and S_t (see Table 4 in “linker paper”). This makes them particularly suitable for undefined environments where AV modeling is not possible (see also Section 4.1.7).

4.1.6 Estimation of errors due to orientational uncertainties (κ^2)

A rigorous procedure to minimize the uncertainties in the orientation factor κ^2 is introduced (see Section 3.5 in “linker paper”). This is necessary because orientational distribution of both D and A is not strictly isotropic even for long C6 linkers (see Section 4.1.2). We used time-resolved fluorescence anisotropy decays of D-only and A-only molecules and the FRET-sensitized acceptor anisotropy decay to determine the residual anisotropies, which allow us to compute probability distributions for possible values of κ^2 (see Figure 3.6 and Section 5.3.9) and estimate errors due to uncertainties in the relative orientation (Table 5 in “linker paper”). It is shown that for short and intermediate linkers κ^2 -related errors are only slightly higher than for long linkers. When using them for quantitative FRET measurements of internally labeled nucleic acids with Alexa488 and Cy5 as a FRET pair it is, therefore, safe to assume κ^2 to be 2/3.

4.1.7 Short dye linkers reduce position uncertainties

We introduce short dye linkers with a propargyl (S_t) or alkenyl (S_d) unit for internal labeling of nucleic acids (see Figure 5.1) to minimize position uncertainties. We show that the nature of the linker strongly affects the radius of the dye's accessible volume (6 to 16 Å, see Table 3 in "linker paper"). The κ^2 -related errors for S_t and S_d (see Table 5 in "linker paper") indicate that, for short linkers and unknown local environments, additional κ^2 -related uncertainties are clearly outweighed by better defined dye positions. However, due to the inhomogeneities of DA distances observable for S_t as a linker for Alexa488 (see Figure 7 in "linker paper"), we advise against its use and recommend the S_d or the I linker instead.

4.1.8 Dye diffusion due to linker movement: additional unpublished data

In this section additional unpublished data is presented which gives further insight into the features of translational dye linker diffusion for long C6 linkers (L). The data shown in this section is not included in "Linker paper", "RT manuscript" or "RNA4WJ manuscript", but is part of a manuscript in preparation for publishing.

4.1.8.1 Dye diffusion measured through fluorescence correlation spectroscopy (FCS)

We studied the FRET labeled RNA2(22-) (see Section 5.1.1.3) with L linkers and a DA separation of 22bp ($R_{mp} = 54$ Å) by full FCS with cw excitation (Felekyan, Kühnemuth et al. 2005). RNA2(22-) has a FRET efficiency of $E \sim 0.4$, which is nearly optimal for detection of FRET fluctuations, as for $E = 0.5$ ($R_{DA} = R_0$) the contrast is highest for changes in R_{DA} (see Eq. 2.4 and Figure 2.1B). We studied the FRET sensitized acceptor signal, because it has the highest relative signal changes due to linker dynamics (from $E_{(-)} = 0.72$ to $E_{(+)} = 0.23$ estimated for $R_{mp} - \sigma_{DA}$ and $R_{mp} + \sigma_{DA}$, respectively (see Tables S2 in the SI and 4 in the main text of "linker paper"). In Figure 4.4A, the red-red correlation curves are presented for FRET (green) and red cw excitation of the acceptor in RNA2(22-). Besides the diffusion, cis-trans isomerization and triplet terms the FCS curve obtained using FRET excitation (green line) clearly exhibits an extra

bunching term in the 10-100 ns range. The absence of a comparable term in the FCS curve obtained using direct excitation of Cy5 (red line in Figure 4.4A) clearly indicates FRET dynamics, which is most likely due to linker flexibility. The fastest bunching term observed in both curves at ~ 8 ns is the only polarization-sensitive one (see Figure 4.4B) and thus represents the global rotation of RNA2 (Kask, Piksarv et al. 1989; Eggeling, Schaffer et al. 2001). It is also in perfect agreement with time resolved anisotropy data (see Table 2 in “linker paper”). The curves in Figure 4.4B were measured at a 10-fold higher power than the direct excitation curve in Figure 4.4A to reduce noise.

Fitting of the FCS curves reveals a relaxation time of $t_{\text{linker}} \sim 150$ ns (see Figure 4.4A), which is specific for the case of FRET excitation (for details and all fit parameters, see Section 5.3.10). Thus, a reasonable estimation of the rate constant for translational diffusion within the accessible volume $k_d \sim 10^7 \text{ s}^{-1}$ can be obtained. For quantitative FRET measurements, the main consequence of this result is that DA distance distributions due to linker dynamics are not significantly averaged out on the timescale of FRET (few ns). This conclusion is also supported by the eTCSPC data as discussed in Section 4.1.3. Moreover, it is safe to say that L linkers are well suited to measure macromolecular dynamics in the range of μs and slower.

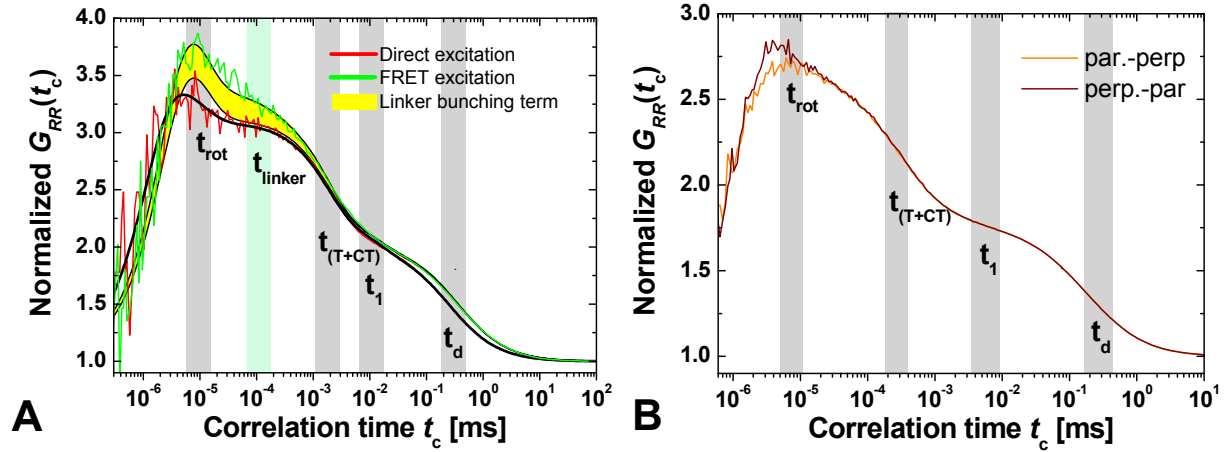


Figure 4.4. (A) FCS curves obtained for the DA labeled RNA2(22-) sample with L linkers using green (i.e. via FRET; green line) or red (red line) cw excitation. The curves are averages of the curves calculated in parallel-perpendicular and in perpendicular-parallel directions. The fits are shown as the black solid lines, respectively, with the fit parameters given in Section 5.3.10. The grey areas visualize the time regimes of the bunching terms for both curves (t_d : diffusion; t_1 : bunching term due to unknown quenching process; $t_{(T+CT)}$: cis-trans isomerization of Cy5 and triplet; and t_{rot} : rotational diffusion). The light green area visualizes the time regime of the bunching term due to linker dynamics in the FRET excitation curve (t_{linker}). The yellow area represents the difference between the fit for the green curve (FRET excitation) and the same curve without the bunching term due to linker dynamics. All curves and fits are normalized for the average number of molecules in the focus to be one ($N = 1$). (B) cross-correlation of the polarized components, calculated in parallel-perpendicular (orange line) and in perpendicular-parallel directions (dark-red line) for direct excitation measured at a 10-fold higher power than the direct excitation curve in (A) to reduce noise. The grey areas visualize the time regimes of the bunching terms (t_d : diffusion, t_1 : bunching term due to an unknown quenching process, $t_{(T+CT)}$: cis-trans isomerization of Cy5 and triplet relaxation and t_{rot} : rotational diffusion).

It is very likely that the obtained translational diffusion rate represents mainly the diffusion along the RNA groove (see Figure 4.1A) because only this motion can be associated with significant changes in R_{DA} and thus E . On the other hand, the anisotropy measurements provide much faster rotational relaxation rates $k_R = 1/\rho_1$ (see Table 2 in "linker paper"), describing fast local reorientations and rotations about the linker in the case of Alexa488, whereby the whole dye's AV is not explored (Figure 4.1A). Thus, it is absolutely reasonable that $k_R \gg k_d$. In summary, we believe that dye motions are best described as shown in Figure 4.1B, that is, by a superposition of fast local fluctuations (largely responsible for the fluorescence depolarization with the rate k_R) and much slower diffusion along the DNA or RNA groove with the rate k_d .

A crude estimation of the diffusion coefficients of the dyes (D_{dye}) can be obtained by assuming that during $t_{linker} = 150$ ns, the dyes explore their accessible volumes with the characteristic width of $2\sigma_{DA}$ (see Table 3 in "linker paper"). From $(2\sigma_{DA})^2 = 2D_{dye}t_{linker}$ (Atkins and de Paula 2001) one calculates $D_{dye} \cong 2 \times 10^{-7} \text{ cm}^2/\text{s}$ which is about an order of magnitude slower than diffusion of comparable free dyes

in water (i.e. $2.9 \times 10^{-6} \text{ cm}^2/\text{s}$ for Rhodamine 6G (Magde, Elson et al. 1974)). Strictly speaking, t_{linker} reflects simultaneous diffusion of both dyes which has been neglected in the above estimation.

4.1.8.2 Dye diffusion modeled by MD simulations

Additionally, an all-atom molecular dynamics (MD) simulation in explicit solvent (Case, Cheatham et al. 2005) for FRET labeled RNA2(22-) (see Section 5.1.1.3), L linkers and Alexa488 and Cy5 and D and A, respectively, was performed to model linker diffusion. A trajectory of 550 ns was analyzed and the autocorrelation of the distance between the centers of mass of the dyes (R_{DA}) was calculated (see Figure 4.5). Fitting with a bi-exponential decay ($y(t_c) = y_0 + A_1 \exp(-t_c/t_1) + A_2 \exp(-t_c/t_2)$) yields a dominant relaxation time $t_2 \sim 60 \text{ ns}$. This is in the same time scale as the experimentally determined relaxation time due to FRET of 150 ns (see Section 4.1.8.1) and further confirms the conclusion that the translational diffusion of the dye through its accessible volume is slow with respect to FRET.

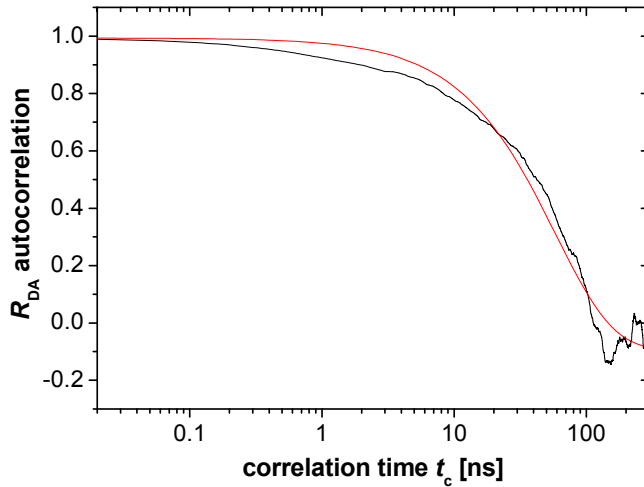


Figure 4.5. Autocorrelation (black line) of the distances between the centers of mass of the dyes (R_{DA}) for Alexa488 and Cy5 coupled to dsRNA (RNA2(22-)) via long C6 linkers modeled by MD simulations. Bi - exponential fit (red line, $y(t_c) = y_0 + A_1 \exp(-t_c/t_1) + A_2 \exp(-t_c/t_2)$) with $y_0 = -0.091$, $A_1 = 6.7 \times 10^{-3}$, $t_1 = 7.02 \times 10^{-4} \text{ ns}$, $A_2 = 1.08$, $t_2 = 58.85 \text{ ns}$.

4.2 FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang

In this work a comprehensive set of tools for FRET-restrained modeling of biomolecules and their complexes is presented (see Figure 3.1). For validation of this approach, we studied the human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT). HIV-1 RT is responsible for transcription of viral RNA into double stranded DNA (Liu, Abbondanzieri et al. 2008; Goette, Rausch et al. 2010). We performed a smFRET study of RT complexed with a 19/35 DNA/DNA primer/template (dp/dt) (see Figure 4.6A and B) for which the protein and the DNA were labeled at eight positions with Alexa488 and at five positions with Cy5, respectively (see Figure 4.6A and Section 5.1.2). The crystal structure of RT in complex with dsDNA has been determined (e.g., ref. (Peletskaya, Kogon et al. 2004)), however, so far, the ssDNA template overhang was not resolved by X-ray crystallography. There is biochemical evidence that the properly bound template overhang plays an important role in translocation of nucleic acid during processive DNA synthesis (Goette, Rausch et al. 2010). It is shown that FRET restrained high-precision structural modeling dramatically improves the precision of FRET-derived structures through explicit consideration of spatial distributions of dye positions, which greatly reduces uncertainties due to flexible dye linkers. Furthermore, it includes the calculation of precision and confidence levels of the models by rigorous error estimation. The individual steps of FRET restrained high-precision structural modeling are depicted in Figure 3.1 and are discussed in detail in Section 3.

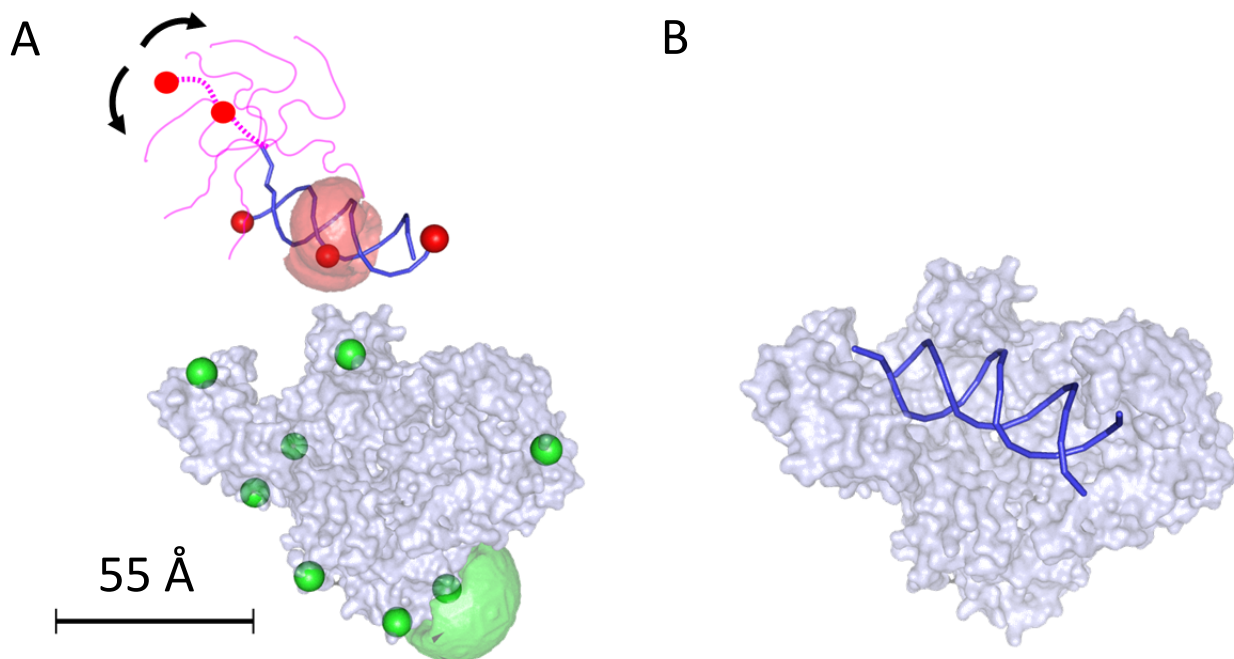


Figure 4.6 (A) Crystal structures of the protein and the (ds)DNA (both from PDB-ID 1R0A). The donor labeling positions (green spheres) are named according to the RT subunit (p51 or p66) and the position of the introduced cysteine mutation to which they were coupled: p51Q6C, p51K173C, p51E194C, p51K281C, p66Q6C, p66T27C, p66E194C, and p66K287C. Five acceptor labeling positions on the primer/template DNA (red spheres) are named according to strand (dp or dt) and the position of the labeled nucleotide. The last paired nucleotides are referred to as position 1. For labeled positions on the template overhang the position is negative with regard to position 1, and for the primer coupled fluorophores the number is positive. Thus, the labeled dp/dt complexes are named dp(1)/dt, dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15) (see Section 5.1.2). As an example, FRET-pairs are named RT(p51E194C):dp/dt(-15). AV clouds are shown for labeling position p51K173C (green) and dp(10) (red). Parameters used for generation of AV clouds: for donor positions: $L_{\text{link}} = 20 \text{ \AA}$, $w_{\text{link}} = 4.5 \text{ \AA}$, $R_{\text{dye}} = 3.5 \text{ \AA}$, for acceptor positions: $L_{\text{link}} = 23 \text{ \AA}$ for dp(10), dp(19), dt(-6) and dt(-15), $L_{\text{link}} = 8.5 \text{ \AA}$ for dp(1), $w_{\text{link}} = 4.5 \text{ \AA}$, $R_{\text{dye}(1)} = 11 \text{ \AA}$, $R_{\text{dye}(2)} = 3 \text{ \AA}$, $R_{\text{dye}(1)} = 1.5 \text{ \AA}$; (C) Structural model of the RT:dp/dt complex obtained by rigid body docking. The structures were rendered via PyMOL (DeLano 2002).

4.2.1 Recovering the position of the rigid dsDNA through FRET restrained rigid body docking

The accuracy of the approach is demonstrated by docking the double-stranded part of dp/pt to RT. As prior knowledge we used the crystal structure of the RT:dp/pt complex from (Peletskaya, Kogon et al. 2004) (PDB-ID: 1R0A, dp/pt without single-stranded overhang), separated dp/dt from the protein and considered them as rigid bodies (starting model, see Section “Step 1” in the “RT manuscript”). 20 distances were measured via smMFD (see Section “Step 3.1” in “RT manuscript”) between the labeling positions on the protein and those on the dsDNA (see Figure 4.6) and their uncertainties were determined (see Section “Step 3.2” in “RT manuscript”). The possible mutual arrangements in

agreement with FRET data were discovered by repeated rigid-body docking of the substructures (see Section “Step 4” in “RT manuscript” and Section 5.3.11). Cluster analysis shows that the solution which agrees best with the FRET data is unique with a confidence of 94 % (see Section “Step 5” in “RT manuscript” and Section 5.3.13) and deviates from the crystal structure by only 0.5 Å (RMSD). This is well within the precision of the model of 2.4 Å which was determined by bootstrapping (see Figure 4.7 and Section “Step 6” in “RT manuscript”) and is smaller by one order of magnitude compared to the uncertainties due to linker effects (~ 20 Å).

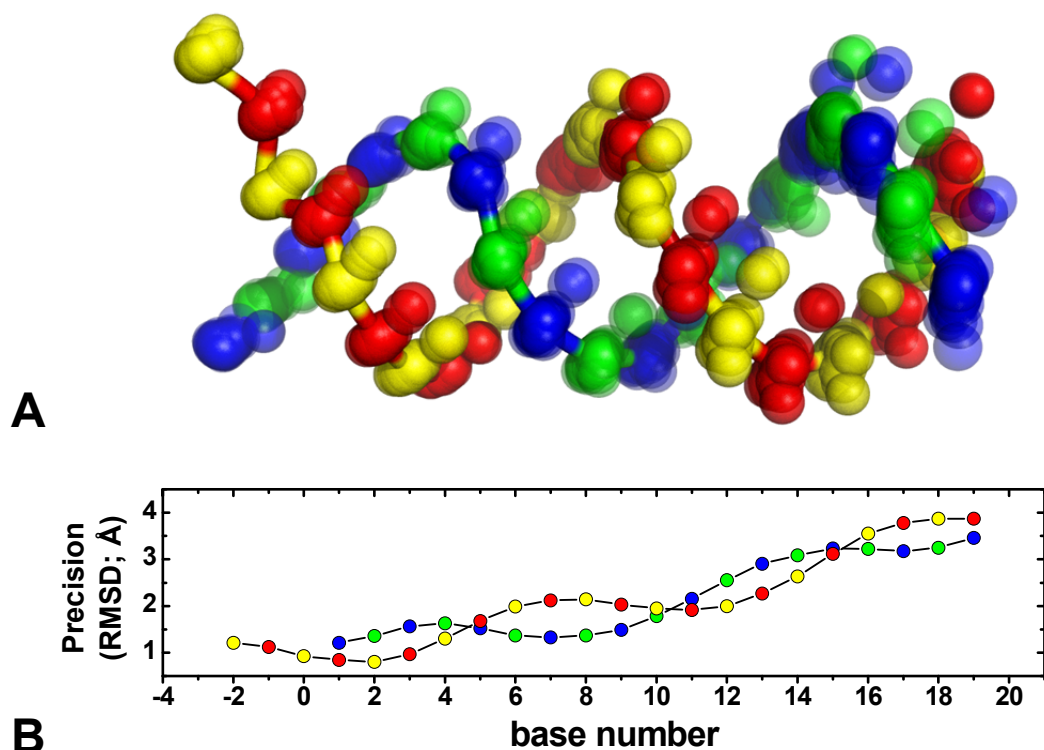


Figure 4.7. Bootstrapping of docked dsDNA dp/dt. (A): Ensemble of structures generated with perturbed distances (bootstrapping, see Section 3.8) for 1 Å clash tolerance. For better comparison the phosphate atoms of the DNA-backbone are colored alternately yellow/red or blue/green for the primer and template strands, respectively. The DNA is oriented as in the figure below. (B): Uncertainty of phosphate atom positions calculated for each P atom using Eq. 3.4 after the refinement steps with 1 Å (circles) clash tolerances. The average RMSD value for all phosphate atoms is 2.4 Å.

4.2.2 Structure determination of the single-stranded template overhang through FRET guided screening of a structural ensemble

The formerly unknown configuration of the flexible single stranded template overhang was determined by FRET-guided “screening” of a large structural ensemble created by molecular dynamics simulations.

We generated a starting model by attaching the single strand to the crystalized DNA such that it sticks out straight from the protein (see Section “Step 1” in “RT manuscript”). A large number of putative structures is then generated by extracting conformations from an MD trajectory (all-atom MD with explicit solvent (Case, Cheatham et al. 2005), see Section S3.4.3 in the SI of the “RT manuscript”). Afterwards the results are filtered with respect to agreement with FRET data and good stereochemical quality (see Section “Extension to flexible parts of the complex: ...” in the “RT manuscript”). Figure 4.8A shows the positions of the N1 atom of the last nucleobase (orange dots) resulting from ten MD trajectories (2855 conformations). They were filtered using 16 distances for which the acceptor dye was attached to the template overhang. The structure that fits best to the FRET data is depicted in magenta together with the approximate 1σ confidence interval indicated by green dots (150 conformations, see Figure 4.8A). The conformational ensemble satisfying FRET data is represented by three major configurations (bold) in Figure 4.8B preferentially located in positively charged regions. The green isopleths for the N1 atom determined by MD simulations (Figure 4.8C) illustrate good agreement between these regions and the structures satisfying FRET restraints. The structures wind around the fingers domain with the 5'-end binding to RT in a loop structure within positively charged protein regions. This finding might have important implications concerning proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis.

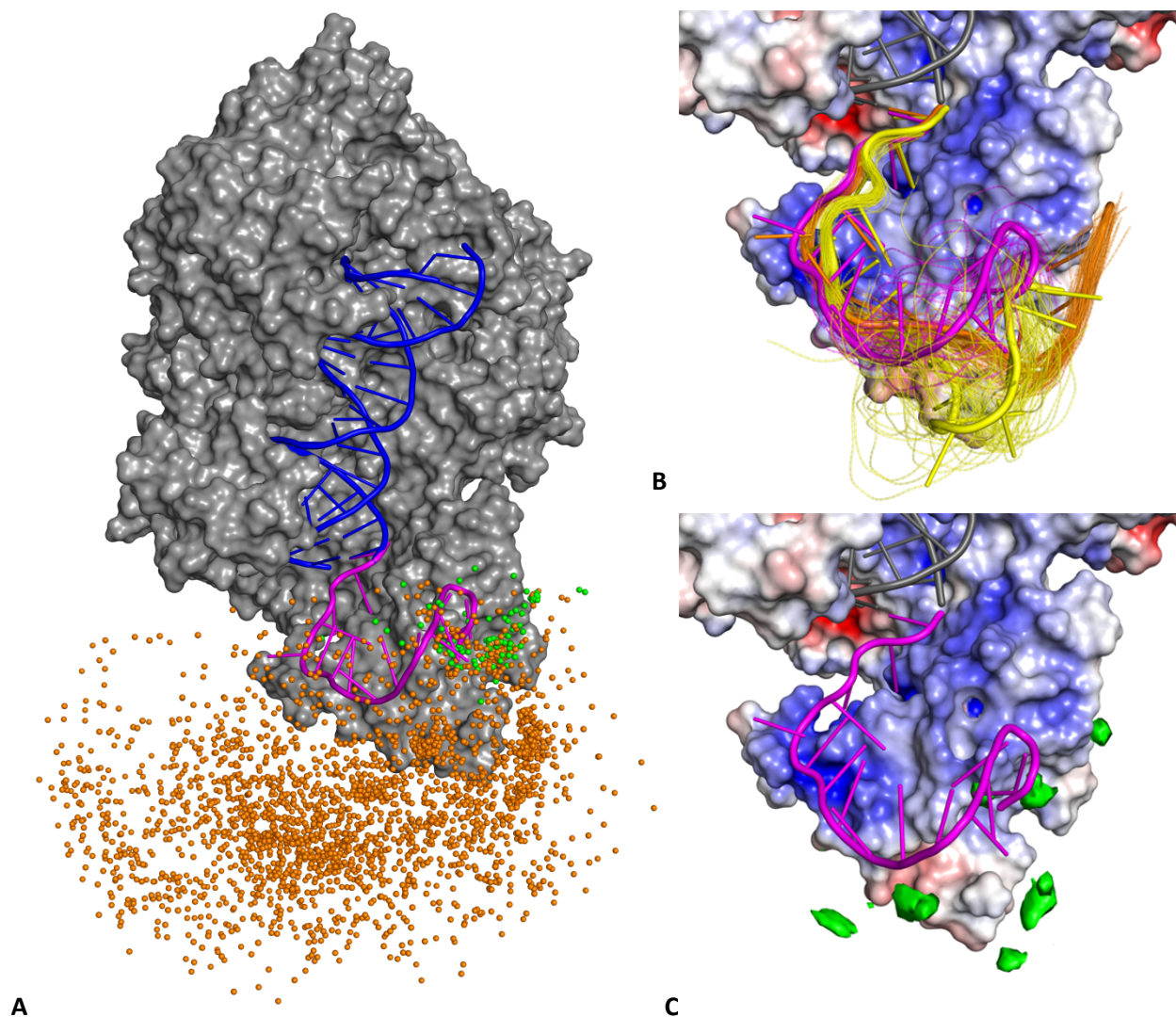


Figure 4.8. Structure of ssDNA dp/dt overhang obtained by screening of MD trajectories. (A) The ensemble of ssDNA overhang structures generated by all-atom MD simulations in explicit solvent (orange – 2855 conformations) and conformations filtered by smTCSPC using a confidence interval of 1σ (green – 150 conformations). Dots represent the N1 atom of the nucleobase of the nucleotide dt(-15). The structure that best fits to the FRET data is shown as a magenta cartoon ($\chi^2=0.88$). (B) Conformations within a confidence interval of 1σ of the smTCSPC data. The electrostatic potential of the protein as determined by APBS has been mapped onto the protein surface using a color scale (dark blue: $6 \text{ k}_B T/e_c$; dark red: $-6 \text{ k}_B T/e_c$). (C) Green isopleths show regions of preferred residence of the N1 atom of the nucleobase of nucleotide dt(-15) as determined from MD simulations. The isopleths encompass regions with at least 2% of the maximal residence likelihood. Note the overall good agreement between these regions and the location of the 1σ confidence interval determined by smTCSPC (panel B). The structures were rendered via PyMOL (DeLano 2002).

4.3 Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling

In this work FRET restrained high-precision structural modeling is applied to an RNA four-way junction (RNA4WJ) based on the hairpin ribozyme (see Figure 4.9). Helical junctions are an essential structural motif in many non-coding RNAs (e.g. spliceosomes (Duckett, Murchie et al. 1995) and ribosomal RNA (Shen and Hagerman 1994; Orr, Hagerman et al. 1998)). However, there are no known structural models of RNA4WJs which are determined entirely by the junction itself and not e.g. by interactions between internal loops or with proteins. The effects caused by sole staking interactions are unknown. The RNA4WJ used here is known to exhibit structural and dynamic heterogeneities (Hohng, Wilson et al. 2004). Furthermore, none of its conformers have been structurally characterized. Transient minor conformers have never been structurally characterized by FRET.

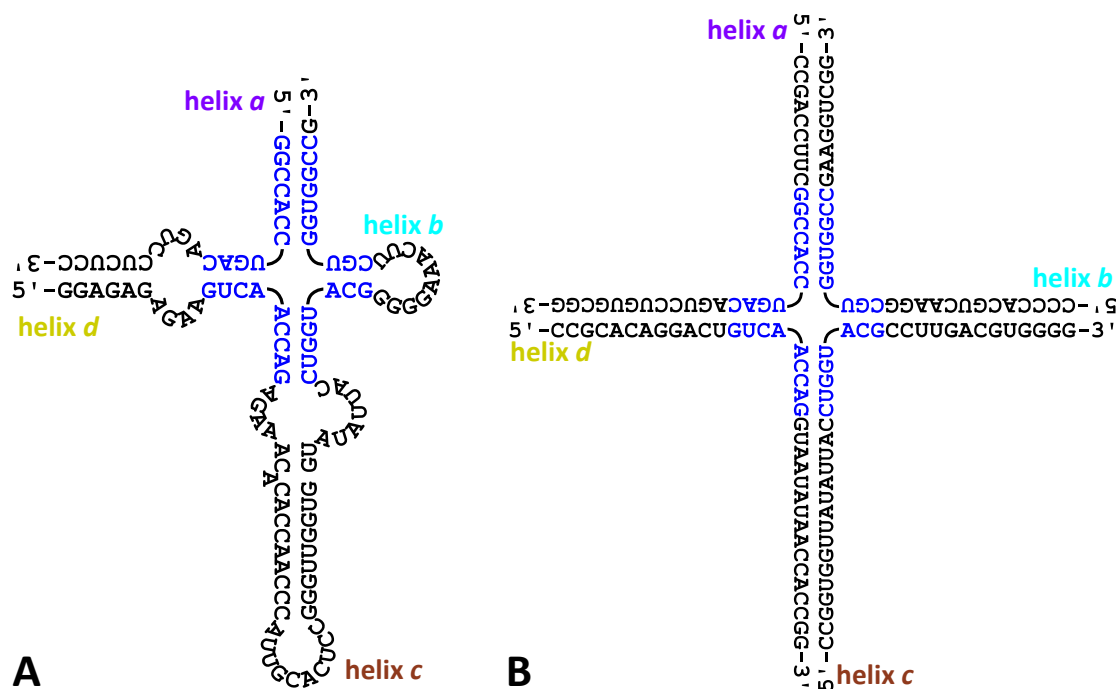


Figure 4.9. Secondary structures of (A) the hairpin ribozyme (Rupert, Massey et al. 2002) (PDB-ID: 1M5K) and (B) the RNA four-way junction used in this work. The matching base pairs are depicted in blue.

First, for an accurate determination of R_{DA} (see Eq. 5.15), the quantum yields Φ_{FA} for all A labeling positions (see Figure 3.2) were measured via eTCSPC (see Section 4.1 in the SI of “RNA4WJ manuscript” and Appendix IV for all eTCSPC data sets). Second, 51 different donor-acceptor-pairs (see Figure 3.2) were measured using single-molecule multi-parameter fluorescence detection (smMFD, see Appendix IV for all measurement data sets). The single-molecule approach allowed for the simultaneous extraction of up to three distances (and their corresponding errors) via PDA (see Sections 3.4 and 3.7 in the SI of “RNA4WJ manuscript” and Appendix IV for all PDA plots). For approximately half of the 51 datasets three FRET states (one major and two minor states) were necessary to reach a satisfying fit quality (see Figure 3.5B and Figure 3.6A). This strongly suggests the presence of at least 3 quasi-static conformers in equilibrium, which has been suggested by previous studies (Hohng, Wilson et al. 2004). For the other half, two states were sufficient (one major and one minor). In these cases, the second “hidden” minor state is overlapped by one of the two visible ones.

4.3.1 Assignment of the minor FRET states to the corresponding conformers

Using all 51 distances from the major FRET populations to model a structure yields conformer $(ad)_a$ (see Figure 4.10A and Figure 4.11A). Considering there are three conformers present, we, therefore, conclude that, out of the four possible conformers for stacked helix pairs (see Figure 4.10A), the ones present in equilibrium are $(ad)_a$ (major), $(ad)_p$ and $(ab)_a$ (minors). The presence of conformer $(ab)_p$ can neither be confirmed nor excluded as for every combination of labeled helix pair, its FRET populations could overlap with those of the other three (see Figure 4.10A). Because of similar equilibrium fractions of the two minor FRET states, they could not be reliably assigned to the corresponding conformers by their amplitudes. Therefore, their distinct Mg^{2+} -affinities were used for the assignment. Upon lowering of the Mg^{2+} -concentration one minor peak significantly decreases in amplitude, whereas, the other one remains stable (see Figure 4.10C and D). This observation can be attributed to fast species interconversion at low Mg^{2+} (Buck, Noeske et al. 2010; Buck, Wacker et al. 2011). A possible kinetic scheme is shown in Figure 4.10B. Considering the time resolution of our experiment, FRET peaks due to species with sub-ms lifetimes cannot be resolved and apparently disappear. Thus, minor peaks can be unambiguously assigned according to their behavior at low Mg^{2+} , which is in principle sufficient for structural modeling. Two exemplary Mg^{2+} titrations are presented in Figure 4.10C and D. At low Mg^{2+} -concentrations, the minor high FRET peak slightly increases in amplitude for $(D)\beta11c_{(A)}\gamma8b$ (blue line in

Figure 4.10D) and significantly decreases for $(D)\beta 27b_{-}(A)\delta 28d$ (green line in Figure 4.10D). Referring to the geometric model of RNA4WJ (Figure 4.10A), the “stable” and “unstable” species resemble $(ab)_a$ and $(ad)_p$ conformers, respectively, which can be shown as follows. The model suggests that, in most of cases, the DA distances should be determined mainly by the angle between the helical axes. For instance, for labeling on helices b and d the minor state yielding the smaller distance can be assigned to $(ad)_p$ (Figure 4.10A). Thus, in the case of $(D)\beta 27b_{-}(A)\delta 28d$ (green FRET state in the upper panel of Figure 4.10C) the minor high FRET peak is due to the $(ad)_p$ state, and for $(D)\beta 11c_{-}(A)\gamma 8b$ (blue FRET state in the lower panel of Figure 4.10C) it is due to $(ab)_a$. 14 titrations were performed which all confirmed the above assignment, that is, the minor FRET peak that disappears at low Mg^{2+} is consistent with the assumed geometry of $(ad)_p$. In addition, Mg^{2+} titrations were performed for all samples with $(A)\alpha 12d$ which is close to the junction, and all samples with labeled guanines $((A)\gamma 24a$ and $(A)\delta 10a)$ where the dye’s mean position is far away from the RNA’s helical axis (see Section 3.2 in the SI of RNA4WJ manuscript and Figure 3.4). In other 19 cases the minor peaks were assigned based on the geometric model (Figure 4.10A). All PDA results are summarized in Tables 4-4 (20 mM $MgCl_2$) and 4-5 (Mg^{2+} - titrations) in the SI of “RNA4WJ manuscript”.

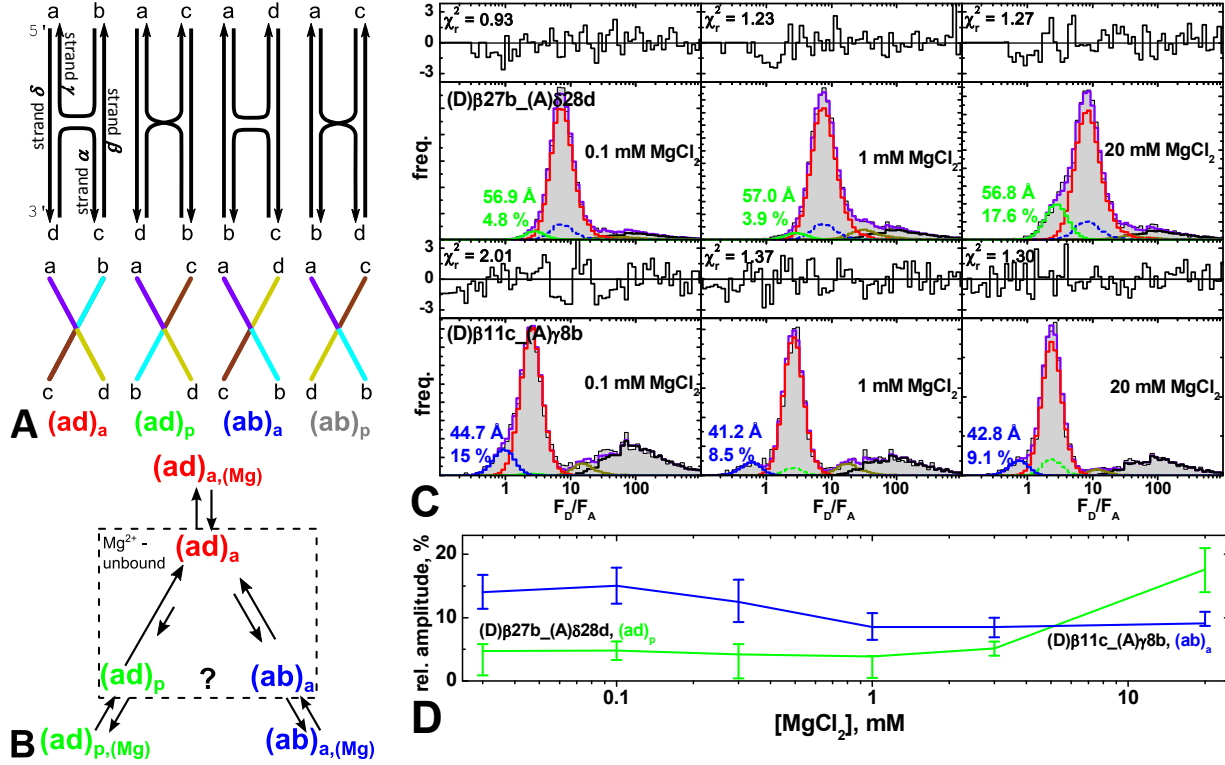


Figure 4.10. (A) Four possible conformers of a four-way junction of nucleic acids for the case of stacked helix pairs. (B) Possible transitions between the three conformers and their respective Mg²⁺-bound and unbound states. (C) PDAs (selected bursts) of Mg²⁺-titrations (0.1 mM, left, 1 mM, middle and 20 mM, right) for samples (D)β27b_(A)δ28d (upper panel) and (D)β11c_(A)γ8b (lower panel). F_D/F_A histograms of experimental datasets (gray areas) are fitted (purple solid lines) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, green for (ad)_a, $\langle R_{DA} \rangle_{E(2)}$, red for (ad)_p and $\langle R_{DA} \rangle_{E(3)}$, blue for (ab)_a), one D-only (black) and one impurity state (dark yellow) (see Table 4-5 in the SI of “RNA4WJ manuscript” for all PDA fit parameters). Weighted residuals are shown in the upper plots, respectively. (D) Relative FRET amplitudes of the state assigned to conformer (ad)_p for (D)β27b_(A)δ28d (green line) and the state assigned to (ab)_a for (D)β11c_(A)γ8b (blue line) measured at various concentrations of MgCl₂.

4.3.2 Rigid body docking and model validation

To generate 3D structural models accurately describing the orientations and positions the junction’s four helices, we, initially, treat them as rigid bodies having perfect A-RNA form. This simplification is justified due to Watson-Crick base pairing throughout the whole molecule. Thus, significant deviations from A-RNA structure are expected only within close proximity to the junction. Additionally to the 51 FRET distance restraints, four strong constraints were introduced representing chemical bonds between neighboring helices at the junction. Unique rigid body models for the major and both minor conformers were obtained by docking the dsRNA helices explicitly taking dye position distributions into account (see Figure 4.4 in the SI of “RNA4WJ manuscript”). Confidence levels of > 99.99 % for (ad)_a, 84 % for (ad)_p and 97 % for (ab)_a were obtained by cluster analysis (see Figure 4.11D) and the models’ quality was

assessed through rigorous error estimation via bootstrapping (Efron 1986), providing RMSD values for all P-atoms (see Figure 4.11E). The achieved precisions ($\langle \text{RMSD} \rangle$) are 1.7 Å for the major and 2.3 and 2.6 Å for minor conformers, respectively, which is significantly better than the uncertainty of dye position with respect to the macromolecule.

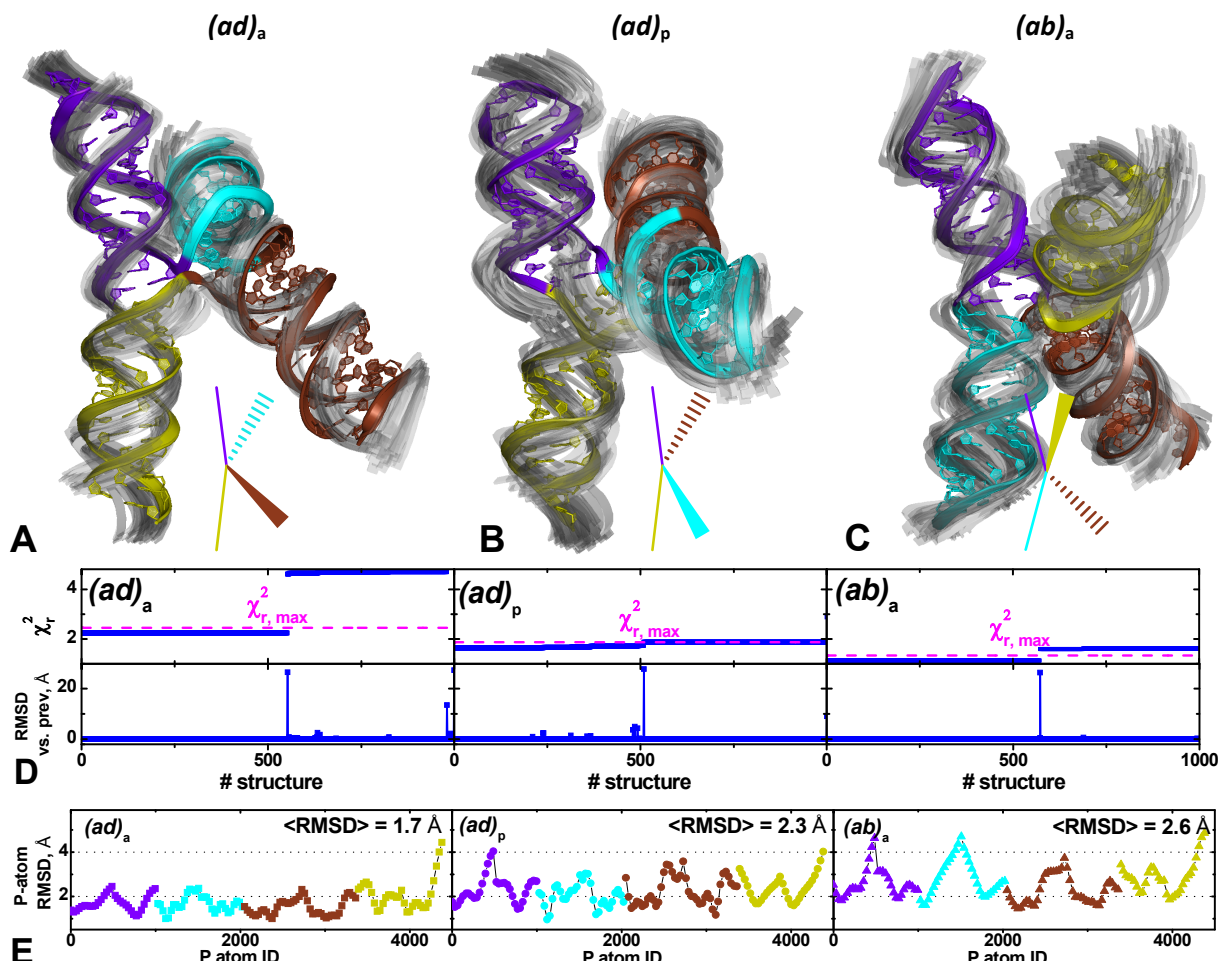


Figure 4.11. (A-C) Structures (cartoon representation) with the lowest χ_r^2 after refinement of the rigid body models by MD simulations (colored) for conformers $(ad)_a$ (A), $(ad)_p$ (B) and $(ab)_a$ (C) overlaid with 100 structures (grey transparent) indicating the uncertainties of the helix positions and orientations resulting from bootstrapping the rigid body model. The structures were rendered via PyMOL (DeLano 2002). Bottom, respectively: Sketches depicting the mutual orientation of helices *a* (purple), *b* (cyan), *c* (brown) and *d* (dark yellow). (D) χ_r^2 (upper panels) and RMSD vs. the previous structure (lower panels) plotted against the structure ID found after docking in ascending order with respect to χ_r^2 for $(ad)_a$ (left), $(ad)_p$ (middle) and $(ab)_a$ (right). The dashed magenta lines represent 84 % confidence thresholds ($\chi_{r,max}^2$, see Section 5.3.13.2): $\chi_{r,max}^2 = 2.46, 1.86$ and 1.34 for $(ad)_a$, $(ad)_p$ and $(ab)_a$, respectively. (E) uncertainty of phosphate atom positions in helices *a* (purple), *b* (cyan), *c* (brown) and *d* (dark yellow) calculated by bootstrapping for the respective solutions with the lowest χ_r^2 after docking for $(ad)_a$ (left, squares), $(ad)_p$ (middle, circles) and $(ab)_a$ (right, triangles). The average RMSD values over all P atoms are 1.7 Å, 2.3 Å, and 2.6 Å, respectively.

4.3.3 Model refinement

4.3.3.1 Molecular dynamics (MD) simulations

The three final rigid body models are used as starting structures for further refinement through all-atom molecular dynamics (MD) simulations in explicit solvent (Case, Cheatham et al. 2005) (see Section 3.12 in the SI of “RNA4WJ manuscript”). Details are described in Section “MD simulations” in “RNA4WJ manuscript” and in Section 3.9. The resulting trajectories were filtered with respect to agreement with FRET data. Figure 4.11A-C show the resulting structures with the lowest value of χ_r^2 , i.e., structures with best agreement with FRET distance restraints. Comparing χ_r^2 between the best MD and rigid body models shows that for conformers $(ad)_a$ and $(ad)_p$, χ_r^2 is significantly reduced from 1.93 and 1.57 to 1.75 and 1.40, respectively, while for $(ab)_a$ it increased from 1.06 to 1.32 (the values of χ_r^2 shown here include only violations of FRET restraints). The models of the RNA4WJ from FRET-filtered molecular dynamics simulations have both the global geometry consistent with the FRET restraints and the local stereochemistry encoded in the MD force field (see Section 3.12 in the SI of “RNA4WJ manuscript”). In particular, all bases at the junction are properly stacked after MD-refinement. The optimization of the local structure is achieved without significant violations of the global geometry as judged by comparison to respective rigid body models (see Section 4.6 in the SI). The RMSDs over all P atoms, excluding those within 6 bps distance from the junction, are 3.17 Å , 2.96 Å and 2.41 Å for $(ad)_a$, $(ad)_p$ and $(ab)_a$, respectively.

4.3.3.2 Coarse-grained modeling

As an alternative approach, a coarse grained modeling program SimRNA (Rother, Rother et al. 2012) was used for the *de novo* folding of the RNA4WJ. SimRNA uses a reduced representation of RNA (only five centers of interaction per residue), a Monte Carlo sampling scheme, and a statistical (knowledge-based) potential to estimate the energy of interactions, and as a result of these simplifications, compared to all-atom MD simulations, it is faster by a factor of ~1000. The formation of helices was enforced by including distance restraints for residues expected to form base pairs (excluding the 6 pairs in each helix closest to the junction). Furthermore, 51 distance restraints were used between the P atoms of the 18

labeled nucleotides with uncertainties according to the RMSD values calculated by bootstrapping for the respective Ps (see Figure 4E). A series of simulations were performed to minimize both the energy of interactions, and to satisfy the restraints. The resulting models were converted to full atom representations and energy-minimized. Unfortunately, at the time of submission of this thesis, not all models were ready. The resulting structure for $(ad)_a$ is shown in Figure 4.6 in the SI of “RNA4WJ manuscript”. The RMSD over all P atoms compared to the according MD model is 3.5 Å. All bases at the junction are properly stacked.

5 Methods and Materials

5.1 Materials

5.1.1 Materials for “linker paper”

5.1.1.1 Dye linkers

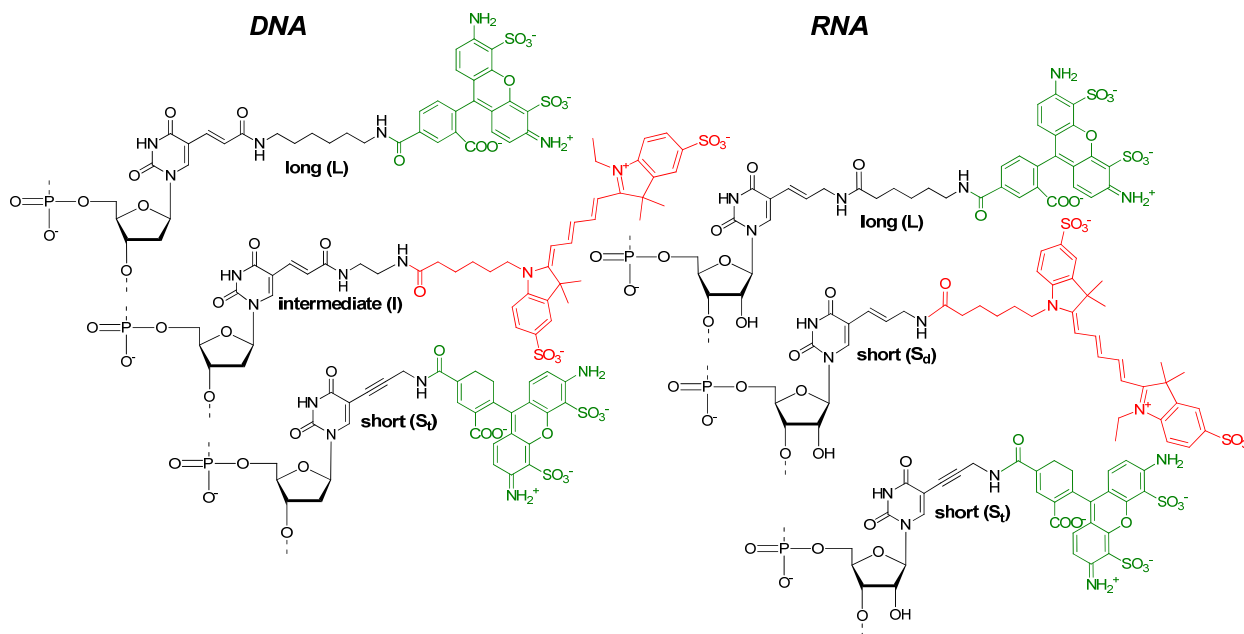


Figure 5.1. Exemplary structures of the linkers S_t (propargyl) and S_d , I and L (all alkenyl) with the dyes (Alexa488 and Cy5) used in DNA (left) and RNA (right). The donor and acceptor dyes are depicted in green and red color, respectively. Donor- and acceptor dyes have been used in combination with each of the shown linkers.

5.1.1.2 Oligonucleotides

Ultrapure labeled “DNA1” with the L linker (see Figure 5.1), “RNA2” and “RNA3” oligonucleotides (PAGE Grade), and all unlabeled counter sequences were purchased from Purimex (Grebenstein, Germany). “DNA1” oligonucleotides with the I Linker (see Figure 5.1) were purchased from IBA (Göttingen,

Germany). All sequences of the DNA and RNA strands are given in Section 5.1.1.3. The linker types L and I for DNA and L for RNA were chosen for reasons of commercial availability, the RNA linker types for reasons of chemical suitability. Throughout this work, if not stated differently, the same linker type is used at the donor (D) and the acceptor (A) position. Details on oligonucleotide synthesis (Bannwarth 1987; Schmidt, Welz et al. 2000) of samples with linkers S_t and S_d (see Figure 5.1) and RNA1 with L as well as on labeling of deoxyoligonucleotides and oligoribonucleotides with Cy5 and Alexa488 can be found in Section 2 of the main text and in Sections S1.1, S1.2 and S1.3 in the SI of “linker paper”.

5.1.1.3 Nucleic acid sequences and labeling positions

5.1.1.3.1 DNA sequences

DNA1 with different linkers L, I and S_t (10 bp separation, linkers pointing towards each other, see Figure 5C in “linker paper”), labeling positions depicted in green and red for Alexa488 and Cy5, respectively.

5' -d (ACT GAT CGT AAG CTA CTG AAG CGT A) -3'
 3' -d (TGA CTA GCA TTC GAT GAC TTC GCA T) -5'

5.1.1.3.2 RNA sequences

RNA1 with different linkers L, S_d and S_t (11 bp separation, linkers pointing towards each other, see Figure 5C in “linker paper”), labeling positions depicted in green and red for Alexa488 and Cy5, respectively.

5' -GGC GGU GCC GAC UGC GAG CUU GCC A-3'
 3' -CCG CCA CGG CUG ACG CUC GAA CGG U-5'

RNA2 and RNA3 all with linkers L, labeling positions depicted in green and red for Alexa488 and Cy5, respectively. The numbers in parenthesis represent the basepair separation between the dyes whereas the + and – signs indicate the effect of the dye displacement towards the 3'-end on the DA distance (see Figure 5C in “linker paper”).

RNA2(16-)

5' -CCG GUG GUU AUA UA CCU GGU ACG CCU UGA CGU GGG G-3'

3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA2(19-)

5' -CCG GUG GUU AUA UUA CCU GGU ACG CCU UGA CGU GGG G-3'

3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA2(22-)

5' -CCG GUG GUU AUA UUA CCU GGU ACG CCU UGA CGU GGG G-3'

3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA3(18+)

5' -CCC CAC GUC AAG GCG UGG UGG CCG AAG GUC GG-3'

3' -GGG GUG CAG UC CGC ACC ACC GGC UUC CAG CC-5'

RNA3(24+)

5' -CCC CAC GUC AAG GCG UGG UGG CCG AAG GUC GG-3'

3' -GGG GUG CAG UUC CGC ACC ACC GGC UUC CAG CC-5'

5.1.2 Materials for “RT manuscript”

5.1.2.1 HIV-RT:dp/pt complex

Sample preparation and labeling was done according to (Rothwell 2002). We used RT mutants containing a single accessible cysteine at positions 6, 27, 194 and 287 on the p66 subunit (RT(p66Q6C), RT(p66T27C), RT(p66E194C) and RT(p66K287C), respectively) and at positions 6, 173, 194 and 281 of the p51 subunit (RT(p51Q6C), RT(p51K173C), RT(p66E194C) and RT(p66K281C), respectively) (Kensch, Restle et al. 2000) (see Figure 4.6B). These cysteines were labeled with the green donor fluorophore Alexa488-C5 maleimide (Kensch, Connolly et al. 2000). The red acceptor dye Cy5 was attached to the primer

strand at positions 1, 10 and 19 (dp(1)/dt, dp(10)/dt and dp(19)/dt, respectively) or the template strand at positions -15 and -6 (dp/dt(-15) and dp/dt(-6), respectively) of a 19/35 DNA/DNA primer/template (see Table 5-1 and Figure 5.3 for sequences, labeling positions, secondary structure and numbering of dp/dt) either by internal labeling with a C6-aminolink with the NHS-ester of Cy5 (for dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15), see Figure 5.2A) or to the 3' end of the primer with the phosphoramidite derivative of Cy5 (for dp(1)/dt, see Figure 5.2B). The sequence of the primer/template is based on the HIV-1 viral primer binding site.

5.1.2.2 dt/dp sequences

Table 5-1. Sequences and labeling positions (red) of primer (dp) and template (dt) strands.

sample	sequence and labeling position
dp	5' -d (TTGTCCCTGTTTCGGGCGCC) -3'
dp(1)	5' -d (TTGTCCCTGTTTCGGGCGC <u>C</u>) -3'
dp(10)	5' -d (TTGTCCCTG <u>T</u> TCGGGCGCC) -3'
dp(19)	5' -d (<u>T</u> TTGTCCCTGTTTCGGGCGCC) -3'
dt	5' -d (TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA) -3'
dt(-6)	5' -d (GGGTTAATC <u>T</u> CTGCATGGCGCCCGAACAGGGACAA) -3'
dt(-15)	5' -d (<u>T</u> GGTTAATCTCTGCATGGCGCCCGAACAGGGACAA) -3'

5.1.2.3 Structures of modified nucleotides

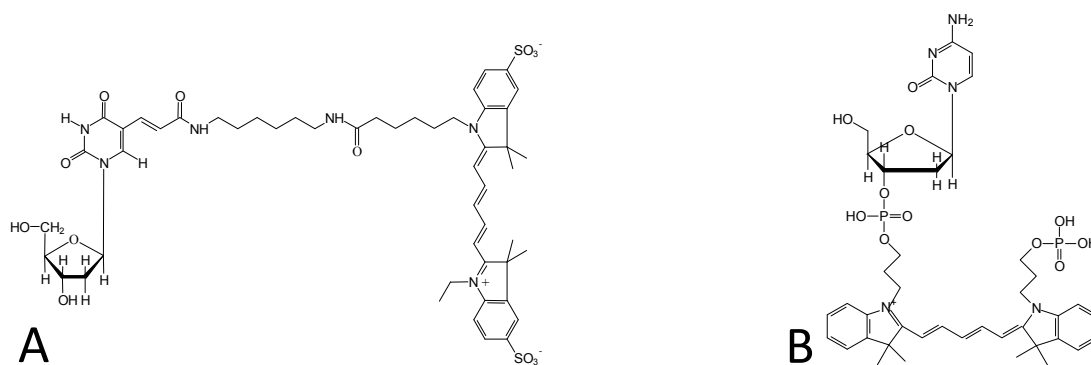


Figure 5.2. Structures of the fluorescent modified nucleotides used. (A) dThymine labeled with C6-aminolink with the NHS-ester of Cy5 used at positions dp(10), dp(19), dt(-6) and dt(-15). (B) 3' labeled dCytosine with the phosphoramidite derivative of Cy5 used for position dp(1).

5.1.2.4 Secondary structure of dp/dt

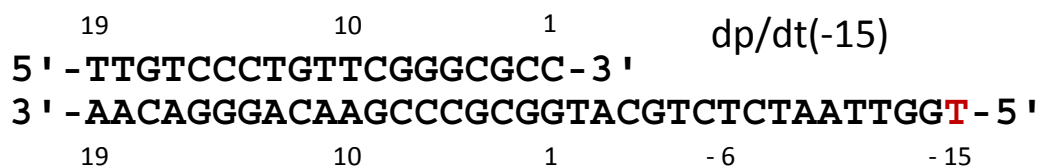


Figure 5.3. Secondary structure of DNA dp/dt(-15) with labeling position (red).

5.1.3 Materials for “RNA4WJ manuscript”

Oligonucleotides. Ultrapure labeled RNA with hexamethylen linkers (PAGE Grade), and all unlabeled sequences were purchased from Purimex (Giebenstein, Germany). See Figure 5.4 for chemical structures of modified U and dG nucleotides with linker and dye and Figure 3.2 for the sequences and labeling positions of the RNA strands.

DNA sequences. Sequences and labeling positions (green for Alexa488 and red for Cy5) for the dsDNA used for calibration of the detection efficiency ratio (see Section 5.3.7).

5'-d(GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TTA CAA CTC CGA AAT AGG CCG)-3'

5'-d(CGG CCT ATT TCG GAG TTG TAA ACA GAG ATC GCC TTA AAC GTT CGC CTA GAC TAG TCC AAG TAT TGC)-3'

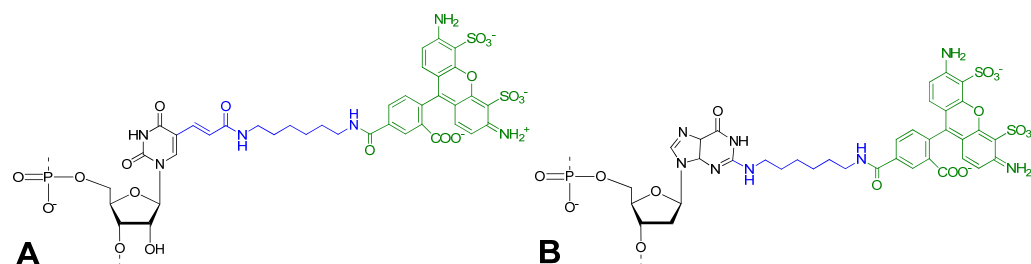


Figure 5.4. Structures of modified U (A) and dG (B) nucleotides with hexamethylen (C6) linkers (blue) and fluorescent dyes Alexa488 (green).

5.2 Experimental conditions

5.2.1 Buffer conditions for RT: dp/dt

The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris-HCl, pH 7.8, 6 mM MgCl₂, and 400 μM ascorbic acid. Because of the extremely low concentrations used in single-molecule studies (50 pM RT and 200 pM dp/dt), high binding affinities are required to ensure stable complex formation. By reducing the KCl concentration from the 50 mM used in previous studies on this system (Wohrl, Krebs et al. 1999) to 10 mM, we were able to reduce the dissociation constant, K_d , dramatically. Data from classical ensemble titrations indicated an upper limit for the K_d value at this salt concentration of <100 pM. The actual single-molecule measurements confirmed that the affinity was high enough to ensure complex formation, because most RT molecules sampled contained bound (labeled) substrate. Ensemble measurements on single nucleotide incorporation under these salt conditions showed similar kinetic behavior to that seen at higher salt, with three kinetic phases being observed (Wohrl, Krebs et al. 1999; Rothwell 2002). Before measurements, a “stock” complex solution was made at a concentration of 250 nM protein and 1 μM dp/dt. This solution was equilibrated on ice for at least 30 min. Immediately

before a measurement, the stock solution was diluted rapidly into the measurement buffer 1:10, followed by a further 1:10 dilution before a final 1:50 dilution to a final concentration of 50 pM protein and 200 pM dp/dt. A droplet of the sample solution (50 μ l) was applied to a coverslip which formed the bottom of a closed chamber with a water-saturated atmosphere. Coating the coverslip with κ -casein led to complete suppression of adsorption of the sample molecules during the measurement time of up to an hour.

5.2.2 Buffer conditions for nucleic acids

5.2.2.1 Hybridization of dsDNA and dsRNA.

The hybridization buffer for the DNA samples contained 20 mM TRIS, 100 mM NaCl and 10 mM MgCl_2 , pH 7.5. For RNA it contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl, and 10 mM MgCl_2 , pH 6.5. The concentration of the DNA or RNA molecules in the buffer ranged between 2 and 10 μ M. For FRET molecules the ratio between the amount of acceptor and donor strand ranged from 1 to 2. For donor and acceptor only molecules the ratio of unlabeled to labeled strand was 3 to 1. The solution was heated up to a temperature of 90°C inside a water bath and was then allowed to cool down to rt. overnight.

5.2.2.2 Hybridization of the RNA four-way junction.

The hybridization buffer contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl, and 20 mM MgCl_2 , pH 6.5. The concentration of donor-labeled (or acceptor labeled for A-only samples) RNA single strand in the buffer ranged between 0.5 and 2 μ M. For FRET molecules the ratio between the amount of donor, acceptor and the two unlabeled strands was 1 : 3 : 4 : 4. For donor and acceptor only molecules the ratio of labeled to the three unlabeled strands was 1 : 4 : 4 : 4. The solution was heated up to a temperature of 85°C inside a thermo-cycler (primus 96 advanced, peqLab, Erlangen, Germany) with 0.1°C/s and was then cooled down to 27°C with 2°C/h. It was then quickly cooled down to 4°C.

5.2.2.3 Measurement buffer

The measurement buffer contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl and between 20 and 0.03 mM MgCl_2 , pH 6.5. Additionally, approximately 0.5 mM of Trolox (Rasnik, McKinney et al. 2006) was added.

5.3 Methods

5.3.1 The AV approach

Dye distributions were modeled using the AV approach (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008) according to (Sindbert, Kalinin et al. 2011). The dyes are approximated by a sphere with an empirical radius of R_{dye} , where the central atom of the fluorophore is connected by a flexible linkage of a certain effective length L_{link} and width w_{link} to the nucleobase. The overall length of the linkage is given by the actual length of the linker and the internal chemical structure of the dye. A geometric search algorithm finds all dye positions within the linkage length from the attachment point which do not cause steric clashes with the macromolecular surface. All allowed positions are considered as equally probable which allows one to define an accessible volume for the dye (AV). To take the three quite different dimensions of fluorophores into account, we used their real physical dimensions for each calculation of a position distribution and performed three independent AV simulations with three different radii $R_{\text{dye}(i)}$ and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. Throughout this work we used $w_{\text{link}} = 4.5 \text{ \AA}$. As values for $R_{\text{dye}(i)}$ we used $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for Alexa488 (D) and $R_{\text{dye}(1)} = 11 \text{ \AA}$, $R_{\text{dye}(2)} = 3 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for Cy5 (A). See Table 5-2 for values used for L_{link} in “linker paper” (see also Figure 5.1). In “RT manuscript” we used $L_{\text{link}} = 20 \text{ \AA}$ for all D positions, $L_{\text{link}} = 23 \text{ \AA}$ for all A labeled dThymines (see Figure 5.2A) and $L_{\text{link}} = 8 \text{ \AA}$ for the labeled triphosphate (dp(1)/dt, see Figure 5.2B). In “RNA4WJ manuscript” we used $L_{\text{link}} = 20 \text{ \AA}$ for all D positions and $L_{\text{link}} = 22 \text{ \AA}$ for all A positions. As attachment atoms for nucleic acids we used “C5” for labeled Uracils and dThymines (see Figure 5.1, Figure 5.2A and Figure 5.4A), “C2” for labeled dGuanines (see Figure 5.4B) and “OP2” for labeled triphosphates (see Figure 5.2B). For labeled amino acids in the RT protein we used C_β (CB) as the attachment atom.

Table 5-2. Values for L_{link} used in “linker paper” (see Figure 5.1)

Linker	L	L	I	I	S_d	S_d	S_t	S_t
Dye	D	A	D	A	D	A	D	A
$L_{\text{link}}, \text{\AA}$	20	22	15	17	11	14	11	16

5.3.2 Definitions of differently averaged DA distances

For a proper interpretation of FRET results it is crucial to consider the distributions of D and A dye positions given by the position vectors \mathbf{R}_D and \mathbf{R}_A , respectively. Since different techniques determine distinct average distances between donor and acceptor dyes, three different quantities have to be defined and distinguished. Additionally, it is shown how to calculate them for sets of possible donor and acceptor positions, $\{\mathbf{R}_{D(i)}\}$ and $\{\mathbf{R}_{A(j)}\}$, respectively, with $i = 1...n$ and $j = 1...m$ which are provided by AV simulations and in which all allowed positions are assumed to be equally probable.

5.3.2.1 The mean distance $\langle R_{DA} \rangle$

$\langle R_{DA} \rangle$ denotes the *mean distance* between the dyes and can be determined by TCSPC measurements (see Eq. 5.8). $\langle R_{DA} \rangle$ is calculated by integrating over all possible positions of the two dyes and the resulting distances ($\langle R_{DA} \rangle = \langle |\mathbf{R}_D - \mathbf{R}_A| \rangle$):

$$\langle R_{DA} \rangle = \left\langle \left| \mathbf{R}_{D(i)} - \mathbf{R}_{A(j)} \right| \right\rangle_{i,j} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \mathbf{R}_{A(j)} - \mathbf{R}_{D(i)} \right| \quad \text{Eq. 5.1}$$

5.3.2.2 The FRET-averaged mean distance $\langle R_{DA} \rangle_E$

$\langle R_{DA} \rangle_E$ is the *FRET-averaged distance* between the dyes. It is calculated from the mean FRET efficiency (Eq. 5.2) using Eq. 5.3.

$$\langle E \rangle = \left\langle \frac{1}{1 + R_{DA}^6 / R_0^6} \right\rangle \quad \text{Eq. 5.2}$$

$$\langle R_{DA} \rangle_E = R_0 (\langle E \rangle^{-1} - 1)^{1/6} \quad \text{Eq. 5.3}$$

In Eq. 5.2 and Eq. 5.3 R_0 is the Förster radius. $\langle R_{DA} \rangle_E$ is determined from time-averaged fluorescence intensity measurements on the single-molecule (see Section 5.3.4) or ensemble level. For AVs it can be calculated with Eq. 5.4:

$$\langle R_{DA} \rangle_E = R_0 \left(\langle E \rangle^{-1} - 1 \right)^{1/6}$$

Eq. 5.4

$$\text{with } \langle E \rangle = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{1}{1 + |\mathbf{R}_{A(j)} - \mathbf{R}_{D(i)}|^6 / R_0^6} \right)$$

5.3.2.3 The distance between the mean positions of the dyes R_{mp}

R_{mp} is the distance between the mean positions of the dyes ($R_{mp} = |\langle \mathbf{R}_D \rangle - \langle \mathbf{R}_A \rangle|$) and is used for the geometric description of FRET based structural models (Woźniak, Schröder et al. 2008). R_{mp} cannot be measured directly via FRET. For AVs it can be defined as follows:

$$R_{mp} = \left| \langle \mathbf{R}_{D(i)} \rangle - \langle \mathbf{R}_{A(j)} \rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \mathbf{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^m \mathbf{R}_{A(j)} \right|$$

Eq. 5.5

5.3.3 Time-resolved polarized fluorescence experiments and data analysis

5.3.3.1 Ensemble time-correlated single-photon-counting (eTCSPC)

Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation sources were either a 470 nm diode laser (LDH-P-C 470, Picoquant, Berlin, Germany) operating at 8 MHz for donor excitation or a 635 nm diode laser (LDH-8-1 126, Picoquant, Berlin, Germany) operating at 10 MHz for direct acceptor excitation. The emission wavelength was set to 520 nm for donor emission and to 665 nm for acceptor emission, respectively. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path) resolution. Additional cut-off filters were used to reduce the contribution of the scattered light (>500 nm for donor and >640 nm for acceptor emission, respectively). All measurements were performed at room temperature. The concentrations of DNA/RNA molecules were kept below 1 μ M.

The G-factor was calculated using steady state anisotropies of the solutions measured with a Fluorolog-3 (Horiba Jobin Yvon, Munich, Germany).

Fluorescence intensity and anisotropy decay curves were fitted using the iterative re-convolution approach (O'Connor and Phillips 1984). The maximum number of counts was typically between 25,000 and 50,000. The fits approximately range from the maximum of the instrument response functions (IRF) to the first time channel with less than 100 detected photons. The fluorescence intensity decays of FRET-labeled molecules (donor and acceptor emission) were fitted globally with the decays of the molecules only labeled with either the donor (donor only, D-only) or the acceptor (acceptor only, A-only) dye. The fluorescence decays $F(t)$ were modeled with up to two fluorescence lifetimes τ_i with the species fractions x_i and a species-averaged fluorescence lifetime $\langle \tau \rangle_x$ (Eq. 5.6)

$$F(t) = x_1 \exp(-t/\tau_1) + x_2 \exp(-t/\tau_2)$$

$$\text{with } \langle \tau \rangle_x = x_1 \tau_1 + x_2 \tau_2$$
Eq. 5.6

Due to its high time resolution TCSPC can also be used to characterize DA distance distributions $p(R_{DA})$. For example, the analysis of donor fluorescence decays $F_D(t)$ (D decay) by Eq. 5.7 recovers $p(R_{DA})$ (Haas, Wilchek et al. 1975; Lakowicz 1999; Haas 2005).

$$F_D(t) = \int_{R_{DA}} p(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA}$$
Eq. 5.7

where $\tau_{D(0)}$ is the donor fluorescence lifetime without acceptor. For simplicity, the distribution $p(R_{DA})$ can be assumed to be Gaussian (as justified by Figure S5A in the Si of “linker paper”). Considering also the presence of Donor-only molecules, the fitting parameters of Eq. 5.8 are then the mean DA distance $\langle R_{DA} \rangle$, the half-width σ_{DA} of the R_{DA} -distribution and also the fraction of Donor-only molecules x_D (in our measurements typically below 10%),

$$F_D(t) = (1 - x_D) \int_{R_{DA}} \frac{1}{\sqrt{2\pi}\sigma_{DA}} \exp\left(-\frac{(R_{DA} - \langle R_{DA} \rangle)^2}{2\sigma_{DA}^2}\right) \exp\left(-\frac{t}{\tau_{D(0)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA}$$

$$+ x_D \exp\left(-\frac{t}{\tau_{D(0)}}\right)$$
Eq. 5.8

The anisotropy decays $r(t)$ were formally described by double or triple exponential decays (rotational correlation times ρ_1 , ρ_2 and ρ_3) with free amplitudes (b_1 , b_2 and b_3) (Eq. 5.12). Applying appropriate

weighting (O'Connor and Phillips 1984) the anisotropy decays were recovered by global fitting of the sum and difference curves according to Eq. 5.9 and Eq. 5.10,

$$F_{\text{sum}}(t) = F_{\parallel}(t) + 2GF_{\perp}(t) = F(t) \quad \text{Eq. 5.9}$$

$$F_{\text{diff}}(t) = F_{\parallel}(t) - GF_{\perp}(t) = F(t)r(t) \quad \text{Eq. 5.10}$$

All species irrespective of the fluorescence lifetime were assumed to exhibit the same anisotropy decay. The fluorescence anisotropy decay $r(t)$ can be described as a product of the separate factors responsible for local dye reorientations and the overall rotation of the macromolecule as (Lakowicz 1999)

$$r(t) = [(r_0 - r_{\infty})\exp(-t/\rho_{\text{local}}) + r_{\infty}]\exp(-t/\rho_{\text{global}}) \quad \text{Eq. 5.11}$$

where r_0 is the fundamental anisotropy, and r_{∞} is the residual anisotropy. In this work, two exponentials were often needed to describe the local fluorophore dynamics. This leads to the following more formal form of Eq. 5.11 with up to three rotational correlation times ρ_i and the corresponding anisotropy amplitudes b_i (Eq. 5.12)

$$r(t) = b_1 \exp(-t/\rho_1) + b_2 \exp(-t/\rho_2) + b_3 \exp(-t/\rho_3) \quad \text{Eq. 5.12}$$

with $r_0 \geq b_1 + b_2 + b_3$

In the case of timescale separation of local and overall motions ($\rho_{\text{overall}} \gg \rho_{\text{local}}$), the times ρ_1 and ρ_2 characterize only local dye reorientations and $r_{\infty} = b_3$ (cf. Eq. 5.11 and Eq. 5.12).

5.3.3.2 TCSPC with single molecule data

Analogous to Section 5.3.3.1, time-resolved polarized fluorescence analysis can also be performed for data acquired via smMFD (see Section 5.3.4). Here, the polarized components $F_{\parallel}(t)$ and $F_{\perp}(t)$ of a fluorescence decay are globally fitted with the following model:

$$\begin{aligned} F_{\parallel}(t) &= F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3 + B_{\parallel} \\ F_{\perp}(t) &= GF(t) \cdot (1 - (1 - 3l_2) \cdot r(t)) / 3 + B_{\perp} \end{aligned} \quad \text{Eq. 5.13}$$

In Eq. 5.13 $F(t)$ is the fluorescence decay typically modeled by 2-exponential relaxation, G is the ratio of detection efficiencies of parallel and perpendicular channels, factors l_1 and l_2 describe polarization mixing in high-NA objectives (Koshioka, Sasaki et al. 1995), and B_{\parallel} and B_{\perp} represent background contributions in parallel and perpendicular detection channels, respectively.

5.3.4 Single-molecule multi-parameter fluorescence detection (smMFD) measurements

The fluorescent donor molecules (Alexa 488) are excited by a linearly polarized, active-mode-locked Argon-ion laser (Innova Saber, Coherent, Santa Clara, CA, USA, 496.5 nm, 73.5 MHz, ~ 300 ps FWHM) or by a 485 nm diode laser (LDH-D-C 485, Picoquant, Berlin, Germany) operating at 64 MHz. The laser light is focused into the dilute solution (< 50 pM) of labeled molecules by a 60x/1.2 water immersion objective. Each molecule generates a brief burst of fluorescence photons as it traverses the detection volume. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into a wavelength ranges below and above 595 nm by using a dichroic beamsplitter (595 DCXR, AHF, Tübingen, Germany). Additionally, red (HQ 720/150 nm or HQ 730/140 nm for Cy5) and green (HQ 533/46 nm or HQ 535/50 nm for Alexa 488 and Rh110) bandpass filters (both made by AHF, Tübingen, Germany) in front of the detectors ensure that only fluorescence photons coming from the acceptor and donor molecules are registered. An estimate of the focal geometry is acquired by determining the diffusion correlation time of 200 ± 13 μ s for Rhodamine 110 and knowing its diffusion coefficient of 0.34 ± 0.03 μ m²/ms. Detection is performed using four avalanche photodiodes (SPCM-AQR-14, Laser Components, Germany or alternatively for the green channels PDM050CTC or τ -SPAD-100, both PicoQuant, Berlin, Germany) . The signals from all detectors are guided through a passive delay unit and two routers to two synchronized time-correlated single photon counting boards (SPC 432 or SPC 132 or SPC 832, Becker and Hickl, Berlin, Germany) connected to a PC. Bursts of fluorescence photons are distinguished from the background of 1-2 kHz by applying certain threshold intensity criteria (Eggeling, Berger et al. 2001). Bursts during which bleaching of the acceptor occurs are excluded from further analysis by applying a criterion regarding the difference in macroscopic times, $|T_G - T_R| < 1$ ms, where T_G and T_R are the average macroscopic times in which all photons have been detected in the green and red channels respectively during one burst (Eggeling, Widengren et al. 2006).

5.3.5 Distance determination via photon distribution analysis (PDA)

As specified in Section 2, we calculate DA distances (R_{DA}) by measuring the fluorescence intensities of D and A (F_D and F_A , respectively). R_{DA} is then given by combining Eq. 2.3 and Eq. 2.4 which yields Eq. 5.14:

$$R_{DA} = \left(\frac{\Phi_{FA}}{\Phi_{FD(0)}} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_0 \quad \text{Eq. 5.14}$$

or equivalently by Eq. 5.15:

$$R_{DA} = \left(\Phi_{FA} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_{0r} \quad \text{Eq. 5.15}$$

Here, we use the reduced Förster radius R_{0r} (Rothwell 2002), which, in contrast to R_0 , does not depend on $\Phi_{FD(0)}$. Throughout this work we used $R_{0r} = 53.97 \text{ \AA}$ (using $R_{0r} = 53.97 \text{ \AA}$ is equivalent to using $R_0 = 52 \text{ \AA}$ and $\Phi_{FD(0)} = 0.8$). In “linker paper” and “RNA4WJ manuscript” Φ_{FA} was determined through eTCSPC measurements (see e.g. Section 4.1 in the SI of “RNA4WJ manuscript”). Throughout “RT manuscript” $\Phi_{FA} = 0.32$ was used. F_D and F_A can be calculated from the signals measured in the green and red detection channels of the smMFD setup (see Figure 3.5 and Section 5.3.4), S_G and S_R , respectively, via Eq. 5.16 and Eq. 5.17:

$$F_D = \frac{F_G}{g_G} = \frac{S_G - \langle B_G \rangle}{g_G} \quad \text{Eq. 5.16}$$

$$F_A = \frac{F_R}{g_R} = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \quad \text{Eq. 5.17}$$

where F_G and F_R are the fluorescence signals in the green and the red signal channels, respectively, α is the crosstalk factor which is determined as the ratio between donor photons detected in the red channels and those detected in the green channels ($\alpha = F_{R(D)} / F_{G(D)}$) for the D only labeled sample, g_G and g_R are the detection efficiencies in the green and red channels, respectively (see Section 5.3.7 for the determination of g_G/g_R), and $\langle B_G \rangle$ and $\langle B_R \rangle$ are the mean background intensities in the green and red channels, respectively.

To accurately predict the shape of S_G/S_R (or equivalently F_D/F_A) histograms in the presence of FRET we use photon distribution analysis (PDA), which explicitly takes into account shot noise, background contributions and additional broadening due to complex acceptor photophysics (Antonik, Felekyan et al. 2006, Kalinin, 2010 #2744; Kalinin, Felekyan et al. 2007; Kalinin, Felekyan et al. 2008). PDA calculates the probability of observing a certain combination of photon counts $P(S_G, S_R)$

$$P(S_G, S_R) = \sum_{F_G+B_G=S_G; F_R+B_R=S_R} P(F)P(F_G, F_R | F)P(B_G)P(B_R) \quad \text{Eq. 5.18}$$

The intensity distribution of the fluorescence only contribution to the signal, $P(F)$, is obtained from the total measured signal intensity distribution $P(S)$ by deconvolution assuming that the background signals B_G and B_R obey Poisson distributions, $P(B_G)$ and $P(B_R)$, with known mean intensities $\langle B_G \rangle$ and $\langle B_R \rangle$. $P(F_G, F_R | F)$ represents the conditional probability of observing a particular combination of green and red fluorescence photons, F_G and F_R , provided the total number of registered fluorescence photons is F , and can be expressed as a binomial distribution (Antonik, Felekyan et al. 2006). Subsequently, $P(S_G, S_R)$ may be further manipulated to generate a theoretical histogram of any FRET-related parameter as discussed elsewhere (Kalinin, Felekyan et al. 2007). The theory of PDA is described in details in (Antonik, Felekyan et al. 2006; Kalinin, Felekyan et al. 2008).

In general, for a model with n FRET states and a D-only fraction, $2n+1$ fit parameters are needed: n mean DA distances, n fractions, and an additional (beyond the shot noise) distribution width σ_{app} expressed as a fraction of the corresponding mean distance. This additional distribution width can be attributed mainly to complex acceptor photophysics and thus can be fitted globally as justified in (Kalinin, Sisamakos et al. 2010). As a result, PDA needs much fewer free parameters than the classical approach of fitting multiple Gaussian peaks, which requires up to $3n+1$ parameters (n mean DA distances, n fractions, $n+1$ peak widths). The fit quality is judged by the reduced chi-squared (χ^2_r) parameter and by visually inspecting weighted residuals plots.

5.3.5.1 PDA model for RT:dp/dt

In “RT manuscript” we typically used a model which accounts for two FRET-populations representing the educt-state (P-E) and the product-state (P-P), a D-only population, and, if necessary, populations accounting for impurities and for a dead-end complex (see Section S3.1 in the SI of “RT manuscript”)

were added as well (see Section S4.2 in the SI of “RT manuscript” for the assignments of the states in MFD plots for all experiments). The FRET states were fitted using the same value for the global relative additional (beyond the shot noise) width σ_{app} . Of the P-E and P-P states the distance of the one with the larger amplitude was assumed to be P-E and, thus, was chosen to be used for structural modeling. For some datasets only one FRET state is visible. In those cases we assume that the product and educt state are overlapping. If the P-P and P-E-state have similar amplitudes (ratio smaller than 4:3) an assignment of the distances to the corresponding states is not possible and the amplitude weighted distance $\langle R \rangle_x$ with the uncertainties err_+ , err_- has been used for modeling (see Table S2 in the SI of “RT manuscript”).

5.3.5.2 PDA model for RNA four-way junctions (RNA4WJs)

In “RNA4WJ manuscript” a model was used accounting for 2 or 3 FRET states and a D-only contribution. Additional broadening of FRET states was accounted for by a global parameter σ_{app} . In most cases an impurity (mostly 1-3%) with an apparent distance of typically 70 - 90 Å had to be taken into account. This state was in most of cases present in respective D-only samples with an amplitude of a few percent. Thus, for n FRET states $2n+1$ to $2n+3$ fit parameters were required depending on whether the impurity state was considered.

In those cases when only 2 FRET states (one major and one minor) were visible, the second “invisible” minor state was assumed to be overlapped by either the visible minor or the visible major state depending on their relative amplitudes. If the amplitude of the visible minor state was larger than 25 %, it was assumed that this FRET state contained both minor states. Thus, two states with the same distance and amplitude were used. Else, the “invisible” minor was assumed to have a relative amplitude of ~ 10 % and the same distance as the major FRET state. See Appendix IV for all PDA plots.

5.3.6 Static FRET line and distribution of possible σ_{DA} -values

The static FRET line represents the expected dependence between FRET indicators derived from intensities (e.g. F_D/F_A) and the fluorescence lifetime of the donor. In the simplest case it is given by the well-known Eq. 5.19 (see also Eq. 2.3):

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad \text{Eq. 5.19}$$

For F_D/F_A it is given by:

$$F_D/F_A = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left(\frac{\tau_{D(0)}}{\tau_{D(A)}} - 1 \right) \right. \quad \text{Eq. 5.20}$$

In reality this relationship does not hold because of distributions of donor-acceptor distances due to flexible dye linkers are not accounted for (see Section S2.7 in the SI of “linker paper”). In addition, non-exponential fluorescence decay of the donor dye (see Table S6 in the SI of “RT manuscript” and Section 4.1 in the SI of “RNA4WJ manuscript”) itself must be considered. These effects can be corrected for as described (Kalinin, Valeri et al. 2010; Kalinin 2012). As there is no analytical expression for the $E(\tau_{D(A)})$ dependence that considers the above effects, a polynomial approximation is used. In this work we used the following approximation:

$$F_D/F_A = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left(\frac{\langle \tau_{D(0)} \rangle_x}{c_3 \langle \tau_{D(A)} \rangle_f^3 + c_2 \langle \tau_{D(A)} \rangle_f^2 + c_1 \langle \tau_{D(A)} \rangle_f + c_0} - 1 \right) \right. \quad \text{Eq. 5.21}$$

Where $\langle \tau_i \rangle_x$ and $\langle \tau_i \rangle_f$ are species and fluorescence averaged mean lifetimes, respectively. See Table S7 in the SI of “RT manuscript” and Section 4.1 in the SI of “RNA4WJ manuscript” for the polynomial coefficients $c_{(i)}$ used and the values assumed for the half-width of the DA distance distribution (σ_{DA}). For highly asymmetric AVs σ_{DA} depends also on the mutual orientation of D and A clouds, which implies that individual σ_{DA} -values should be used for different samples and/or even FRET states. Considering various possible orientations of calculated dyes’ AVs, we estimated that σ_{DA} can vary between ca. 5.5 and 12 Å being somewhat correlated with $\langle R_{DA} \rangle$ (Figure 5.5). To fit static FRET lines to the observed FRET states for the RNA4WJ, values of σ_{DA} between 6 and 9 Å were required, which is within the expected range.

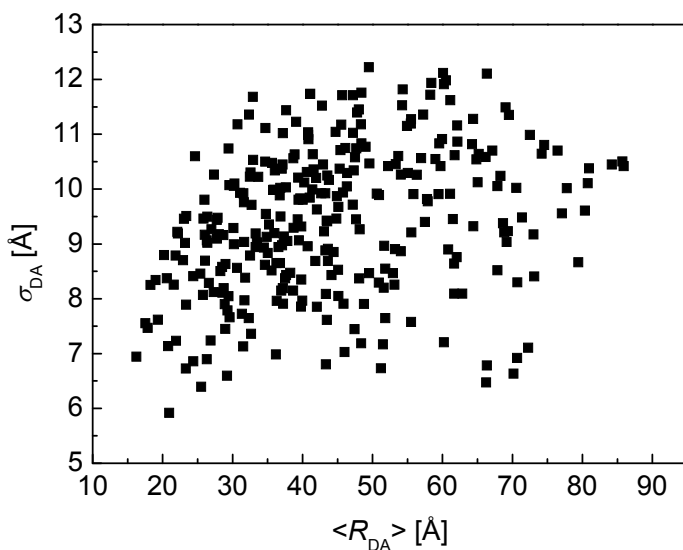


Figure 5.5. Possible values of σ_{DA} calculated for randomly positioned and oriented AVs of Alexa488 and Cy5 initially calculated for positions (D) β 8c and (A) δ 23d of the RNA4WJ (see Figure 3.2), respectively.

5.3.7 Determination of detection efficiency ratio g_G/g_R

In PDA, to be able to convert model distances into probabilities of observing green photons, the detection efficiency ratio g_G/g_R is needed (g_G and g_R stand for the detection efficiencies of “green” and “red” channels, respectively). These values are calculated for each measurement session by requiring that the linker-corrected static FRET line (Kalinin, Valeri et al. 2010) (see Eq. 5.27) goes through the observed FRET populations in a 2D histogram of F_D/F_A vs $\tau_{D(A)}$. In “RT manuscript”, the actual measurements of the protein were used for the calibration of g_G/g_R , in “RNA4WJ manuscript”, a measurement of a FRET labeled dsDNA (see Section 5.1.3) was performed. For the calculation of the static FRET line, the respective fluorescence decays of the D-only labeled RT:dp/dt sample (see Table S6 in the SI of “RT manuscript”) or, for the dsDNA used in “RT manuscript”, $\tau_{D(0)} = 4.1$ ns (mono-exponential decay) was used. In both cases it was assumed that $\sigma_{DA} = 6$ Å.

5.3.8 Confidence intervals for fit parameters in PDA

Confidence intervals estimation for multiple fit parameters is performed as follows. All free fit parameters are varied simultaneously in a random manner. The χ_r^2 -value is calculated at 100000 random points yielding 100-1000 points with χ_r^2 -values below $\chi_{r,\min}^2 + (2 / N_{\text{bins}})^{1/2}$ (here N_{bins} is the number of histogram bins, and $\chi_{r,\min}^2$ is the reduced chi-squared of the best fit). The range where such fits are possible is assigned as 1σ confidence interval. Whereas one could calculate χ_r^2 thresholds more strictly from the chi-squared distribution (Soong 2004), in practice $\chi_{r,\min}^2$ is often affected by experimental imperfections and can be considerably larger than one. For this reason, we prefer the simple test mentioned above which relates χ_r^2 values to that of the best fit.

5.3.9 Estimation of the κ^2 related distance uncertainties

To minimize the uncertainty in κ^2 (see Section 2), we use the residual anisotropies that result from the measurements of the donor-only, the acceptor-only and the FRET sensitized acceptor anisotropy decays ($r_{\infty,D}$, $r_{\infty,A}$ and $r_{\infty,A(D)}$, respectively) as determined from TCSPC measurements (see Eq. 5.12). First, we calculate the second-rank order parameters $S_D^{(2)}$ and $S_A^{(2)}$ by

$$\sqrt{\frac{r_{\infty,D}}{r_0}} = -S_D^{(2)} \text{ and } \sqrt{\frac{r_{\infty,A}}{r_0}} = S_A^{(2)} \quad \text{Eq. 5.22}$$

where r_0 is the fundamental anisotropy. $S_D^{(2)}$ and $S_A^{(2)}$ are defined differently (inverse sign) due to the different orientation of the transition dipole moments with respect to the linker for Alexa488 and Cy5 (see Figure 5.1, for further details see Section 3.1.2 of “linker paper”). The range of all possible values for κ^2 is calculated by Eq. 5.23 (Dale, Eisinger et al. 1979; van der Meer, Cooke et al. 1994):

$$\begin{aligned} \kappa^2 = & \frac{2}{3} + \frac{2}{3} S_D^{(2)} S^{(2)}(\theta_D) + \frac{2}{3} S_A^{(2)} S^{(2)}(\theta_A) \\ & + \frac{2}{3} S_D^{(2)} S_A^{(2)} [S^{(2)}(\varphi) + 6S^{(2)}(\theta_D) S^{(2)}(\theta_A) + 1 + 2S^{(2)}(\theta_D) + 2S^{(2)}(\theta_A) - 9 \cos \theta_D \cos \theta_A \cos \varphi] \end{aligned} \quad \begin{array}{l} \text{Eq.} \\ 5.23 \end{array}$$

In Eq. 5.23 θ_D and θ_A are the angles between the symmetry axes of the rotations of the dyes and the distance vector \vec{R}_{DA} , while φ is the angle between the symmetry axes (see Figure 2.2). $S^{(2)}$ is defined formally by:

$$S^{(2)}(\varphi) = \frac{1}{2}(3\cos^2\varphi - 1) = \frac{r_{\infty,A(D)}}{r_0 S_D^{(2)} S_A^{(2)}},$$

$$S^{(2)}(\theta_D) = \frac{1}{2}(3\cos^2\theta_D - 1) \quad \text{Eq. 5.24}$$

$$\text{and } S^{(2)}(\theta_A) = \frac{1}{2}(3\cos^2\theta_A - 1)$$

κ^2 -values (Eq. 5.23) cannot be calculated unambiguously because, in general, the angles θ_D and θ_A are not experimentally accessible. However, for an experimentally determined φ (Eq. 5.24) it is possible to define a range of possible values for θ_D and θ_A (see Figure 2.2) by

$$0 < \theta_D < \pi/2$$

$$|\varphi - \theta_D| < \theta_A < \min(\varphi + \theta_D, \pi/2) \quad \text{Eq. 5.25}$$

For example: if $\varphi = 0$, θ_D and θ_A have to be equal (see also (Ivanov, Li et al. 2009)). With known φ and a corresponding range of possible values for θ_D and θ_A one can calculate the range of κ^2 -values (Eq. 5.23) compatible with the experimental data ($r_{\infty,D}$, $r_{\infty,A}$ and $r_{\infty,A(D)}$).

As an example a distribution of possible κ^2 -values is calculated here for dsRNA and C6-linkers (see “linker paper”): The measurements of the donor-only, the acceptor-only and the FRET sensitized acceptor anisotropy decays yield $r_{\infty,D} = 0.05$, $r_{\infty,A} = 0.13$, and $r_{\infty,A(D)} = 0.012$ (see Tables 2 and 5 in “linker paper”). Eq. 5.22 gives $S_D^{(2)} = -0.365$ and $S_A^{(2)} = 0.577$ and Eq. 5.24 gives $S^{(2)}(\varphi) = -0.150$ and $\varphi = 1.067$ ($= 61.1^\circ$). For this value of φ Eq. 5.36 gives a range of possible values for θ_D and θ_A , which can further be used to calculate possible ranges of values for $S^{(2)}(\theta_D)$ and $S^{(2)}(\theta_A)$ (Eq. 5.24). These can then be used to calculate possible values for κ^2 according to Eq. 5.23 (see Figure 3.6B). See Table 5 in “linker paper” for the resulting values for κ_{\min}^2 , κ_{\max}^2 and κ_{mean}^2 and for the resulting accuracy and precision for R_{DA} .

5.3.10 Fitting of correlation curves

The correlation curves shown in Figure 4.4A are fitted using Eq. 5.26,

$$G(t_c) = 1 + \frac{1}{N} \cdot \frac{1}{1 + \frac{t_c}{t_d}} \cdot \frac{1}{\sqrt{\left(1 + \frac{t_c}{(z_0/w_0)^2 \cdot t_d}\right)}} \cdot \left(1 - A_1 + A_1 \cdot e^{-\frac{t_c}{t_1}} - A_2 + A_2 \cdot e^{-\frac{t_c}{t_2}} - A_3 + A_3 \cdot e^{-\frac{t_c}{t_3}} - A_4 \cdot e^{-\frac{t_c}{t_4}}\right) \quad \text{Eq. 5.26}$$

$$\times \left(1 + R \cdot \left(\frac{1}{1+C} \exp(-t_c/t_{rot}) + \frac{C}{1+C} \cdot \exp(-t_c/(s \cdot t_{rot}))\right)\right)$$

The obtained fit parameters and their assigned meanings are summarized in Table 5-3.

Table 5-3. Fit parameters and their assigned meanings for the correlation curves for acceptor excitation though FRET and though direct excitation, respectively.

Parameter	Meaning	FRET excitation	Direct excitation
$N/(1-A_1-A_2-A_3)$	Number of all molecules	1.54	1.63
t_d [ms]	Diffusion time	0.36	0.25
z_0/w_0	Axial ratio	13	13
A_1	unknown quenching process	0.13 (global)	0.13 (global)
t_1 [μs]		12 (global)	12 (global)
$A_{(T+CT)}$	Cy5 cis-trans isomerization and triplet	0.39 (global)	0.39 (global)
$t_{(T+CT)}$ [μs]		1.7 (global)	1.7 (global)
A_{linker}	Linker dynamics	0.06	0 (fixed)
t_{linker} [ns]		162	---
A_{rot}	Rotational diffusion	0.42	0.19
C		-0.17 (fixed)	-0.17 (fixed)
s		0.3 (fixed)	0.3 (fixed)
t_{rot} [ns]		9.6 (fixed)	9.6 (fixed)
A_5	Photon anti-bunching	0.98 (fixed)	0.98 (fixed)
t_5 [ns]		2.5	1.2

5.3.11 Rigid body docking: implementation details

5.3.11.1 Mechanical model.

As discussed in Section 3.6 (see also Section “Step 4” in “RT manuscript”), the purpose of rigid body docking is to minimize the weighed deviation between n experimentally obtained distances $\{R_{DA}\}$ and corresponding model distances $\{R_{model}\}$, given the uncertainties $\{\Delta R_{DA}\}$:

$$\chi_E^2 = \sum_n \frac{(R_{DA} - R_{model})^2}{(\Delta R_{DA})^2} \quad \text{Eq. 5.27}$$

In Eq. 5.27, $\{R_{DA}\}$ can represent a set of R_{mp} or $\langle R_{DA} \rangle_E$ values. The first option is easier to implement in combination with rigid body dynamics (see Section 5.3.12 for the calculation of an R_{mp} to $\langle R_{DA} \rangle_E$ conversion function), whereas directly calculating deviations between experimental and model $\langle R_{DA} \rangle_E$ is more appropriate for structure screening (see Section “Extension to flexible parts of the complex: ...” in “RT manuscript”). Now we notice that the right-hand part of Eq. 5.27 is equivalent to the energy of a network of mean dye positions (points in space), connected with n springs with relaxed lengths of $\{R_{DA}\}$ and corresponding spring constants $k = 2/(\Delta R_{DA})^2$. The coordinates of mean dye positions are obtained by MD or AV simulations and then fixed with respect to corresponding labeled macromolecules, which are treated as rigid bodies. Minimizing χ_E^2 in Eq. 5.27 is equivalent to relaxing this rigid body system. Here and further in this section we use reduced energy which results in unusual units for k [$1/\text{\AA}^2$] and other familiar quantities. All distances are expressed in \AA .

In addition to FRET restraints, clashes between different subunits are prevented by considering clash contributions to the total “energy”, χ_{clash}^2 , which are calculated, equivalently to Eq. 5.27, by Eq. 5.28:

$$\chi_{clash}^2 = \sum_{i,j} \begin{cases} 0, & r_{ij} \geq r_{wi} + r_{wj} \\ (r_{wi} + r_{wj} - r_{ij})^2 / r_{ctol}^2, & r_{ij} < r_{wi} + r_{wj} \end{cases} \quad \text{Eq. 5.28}$$

where r_{ij} is the distance between atoms i and j which belong to different subunits, r_{wi} and r_{wj} are their van der Waals radii, and r_{ctol} is the pre-defined clash tolerance. For the RT:dp/dt complex (double-stranded part of dp/dt) $r_{ctol} = 6 \text{\AA}$ was used during initial search run and $r_{ctol} = 2$ and 1\AA during the first and second refinement runs, respectively (see Section 4.3.2 and Section “Step 4” in “RT manuscript”). For the RNA4WJ $r_{ctol} = 0.15 \text{\AA}$ was used during initial search run and $r_{ctol} = 0.05 \text{\AA}$ during the refinement run (see e.g. Section 3.6). More realistic potentials (e.g. the Lennard-Jones potential (Atkins and de Paula 2001)) can be also used here instead of the harmonic potential in Eq. 5.28. The simplified approach (Eq.

5.28) is justified by low accuracy of FRET data (a few Å) as compared to possible violations of van der Waals radii ($r_{wi} + r_{wj} - r_{ij} < 0.1$ Å). Anyhow, by choosing a sufficiently small r_{ctol} the atoms can be made as “hard” as desired, in which case the contribution of χ_{clash}^2 becomes negligible. The reduced chi-squared parameter to be minimized is then given by Eq. 5.29

$$\chi_r^2 = (\chi_E^2 + \chi_{clash}^2)/(n - p) \rightarrow \min \quad \text{Eq. 5.29}$$

where p is the number of degrees of freedom, which is equal to $6 \times (\text{number of bodies} - 1)$. For the RNA4WJ we used $n = 51$ and $p = 9$ (see e.g. Section 3.6) and for the RT:dp/dt complex (double-stranded part of dp/dt) we used $n = 20$ and $p = 6$ (see Section 4.3.2 and Section “Step 4” in “RT manuscript”).

5.3.11.2 Time evolution of the rigid body system

The position and orientation of each subunit at any time t is described by a coordinate vector of its center of mass $\mathbf{x}(t)$ and a rotational matrix $\mathbf{Q}(t)$. For the relaxation of a system of structural subunits we used the Verlet algorithm (Martyna 1994) with damping to model translational and rotational motions. If not stated otherwise, all distances are expressed in Å, and masses are expressed in Da. Translational movement is described by:

$$\mathbf{x}(t + \Delta t) = (2 - \nu)\mathbf{x}(t) - (1 - \nu)\mathbf{x}(t - \Delta t) + \mathbf{F}\Delta t^2/m \quad \text{Eq. 5.30}$$

In Eq. 5.30, Δt is the simulation time step, the factor ν accounts for viscosity (see Eq. 5.35), \mathbf{F} is the total force acting on the center of mass of a subunit (see Figure 5.6), and m stands for its mass. The “forces” are derived from violations of FRET distances, clashes between subunits, and optionally other constraints (e.g. flexible chemical linkages between subunits). Since here we are only interested in finding energy minima (rather than in investigating trajectories), we use a simplified representation of the moment of inertia (I) by assuming the same (mean) value of I for all axes.

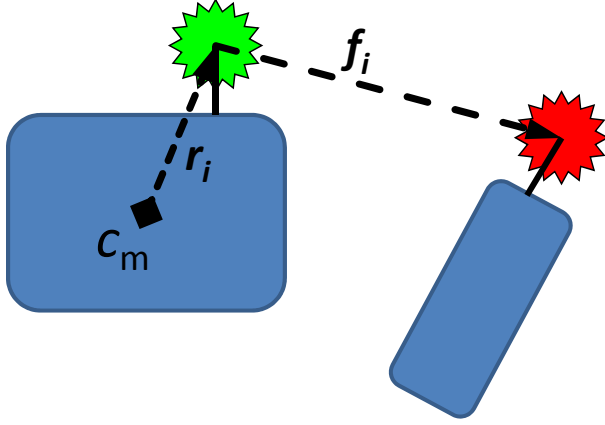


Figure 5.6. Sketch depicting the derivation of the total force F and the total Torque T acting on a subunit. Vector f_i describes the force between two labeling positions (green and red), its absolute value is determined by $(R_{DA, i} - R_{model, i})^2 / (\Delta R_{DA, i})^2$ (see Eq. 5.27). Forces due to clashes are calculated accordingly (see Eq. 5.28). Then, $F = \sum f_i$. The vector r_i connects the center of mass of one body c_m and the point where the force f_i is applied. Therefore, $T = \sum r_i \times f_i$.

In analogy with Eq. 5.30, we obtain the rotational movement by:

$$Q(t + \Delta t) = Q\left(\frac{T}{|T|}, \theta\right) Q(w, -\nu\theta_w) Q(t) Q^{-1}(t - \Delta t) Q(t) \quad \text{Eq. 5.31}$$

In Eq. 5.31, T is the total torque vector (see Figure 5.6); the angle θ is given by

$$\theta = \Delta t^2 |T| / I \quad \text{Eq. 5.32}$$

the rotational matrices with parameters represent a rotation by angle θ about an axis u ,

$$Q(u, \theta) = \begin{pmatrix} c + u_x^2 d & u_x u_y d - u_z s & u_x u_z d + u_y s \\ u_x u_y d + u_z s & c + u_y^2 d & u_y u_z d - u_x s \\ u_x u_z d - u_y s & u_y u_z d + u_x s & c + u_z^2 d \end{pmatrix} \quad \text{Eq. 5.33}$$

with: $u = (u_x, u_y, u_z)$; $c = \cos\theta$; $d = 1 - \cos\theta$; $s = \sin\theta$

and w and θ_w fulfil Eq. 5.34

$$Q(w, \theta_w) = Q(t) Q^{-1}(t - \Delta t) \quad \text{Eq. 5.34}$$

To minimize oscillations and to improve convergence, viscosity factor ν must be chosen so that the system is close to being critically damped. ν is initially estimated by Eq. 5.35

$$\nu = 2\Delta t \sqrt{K/M} \quad \text{Eq. 5.35}$$

with an option of additional fine-tuning to improve convergence. In Eq. 5.35 K is the sum of all spring constants and M is the total mass of all subunits.

Each rigid body simulation is run until the following criteria are fulfilled: kinetic energy $E_k < 0.001$; $|\mathbf{F}| < 0.001$; $|\mathbf{T}| < 0.02$ (in the units described above). After the initial search procedure, typically more than 99% of the structures reach convergence as defined by these criteria. We should also mention that more advanced optimization procedures (Marushchak, Grenklo et al. 2007) can be also applied at this step. However, in our experience, even with rigid body dynamics >95% convergence probability is achieved in most of cases.

5.3.12 R_{mp} to $\langle R_{DA} \rangle_E$ conversion function

To avoid repeated calculations of AVs during the iterative structure optimization (see Section 5.3.11) we convert $\langle R_{DA} \rangle_E$ into distances between the mean positions of the dyes $R_{mp} = |\langle \mathbf{R}_D \rangle - \langle \mathbf{R}_A \rangle|$. In this section we show how an R_{mp} to $\langle R_{DA} \rangle_E$ conversion function is calculated for Alexa488 and Cy5 as D and A, respectively, attached via C6 linkers to a ds A-RNA (see “RNA2” in Section 5.1.1.3). A series of AVs (see Section 5.3.1) is generated for D and A dyes separated by -7 to 30 basepairs. For each pair of AVs we calculated R_{mp} and $\langle R_{DA} \rangle_E$ (see Section 5.3.2). Corresponding R_{mp} and $\langle R_{DA} \rangle_E$ values are plotted in Figure 5.7. The solid red line represents a 3rd order polynomial approximation to the $\langle R_{DA} \rangle_E - (R_{mp})$ dependence given by $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$. For comparison, Figure 5.7 contains also polynomial approximation to the $\langle R_{DA} \rangle_E - (R_{mp})$ dependence resulting for the RT:dp/dt complex.

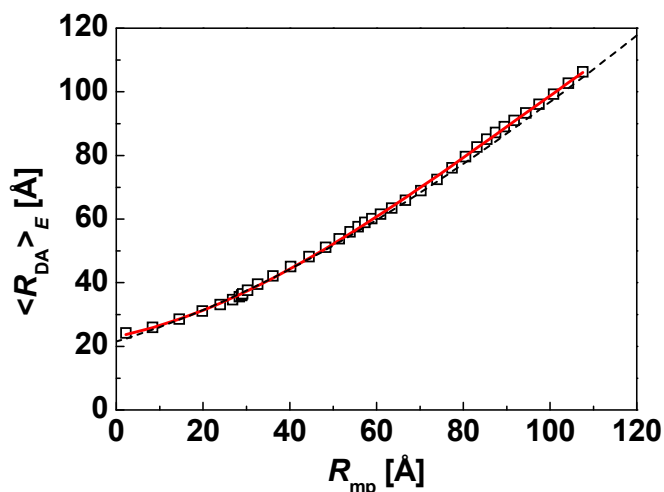


Figure 5.7. R_{mp} to $\langle R_{DA} \rangle_E$ conversion function for a ds A-RNA and Alexa488 and Cy5 as D and A, respectively, attached via C6 linkers (open squares), its 3rd order polynomial approximation $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$ (red line) and a 3rd order polynomial approximation $\langle R_{DA} \rangle_E = -0.08 \times 10^{-5} R_{mp}^3 + 4.191 \times 10^{-3} R_{mp}^2 + 0.414 R_{mp} + 21.51$ (black dashed line) for a R_{mp} to $\langle R_{DA} \rangle_E$ conversion function for the RT:dp/dt complex (see Sections 5.1.2 and 5.3.1 for labeling positions and the parameters used for AV modeling, respectively).

5.3.13 Model discrimination via cluster analysis

5.3.13.1 Cluster plots

The generation of cluster plots for the rigid body docking results is explained for the double stranded dp/dt and RT (see Section “Step 5” in the main text and Section S3.5.2 in the SI of “RT manuscript”). The obtained solutions are sorted by χ_r^2 and plotted as shown in Figure 5.8A (here, resulting structures after “search” run with 6 Å clash tolerance are shown). Obviously, there are groups of very similar solutions (clusters), which are separated by steps in the χ_r^2 plot and corresponding peaks in the RMSD plot (Figure 5.8C). To generate a cluster plot as shown in Figure 5.8D, we applied thresholds to $\Delta\chi_r^2$ and RMSD plots as shown by red dashed lines in Figure 5.8B and Figure 5.8C. Solutions for which both $\Delta\chi_r^2$ and RMSD fell below these thresholds were grouped with previous structures. As a result, several clusters of similar structures were obtained as shown in Figure 5.8D using different symbol sizes.

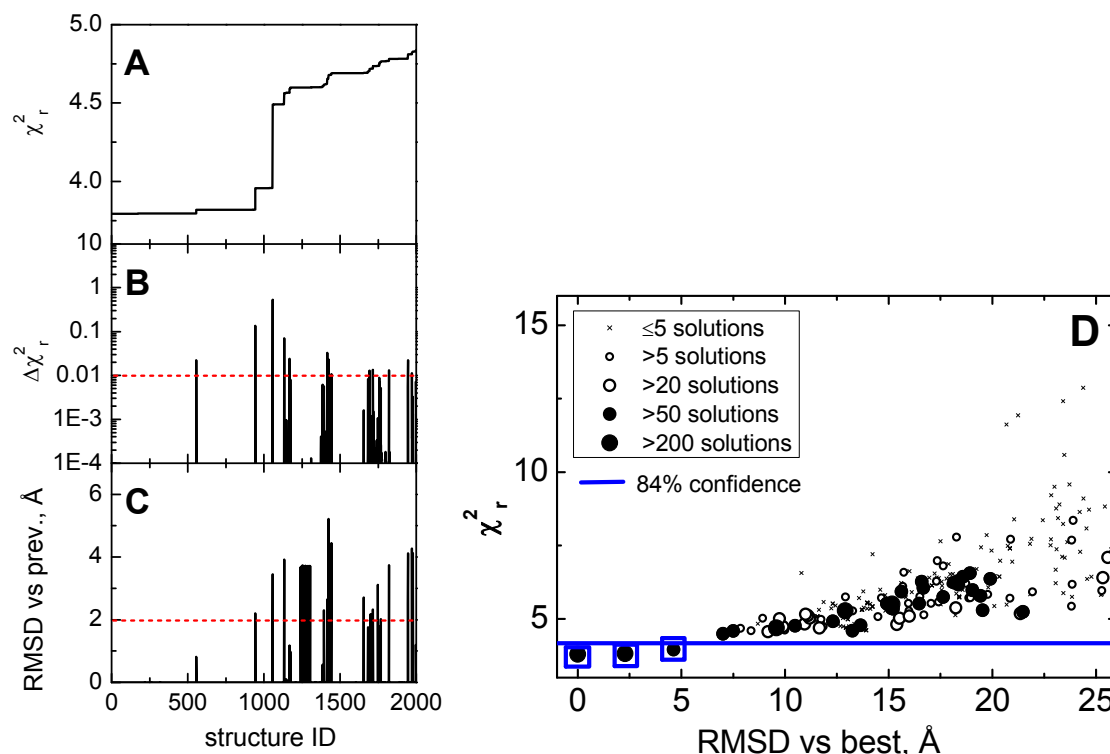


Figure 5.8. Rigid body docking of RT and dp/dt (both from crystal structure PDB-ID 1R0A). (A) χ_r^2 for the 2000 best solutions. (B) Increase of χ_r^2 compared to the previous structure. (C) RMSD between current and previous structures. Best 2000 of 10000 structures generated during the search phase with $r_{\text{ctol}} = 6 \text{ \AA}$ (see Eq. 5.28) are shown. Red dashed lines indicate clustering thresholds. (D) Resulting clusters: The symbol size reflects the number of solutions in the shown cluster. Horizontal line represents confidence threshold as defined by $\chi_r^2 < \chi_{r,\text{min}}^2 + 0.378$ (blue solid line, ~84% confidence, ~1.4 σ , see Eq. 5.36). The threshold is derived from the chi-squared distribution with 14 degrees of freedom ($\chi^2 < 19.3$).

5.3.13.2 Model discrimination of docking/screening results.

Solutions are considered ambiguous if the respective χ_r^2 values do not differ significantly. We typically apply a threshold given by Eq. 5.36:

$$\chi_r^2 < \chi_{r,\text{min}}^2 + [2/(n-p)]^{1/2} = \chi_{r,\text{max}}^2 \quad \text{Eq. 5.36}$$

where n is the number of distance constraints and p is the number of degrees of freedom. The following values for n and p were used in this work: $n = 20$ for the rigid body model of the RT:dp/dt complex (see Section 4.2.1), 16 for screening MD models of the ssDNA template overhang (see Section 4.2.2) and 51 for the RNA4WJ (see Section 4.3). $p = 6$ for the rigid body model of the RT:dp/dt complex (see Section 4.2.1), 0 for screening MD models of the template overhang (see Section 4.2.2) and 9 for the RNA4WJ

(see Section 4.3). $\chi^2_{r,\max}$ roughly corresponds to the variance of the chi-squared distribution of $2 \times (\text{degrees of freedom})$ (Soong 2004) (blue line in Figure 5.8D). The fact that $\chi^2_{r,\min}$ is often (especially for the RT:dp/dt complex) larger than one is attributed to systematic experimental errors and to possible violations of the AV and/or rigid body models. Other criteria defining different levels of significance can be applied here in a straightforward way.

6 Conclusions and Outlook

The method introduced in this work opens up new possibilities in structural biology which are not accessible to other methods. As we have shown, accurate structural modeling becomes applicable to heterogeneous and dynamic molecules. It was even possible to establish models for two coexisting minor conformers in parallel with precisions of ~ 2 Å.

In “linker paper” it was shown that the modeling of dye position distributions is the key to overcome the problem of uncertainties due to flexible dye linkers. For this we used the relatively simple and fast accessible volume (AV) algorithm (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008). However, to apply AV modeling, certain prerequisites need to be fulfilled: (1) The environment of the dye must be well defined. (2) There must be only steric interactions between the fluorophore and the macromolecule. (3) The dye diffuses freely in the sterically allowed volume. For nucleic acids we were able to experimentally verify (2) and (3) for most of the linkers used in “linker paper”. It was furthermore shown that for most linkers the translational motions of the dyes are completely averaged out in the timescale of ms while they are slow on the timescale of the fluorescence lifetimes. Fitting distance distributions to eTCSPC data (ns time resolution) yielded widths similar to the ones predicted by AV. Furthermore, dye position distributions modeled by MD are consistent with AV results. In a benchmark study on dsRNA the accuracy of AV modeling could be verified. An RMSD of 1.3 Å between predicted and measured distances was achieved, however, only when properly averaging over dye position distributions. In summary, for known dye environments, we have proven the applicability of the AV approach for structural modeling of nucleic acids. For proteins the assumptions of the AV approach are not necessarily given. However, its use is still justified as long as oversampling with sufficient FRET distance restraints compensates for systematic errors. To solve the second major limitation to the accuracy of FRET, κ^2 -related uncertainties, we established a rigorous procedure for their estimation and minimization. We further introduced short dye linkers to minimize position uncertainties. It was shown that the nature of the linker strongly affects the radius of the dye’s accessible volume (6 to 16 Å). For short linkers additional κ^2 -related uncertainties are clearly outweighed by better defined dye positions. This makes them particularly suitable for undefined environments.

In “RT manuscript” a comprehensive toolbox for FRET-restrained modeling of biomolecules and their complexes is introduced for quantitative applications in structural biology. A dramatic improvement of the precision is achieved through proper consideration of dye position distributions. A procedure was

developed to characterize the uniqueness and precision of FRET-restrained models, based on a precise spectroscopic estimation of “input” uncertainties. In combination with advanced computer simulations the method allows for a detailed molecular description of the proposed structure models. The accuracy of this approach is demonstrated by docking the double-stranded part of a DNA/DNA 19/35 primer/template to HIV-1 reverse transcriptase. The derived model agrees with the known X-ray structure with an RMSD of 0.5 Å. Moreover, the formerly unknown configuration of the flexible single strand template overhang was determined by FRET-guided “screening” of a large structural ensemble created by molecular dynamics simulations. We found a preferential structure with the 5'-end of the overhang bound to the fingers domain of RT, which might have important implications concerning proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis.

In “RNA4WJ manuscript” it was shown that FRET restrained high-precision structural modeling was able to establish highly accurate models of the three coexisting conformers of the RNA4WJ. Although, compared to e.g. NMR spectroscopy, only few (51) distance restraints were used, FRET-restrained MD and coarse-grained simulations provided meaningful full atom models. The dynamic and heterogeneous behavior of the RNA4WJ makes structure determination of the RNA4WJ with traditional methods of structural biology very difficult. It is, furthermore, the first time that coexisting transient minor conformers are structurally solved by FRET. Although the sequence of the RNA4WJ was significantly altered compared to the natural hairpin ribozyme, we believe that the present results are nevertheless structurally relevant because the sequence in the junction region was left unchanged; also, because of the absence of interactions between the arms, the overall conformations are entirely determined by the junction itself. Thus, for the first time, the structure of an RNA4WJ was determined in a state uninfluenced by, e.g., interactions between the arms or with other macromolecules.

Perhaps the most important feature in our method is to explicitly account for dye position distributions, which we determined with the simple AV approach. Thus, one obvious possibility to improve the method presented in this thesis would be to use a more sophisticated technique such as MD simulations to obtain dye distributions. This way, effects like sticking of the dye could be taken into account. Furthermore, the assumption of equally distributed probability throughout the AV is questionable. Position distributions resulting from MD would be more realistic and would yield more accurate DA distance predictions during structural modeling. Another major advantage would be the accurate prediction of translational and rotational diffusion rates of the dye. When measuring transition rates due to intrinsic motion of the molecule via filtered fluorescence correlation spectroscopy (Felekyan,

Kalinin et al. 2012) this would allow to separate molecule and linker dynamics and would yield more accurate results. Thus, conformational control of biomolecular function in complex systems could be studied within a nanosecond to minute time range and associated to detailed dynamic structures without spatial averaging. In Section 4.1.8.2 preliminary results of an MD simulation for translational movements of Alexa488 and Cy5 on dsRNA are presented. The MD trajectory can, of course, be further analyzed with respect to the other above mentioned features, such as rotational movement and dye sticking.

As demonstrated in this work, experimental restraints from FRET can be used in computational modeling of structures by post-filtering of, e.g., MD-derived ensembles. However, as a mean to improve the method, AV modeling could be implemented into a computational method. Using simple starting structures based on secondary structure elements (e.g. nucleic acid sequence) FRET restraints could then be used as a guiding potential which leads to relaxed solutions unbiased by, e.g., rigid body assumptions. Moreover, FRET-restrained high-precision structural modeling is also applicable to structurally heterogeneous and flexible proteins, whose overall structures are notoriously difficult to determine. It could, for example, be used to discriminate between several possible solutions determined by homology or coarse grained simulations.

7 Acknowledgements

I would like to thank all the people who supported and helped me during my time as a PhD student and, thus, contributed to this project:

First, I thank **Prof. Dr. Claus Seidel** who gave me the opportunity to work on this very interesting project, to use the excellent equipment in his labs and to benefit from his and his group's rich experience in the field of fluorescence spectroscopy, and, furthermore, for his enthusiastic support and advice and for fruitful discussions.

Second, I am especially grateful to **Dr. Stanislav Kalinin** who has been a great colleague and an even better supervisor. Thank you for always having time to for my (many) questions and for the many pushes in the right direction.

Hayk Vardanyan for performing many of the measurements shown in this thesis and for being a great guy to share the office with.

Dr. Stefan Marawske, Thomas Peulen and **Denis Dörr** for proofreading this thesis.

The whole **Seidel group** for always being supportive and helpful.

Dr. Ralf Kühnemuth and **Dr. Suren Felekyan** for the measurements and data analysis in Section 4.1.8.1.

My second supervisor **Prof. Dr. Dieter Willbold**.

The NRW Research School Biostruct for funding and especially BioStruct's coordinators **Dr. Christian Dumpitak** and **Dr. Cordula Kruse** for always being helpful and supportive.

My wife **Rike** for the support at home.

8 References

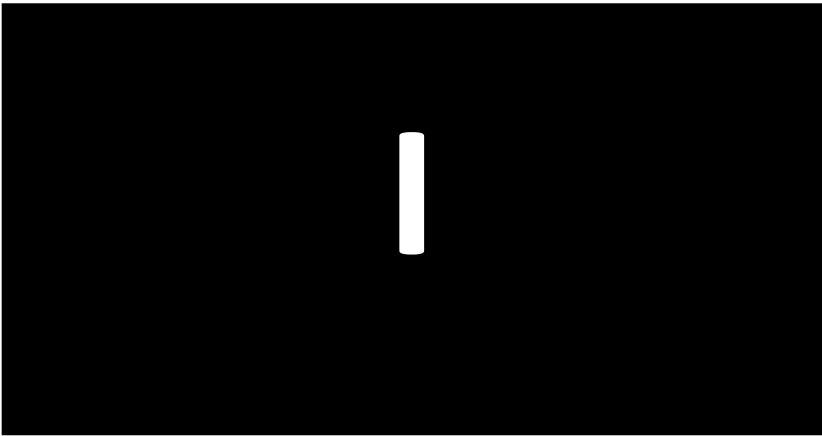
- Andrecka, J., R. Lewis, et al. (2008). "Single-molecule tracking of mRNA exiting from RNA polymerase II." Proceedings of the National Academy of Sciences of the United States of America **105**(1): 135-140.
- Andrews, D. L. (1989). "A Unified Theory of Radiative and Radiationless Molecular-Energy Transfer." Chemical Physics **135**(2): 195-201.
- Andrews, D. L. and D. S. Bradshaw (2004). "Virtual photons, dipole fields and energy transfer: a quantum electrodynamical approach." European Journal of Physics **25**(6): 845-858.
- Antonik, M., S. Felekyan, et al. (2006). "Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis." Journal of Physical Chemistry B **110**(13): 6970-6978.
- Atkins, P. W. and J. de Paula (2001). Atkins' Physical Chemistry. Oxford, U.K., Oxford University Press.
- Balci, H., S. Arslan, et al. (2011). "Single-Molecule Nanopositioning: Structural Transitions of a Helicase-DNA Complex during ATP Hydrolysis." Biophysical Journal **101**(4): 976-984.
- Bannwarth, W. (1987). "GENE TECHNOLOGY - A CHALLENGE FOR A CHEMIST." Chimia **41**(9): 302-317.
- Best, R. B., K. A. Merchant, et al. (2007). "Effect of flexibility and cis residues in single-molecule FRET studies of polyproline." Proceedings of the National Academy of Sciences of the United States of America **104**(48): 18964-18969.
- Borgia, M. B., A. Borgia, et al. (2011). "Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins." Nature **474**(7353): 662-U142.
- Braslavsky, S. E., E. Fron, et al. (2008). "Pitfalls and limitations in the practical use of Förster's theory of resonance energy transfer." Photochemical & Photobiological Sciences **7**(12): 1444-1448.
- Brunger, A. T., P. Strop, et al. (2011). "Three-dimensional molecular modeling with single molecule FRET." Journal of Structural Biology **173**(3): 497-505.
- Buck, J., J. Noeske, et al. (2010). "Dissecting the influence of Mg(2+) on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain." Nucleic Acids Research **38**(12): 4143-4153.
- Buck, J., A. Wacker, et al. (2011). "Influence of ground-state structure and Mg(2+) binding on folding kinetics of the guanine-sensing riboswitch aptamer domain." Nucleic Acids Research **39**(22): 9768-9778.
- Cai, Q., A. K. Kusnetzow, et al. (2007). "Nanometer distance measurements in RNA using site-directed spin Labeling." Biophysical Journal **93**(6): 2110-2117.
- Case, D. A., T. E. Cheatham, et al. (2005). "The Amber biomolecular simulation programs." Journal of Computational Chemistry **26**(16): 1668-1688.
- Choi, U. B., P. Strop, et al. (2010). "Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex." Nature Structural & Molecular Biology **17**(3): 318-U384.
- Chowdhury, S., C. Maris, et al. (2006). "Molecular basis for temperature sensing by an RNA thermometer." EMBO Journal **25**(11): 2487-2497.
- Dale, R. E., J. Eisinger, et al. (1979). "Orientational freedom of molecular probes - Orientation factor in intra-molecular energy transfer." Biophysical Journal **26**(2): 161-193.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. Palo Alto, CA, USA, DeLano Scientific.
- Deniz, A. A., M. Dahan, et al. (1999). "Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations." Proc.Natl.Acad.Sci.USA **96**(7): 3670-3675.

- Dolghih, E., W. Ortiz, et al. (2009). "Theoretical Studies of Short Polypyrrole Systems: Recalibration of a Molecular Ruler." Journal of Physical Chemistry A **113**(16): 4639-4646.
- Dolghih, E., A. E. Roitberg, et al. (2007). "Fluorescence resonance energy transfer in dye-labeled DNA." Journal of Photochemistry and Photobiology A-Chemistry **190**(2-3): 321-327.
- Duckett, D. R., A. I. H. Murchie, et al. (1995). "The Global Folding of 4-Way Helical Junctions in Rna, Including That in U1 Snrna." Cell **83**(6): 1027-1036.
- Efron, B. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion." Annals of Statistics **14**(4): 1301-1304.
- Eggeling, C., S. Berger, et al. (2001). "Data registration and selective single-molecule analysis using multi-parameter fluorescence detection." Journal of Biotechnology **86**(3): 163-180.
- Eggeling, C., J. Schaffer, et al. (2001). "Homogeneity, transport, and signal properties of single Ag particles studied by single-molecule surface-enhanced resonance Raman scattering." Journal of Physical Chemistry A **105**(15): 3673-3679.
- Eggeling, C., J. Widengren, et al. (2006). "Analysis of Photobleaching in Single-Molecule Multicolor Excitation and Förster Resonance Energy Transfer Measurements." Journal of Physical Chemistry A **110**(9): 2979-2995.
- Esnouf, R., J. Ren, et al. (1995). "Mechanism of inhibition of HIV-1 reverse transcriptase by non nucleoside inhibitors." Nature Structural Biology **2**: 303-308.
- Fedor, M. J. (1999). "Tertiary structure stabilization promotes hairpin ribozyme ligation." Biochemistry **38**(34): 11040-11050.
- Felekyan, S., S. Kalinin, et al. (2012). "Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules." ChemPhysChem **13**: 1036-1053.
- Felekyan, S., R. Kühnemuth, et al. (2005). "Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection." Review of Scientific Instruments **76**(8): 083104.
- Förster, T. (1948). "Zwischenmolekulare Energiewanderung und Fluoreszenz." Annalen der Physik **437**(2): 55-75.
- Goette, M., J. W. Rausch, et al. (2010). "Reverse transcriptase in motion: Conformational dynamics of enzyme-substrate interactions." Biochimica et Biophysica Acta - Proteins and Proteomics **1804**(5): 1202-1212.
- Gonçalves, M. S. T. (2009). "Fluorescent Labeling of Biomolecules with Organic Probes." Chemical Reviews **109**(1): 190-212.
- Ha, T., T. Enderle, et al. (1996). "Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor." Proc.Natl.Acad.Sci.USA. **93**: 6264-6268.
- Haas, E. (2005). "The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements." ChemPhysChem. **6**(5): 858-870.
- Haas, E., M. Wilchek, et al. (1975). "Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer." Proceedings of the National Academy of Sciences of the United States of America **72**(5): 1807-1811.
- Henzler-Wildman, K. and D. Kern (2007). "Dynamic personalities of proteins." Nature **450**(7172): 964-972.
- Hoefling, M., N. Lima, et al. (2011). "Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polypyrroles through a Hybrid Atomistic Simulation and Monte Carlo Approach." Plos One **6**(5): e19791
- Hohng, S., T. J. Wilson, et al. (2004). "Conformational flexibility of four-way junctions in RNA." Journal of Molecular Biology **336**(1): 69-79.

- Ivanov, V., M. Li, et al. (2009). "Impact of Emission Anisotropy on Fluorescence Spectroscopy and FRET Distance Measurements." Biophysical Journal **97**(3): 922-929.
- Jacobo-Molina, A., J. Ding, et al. (1993). "Crystal Structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA." Proc.Natl.Acad.Sci.USA. **90**: 6320-6324.
- Kalinin, S., S. Felekyan, et al. (2007). "Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer." Journal of Physical Chemistry B **111**(34): 10253-10262.
- Kalinin, S., S. Felekyan, et al. (2008). "Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis." Journal of Physical Chemistry B **112**(28): 8361-8374.
- Kalinin, S., E. Sisamak, et al. (2010). "On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits." Journal of Physical Chemistry B **114**: 6197–6206.
- Kalinin, S., A. Valeri, et al. (2010). "Detection of structural dynamics by FRET: A photon distribution and fluorescence lifetime analysis of systems with multiple states." J. Phys. Chem. B **114**: 7983-7995.
- Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. (2012). "FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang."
- Kask, P., P. Piksarv, et al. (1989). "Separation of the rotational contribution in fluorescence correlation experiments." Biophysical Journal **55**: 213-220.
- Kensch, O., B. A. Connolly, et al. (2000). "HIV-1 Reverse Transcriptase-Pseudoknot RNA Aptamer Interaction Has a Binding Affinity in the Low Picomolar Range Coupled with High Specificity." Journal of Biological Chemistry **275**(24): 18271-18278.
- Kensch, O., T. Restle, et al. (2000). "Temperature-dependent Equilibrium between the Open and Closed Conformation of the p66-Subunit of HIV-1 Reverse Transcriptase Revealed by Site-directed Spin Labelling." Journal of Molecular Biology **301**: 1029-1039.
- Kohlstaedt, L. A., J. Wang, et al. (1992). "Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor." Science **256**(5065): 1783-1790.
- Koshioka, M., K. Sasaki, et al. (1995). "Time-Dependent Fluorescence Depolarization Analysis in Three Dimensional Microscopy." Applied Spectroscopy **49**(2): 224-228.
- Kühnemuth, R. and C. A. M. Seidel (2001). "Principles of Single Molecule Multiparameter Fluorescence Spectroscopy." Single Molecules **2**(4): 251-254.
- Lafontaine, D. A., D. G. Norman, et al. (2002). "The global structure of the VS ribozyme." EMBO Journal **21**(10): 2461-2471.
- Lakowicz, J. R. (1999). Principles of Fluorescence Spectroscopy. New York, Kluwer Academic/ Plenum Publishers.
- Lakowicz, J. R. (2006). Principles of Fluorescence Spectroscopy. New York, Springer.
- Liu, S. X., E. A. Abbondanzieri, et al. (2008). "Slide into action: Dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates." Science **322**(5904): 1092-1097.
- Magde, D., E. L. Elson, et al. (1974). "Fluorescence Correlation Spectroscopy. II. An Experimental Realization." Biopolymers **13**: 29-61.
- Margittai, M., J. Widengren, et al. (2003). "Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1." Proc.Natl.Acad.Sci.USA. **100**: 15516-15521.
- Martyna, G. J., Tobias, D.J., Klein M.L. (1994). "Constant pressure molecular dynamics algorithms." Journal of Chemical Physics **101**: 4177-4189.
- Marushchak, D., S. Grenklo, et al. (2007). "Fluorescence depolarization studies of filamentous actin analyzed with a genetic algorithm." Biophysical Journal **93**(9): 3291-3299.

- McCann, J. J., L. Q. Zheng, et al. (2011). "Domain Orientation in the N-Terminal PDZ Tandem from PSD-95 Is Maintained in the Full-Length Protein." Structure **19**(6): 810-820.
- Mekler, V., E. Kortkhonja, et al. (2002). "Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex." Cell **108**(5): 599-614.
- Moerner, W. E. and L. Kador (1989). "Optical Detection and Spectroscopy of Single Molecules in a Solid." Physical Review Letters **62**(21): 2535-2538.
- Murchie, A. I. H., J. B. Thomson, et al. (1998). "Folding of the Hairpin Ribozyme in Its Natural Conformation Achieves Close Physical Proximity of the Loops." Molecular Cell **1**: 873-881.
- Muschielok, A., J. Andrecka, et al. (2008). "A nano-positioning system for macromolecular structural analysis." Nature Methods **5**(11): 965-971.
- O'Connor, D. V. and D. Phillips (1984). Time-correlated Single Photon Counting. New York, Academic Press.
- Olofsson, M., S. Kalinin, et al. (2006). "Tryptophan-BODIPY: A versatile donor-acceptor pair for probing generic changes of intraprotein distances." Physical Chemistry Chemical Physics **8**(26): 3130-3140.
- Orr, J. W., P. J. Hagerman, et al. (1998). "Protein and Mg²⁺-induced conformational changes in the S15 binding site of 16 S ribosomal RNA." Journal of Molecular Biology **275**(3): 453-464.
- Orrit, M. and J. Bernard (1990). "Single Pentacene Molecules Detected by Fluorescence Excitation in a p-Terphenyl Crystal." Physical Review Letters **65**(21): 2716-2719.
- Peletskaya, E. N., A. A. Kogon, et al. (2004). "Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA." Journal of Virology **78**(7): 3387-3397.
- Pley, H. W., K. M. Flaherty, et al. (1994). "3-Dimensional Structure of a Hammerhead Ribozyme." Nature **372**(6501): 68-74.
- Rasnik, I., S. A. McKinney, et al. (2006). "Nonblinking and longlasting single-molecule fluorescence imaging." Nature Methods **3**(11): 891-893.
- Rother, K., M. Rother, et al. (2012). Template-based and template-free modeling of RNA 3D structure: Inspirations from protein structure modeling. In RNA 3D structure analysis and prediction. N. B. Leontis and E. Westhof. Berlin, Springer-Verlag.
- Rothwell, J. P. (2002). Structural Investigation of the HIV-1 RT using single pair Fluorescence Energy Transfer. Max-Planck-Institut für Molekulare Physiologie. Dortmund, University of Dortmund.
- Rothwell, P. J., S. Berger, et al. (2003). "Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes." Proceedings of the National Academy of Sciences of the United States of America **100**: 1655-1660.
- Rupert, P. B., A. P. Massey, et al. (2002). "Transition state stabilization by a catalytic RNA." Science **298**(5597): 1421-1424.
- Sabir, T., G. F. Schroder, et al. (2011). "Global Structure of Forked DNA in Solution Revealed by High-Resolution Single-Molecule FRET." Journal of the American Chemical Society **133**(5): 1188-1191.
- Sakon, J. J. and K. R. Weninger (2010). "Detecting the conformation of individual proteins in live cells." Nature Methods **7**(3): 203-U256.
- Schmidt, C., R. Welz, et al. (2000). "RNA double cleavage by a hairpin-derived twin ribozyme." Nucleic Acids Research **28**(4): 886-894.
- Scott, W. G., J. T. Finch, et al. (1995). "The Crystal-Structure of an All-Rna Hammerhead Ribozyme - a Proposed Mechanism for Rna Catalytic Cleavage." Cell **81**(7): 991-1002.
- Shen, Z. and P. J. Hagerman (1994). "Conformation of the central, 3-helix junction of the 5 S ribosomal-RNA of *Sulfolobus acidocaldarius*." Journal of Molecular Biology **241**(3): 415-430.
- Shera, E. B., N. K. Seitzinger, et al. (1990). "Detection of single fluorescent molecules." Chemical Physics Letters **174**(6): 553-557.

- Sindbert, S., S. Kalinin, et al. (2011). "Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity." Journal of the American Chemical Society **133**(8): 2463-2480.
- Sisamakris, E., A. Valeri, et al. (2010). "Accurate single-molecule FRET studies using multiparameter fluorescence detection." Methods in Enzymology **475**: 456-514.
- Soong, T. T. (2004). Fundamentals of probability and statistics for engineers. West Sussex, England, John Wiley & Sons.
- Stryer, L. (1978). "Fluorescence energy transfer as a spectroscopic ruler." Annual Review of Biochemistry **47**: 819-846.
- Tan, E., T. J. Wilson, et al. (2003). "A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate." Proceedings of the National Academy of Sciences of the United States of America **100**(16): 9308-9313.
- Tokuriki, N. and D. S. Tawfik (2009). "Protein Dynamism and Evolvability." Science **324**(5924): 203-207.
- van der Meer, B. W., G. Cooke, et al. (1994). Resonance Energy Transfer: Theory and Data. New York, VCH Publishers.
- VanBeek, D. B., M. C. Zwier, et al. (2007). "Fretting about FRET: Correlation between kappa and R." Biophysical Journal **92**(12): 4168-4178.
- Walter, N. G., J. M. Burke, et al. (1999). "Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction." Nature Structural Biology **6**(6): 544-549.
- Weiss, S. (1999). "Fluorescence spectroscopy of single biomolecules." Science **283**(5408): 1676-1683.
- Wohrl, B. M., R. Krebs, et al. (1999). "Refined model for primer/template binding by HIV-1 reverse transcriptase: Pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate." Journal of Molecular Biology(2): 333-344.
- Woźniak, A. K., G. Schröder, et al. (2008). "Single molecule FRET measures bends and kinks in DNA." Proc.Natl.Acad.Sci.USA. **105**: 18337-18342.
- Zhao, Z. Y., T. J. Wilson, et al. (2000). "The folding of the hairpin ribozyme: Dependence on the loops and the junction." Rna-a Publication of the Rna Society **6**(12): 1833-1846.




Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity

Simon Sindbert,^{†,||} Stanislav Kalinin,^{*,†,||} Hien Nguyen,[‡] Andrea Kienzler,[§] Lilia Clima,[§] Willi Bannwarth,[§] Bettina Appel,[‡] Sabine Müller,^{*,‡} and Claus A. M. Seidel^{*,†}

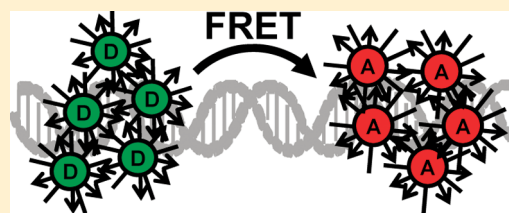
[†]Institut für Physikalische Chemie, Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstrasse 1, Geb 26.32, 40225 Düsseldorf, Germany

[‡]Institut für Biochemie, Bioorganische Chemie, Ernst-Moritz-Arndt-Universität Greifswald, Felix-Hausdorff-Strasse 4, 17487, Greifswald, Germany

[§]Fakultät für Chemie und Biochemie, Albert-Ludwigs Universität Freiburg, AK Bannwarth, Albertstrasse 21, 79104, Freiburg, Germany

 Supporting Information

ABSTRACT: In Förster resonance energy transfer (FRET) experiments, the donor (D) and acceptor (A) fluorophores are usually attached to the macromolecule of interest via long flexible linkers of up to 15 Å in length. This causes significant uncertainties in quantitative distance measurements and prevents experiments with short distances between the attachment points of the dyes due to possible dye–dye interactions. We present two approaches to overcome the above problems as demonstrated by FRET measurements for a series of dsDNA and dsRNA internally labeled with Alexa488 and Cy5 as D and A dye, respectively. First, we characterize the influence of linker length and flexibility on FRET for different dye linker types (long, intermediate, short) by analyzing fluorescence lifetime and anisotropy decays. For long linkers, we describe a straightforward procedure that allows for very high accuracy of FRET-based structure determination through proper consideration of the position distribution of the dye and of linker dynamics. The position distribution can be quickly calculated with geometric accessible volume (AV) simulations, provided that the local structure of RNA or DNA in the proximity of the dye is known and that the dye diffuses freely in the sterically allowed space. The AV approach provides results similar to molecular dynamics simulations (MD) and is fully consistent with experimental FRET data. In a benchmark study for dsA-RNA, an rmsd value of 1.3 Å is achieved. Considering the case of undefined dye environments or very short DA distances, we introduce short linkers with a propargyl or alkenyl unit for internal labeling of nucleic acids to minimize position uncertainties. Studies by ensemble time correlated single photon counting and single-molecule detection show that the nature of the linker strongly affects the radius of the dye's accessible volume (6–16 Å). For short propargyl linkers, heterogeneous dye environments are observed on the millisecond time scale. A detailed analysis of possible orientation effects (κ^2 problem) indicates that, for short linkers and unknown local environments, additional κ^2 -related uncertainties are clearly outweighed by better defined dye positions.



1. INTRODUCTION

Measuring distances within biomolecules via Förster resonance energy transfer (FRET) has been a very useful technique in the field of structural biology for decades.^{1–5} It is based on the fact that an excited fluorescent dye (donor) can transfer energy to another dye (acceptor) if the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor. The efficiency of this energy transfer strongly depends on the distance between the dyes^{3,6–9} allowing for donor–acceptor distance (R_{DA}) measurements in the range of about 20–100 Å. In the past years, FRET measurements on single molecules (smFRET) have become possible.^{10,11} smFRET largely overcomes many problems of ensemble FRET, including species and time averaging, incomplete or unspecific labeling, as well as position-dependent donor or acceptor quenching artifacts.^{12,13}

However, two problems commonly arise when performing quantitative FRET measurements. First, the fluorescent dyes are typically attached to the biomolecule via long flexible linkers, for example, the “standard” C6 (hexamethylen) linker. The overall length of the linkage from the attachment point to the center of the chromophore is, thus, given by the length of the linker and the internal chemical structure of the dye and amounts to up to 20 Å. This yields a significant uncertainty in dye position and quenching environment.^{14–18} Second, the FRET efficiency also depends on the relative orientation of the transition dipole moments of the two dyes,^{3,7,8,19–21} which is expressed by the orientation factor κ^2 . It can range from 0 to 4 and has a strong influence on the measured FRET efficiencies.

Received: June 29, 2010

Published: February 3, 2011

The effect of long dye linkers and the problems they can cause for quantitative FRET measurements are well-known.^{2,22–24} There are approaches to circumvent this problem through modeling of the dye and linker motions,^{25–29} and, thus, to calculate the accessible volume and mean position of the dye. This approach, however, works only for well-defined dye environments such as a straight double helix of double-stranded (ds) DNA or RNA. In many biomolecules, for example, complex RNA structures, the local structure is unknown, which makes such simulations impossible. Furthermore, there are approaches to minimize the dye position uncertainties, thus making modeling less important. Nucleobases can be replaced by fluorescent analogues,^{30,31} and fluorophores can be covalently linked as end-caps of DNA duplexes³² or stacked at the ends of the helices in the manner of additional base pairs.²¹ However, this strongly restricts dye reorientation, and κ^2 cannot be assumed to be 2/3.

In this work, we suggest alternative procedures of obtaining highly accurate FRET-based structural information using internally labeled nucleic acids with dye linkers of different length and flexibility. Depending on whether the structure of the local environment of the dye is known, two cases have to be considered. If a simulation is possible, we demonstrate that an easily applicable accessible volume (AV) simulation method^{28,29} provides realistic dye position distributions, which are consistent MD data and experimental smFRET results. For the case of an unknown local structure, we introduce alternative short dye linkers. They significantly decrease dye position uncertainty while allowing the fluorophores to rotate somewhat freely. In particular, short linkers are also expected to be more suitable for measuring short distances because, in this range, the length of long linkers becomes comparable to the absolute distances between donor and acceptor dye (R_{DA}). To consider the implications of using long and short dye linkers on quantitative FRET measurements, we present systematic studies for DNA and RNA to demonstrate the influence of the linker flexibility and length on the fluorescence properties of two representative dyes.

In section 2, we introduce the new short dye linkers for labeling of nucleic acids. We synthesized the modified nucleoside phosphoramidites shown in Figure 1 and incorporated them into oligonucleotides of defined sequence (section S1.1 in the Supporting Information). The alkenyl linkers (Figure 1A and B) were introduced by Heck chemistry,^{33,34} and the propargyl linkers (Figure 1C and D) were introduced by Sonogashira coupling^{33,35,36} starting from 5-iodo-2'-deoxyuridine or 5-iodouridine.

In section 3.1, we studied the influence of the linker flexibility and length on the individual fluorescence properties (fluorescence quantum yield, lifetime, and anisotropy) of Alexa488 and Cy5 dyes for DNA and RNA.

In sections 3.2 and 3.3, we characterized experimentally and theoretically the broadening of interdye distances due to dye linker motions. We performed quantitative FRET distance measurements using internally labeled dsDNA and dsRNA as test systems using single-molecule multiparameter fluorescence detection (smMFD) and ensemble time-correlated single-photon-counting (eTCSPC) techniques. We found that for both types of FRET measurements the modeling of the dye position and the R_{DA} distribution is essential for the quantitative interpretation of the observed FRET efficiencies. However, for short dye linkers, these corrections are much less important. For the simulation of the environment of fluorescent dye positions, we further developed simple geometric computations to calculate the sterically accessible volume (AV), which has been proposed to predict

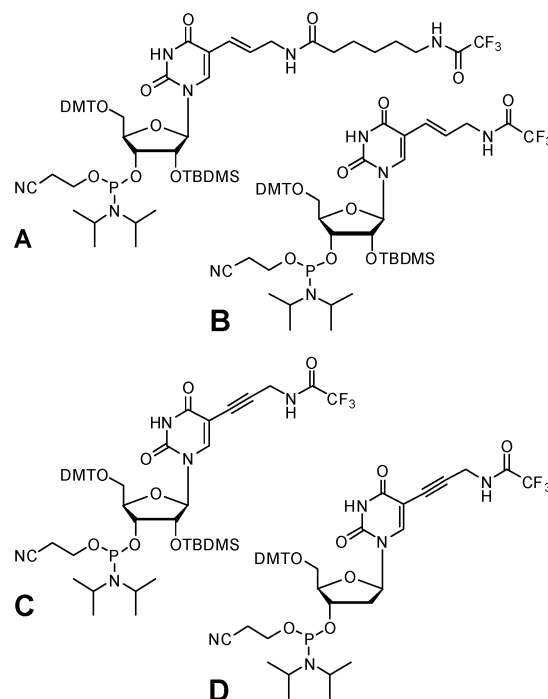


Figure 1. Modified RNA (A, B, and C) and DNA (D) building blocks with trifluoroacetamide-protected amino groups, which contain linkers of different length and flexibility.

possible positions for EPR and FRET labels.^{28,29} For defined environments, the AV approach provides dye position distributions that are very similar to those obtained by molecular dynamics (MD) simulations.²⁵ However, in contrast to the previous work,^{29,37} we do not assume the dye to have a fixed static position within its AV but consider all accessible dye positions as equally populated. We demonstrate that this approach provides a far better approximation of the dye behavior. Furthermore, we account for the three different dimensions of the fluorophores. We applied this methodology in a FRET benchmark study for ds A-RNA with long C6 dye linkers, where a very good rmsd value of 1.3 Å was achieved. This demonstrates that long dye linkers can be safely used in FRET experiments, if three conditions are fulfilled: (1) the local structure of RNA or DNA in the proximity of the dye is known, (2) the R_{DA} distance is larger than the sum of the linkage lengths and larger than $\sim 0.7 \times$ Förster radius (that is, $R_{DA} > 35\text{--}40$ Å for our dyes), and (3) there are no stacking interactions between the dye and the nucleic acid. Using a multiparameter fluorescence detection setup³⁸ allows one to easily test for the presence of such interactions.

In section 3.4, we studied the influence of the linker type on the additional broadening of R_{DA} distributions due to a very slow ($>$ milliseconds) interchange between distinct dye environments. We can show that this effect becomes significant for short and stiff dye linkers.

In section 3.5, we introduce a rigorous procedure to minimize the uncertainties in the orientation factor κ^2 , which is necessary because orientational distribution of both D and A is not strictly isotropic even for the longest linkers. We used time-resolved fluorescence anisotropy decays of D-only and A-only molecules and the FRET-sensitized acceptor anisotropy decay to determine the residual anisotropies, which allow us to compute probability distributions for possible values of κ^2 and estimate errors due to

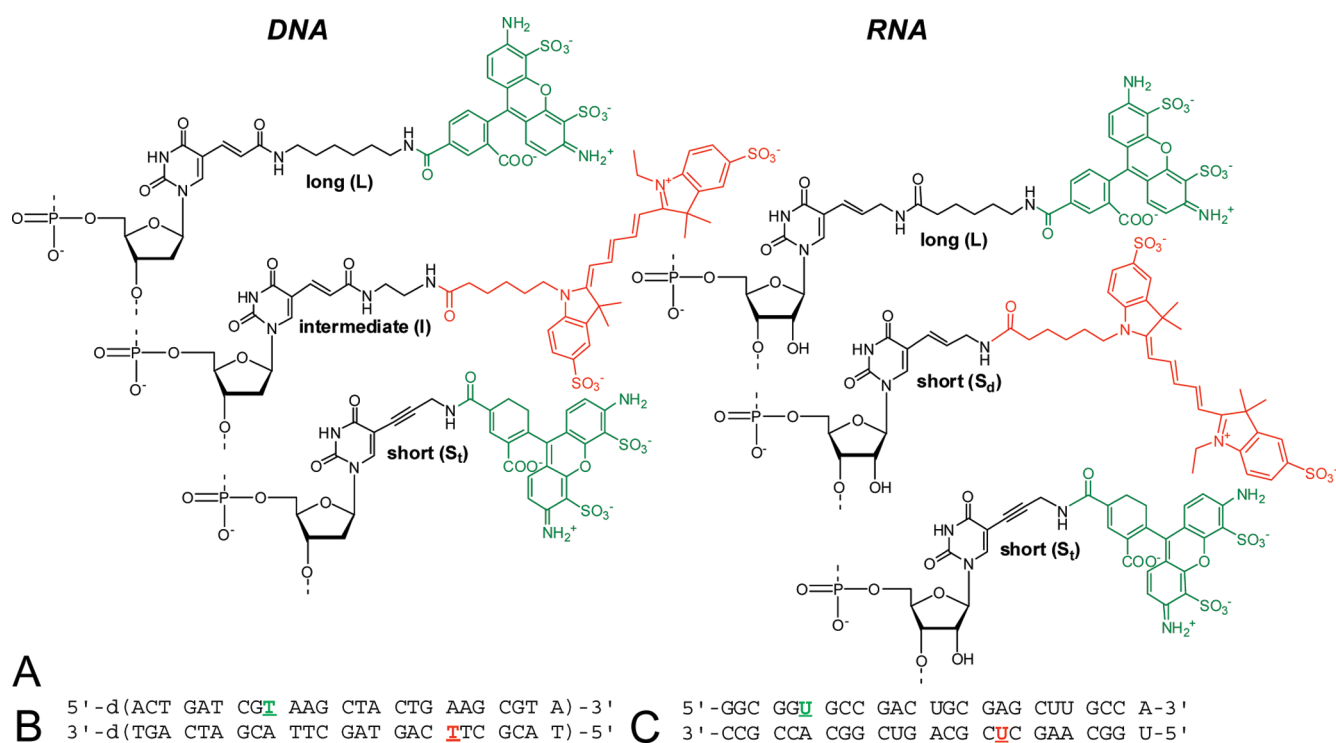


Figure 2. (A) Exemplary structures of the linkers S_t (propargyl) and S_d , I, and L (all alkenyl) with the dyes (Alexa488 and Cy5) used in DNA (left) and RNA (right). The donor and acceptor dyes are depicted in green and red color, respectively. Donor and acceptor dyes have been used in combination with each of the shown linkers. (B,C) Duplex structures of DNA1 (B) and RNA1 (C). Labeling positions are depicted in green and red for Alexa488 and Cy5, respectively.

uncertainties in the relative orientation of the dyes. Even if the residual anisotropies increase significantly for shorter linkers, it turns out that κ^2 errors increase only slightly. Thus, short and flexible linkers can be recommended for unknown local environments, because additional κ^2 -related uncertainties are clearly outweighed by better defined dye positions.

2. MATERIALS AND METHODS

Synthesis of Labeled DNA and RNA. General. All reactions were carried out in dry solvents under argon atmosphere. All solvents and reagents were purchased from commercial sources and used as supplied unless otherwise stated. The solvents used in palladium coupling reactions were freed from oxygen. All products were visualized on TLC plates (aluminum sheets coated with silica gel 60 F 254, 0.2 mm thickness) at 254 nm ultraviolet light. Column chromatography was performed using silica gel type 60 ACC 35–70 μm . ^1H , ^{13}C , and ^{31}P NMR spectra were measured either on AC 250, ARX 300, AM 400, DRX 400, DRX 500, and DRX 600 systems from Bruker or on a Mercury VX 300 system from Varian using CDCl_3 , CD_3CN , or $[d_6]\text{DMSO}$ as solvent. UV spectra were recorded on a Perkin-Elmer-Lambda-35-UV/vis spectrometer or on a Varian Nano Drop ND-1000 spectrophotometer.

Linkers. The alkynyl linker, *N*-propargyltrifluoroacetamide (S_t), was synthesized according to Stockwell;³⁹ the alkenyl linkers, *N*-allyltrifluoroacetamide (S_d) and *N*-allyl-6-(*N*-trifluoroacetamido)hexanamide (L), were synthesized as described by Dey and Sheppard³⁶ and Meller and Brown.⁴⁰ See Figures 1 and 2 for the respective structures.

5-(3-Trifluoroacetamidopropargyl)-2'-deoxyuridine. A two-neck flask was charged with 5'-iodo-uridine (2.00 g, 5.65 mmol), Pd/C (304 mg, 0.282 mmol), CuI (214 mg, 1.13 mmol, 20 mol %), and Amberlite IRA 67 (5.14 g). The compounds were dried under high vacuum and afterward kept under argon. Propargyltrifluoroacetamide

(1.72 g, 11.3 mmol, 2.3 equiv) and 56 mL of anhydrous DMF were added. Anhydrous argon was bubbled through this suspension for 5 min. The apparatus was again degassed and flooded with anhydrous argon. The reaction mixture was stirred at 50 °C for 17 h, and afterward cooled to room temperature and filtered over Celite. The Celite was washed with 50 mL of $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (5:1). The solvent of the combined organic phases was removed under reduced pressure. Remaining DMF was removed by condensation at 40 °C and 4×10^{-2} mbar. The crude product was purified by column chromatography on a Büchi Sepacore chromatography system (5 \times 15 cm, F1–6, 20 mL; F7–30, 10 mL; $\text{CHCl}_3/\text{MeOH}$ = 8.25:1.75). Fractions 7–30 were combined, and the solvent was removed under reduced pressure. The desired product was obtained as slightly brown foam. Yield: 79%. ^1H NMR (400 MHz, in $[d_6]\text{DMSO}$): δ = 2.08–2.14 (m, 2H, 2'-H), 3.51–3.63 (m, 2H, 5'-H), 3.78 (td, $^3J_{4',5'} = 3.4$ Hz, $^3J_{4',3'} = 3.6$ Hz, 1H, 4'-H), 4.20–4.22 (m, 3H, NH-CH₂ + 3'-H), 5.06 (t, $^3J_{5',\text{OH},5'} = 5.1$ Hz, 1H, 5'-OH), 5.22 (d, $^3J_{3',\text{OH},3'} = 4.3$ Hz, 1H, 3'-OH), 6.09 (t, $^3J_{1',2'} = 6.7$ Hz, 1H, 1'-H), 8.17 (s, 1H, 6-H), 10.04 (m, 1H, NH-CH₂), 11.6 (s, 1H, 3-NH). ^{13}C NMR (100 MHz, in $[d_6]\text{DMSO}$): δ = 161.5 (4-C), 156.0 (q, $^3J_{5',6'}F = 36.6$ Hz, (C=O)CF₃), 149.4 (2-C), 144.1 (6-C), 115.8 (quart, $^2J_{6',6''}F = 288$, CF₃), 97.6 (5-C), 87.6 (4'-C), 87.4 (1''-C), 84.8 (1'-C), 75.4 (2''-C), 70.2 (3'-C), 61.0 (5'-C), 40.2 (2'-C superposed by DMSO signals), 29.4 (3''-C). MS (ESI): m/z (%) = 798.7 (13), 777.7 (11), 776.7 (37) [$2\text{M} + \text{Na}^+$], 701.9 (14), 550.9 (12), 399.8 (100) [$\text{M} + \text{Na}^+$].

5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-deoxyuridine-3'-[(2-cyano-ethyl)-*N,N'*-diisopropylaminophosphoramidite] (Figure 1, D). 5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-deoxyuridine was prepared according to the standard protocol for 5'-O-dimethoxytritylation of 2'-deoxynucleosides.⁴¹ 5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-deoxyuridine (430 mg, 0.6 mmol) and bis-diisopropylammonium tetrazolidine (82 mg, 0.5 mmol) were dried three times azeotropically with 5 mL of anhydrous acetonitrile each. The remainder

was then dissolved in anhydrous CH_2Cl_2 (5 mL) and treated with 2-cyanoethoxy-bis-(*N,N*-diisopropylamino)phosphine (530 μL , 1.7 mmol). The reaction mixture was stirred for 3 h under argon at room temperature. Next, the mixture was poured into a degassed saturated NaHCO_3 solution and extracted three times with degassed CH_2Cl_2 . The combined organic layers were dried over Na_2SO_4 , and the solvent was removed under reduced pressure. The crude material was purified by short column chromatography on deactivated silica gel (1% Et_3N). The product was eluted with cyclohexane/ AcOEt (5:7) and was obtained as a slightly beige foam. Yield: 83%. ^1H NMR (300 MHz, CDCl_3 , both diastereomers): δ = 1.01–1.25 (m, 12H, $2 \times \text{CH}(\text{CH}_3)_2$), 2.34–2.61 (m, 4H, CH_2CN , 2'-H), 3.32–3.82 (m, 12H, 5'-H OCH_2 , $2 \times i\text{-Pr}-\text{CH}$, $2 \times \text{OCH}_3$), 3.87–3.91 (m, 1H, 4'-H), 4.14 (m, 2H, 9-H), 4.57–4.59 (m, 1H, 3'-H), 6.34 (t, 3J = 6.6 Hz, 1H, 1'-H), 6.76–6.91 (m, 4H, ar.), 7.61–7.70 (m, 9H, ar.), 8.21, 8.22 ($2 \times$ s, 1H, 6-H). ^{31}P NMR (120 MHz, CDCl_3) δ = 149.1, 150.0. MS (ESI): m/z (%) = 303.2 (14) [DMT^+], 901.9 (100) [$\text{M} + 22^+$].

5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyl Uridine. 5-Ioduridine was 5'-O-tritylated and 2'-O-silylated according to standard procedures described in the literature.^{42,43} 5-Iodo-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyl-uridine (300 mg, 0.38 mmol) was dissolved in 3.0 mL of DMF, and 228 μL of freshly distilled triethylamine was added. *N*-Propargyltrifluoroacetamide (182 mg, 1.1 mmol), $\text{Pd}(\text{PPh}_3)_4$ (44.0 mg, 0.066 mmol), and $\text{Cu}(\text{I})$ iodide (14.4 mg, 0.076 mmol) were added to the solution. The reaction mixture was stirred for 8 h in the dark at room temperature, concentrated in vacuo, and a solution of 5% Na_2EDTA was added to the residue. The crude product was extracted with ethyl acetate (3×20 mL). The combined organic layers were collected, dried over Na_2SO_4 , and concentrated in vacuo. The crude product was purified by column chromatography (hexane/ethylacetate 70:30) to give 5'-O-(4,4'-dimethoxytrityl)-2'-O-tert-butylidimethylsilyl-5-(3-trifluoroacetamidoprop-1-ynyl)uridine as a pale yellow solid. Yield: 66%. ^1H NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference): δ (ppm) = 0.03, 0.05 (2s, 6H, $\text{Si}(\text{CH}_3)_2$), 0.88 (s, 9H, $\text{C}(\text{CH}_3)_3$), 3.08, (2s, 2H, $\text{H}-\text{S}'$, $\text{H}-\text{S}''$), 3.74 (s, 6H, $2 \times \text{OCH}_3$), 3.93 (d, 3J = 5.4 Hz, 2H, $\text{N}-\text{CH}_2$), 4.01–4.09 (m, 2H, $\text{H}-4'$, $\text{H}-3'$), 4.34 (t, 3J = 4.6 Hz, 1H, $\text{H}-2'$), 5.17 (d, 3J = 6.0 Hz, 1H, $\text{OH}-3'$), 5.74 (d, 3J = 5.7 Hz, 1H, $\text{H}-1'$), 6.89 (dd, 2J = 9.0 Hz, 3J = 3.3 Hz, 4H, ar.), 7.23–7.43 (m, 9H, ar.), 7.97 (s, 1H, $\text{H}-6'$), 9.97 (t, 3J = 5.4 Hz, 1H, $\text{N}-\text{H}$), 11.74 (s, 1H, $\text{N}-\text{H}$). ^{13}C NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{C} of the solvent at 39.61 ppm as internal reference): δ (ppm) = -5.14, -4.78, 17.92, 25.63, 29.31, 55.00, 59.75, 63.03, 69.62, 74.87, 75.56, 83.28, 85.92, 87.41, 88.86, 98.08, 113.22, 113.30, 115.76 (quart, 2J = 289 Hz, CF_3), 126.65, 127.42, 127.93, 129.67, 129.73, 134.96, 135.54, 143.34, 144.84, 149.46, 155.93 (quart, 3J = 37 Hz, $\text{C}=\text{O}$), 158.1, 161.41.

5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyl-uridine-3'-[(2-cyanoethyl)-*N,N'*-diisopropylamino-phosphoramidite] (Figure 1, C). 5-(3-Trifluoroacetamidopropargyl)-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyl uridine (296 mg, 0.36 mmol) was coevaporated with 5 mL of dichloromethane containing 10% pyridine. The nucleoside was kept under vacuum overnight. Dry ethyl diisopropyl amine (0.32 mL, 4×0.36 mmol, freshly distilled just before use) was added, followed by 2 mL of dry dichloromethane. 2-Cyanoethyl-*N,N'*-diisopropylamino-chlorophosphoramidite (0.12 mL, 1.5×0.36 mmol) was added dropwise to the solution. After 3 h, 0.2 mL of dry methanol was added. The reaction mixture was diluted with 150 mL of ethyl acetate containing 10% triethylamine, washed with a saturated solution of Na_2CO_3 (1×10 mL), and with a saturated solution of potassium chloride (1×10 mL), dried over Na_2SO_4 , and concentrated in vacuo. After purification by short column chromatography (hexane/ethyl acetate/triethyl amine 60:30:10), the product was obtained as a pale yellow foam. Yield: 80%. ^1H NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference, both diastereomers): δ (ppm) = 0.01, 0.03, 0.05, 0.09 (4s, 12H, $\text{Si}-\text{CH}_3$), 0.83, 0.86 (2s, 18H, $\text{C}(\text{CH}_3)_3$), 1.18–1.53 (m, 12H,

$\text{CH}(\text{CH}_3)_2$), 2.77 (t, 3J = 5.8 Hz, 1H, CH_2-CN), 2.88 (t, 3J = 5.8 Hz, 1H, CH_2-CN), 3.48–3.61 (m, 2H, $\text{CH}(\text{CH}_3)_2$), 3.74 (s, 6H, OCH_3), 3.97 (t, 3J = 5.7 Hz, 2H, OCH_2), 4.16–4.22 (m, 2H, $\text{H}-4'$, $\text{H}-3'$), 4.49–4.56 (m, 1H, $\text{H}-2'$), 5.78 (d, 3J = 5.7 Hz, 1H, $\text{H}-1'$), 5.84 (d, 3J = 6.3 Hz, 1H, $\text{H}-1'$), 6.87–6.92 (m, 4H, ar.), 7.27–7.43 (m, 9H, ar.), 7.99, 8.01 (2s, 1H, $\text{H}-6$), 9.97–9.99 (m, 1H, NH), 11.75 (br s, 1H, NH). ^{31}P NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{P} of H_3PO_4 at 0.0 ppm as the external reference): δ (ppm) = 148.24, 149.27.

5-(3-Trifluoroacetamidopropenyl)-uridine. 5-Ioduridine (740 mg, 2 mmol) was dissolved in 7 mL of DMF. To the resulting solution were added sodium acetate buffer (7.1 mL, 0.1 M, pH 5.2) and *N*-allyltrifluoroacetamide (2 mL, 17 mmol). A solution of $\text{Na}_2[\text{PdCl}_4]$ (658 mg, 2.2 mmol) in DMF (7 mL) was added while stirring vigorously. The reaction flask was placed in an oil bath at 80 $^\circ\text{C}$ for 8 h. The precipitated palladium was filtered off through Celite. The filtrate was concentrated in vacuo to a viscous brown oil. The crude product was purified by column chromatography (dichloromethane/methanol 98:2) to obtain a white powder. Yield: 58%. ^1H NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference): δ (ppm) = 3.55–3.70 (m, 2H, $\text{H}-\text{S}'$, $\text{H}-\text{S}''$), 3.82–3.89 (m, 3H, CH_2 , $\text{H}-4'$), 3.99 (1H, br s, $\text{H}-3'$), 4.07 (m, 1H, $\text{H}-2'$), 5.09 (br s, 1H, OH), 5.22 (br s, 1H, OH), 5.41 (d, 1H, OH), 5.77 (d, 3J = 4.8 Hz, 1H, $\text{H}-1'$), 6.18 (d, 2J = 15.9 Hz, 1H, $=\text{CH}$), 6.46 (t, 2J = 15.9 Hz, 3J = 6.1 Hz, 1H, $=\text{CH}$), 8.12 (s, 1H, $\text{H}-6$), 9.69 (t, 3J = 5.4 Hz, 1H, NH). ^{13}C NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{C} of the solvent at 39.61 ppm as internal reference): δ (ppm) = 41.59, 60.51, 69.51, 73.73, 84.75, 88.13, 109.86, 115.97 (quart, 2J = 288 Hz, CF_3), 123.92, 124.23, 138.20, 149.90, 156.06 (quart, 3J = 36 Hz, $\text{C}=\text{O}$), 162.25.

5-[3-(6-Trifluoroacetylaminohexanamido)propenyl]uridine. 5-Ioduridine (740 mg, 2 mmol) was dissolved in 7 mL of DMF. NaOAc buffer (7.1 mL, 0.1 M, pH 5.2) and *N*-allyl-6-(*N*-trifluoroacetylaminohexanamido)hexanamide (3.85 g, 14 mmol) were added to the solution, and a mixture of $\text{Na}_2[\text{PdCl}_4]$ (172 mg, 0.59 mmol) in DMF (2.5 mL) was added while stirring vigorously. The reaction flask was placed in an oil bath at 83 $^\circ\text{C}$. After 2 h, another portion of $\text{Na}_2[\text{PdCl}_4]$ was added. After 8 h, the precipitated palladium was filtered off through Celite. NaBH_4 (2×12 mg) was added to the filtrate while vigorously stirring. The resulting yellowish solution was filtered through Celite, and the solvents were evaporated to give a viscous yellow oil. The crude product was purified by column chromatography, giving a white solid. Yield: 50%. ^1H NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference): δ (ppm) = 1.20–1.31 (m, 2H, CH_2), 1.42–1.55 (m, 4H, CH_2-CH_2), 2.09 (t, 3J = 7.4 Hz, 2H, CH_2), 3.13–3.19 (m, 2H, CH_2), 3.54–3.60 (m, 2H, HS' , HS''), 3.74 (t, 3J = 5.4 Hz, 2H, CH_2), 3.82–3.87 (m, 1H, $\text{H}-4'$), 3.99 (q, 3J = 4.7 Hz, 1H, $\text{H}-3'$), 4.04–4.09 (m, 1H, $\text{H}-2'$), 5.07 (d, 3J = 5.1 Hz, 1H, OH), 5.2 (t, 3J = 4.8 Hz, 1H, OH), 5.39 (d, 3J = 5.5 Hz, 1H, OH), 5.78 (d, 3J = 4.9 Hz, 1H, $\text{H}-1'$), 6.12 (d, 2J = 16.0 Hz, 1H, $=\text{CH}$), 6.40 (t, 2J = 15.9 Hz, 3J = 5.8 Hz, 1H, $=\text{CH}$), 7.96 (t, 3J = 5.6 Hz, 1H, NH), 8.08 (s, 1H, $\text{H}-6$), 9.39 (br s, 1H, NH), 11.42 (br s, 1H, NH). ^{13}C NMR (300 MHz, $[d_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{C} of the solvent at 39.61 ppm as internal reference): δ (ppm) = 25.11, 25.79, 28.04, 35.06, 40.80, 43.20, 60.55, 69.57, 73.68, 84.75, 88.04, 110.32, 115.95 (quart, 2J = 289 Hz, CF_3), 122.09, 126.90, 137.34, 149.79, 156.10 (quart, 3J = 36 Hz, $\text{C}=\text{O}$), 162.06, 171.69.

Both compounds, 5-(3-trifluoroacetamidopropenyl)uridine and 5-[3-(6-trifluoroacetylaminohexanamido)propenyl]uridine, were 5'-O-dimethoxytritylated and 2'-O-silylated with TBDMSCl according to standard protocols.^{42,43}

5-(3-Trifluoroacetamidopropenyl)-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyl-uridine-3'-[(2-cyanoethyl)-*N,N'*-diisopropylaminophosphoramidite] (Figure 1, B). 5-(3-Trifluoroacetamidopropenyl)-5'-O-dimethoxytrityl-2'-O-tert-butylidimethylsilyluridine (243 mg, 0.3 mmol) was coevaporated with dichloromethane (3×5 mL) containing 10% pyridine. The nucleoside was kept under vacuum overnight. Dry ethyl diisopropyl amine (0.27 mL, 1.2 mmol, freshly distilled over CaH_2 just before

used) was added, followed by 2 mL of dry dichloromethane. 2-Cyanoethyl-*N,N'*-diisopropylamino-chlorophosphoramidite (0.1 mL, 0.45 mmol) was added dropwise to the solution. After 3 h, 0.1 mL of dry methanol was added. After 15 min, the reaction mixture was diluted with 100 mL of ethyl acetate (prewashed with Na_2CO_3) containing 1% NEt_3 , washed with a saturated solution of Na_2CO_3 (20 mL), dried over Na_2SO_4 , and concentrated in vacuo to remove the solvents. The residue was purified by short column chromatography (hexane/ethyl acetate 60:40 \rightarrow 50:50, 1% triethylamine) to get the product as a white foam. Yield: 87%. ^1H NMR (300 MHz, $[\text{d}_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference, both diastereomers): δ (ppm) = 0.01, 0.03, 0.05, 0.08 (4s, 6H, Si- CH_3), 0.83, 0.85 (2s, 9H, $\text{C}(\text{CH}_3)_3$), 1.09–1.12 (m, 12H, $\text{CH}(\text{CH}_3)_2$), 2.77 (t, $^3J = 6.0$ Hz, 2H, CH_2), 3.52–3.67 (m, 4H, HS' , HS'' , $\text{NCH}(\text{CH}_3)_2$), 3.73 (s, 6H, OCH_3), 3.76–3.80 (m, 2H, OCH_2), 4.16–4.23 (m, 2H, $\text{H-4}'$, $\text{H-3}'$), 4.48–4.55 (m, 1H, $\text{H-2}'$), 5.65, 5.66 (2d, $^2J = 15.8$ Hz, 1H, $=\text{CH}$), 5.84, 5.88 (2d, $^2J = 5.96$ Hz, 6.38 Hz, 1H, $\text{H-1}'$), 6.25–6.36 (m, 1H, $=\text{CH}$), 6.87–6.91 (m, 4H, *ar*), 7.24–7.42 (m, 9H, *ar*), 7.68, 7.69 (2s, 1H, H-6), 9.56 (br s, 1H, NH), 11.59 (br s, 1H, NH). ^{31}P NMR (300 MHz, $[\text{d}_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{P} of H_3PO_4 at 0.0 ppm as the external reference): δ (ppm) = 148.27, 149.23.

5-[3-(6-Trifluoroacetylaminohexanamido)-propenyl]-5'-*O*-dimethoxytrityl-2'-*O*-*tert*-butyldimethylsilyluridine-3'-[(2-cyanoethyl)-*N,N'*-diisopropylaminophosphoramidite] (Figure 1, A). 5-[3-(6-Trifluoroacetylaminohexanamido)propenyl]-5'-*O*-dimethoxytrityl-2'-*O*-*tert*-butyldimethylsilyluridine (278 mg, 0.3 mmol) was coevaporated with dry pyridine (3 \times 5 mL) and dichloromethane (3 \times 5 mL). The nucleoside was kept under vacuum overnight. Dry ethyl diisopropyl amine (0.27 mL, 1.2 mmol, freshly distilled over CaH_2 just before used) was added, followed by 2 mL of dry dichloromethane. 2-Cyanoethyl-*N,N'*-diisopropylamino-chlorophosphoramidite (0.1 mL, 0.45 mmol) was added dropwise to the solution. After 2 h, another 0.1 equiv of phosphitylating reagent was added. After 2 h, 0.1 mL of dry methanol was added, and after 15 min the reaction mixture was diluted with 100 mL of ethyl acetate (prewashed with Na_2CO_3) containing 1% triethylamine, washed with a saturated solution of Na_2CO_3 (20 mL), dried over Na_2SO_4 , and concentrated in vacuo. The residue was purified by short column chromatography (hexane/ethyl acetate 60:40 \rightarrow 50:50, 1% triethylamine) to obtain a white foam. Yield: 82%. ^1H NMR (300 MHz, $[\text{d}_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{H} of the solvent at 2.5 ppm as internal reference, both diastereomers): δ (ppm) = 0.00, 0.02, 0.05, 0.08 (2s, 6H, Si(CH_3) $_2$), 0.83, 0.85 (2s, 9H, $\text{C}(\text{CH}_3)_3$), 1.11 (d, $^3J = 6.9$ Hz, 12H, $\text{CH}(\text{CH}_3)_2$), 1.15–1.19 (m, 2H, CH_2), 1.43–1.48 (m, 2H, CH_2), 1.99–2.02 (m, 2H, CH_2), 2.75, 2.86 (2t, $^3J = 5.8$ Hz, 2H, CH_2CN), 3.12–3.18 (m, 2H, CH_2), 3.24–3.30 (m, 2H, $\text{H-5}'$, $\text{H-5}''$), 3.46–3.59 (m, 2H, CH), 3.72–3.74 (m, 8H, OCH_3 , CH_2), 3.77–3.84 (m, 2H, OCH_2), 4.14–4.16 (m, 1H, $\text{H-4}'$), 4.20–4.26 (m, 1H, $\text{H-3}'$), 4.47–4.57 (m, 1H, $\text{H-2}'$), 5.63, 5.65 (2d, $^2J = 15.8$ Hz, 1H, $=\text{CH}$), 5.86, 5.90 (2d, $^3J = 6.4$ Hz, 1H, $\text{H-1}'$), 6.17, 6.27 (m, 1H, $=\text{CH}$), 6.87–6.91 (m, 4H, *ar*), 7.23–7.43 (m, 9H, *ar*), 7.64, 7.67 (2s, 1H, H-6), 7.76–7.81 (m, 1H, NH), 9.39 (m, 1H, NH), 11.6 (br s, 1H, NH). ^{31}P NMR (300 MHz, $[\text{d}_6]\text{DMSO}$, 25 $^\circ\text{C}$, δ_{P} of H_3PO_4 at 0.0 ppm as the external reference): δ (ppm) = 149.39, 148.22.

Oligonucleotides. Details on oligonucleotide synthesis^{44,45} as well as on labeling of deoxyoligonucleotides and oligoribonucleotides with Cy5 and Alexa488 can be found in sections S1.1, S1.2, and S1.3 of the Supporting Information. Ultrapure labeled DNA1 with the L linker, RNA2 and RNA3 oligonucleotides (PAGE grade), and all unlabeled counter sequences were purchased from Purimex (Grebenstein, Germany). DNA1 oligonucleotides with the I Linker were purchased from IBA (Göttingen, Germany). All sequences of the DNA and RNA strands are given in section S1.4 of the Supporting Information. Additionally, the sequences and labeling positions of DNA1 and RNA1 are illustrated in Figure 2B and C. The linker types L and I for DNA and L for RNA were chosen for reasons of commercial availability, and the RNA linker types for reasons of chemical suitability. Throughout this work, if not stated differently, the same linker type is used at the donor (D) and the acceptor (A) positions.

Procedures . Hybridization of DNA and RNA. The hybridization buffer for the DNA samples contained 20 mM TRIS, 100 mM NaCl, and 10 mM MgCl_2 , pH 7.5. For RNA, it contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl, and 10 mM MgCl_2 , pH 6.5. The concentration of the DNA or RNA molecules in the buffer ranged between 2 and 10 μM . For FRET molecules, the ratio between the amount of acceptor and donor strand ranged from 1 to 2. For donor- and acceptor-only molecules, the ratio of unlabeled to labeled strand was 3:1. The solution was heated to a temperature of 90 $^\circ\text{C}$ inside a water bath and was then allowed to cool to room temperature overnight.

Time-Resolved Polarized Fluorescence Experiments and Data Analysis. Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed using pulsed laser excitation. Fluorescence intensity and anisotropy decay curves were fitted using the iterative reconvolution approach.⁴⁶ The fits approximately range from the maximum of the instrument response functions (IRF) to the first time channel with less than 100 detected photons. The fluorescence intensity decays of FRET-labeled molecules (donor and acceptor emission) were fitted globally with the decays of the molecules only labeled with either the donor (donor only, D-only) or the acceptor (acceptor only, A-only) dye. The fluorescence decays were modeled by single or double exponential decays or by assuming a Gaussian distribution of distances (section 3.3). Alternatively, the fluorescence decays were deconvoluted by using the maximum entropy method.^{47,48} The anisotropy decays were recovered by globally fitting the sum ($F_{\parallel} + 2G_{\perp}$) and difference ($F_{\parallel} - G_{\perp}$) curves (F_{\parallel} , F_{\perp} , fluorescence signals in parallel and perpendicular polarization planes relative to the vertically polarized excitation light, respectively; G , ratio of the sensitivities of the detection system for vertically and horizontally polarized light). The anisotropy decays $r(t)$ were modeled by double or triple exponential decays (rotational correlation times ρ_1 , ρ_2 , and ρ_3) with free amplitudes (b_1 , b_2 , and b_3). For further details, see section S1.5 in the Supporting Information.

Single-Molecule Fluorescence Measurements. Multiparameter fluorescence detection (MFD) measurements were performed as described in refs 12,13,49. Each molecule generates a brief burst of fluorescence photons as it traverses the detection volume. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into wavelength ranges using a dichroic beamsplitter. Bursts of fluorescence photons are distinguished from the background of 1–2 kHz by applying certain threshold intensity criteria.⁵⁰ For further details, see section S1.5 in the Supporting Information.

Measurements of Fluorescence Quantum Yields. Φ_{F} determination was performed according to ref 7. Rhodamine 700 in ethanol ($\Phi_{\text{F}} = 0.38$)⁵¹ and Rhodamine 110 ($\Phi_{\text{F}} = 0.95$) were used as reference dyes for Cy5 and Alexa488, respectively. Correction for the solvent refractive index was performed as described in ref 7.

3. RESULTS AND DISCUSSION

For highly accurate FRET measurements, several linker and dye effects must be taken into account. First, the DNA and RNA microenvironment affects the dyes' photophysics and local motions. Second, all observable FRET parameters depend on spatial distributions of donor and acceptor positions. The final goal is to restrict the dye motions to achieve a well-defined dye localization in FRET experiments. If, on the other hand, the dye reorientation is restricted, additional orientational effects (κ^2 effects) finally influence the FRET efficiency, which we must learn to take into account.

3.1. Characterization of the Local Environment of the Dyes. As local quenching processes and restricted mobilities will complicate FRET analysis, we investigated how the nature and length of the linker influence the fluorescence properties of the dye for internally labeled dsDNA and dsRNA. Throughout

this work, we used Alexa488 as a donor and Cy5 as an acceptor dye (Figure 2). For internal postlabeling of the nucleic acids, we use the NHS-ester of Cy5, which, in contrast to phosphoramidite derivatives of Cy5 (having the same name, which leads to confusions; see Figure S2 in the Supporting Information), contains two sulfonic acid groups to prevent dye sticking. Three different linker types for the uracil or thymine base with decreasing stiffness and increasing length were used: (i) short stiff linkers with four backbone atoms (S_t and S_d); (ii) a linker of intermediate length (I) with seven backbone atoms; and (iii) long flexible linkers (L) with 11 backbone atoms, which are most frequently used in the scientific community (usually referred to as “C6-amino linker”) (Figure 2).

3.1.1. Analysis of Local Quenching in DNA and RNA.

We analyzed local quenching by fluorescence quantum yield and lifetime measurements. Figure 3A shows typical ensemble fluorescence lifetime measurements by eTCSPC (see Figure S1A and S1B in section S2.1 of the Supporting Information for the decays with the complementary FRET dye). We describe the fluorescence decays $F(t)$ of single-labeled dsDNA and dsRNA by up to two fluorescence lifetimes τ_i with the species fractions x_i and a species-averaged fluorescence lifetime $\langle\tau\rangle_x$ (eq 1).

$$F(t) = x_1 \exp(-t/\tau_1) + x_2 \exp(-t/\tau_2) \quad \text{with} \\ \langle\tau\rangle_x = x_1\tau_1 + x_2\tau_2 \quad (1)$$

The results of the fluorescence lifetime analysis are summarized in Table 1. In most cases, the $\langle\tau\rangle_x$ values are proportional to the fluorescence quantum yields of the donor and the acceptor (Φ_{FD} and Φ_{FA} , respectively; see Table 1), which indicates purely dynamic quenching (see Table S1 in the Supporting Information).

The fluorescence of the rhodamine dye Alexa488 can be in principle quenched by nucleobases, which results in a multiexponential fluorescence decay. Photoinduced electron transfer (PET) or proton-coupled electron transfer between organic fluorophores and suitable electron-donating moieties, such as the nucleobase guanine, can quench fluorescence upon van der Waals contact.^{15,52–54} PET quenching has been used as reporter for monitoring conformational dynamics in oligonucleotides.⁵⁴ It is striking that no quenching effects are observed for DNA, but there is quenching for RNA (see $\langle\tau\rangle_x$ in Table 1). This is in line with the spatial distribution of the dyes (see further). For RNA, the dye is closer to the nucleobases, because it is still partially inside the major groove, whereas in DNA it is primarily outside.

In contrast to Alexa488, Cy5 has its own additional linker with 6 atoms between the chromophore and the reactive coupling group (Figure 2), which increases the distance to the nucleobases, so that only slight fluorescence lifetime differences between DNA and RNA are noticeable.

The fluorescence properties of the Cyanine dye Cy5 are less affected by PET quenching⁵⁴ but rather more by trans \rightarrow cis photoisomerization,^{55–58} which is influenced by specific solvent effects^{59,60} and sterical constraints set by the local environment.⁵⁸ In water, free Cy5 shows a single exponential fluorescence relaxation with a lifetime of 0.9 ns,^{60,61} whereas our measurements yielded for each linker biexponential fluorescence decays of Cy5. The second lifetime was usually similar to that in water, and the first lifetime is significantly larger. In view of the above findings for the photoisomerization of Cyanine dyes, the multiexponential decay of Cy5 is most likely due to the heterogeneous DNA or RNA microenvironment and not necessarily due to significant

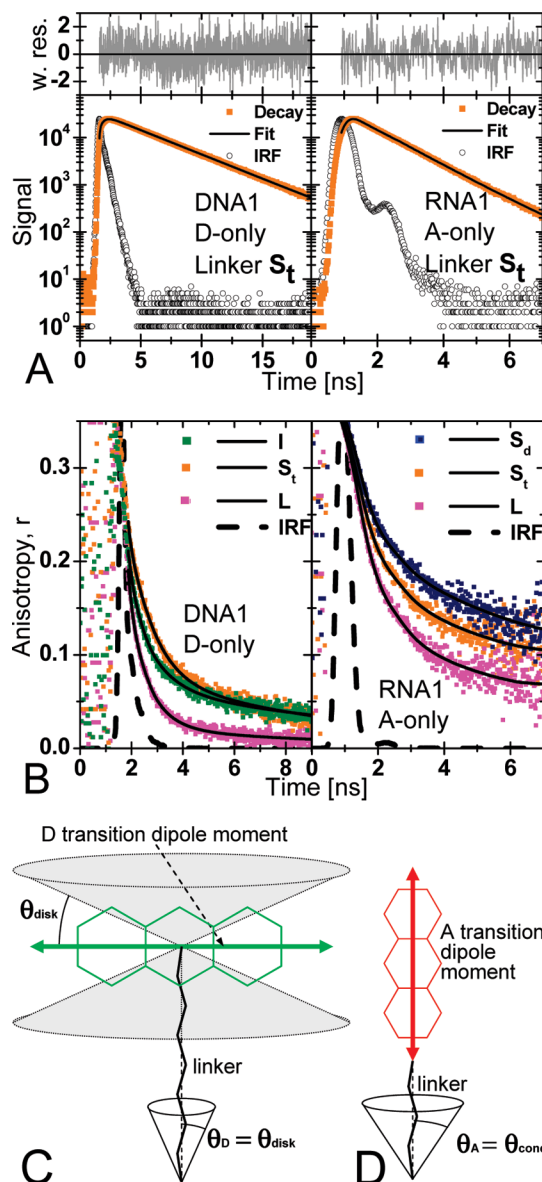


Figure 3. (A,B) Exemplary presentation of DNA1 (left panel) and RNA1 (right panel) eTCSPC measurements of single labeled nucleic acids: (A) fluorescence decay curves (weighted residuals are presented above each plot) and (B) fluorescence anisotropy decays with the re-scaled IRFs curves. The fit results are listed in Tables 1 and 2. (C,D) Schematic sketch for the orientation of the transition dipole moments of the dyes with respect to the linker axes. The transition dipole moment of Alexa488 is assumed to be perpendicular to the linker axis; the linker can wobble within a cone with the opening half angle θ_D . (D) The transition dipole moment of Cy5 is assumed to be parallel to the linker axis; the linker can wobble within a cone with the opening half angle θ_A .

sticking of the dye to DNA or RNA. This interpretation of multiple microenvironments is supported by our recent single-molecule studies on Cy5-labeled dsDNA,⁶¹ where at least two Cy5 states with distinct fluorescence lifetimes and anisotropies have been found. The steady-state anisotropy increases with the lifetime indicating a more restricted environment, which reduces the rate for cis–trans isomerization as well as the local linker wobbling motion (for more details, see section 3.1.2). Moreover, it is remarkable that for all linkers with close proximity to the allyl-unit (I for DNA

Table 1. Fluorescence Lifetimes (τ) and Quantum Yields (Φ_F) of Single Dyes Coupled to ds Nucleic Acids^a

linker	D-only				A-only			
	DNA1							
	$\langle\tau\rangle_x$, ns	Φ_{FD}	τ , ns		$\langle\tau\rangle_x$, ns	Φ_{FA}	τ_1 , ns (x_1)	τ_2 , ns (x_2)
L	4.13	0.92	4.13		1.16	0.38	2.34 (11%)	1.01 (89%)
I	4.16	1.03	4.16		1.29	0.40	2.41 (16%)	1.08 (84%)
S _t	4.20	0.98	4.20		1.17	0.21	2.18 (12%)	1.03 (88%)

linker	RNA1							
	$\langle\tau\rangle_x$, ns	Φ_{FD}	τ_1 , ns (x_1)	τ_2 , ns (x_2)	$\langle\tau\rangle_x$, ns	Φ_{FA}	τ_1 , ns (x_1)	τ_2 , ns (x_2)
L	3.70	0.83	4.08 (90%)	0.27 (10%)	1.10	0.36	1.22 (73%)	0.79 (27%)
S _d	3.28	0.70	3.89 (81%)	0.67 (19%)	1.20	0.38	1.45 (51%)	0.93 (49%)
S _t	3.60	0.84	3.99 (88%)	0.71 (12%)	1.10	0.36	1.21 (77%)	0.72 (23%)

^aTypical errors: single τ , ± 0.02 ns; double τ_i (major component with $x_i \approx 80$ –90%), ± 0.03 ns; τ_j (minor component with $x_j \approx 10$ –20%), ± 0.4 ns; Φ_{FD} , ± 0.05 ; Φ_{FA} , ± 0.03 .

and S_d for RNA) the mean fluorescence lifetimes $\langle\tau\rangle_x$ of Cy5 deviate from those of the other linkers, which indicates a slightly different mean environment.

Let us finally mention the only exception from the strict correlation between the species-averaged fluorescence lifetime $\langle\tau\rangle_x$ and the fluorescence quantum yield listed in Table 1 and Table S1. We surprisingly observe additional static Cy5 quenching for the S_t linker in DNA, which is also clearly detectable in ensemble- and sm-FRET experiments and will be discussed in detail in section 3.4.

3.1.2. Linker Motions in DNA and RNA. Figure 3B shows typical time-resolved ensemble measurements of fluorescence anisotropies (see also Figure S1C and S1D in section S2.1 of the Supporting Information for decays with the complementary FRET dye). The fluorescence anisotropy decays $r(t)$ with the fundamental anisotropy r_0 were formally characterized by up to three rotational correlation times ρ_i with the anisotropy amplitudes b_i (eq 2):

$$r(t) = b_1 \exp(-t/\rho_1) + b_2 \exp(-t/\rho_2) + b_3 \exp(-t/\rho_3) \quad \text{with} \quad r_0 \geq b_1 + b_2 + b_3 \quad (2)$$

As the dye motion is partially restricted by the nucleic acids, the longest correlation time reflects to a significant extent the overall tumbling motion of the molecule (global motion; for more details, see section S1.5 of the Supporting Information), and its amplitude corresponds to the residual anisotropy r_{∞} , which allows the determination of second-rank order parameters $S^{(2)}$ (eqs 3 and 4). The average anisotropy corresponds to the steady-state anisotropy r_s , which is also measured in ensemble or single-molecule experiments by multiparameter fluorescence detection (MFD).

To rationalize dye motion, we must consider that the orientation of the transition dipole moment with respect to the linker is different for D and A. On the basis of the chemical structure of the linked donor dye Alexa488 in Figure 2, we assume the transition dipole moment to be approximately perpendicular to the linker axis as depicted in Figure 3C. Irrespective of the linker motions, rotations of the transition dipole about the linker axis significantly depolarize the donor fluorescence. In addition, if the linker can wobble within a cone with the opening half angle θ_D , the transition dipole of the donor can explore the space within a

“disk” with the opening half angle $\theta_{\text{disk}} = \theta_D$ (Figure 3C). The second-rank order parameter $S_D^{(2)}$ is given by eq 3.

$$\text{donor} : \frac{1}{2} \cos^2 \theta_{\text{disk}} = \sqrt{\frac{r_{\infty,D}}{r_0}} = -S_D^{(2)} \quad (3)$$

Thus, anisotropy senses both the linker and the dye rotations. Even if the linker cannot wobble (i.e., is totally stiff), the dye can still rotate about the linker, and a rather low residual anisotropy $r_{\infty,D} = 1/4(r_0)$ is expected.

In contrast, the transition dipole moment of the acceptor dye Cy5 is more parallel to the linker axis (Figure 2), and the linker together with the dye can wobble within a cone with the opening half angle $\theta_{\text{cone}} = \theta_A$ as depicted in Figure 3D. The motion is characterized by the second-rank order parameter $S_A^{(2)}$ in eq 4.⁸

$$\text{acceptor} : \frac{1}{2} \cos \theta_{\text{cone}} (1 + \cos \theta_{\text{cone}}) = \sqrt{\frac{r_{\infty,A}}{r_0}} = S_A^{(2)} \quad (4)$$

Thus, anisotropy senses predominantly the linker motions as the dye rotates about the linker producing little or no fluorescence depolarization. If the linker cannot wobble, the dye rotates only parallel to the linker, which results in a very high residual anisotropy $r_{\infty,A} \approx r_0$. The approximation of the Cy5 motion by eq 4 is supported by the fact that $r_{\infty,A}$ is always >0.1 (Table 2), which is inconsistent with eq 3.

The analysis of the fluorescence anisotropies together with the rotational correlation times for D-only and A-only labeled DNA and RNA are compiled in Table 2.

As, in contrast to Cy5, the donor dye Alexa488 has no additional internal linker between the chromophore and the reactive coupling group, it is most suited to study the influence of the different nucleobase linkers. In RNA, the wobbling motion of the propargyl linker S_t and of the propenyl linker S_d is negligible (eq 3, $\theta_{\text{disk}} \approx 0^\circ$), because the linker is stiff and short and the major groove of the RNA is very deep and narrow. The major groove of DNA is wider, and thus a small linker wobbling motion is observed ($\theta_{\text{disk}} = 16^\circ$ and 23° for the S_t and I linkers, respectively). If the linkers become longer and more flexible, the linker wobbling should be limited not by the size of the dye but rather by the opening angle of the groove. Because Cy5 has its own

Table 2. Rotational Correlation Times ρ_i , Obtained for Donor-Only ($r_0 = 0.375$) and Acceptor-Only ($r_0 = 0.390$) DNA1 and RNA1 Samples^a

linker	D-only					A-only			
	r_s	ρ_1 , ns (b_1)	ρ_2 , ns (b_2)	$\rho_{3(\text{global})}$, ns ($b_3 = r_{\infty, D}$) ^b	θ_{disk}	r_s	ρ_1 , ns (b_1)	$\rho_{2(\text{global})}$, ns ($b_2 = r_{\infty, A}$) ^b	θ_{cone}
DNA1									
L	0.043	0.17 (0.23)	0.76 (0.12)	7.9 [4–∞] (0.02)	45°	0.208	0.51 (0.26)	>60 [60–∞] (0.13)	47°
I	0.077	0.20 (0.21)	1.0 (0.10)	11.7 [8–25] (0.07)	23°	0.234	0.57 (0.22)	>60 [40–∞] (0.17)	41°
S _t	0.085	0.11 (0.13)	0.86 (0.16)	8.7 [6–11] (0.08)	16°	0.206	0.60 (0.25)	>60 [20–∞] (0.12)	47°
RNA1									
L	0.065	0.27 (0.22)	1.5 (0.09)	6.6 [4–60] (0.05)	32°	0.211	0.68 (0.26)	8.1 [5–16] (0.13)	47°
S _d	0.128	0.38 (0.14)	2.2 (0.10)	9.5 [7–∞] (0.11)	~0°	0.254	0.66 (0.18)	10.9 [8–15] (0.21)	35°
S _t	0.117	0.30 (0.15)	2.1 (0.09)	13.0 [8–15] (0.10)	~0°	0.237	0.70 (0.21)	10.7 [8–19] (0.17)	40°

^aTypical errors: r_s , ± 0.002 ; ρ (major component), ± 0.1 ns; r_{∞} , ± 0.015 . ^bFor the longest (global) correlation time, 1σ confidence intervals are shown in squared brackets.

flexible 6 atom linker, the nucleobase linker effects are much weaker. Consistently, the L linker has the largest flexibility for both DNA and RNA. In DNA, the cone angles of both dyes, θ_{disk} and θ_{cone} , are equal, which supports the idea that the wobbling angle is only limited by the opening angle of the major groove. In RNA, the results still differ slightly because of the different dye linker lengths and the greater depth of the major groove. The experimental residual anisotropies nicely agree with the half opening angles of approximately 45° and 30° observed for the grooves of DNA and RNA, respectively (Table 2). As fluorescence lifetime measurements and single-molecule studies⁶¹ indicate heterogeneous microenvironments of Cy5, it is important to note that the additional state with the longer lifetime is to some extent less mobile. The steady-state anisotropy r_s of this state is about 0.2,⁶¹ indicating that the dye is not stuck but has sufficient rotational freedom (having in mind large values of $\rho_{(\text{global})}$, $r_s \approx r_0$ would be expected for an immobile species).

To better understand the longest (global) rotational correlation time $\rho_{(\text{global})}$ (Table 2), we performed simulations of DNA1 and RNA1 rotations using the HydroPro software.⁶² For DNA1 and RNA1, respectively, three correlation times of 10, 22, and 35 ns (DNA1) and 12, 23, and 32 ns (RNA1) are predicted. In most cases, experimentally obtained values of $\rho_{(\text{global})}$ (Table 2) are similar to the shortest predicted correlation time, which represents the rotation about the helical axis. A notable exception is Cy5 attached to DNA1, which shows systematically longer values of $\rho_{(\text{global})}$. This fact suggests that the preferential orientation of the Cy5 transition dipole is nearly parallel to the helical axis of DNA, so that it senses the other slower rotational motions.

3.2. FRET Benchmark Study. In this and the following sections, we will show that it is crucial for a proper interpretation of FRET results to consider the distributions of D and A dye positions given by the vectors \mathbf{R}_D and \mathbf{R}_A , respectively. Because different techniques determine distinct average distances between donor and acceptor dyes, we have to define and distinguish three different quantities.

- (i) $\langle R_{DA} \rangle$ denotes the mean distance between the dyes and can be determined by eTCSPC measurements. $\langle R_{DA} \rangle$ is calculated by integrating over all possible positions of the two dyes and the resulting distances ($\langle R_{DA} \rangle = \langle |\mathbf{R}_D - \mathbf{R}_A| \rangle$).

- (ii) $\langle R_{DA} \rangle_E$ is the FRET-averaged distance between the dyes. It is calculated from the mean FRET efficiency (eq 5A) using eq 5B.

$$\langle E \rangle = \left\langle \frac{1}{1 + R_{DA}^6/R_0^6} \right\rangle \quad (5A)$$

$$\langle R_{DA} \rangle_E = R_0 (\langle E \rangle^{-1} - 1)^{1/6} \quad (5B)$$

In eq 5, R_0 is the Förster radius. $\langle R_{DA} \rangle_E$ is determined from time-averaged fluorescence intensity measurements on the single-molecule (section 3.2.2) or ensemble level.

- (iii) R_{mp} is the distance between the mean positions of the dyes ($R_{mp} = |\langle \mathbf{R}_D \rangle - \langle \mathbf{R}_A \rangle|$) and is used for the geometric description of FRET-based structural models.²⁵ As shown below, R_{mp} cannot be measured directly via FRET. The detailed calculation of $\langle R_{DA} \rangle$, $\langle R_{DA} \rangle_E$, and R_{mp} is described in section S2.2 (eqs S3–S5) in the Supporting Information.

3.2.1. Calculation of the Volume Accessible to the FRET Dyes

The AV Approach. The prediction of the FRET dye positions with respect to the macromolecule of interest is absolutely essential for the interpretation of quantitative FRET measurements, especially when the dyes are attached via long flexible linkers. If the local structure of a macromolecule is known or can be predicted, the dye positions have been successfully computed by molecular dynamics (MD) simulations.^{25–27} However, MD simulations are time-consuming and too complex for everyday use. Recently, alternative methods based on simple geometric computations have been proposed to predict possible positions for EPR and FRET labels.^{28,29}

As sketched in Figure 4A, these methods approximate the dye by a sphere with an empirical radius of R_{dye} , where the central atom of the fluorophore (see Table 3 for definition) is connected by a flexible linkage of a certain effective length L_{link} and width w_{link} to the nucleobase. The overall length of the linkage is given by the actual length of the linker and the internal chemical structure of the dye. A geometric search algorithm finds all dye positions within the linkage length from the attachment point, which do not cause steric clashes with the macromolecular surface. All allowed positions are considered as equally probable, which allows one to define an accessible volume for the dye (AV).

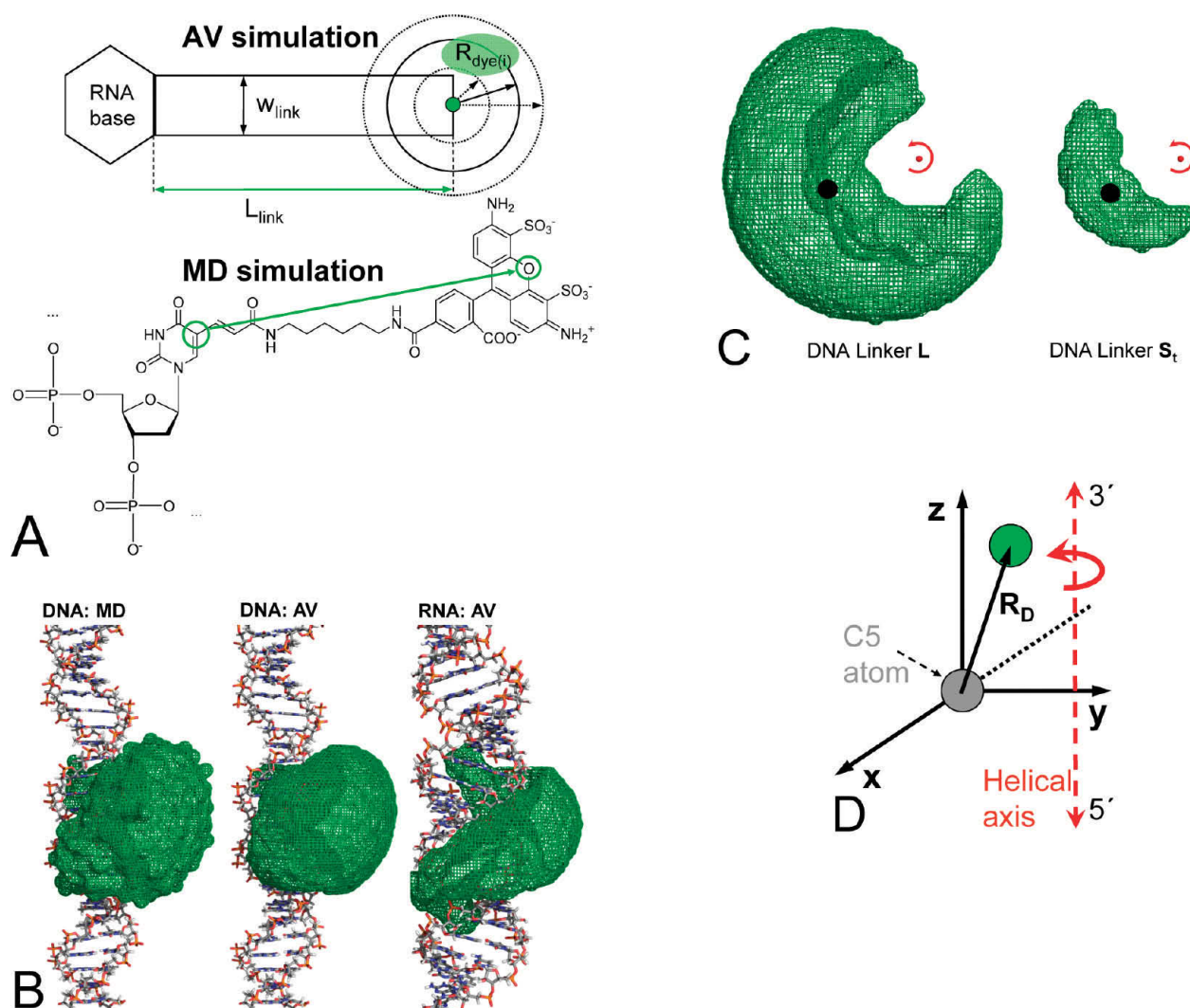


Figure 4. (A) Schematic comparison of AV and MD simulations of dye positions. For AV simulations, we used $w_{\text{link}} = 4.5 \text{ \AA}$, and $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$, and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for the three dimensions of Alexa488. See Table 3 for values for L_{link} . (B) Molecular dynamics (left) and AV (middle, right) simulations of possible positions of Alexa488 attached to DNA1 (B-form) (left, middle) or RNA1 (A-form) (right) via linker L. The structures are rendered via PyMOL.⁶⁵ (C) Top view of Alexa488 position distributions simulated by the AV approach for L (left) and S_t (right) linkers. The helical axis of DNA and the mean position of the dye are shown as red and black spheres, respectively. (D) Coordinate system used to define the mean dye positions $\langle R_D \rangle$ and $\langle R_A \rangle$ in Table 3. The red dashed arrow indicates the helical axis of the nucleic acid.

Spatial Requirements Are Better Described by a Complex Fluorophore Shape. For RNA, AV simulations using the empirical dye radius $R_{\text{dye}} = 3.5 \text{ \AA}$ ²⁹ result in two separated dye clouds (see Figure S3 in the Supporting Information). It is obvious that the space between the obtained clouds should also be accessible for the planar chromophores used here. This demonstrates that, for some sterically demanding local environments, for example, in RNA, it is important to take the three quite different dimensions of a fluorophore into account.

Therefore, for each calculation of a position distribution, we used the real physical dimensions of the fluorophore and performed three independent AV simulations with three different radii $R_{\text{dye}(i)}$ and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. Throughout this work, we used for Alexa488 $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$, and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ and for Cy5 $R_{\text{dye}(1)} = 11 \text{ \AA}$, $R_{\text{dye}(2)} = 3 \text{ \AA}$, and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$. It turned out that these “mixed” AV simulations are necessary to accurately predict

dye distributions for RNA; yet for DNA, the effect of the dye radius is much less pronounced.

The fact that the dyes are assumed as spheres makes it impossible to take into account aberrations due to asymmetric structures (e.g., for Cy5, Figure 2). However, in our case, this results only in an angular shift, which is easy to correct for as we have the possibility of comparison to MD data.

The AV method is clearly not applicable when dyes show considerable interactions (such as sticking) with DNA or RNA.²¹ However, in this work, we employ internal labeling of DNA and RNA, which minimizes interactions of the dyes with DNA and RNA. We use NHS-ester of Cy5, which, in contrast to phosphoramidite derivatives of Cy5 (see Figure S2 in the Supporting Information), contains two sulfonic acid groups. The negative charges of Cy5 and Alexa488 largely prevent dye sticking. With the exception of dyes with S_t linkers (see section 3.4), there is no evidence for the presence of long-lived dye heterogeneities, which justifies the use of the AV method in this work. It is worth

Table 3. Estimation of Mean Positions for Alexa488 and Cy5 Using MD and AV Simulations

method	linker	L_{link} , Å (D)	$\langle R_D \rangle$, Å (x, y, z) ^a			$ \langle R_D \rangle $, Å ^a	σ_D , Å	L_{link} , Å (Å)	$\langle R_A \rangle$, Å (x, y, z) ^a			$ \langle R_A \rangle $, Å ^a	σ_A , Å
DNA1													
MD	L		7.4	−3.8	4.4	9.4	11.4		5.0	−7.1	2.7	9.1	12.9
AV ^b	L	20	6.9	−4.1	2.2	8.3	13.1	22	6.5	−4.0	2.3	7.9	15.2
	I	15	6.0	−4.3	3.1	8.0	9.0	17	6.2	−4.0	2.7	7.9	11.0
	S _t	11	4.3	−4.7	3.2	7.1	5.7	16	6.1	−4.0	2.9	7.8	10.2
RNA1													
AV ^b	L	20	−0.1	−8.5	4.5	9.6	11.3	22	−0.8	−8.5	4.2	9.5	12.8
	S _d	11	−4.0	−2.9	3.9	6.3	4.9	14	−3.9	−3.4	3.7	6.3	6.9
	S _t	11	−4.0	−2.9	3.9	6.3	4.9	16	−2.8	−4.9	3.9	6.9	8.6

^a Between the C5-atom of the base (origin) and for Alexa488 the O atom at position 10 of the xanthene ring or for Cy5 the C atom at position 3 of the pentamethine chain; see Figure 4D. ^b For all AV simulations, we used the same linkage width $w_{\text{link}} = 4.5$ Å. The distinct linkage lengths L_{link} were determined from the most extended conformations.

mentioning that strong interactions between a dye and DNA or RNA are probably impossible to adequately model even by MD simulations, because millisecond time scales are currently not accessible to MD.⁶³ In this work, we developed an improved accessible volume simulation procedure based on the algorithm “Model Satellite Prior” implemented in the “FRETnpsTools” program^{29,37} and then performed our own distance calculations as described. The pdb files of the macromolecules were generated with the Nucleic Acid Builder (NAB) software, which is part of AmberTools.⁶⁴

Comparison of AV with MD. We tested the suitability of the AV approach in two steps: (i) in this section, the predictions of the AV approach are compared to the results of the MD simulation from²⁵ and (ii) in section 3.2.2, the AV approach is used to model the dye position distributions in a FRET benchmark study.

Figure 4B shows distributions of possible positions of Alexa488 attached to DNA and RNA via the L linker simulated by MD (left) and AV (middle and right). Each AV simulation needs five input parameters: three dye radii $R_{\text{dye}(1,2,3)}$ as defined above, w_{link} , and L_{link} . We used typical parameters for the linkage width ($w_{\text{link}} = 4.5$ Å) from ref 29. The linkage lengths (L_{link}) were estimated from the fully extended conformations of each linker using the Hyperchem software⁶⁶ and are listed in Table 3.

Considering the L linker, the outer border of the volume accessible to Alexa488 attached to DNA is displayed as a green net in Figure 4B. The volume calculated by AV closely resembles the distribution predicted by MD. In comparison to MD data, AV predicts the mean position (of Alexa488 ($\langle R_D \rangle$)) (O atom at position 10 in the xanthene ring) with respect to the C5 atom of the uracil (Figure 4D) with a deviation of 2.2 Å (Table 3). As expected, due to the asymmetric structure of Cy5, the distribution of its positions simulated using the AV approach agrees less well with MD data (3.5 Å deviation between the respective mean positions ($\langle R_A \rangle$) defined by the C atom at position 3 in the pentamethine chain). However, the z-displacement (2.3 Å; see Figure 4D) and the distance from ($\langle R_A \rangle$) to the helical axis (11.5 Å) are similar to MD values (2.7 and 11.8 Å, respectively). Therefore, the main difference between the mean positions of Cy5 predicted by MD and AV is a small angular displacement along the xy-plane of $\sim 20^\circ$. Thus, in the following, we apply this additional shift for all Cy5 positions predicted by AV.

Regardless of the linker and the fluorophore, the z-coordinates of the mean positions of the dyes with respect to the C5 atom of

the nucleobase ($\langle R_D \rangle$ and $\langle R_A \rangle$) are always positive (Figure 4D). This means that they are always shifted toward the 3'-end of DNA or RNA, which has a significant effect on DA distances (inset in Figure 5C). For L linkers, the displacement between the dyes in z-direction due to linkers (Δz_{link}) is 7–8 Å (Table 3). Thus, for structure determinations or when choosing labeling positions, it is crucial to take this displacement into account. Additionally, as opposed to DNA, the base pair plane in A-RNA is not perpendicular to the helical axis so that the C5 atoms of opposing nucleobases are displaced against each other by $\Delta z_{\text{C5}} = 1.6$ Å toward the 3'-ends of the respective strands. In summary, the z-displacement of the dyes is determined by three factors and is given by $\Delta z = \Delta z_{\text{bp}} \Delta n + \Delta z_{\text{link}} + \Delta z_{\text{C5}}$ where the basepair separation Δn is counted from D to A toward the 5'-end of DNA or RNA (i.e., Δn is negative if A is closer to the 3' end).

For both DNA and RNA, the attached dyes point into the major groove, and the volume follows the helical twist of the groove, which is visible best for RNA (Figure 4B). The groove is deeper and narrower for RNA, so that the dye motion is expected to be more restricted. This is consistent with the experimental anisotropy decays (Section 3.1.2).

In the case of DNA, the calculated mean dye positions are remarkably insensitive to linkage length (Table 3), which is probably due to the fact that, with increasing linker length, the accessible volume expands in all directions as illustrated in Figure 4C. On the other hand, the distance distribution shape changes significantly with linkage length used for the simulations (Figure 4C). As discussed below, all observable FRET parameters can strongly depend on the width of the R_{DA} distribution. In other words, the knowledge of the mean dye positions is usually insufficient to predict the FRET efficiency. Thus, and to compare AV with MD results, the standard deviations of the simulated position distributions (σ_D and σ_A) are also included in Table 3. For the L linkers, the height (projection to the z-axis, Figure 4D) of the accessible volume corresponds to ± 5 base pairs, which results in significant problems for short DA distances.

3.2.2. Single-Molecule FRET Measurements of RNA Prove the Accuracy of the AV Model. For RNA (Figure 4B, right), no MD data are available to calibrate the AV parameters. To test the predictions of the AV approach nevertheless, we experimentally determined five DA distances within two internally labeled dsRNAs (RNA2 and RNA3, sequences and labeling positions can be found in section S1.4 in the Supporting Information) via

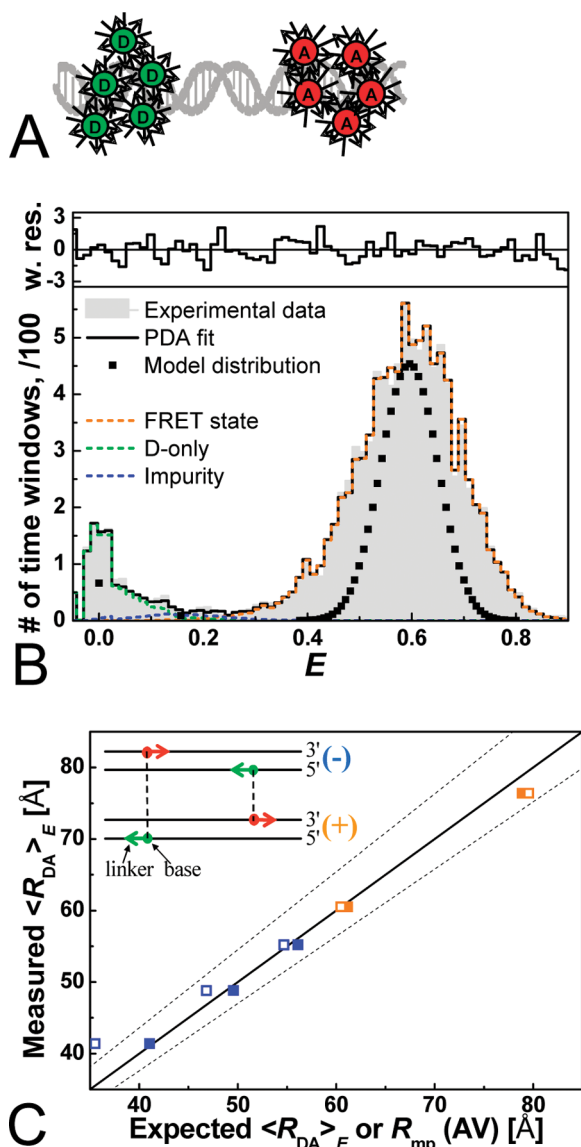


Figure 5. (A) Sketch illustrating the dynamic orientation averaging of FRET: the diffusion of the dyes in the accessible volume with the characteristic diffusion rate constant k_d and the reorientation fluctuations with the rate k_R ($k_R = 1/\rho_1$); $k_R \gg k_{FT} \gg k_d$ (k_{FT} : FRET rate constant). (B) PDA of RNA2(19–) (selected bursts). FRET efficiency histogram of experimental data (gray area) is fitted (black solid line) using the following parameters: $\langle R_{DA} \rangle_E = 48.8$ Å; $\sigma_{app} = 1.9$ Å; 12.5% of D-only; 1.6% of impurities with apparent $R_{DA} = 68.8$ Å (also present in D-only samples); $\chi^2_r = 0.91$. Weighted residuals are shown in the upper plot. PDA parameters: time window $\Delta t = 1$ ms; mean background intensities in the green and red detection channels $\langle B_G \rangle = 1.23$ kHz; $\langle B_R \rangle = 0.49$ kHz; spectral crosstalk, 1.7%; $\Phi_{FD(0)} = 0.8$; $\Phi_{FA} = 0.29$; Alexa488–Cy5 Förster radius $R_0 = 52$ Å; green/red detection efficiency ratio: 0.78. (C) DA distances $\langle R_{DA} \rangle_E$ measured by smFRET (compiled also in Table S2 in section S2.5 of the Supporting Information) plotted versus expected (simulated using AV) distances $\langle R_{DA} \rangle_E$ (■) and R_{mp} (□). The following polynomial approximation of a $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ conversion function was used: $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$. The solid line represents equal expected and measured distances. The statistical experimental errors are smaller than the symbol size. The dashed lines represent the expected uncertainties due to possible errors of κ^2 (section 3.5) given by the typical precision of 5.1% in Table 5. Inset: Sketch illustrating that for a given basepair separation between D and A, the linker orientation may lead to a decrease (–, blue) or an increase (+, orange) in R_{DA} .

smFRET. Moreover, we used these two sequences to demonstrate also the effect of a dye displacement toward the 3' end of the nucleic acid caused by the linker (inset of Figure 5C). In RNA2, both dyes are close to the 5' ends, so that they are displaced toward each other (– effect). In RNA3, both dyes are close to the 3' ends and are thus displaced in opposite directions (+ effect). To avoid any possible orientational artifacts in this benchmark study, we investigated RNAs, where the dyes were attached by L linkers.

From FRET Data $\langle R_{DA} \rangle_E$ to Structural Information R_{mp} . As we will show in sections 3.3 and 3.4, distinct time averaging regimes, from nano- to milliseconds, need to be considered for the interpretation of FRET data acquired with different experimental techniques. For this section, it is sufficient to know that all fluorophore positions of molecules with L linkers will be averaged during the millisecond dwell time of the single molecule in the confocal observation volume. Thus, the information on the width of the position distribution is lost; that is, the mean FRET efficiency $\langle E \rangle$ is observed. Because FRET efficiencies and distances are averaged differently, that is, $E(\langle R_D \rangle - \langle R_A \rangle) \neq \langle E(|R_D - R_A|) \rangle$, simulated mean DA position distances (R_{mp}) cannot be directly compared to the experimental DA distance values $\langle R_{DA} \rangle_E$ (eq 5). As $\langle R_{DA} \rangle_E \neq R_{mp}$, one cannot measure R_{mp} directly. However, for structure determination, it is necessary to obtain distances that allow a comparison between measurement and simulation. For solving this problem, several features must be considered. The easiest solution (algorithm 1) is to calculate R_{mp} from measured $\langle R_{DA} \rangle_E$ or E by applying a $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ conversion function (for details, see ref 25). However, especially for small distances, this function can be ambiguous because the slope becomes very small (Figure S4 in the Supporting Information). Moreover, $\langle R_{DA} \rangle_E$ depends not only on R_{mp} but also to some extent on the mutual orientation of dye clouds. Thus, for a safe solution (algorithm 2), it is generally advantageous to calculate the theoretical $\langle R_{DA} \rangle_E$ values and directly compare with experimental data. Because the R_{DA} distribution is directly obtained from AV simulations, theoretical $\langle R_{DA} \rangle_E$ can be calculated according to eq 5. On the other hand, due to its high speed, the easy approach with the $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ conversion function is more practical for structure modeling. For instance, if an iterative algorithm is applied to structure optimization, remodeling of the dyes' AVs and calculating $\langle R_{DA} \rangle_E$ at each iteration step may become very time-consuming, so that this cannot be done in all steps.

In eq 5A, two assumptions are made, which are sketched in Figure 5A and referred to as dynamic orientation averaging of FRET. First, the diffusion of the dyes in the sterically allowed volume with the characteristic diffusion rate constant k_d is much slower than the FRET rate constant k_{FT} . Here, by FRET rate, we mean the formal overall kinetic rate of donor quenching via FRET (typically some ns^{-1}), which can be directly measured using TCSPC. In other words, the distribution of individual DA distances $R_{DA,i}$ is quasi-static on the FRET time scale. Second, the local reorientation fluctuations with the rate constant of $k_R = 1/\rho_1$ are fast so that a mean effective orientation factor κ^2 (due to the restriction of dye motions it is not necessarily equal to 2/3) can be used. In this work, we will experimentally check the validity and error limits of assuming the “standard” case, that is, isotropic average with fast and unrestricted dye rotation ($k_R \gg k_{FT} \gg k_d$ and the wobble half angle 90° , resulting in $\kappa^2 = 2/3$).

Photon Distribution Analysis (PDA). When photon bursts of freely diffusing molecules are analyzed, it is mandatory to take into account that the obtained FRET efficiency histograms are affected by the stochastic nature of photon emission and detection

(shot noise) and other sources of dynamic or static heterogeneities. The use of an exact description for the theoretical shot noise distribution, photon distribution analysis (PDA),^{67,68} allows us to separate shot noise from inhomogeneous broadening and calculate the FRET-averaged DA distances $\langle R_{DA} \rangle_E$ (for further details on PDA, see section S2.4 in the Supporting Information). A Förster Radius $R_0 = 52$ Å of Alexa488–Cy5 (assuming an orientation factor $\kappa^2 = 2/3$, justification given in section 3.5) is used to calculate the DA distances with eq 5B. As an example, we present a smFRET histogram of the labeled sample RNA2(19–) in Figure 5B. The PDA analysis clearly shows the presence of the three species: (I) major population (85.9%) of the expected FRET species with $\langle R_{DA} \rangle_E = 48.8$ Å (orange line); (II) 12.5% of D-only (green line); and (III) 1.6% of impurities with an apparent $R_{DA} = 68.8$ Å (blue line; also present in donor-only samples). Moreover, a fixed DA distance is not sufficient for the FRET species, and a Gaussian distance distribution with an apparent distribution half width (σ_{app}) has to be used instead. The recovered σ_{app} is about 4–5% of the mean distance $\langle R_{DA} \rangle_E$ and can be attributed mainly to acceptor photophysics.⁶¹ Thus, σ_{app} must not be confused with real physical distance distributions (modeled or recovered by donor fluorescence lifetime analysis of eTCSPC). However, additional broadening leading to considerably larger values for σ_{app} can be due to distance heterogeneities as seen later (section 3.4).

In Figure 5C, the experimental FRET-averaged distances $\langle R_{DA} \rangle_E$ of all five molecules are plotted versus the calculated values for R_{mp} (open symbols) and $\langle R_{DA} \rangle_E$ (full symbols). We assumed a perfect A-RNA (parameters are given in Table S3 in the Supporting Information) for the AV model. As we do not need to find an unknown target structure, it is sufficient to calculate theoretical values of $\langle R_{DA} \rangle_E$ by an $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ conversion function (algorithm 1).

We first calculated the coordinates of the mean D- and A-positions, which are given in Table 3 (alternative representations of $\langle R_D \rangle$ and $\langle R_A \rangle$ are given in Table S3B in the Supporting Information). Next, R_{mp} values were calculated and converted to $\langle R_{DA} \rangle_E$ by using the $R_{mp} \leftrightarrow \langle R_{DA} \rangle_E$ conversion function given in the caption shown in Figure S4 (Supporting Information). In Figure 5C, the solid line with a slope of 1 indicates perfect agreement between theory and experiment. Notably, only FRET-averaged (eq 5) DA distances $\langle R_{DA} \rangle_E$ describe the FRET experiment correctly for the whole distance range; that is, the theory must include the distribution of dye positions. Especially for short distances, there are large deviations between the measured FRET distances ($\langle R_{DA} \rangle_E$) and the distances between the modeled mean positions of the dyes (R_{mp} , □), whereas the theoretical and experimental values for $\langle R_{DA} \rangle_E$ agree very well, which is also indicated by an rmsd value of 1.3 Å. The agreement is even better than expected for the given position and orientational uncertainties (dashed lines in Figure 5C; see Table 3 and section 3.5), which suggests that for the five samples studied here κ^2 is very close to 2/3, as discussed in section 3.5. Moreover, the specific linker displacement effect of the dye relative to C5 of the pyrimidine base is described correctly. This effect results in an average z -shift of the mean position of the dye of ~ 4.3 Å ($= 1/2 \Delta z_{link} + 1/2 \Delta z_{C5}$) toward the 3' end (Figure 5C). Thus, simplified modeling of dye positions by the AV method and the described correction for systematic errors is usually sufficient and agrees well with both MD and our experimental data obtained for DNA and RNA.

So far, we have neglected that the refractive index of the macromolecule should be taken into account for Förster radius calculations.^{69,70} This small effect may become necessary

to consider as the accuracy of FRET increases to angström resolution.

3.3. Influence of DA Distance Distributions on Ensemble Time-Resolved FRET Measurements. In the case of short DA distances with high FRET, uncertainties in dye position represent a large fraction of the absolute R_{DA} values, which may lead to significant systematic errors. As shown in Table 3, the “effective” linkage lengths ($|\langle R_D \rangle|$ and $|\langle R_A \rangle|$) are expected to range typically between 8 and 10 Å. Moreover, distance distribution half widths become comparable to R_{DA} . To characterize the resulting R_{DA} distributions also experimentally, we measured donor and acceptor fluorescence decays in the presence of FRET for labeled DNA and RNA samples with small separations between the dyes (10 and 11 bps for DNA1 and RNA1, respectively), by using eTCSPC. Experimental eTCSPC data in Figure 6 obtained for DNA1 samples could not be fitted using a single FRET rate, that is, with a single fluorescence lifetime to describe the FRET state (see Figure 6, upper residuals plot). We attribute the complex donor decay to a distribution of donor–acceptor distances, which is likely mainly due to the flexibility of the dye linkers. In view of the averaging regime, which has been discussed in section 3.2.2, fluorescence lifetime measurements allow one to obtain snapshots of heterogeneities, which live longer than the fluorescence lifetime (in our case 4 ns). In other words, due to its high time resolution, TCSPC can be used to characterize DA distance distributions $p(R_{DA})$. For example, the analysis of donor fluorescence decays $F_D(t)$ (D decay) by eq 6 recovers $p(R_{DA})$.^{7,24,71}

$$F_D(t) = \int_{R_{DA}} p(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}}[1 + (R_0/R_{DA})^6]\right) dR_{DA} \quad (6)$$

where $\tau_{D(0)}$ is the donor fluorescence lifetime without acceptor. For simplicity, the distribution $p(R_{DA})$ can be assumed to be Gaussian. Considering also the presence of donor-only molecules, the fitting parameters of eq 7 are then the mean DA distance $\langle R_{DA} \rangle$, the half-width σ_{DA} of the R_{DA} distribution, and also the fraction of donor-only molecules x_D (in our measurements typically below 10%):

$$F_D(t) = (1 - x_D) \int_{R_{DA}} \frac{1}{\sqrt{2\pi}\sigma_{DA}} \exp\left(-\frac{(R_{DA} - \langle R_{DA} \rangle)^2}{2\sigma_{DA}^2}\right) \exp\left(-\frac{t}{\tau_{D(0)}}[1 + (R_0/R_{DA})^6]\right) dR_{DA} + x_D \exp\left(-\frac{t}{\tau_{D(0)}}\right) \quad (7)$$

To test the accuracy of the AV model, we compared the mean and the half-width of the distribution $p(R_{DA})$ obtained by fitting eq 7 to experimental TCSPC data, with those predicted by the AV simulations. A good agreement between experimental and simulated data was found (see section 3.3.1). To justify the use of a Gaussian distribution in eq 7, we simulated a DA distribution of a FRET experiment using two AV position distributions (see Figure S5A in the Supporting Information). Interestingly, this DA distribution $p(R_{DA})$ is very well described by a Gaussian distribution. The validity of the approximation of $p(R_{DA})$ by a single Gaussian distribution can be checked experimentally as demonstrated in section 3.4.

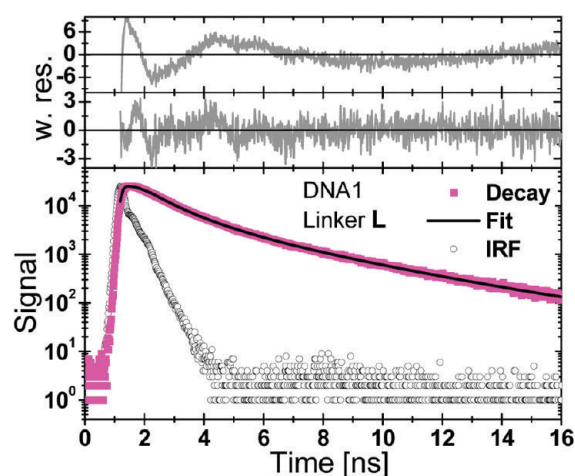


Figure 6. eTCSPC measurement of the fluorescence decay of Alexa488 obtained for the DNA1 FRET sample with L linker. Experimental data (magenta ■), instrument response function (IRF, ○), and the fit assuming a Gaussian distribution of distances (eq 7; bottom panel, solid black line) are shown. Weighted residuals are presented in the upper plots; (top) formal biexponential fluorescence decay with a single FRET state corresponding to $\tau_{D(A)} = 1.0$ ns (87.5%), and a donor-only decay with $\tau_{D(0)} = 4.1$ ns (12.5%); $\chi^2_r = 10.3$; (middle) Gaussian distribution of distances (parameters are given in Table 4) and donor-only decay (eq 7) with $\tau_{D(0)} = 4.1$ ns (6.7%); $\chi^2_r = 1.35$. The fit ranges from the maximum of the IRF to the first time channel with less than 100 detected photons.

The extension of eqs 6 and 7 to the case of multiexponential fluorescence relaxation of D-only can be easily made (see ref 72 and section S2.6 in the Supporting Information).

Moreover, the characteristic rise time of the acceptor fluorescence (A rise) also contains information on the FRET rates^{72–74} (for details, see section S2.6 and Figure S5B in the Supporting Information).

3.3.1. Measured $\langle R_{DA} \rangle$ Distances Do Not Compare to Simulated R_{mp} . The simulated and measured distances are presented in Table 4. Clearly, the values for distances between mean dye positions (R_{mp}) simulated by the AV method do not agree with the distances measured by eTCSPC and PDA. To explain these deviations, one should note that, like for $\langle R_{DA} \rangle_E$ (section 3.2.2), the mean distance $\langle R_{DA} \rangle$ measured by eTCSPC (eqs 6 and 7) is not equal to R_{mp} but is rather given by eq 8:

$$\langle R_{DA} \rangle = \langle |R_D - R_A| \rangle \geq |\langle R_D \rangle - \langle R_A \rangle| \quad (8)$$

In other words, $\langle R_{DA} \rangle \geq R_{mp}$, so that a difference of >5 Å can be seen for broader distance distributions (Table 4). For instance, the distance between the centers of completely overlapping dye clouds is zero ($R_{mp} = 0$), whereas the average distance between individual dye positions is not ($\langle R_{DA} \rangle > 0$). Thus, a realistic modeling of dye position distributions is needed not only for PDA but also for a proper interpretation of time-resolved fluorescence decays. If one compares $\langle R_{DA} \rangle$ with $\langle R_{DA} \rangle_E$, the latter is typically weighted toward R_0 , that is, $\langle R_{DA} \rangle_E \geq \langle R_{DA} \rangle$ for $\langle R_{DA} \rangle < R_0$ and $\langle R_{DA} \rangle_E \leq \langle R_{DA} \rangle$ for $\langle R_{DA} \rangle > R_0$.

Analysis of DNA. For DNA samples with L and I linker, the agreement between the simulated and experimental $\langle R_{DA} \rangle$ values obtained by eTCSPC and PDA is good, and remaining deviations are well within the expected position (Table 3) and the κ^2 -related errors (section 3.5). $\langle R_{DA} \rangle$ and σ_{DA} were also extracted by fitting the acceptor rises (Table 4). The good agreement between

Table 4. DA Distance Distribution Parameters Calculated from eTCSPC and smFRET Data in Comparison with AV Simulations^a

linker	eTCSPC experiment		PDA		AV simulation	
	$\langle R_{DA} \rangle$ (Å), σ_{DA} (Å) (D decay)	$\langle R_{DA} \rangle$ (Å), σ_{DA} (Å) (A rise)	$\langle R_{DA} \rangle_E$ (Å), σ_{app} (Å)	R_{mp} (Å)	$\langle R_{DA} \rangle$ (Å), σ_{DA} (Å)	
DNA1						
L	39.5 (7.5)	42.3 (6.8)	42.5 (1.5)	35.5	39.9 (8.8)	
I	40.8 (5.9)	40.8 (8.2)	41.5 (1.5)	33.9	36.3 (6.0)	
S _t	38.4 (11.2)	45.6 (7.7)	45.9 (2.9)	32.7	35.6 (4.8)	
RNA1						
L	41.5 (7.3)	43.2 (6.3)	42.8 (1.6)	27.5	31.6 (6.3)	
S _d	35.1 (6.0)	34.5 (6.8)	36.6 (1.0)	22.7	23.9 (5.1)	
S _t	34.4 (7.6)	34.6 (11.2)	36.5 (4.7)	23.0	24.4 (5.3)	

^aThe dye–dye separation is 10 and 11 basepairs for DNA and RNA, respectively.

the values calculated from donor and acceptor decay curves shows that donor quenching is indeed due to FRET and not due to, for example, local quenching artifacts. For S_t linkers, systematic deviations are seen, and the R_{DA} distributions are surprisingly broad ($\sigma_{DA} = 11.2$ Å), which is clearly unexpected for the shortest linker. We will show below that a rather irregular distribution of stable conformations of one or both dyes must exist for S_t linkers, which cannot be described with a single Gaussian peak as assumed in eq 7.

Analysis of RNA. In the case of RNA, a similarly broad R_{DA} distribution can be observed for the acceptor rise for the S_t linker. Furthermore, the agreement between theory and experiment is generally worse than for DNA. The main reason is that the DA distances are quite short (less than one-half of the Förster radius of 52 Å). At such short distances, the point dipole approximation does not hold^{8,20,75,76} and should be replaced accordingly, which is beyond the scope of the work. Furthermore, dye–dye interactions⁷⁷ cannot be excluded. However, let us point out that the measurements with short linkers (S_d or S_t) provide distances that are much closer to the structurally relevant distance between the C5 atoms of the labeled nucleobases (30 Å) than for L linkers.

3.3.2. Translational Linker Movements Are Slow on the Time Scale of FRET. eTCSPC data provide additional evidence for quasi-static distance distributions on the time scale of FRET as has been postulated before.²⁵ Considering L, I, and S_d linkers, the fit of R_{DA} distributions to experimental fluorescence decays typically yields distribution half-widths σ_{DA} comparable to simulated ones (Table 4). This indicates that DA distributions due to linker motions are not significantly averaged out on the nanosecond time scale. There is no contradiction between this finding and subns anisotropy decay times: for example, Alexa488 readily rotates about the linker axis, which involves strong fluorescence depolarization but very little distance fluctuations. In the case of low FRET, there might be more averaging of distances; however, taking into account large absolute values of R_{DA} , only minor relative errors could be expected because the differences in E for dynamic and static averaging are small for large distances (for low FRET, $\langle R_{DA} \rangle \approx R_{mp}$ irrespective of the averaging regime).

3.4. FRET Broadening on the Millisecond Time Scale. To understand the surprisingly broad R_{DA} distributions for the S_t

linkers (section 3.3), we performed single-molecule experiments on freely diffusing DNA1 and RNA1 molecules to check whether the distributions are averaged out on the millisecond time scale. Figure 7A–D shows 2D probability histograms of FRET efficiency E versus the donor lifetime $\tau_{D(A)}$, where the corresponding 1D parameter histograms are given as projections. The number of molecules (fluorescence bursts) in each bin is gray scale shaded from white (lowest) to black (highest). All 2D plots show two distinct peaks: (I) at $E \approx 0$ and $\tau_{D(A)} \approx 4$ ns, due to D-only populations, and (II) at $E \approx 0.7$ – 0.8 and $\tau_{D(A)} \approx 1$ – 2 ns attributed to FRET subpopulations. In all plots, solid lines indicate the expected E versus $\tau_{D(A)}$ dependence for dynamic FRET taking fast linker dynamics into account⁷⁸ (eqs S12 and S13 in section S2.6 of the Supporting Information). DNA and RNA samples with L linkers show homogeneous uncorrelated E – $\tau_{D(A)}$ distributions of the FRET population (Figure 7A and B), which are distributed approximately horizontally due to independent shot noise distributions of E ⁶⁸ and $\tau_{D(A)}$.⁷⁹ All samples with I and S_d linkers exhibit similar patterns (see Figure S6 for RNA1 with S_d linker in section S2.7 of the Supporting Information).

On the contrary, DNA1 and RNA1 with S_t linkers have broad and asymmetric 2D distributions of E – $\tau_{D(A)}$ (Figure 7C and D). Moreover, for S_t linkers, a correlation between E and $\tau_{D(A)}$ within the FRET subpopulation is apparent from 2D plots, indicating a heterogeneous distribution of DA distances on the millisecond time scale. This fact is consistent with the eTCSPC data (Table 4), which show unexpectedly broad R_{DA} distributions obtained exclusively for S_t linkers. The comparison of two complementary ds RNA1 FRET samples with different linkers for D and A, respectively, yields that this is mainly caused by the donor. Only the sample labeled with linker S_t at the donor and linker L at the acceptor position shows E broadening, whereas E is narrow for the RNA1 sample with the opposite linker combination of the dyes (Figure S7 in section S2.7 of the Supporting Information).

PDA. To quantify these visual effects, we performed PDA of the smFRET data shown in Figure 7A–D. The mean FRET-averaged DA distances $\langle R_{DA} \rangle_E$ recovered by PDA are typically similar to $\langle R_{DA} \rangle$ values found by eTCSPC (Table 4). A large deviation is observed only for DNA1 with S_t linkers, indicating distinct dye subpopulations. Moreover, the apparent R_{DA} distribution half-widths found by PDA (σ_{app}) are significantly larger for S_t linkers (Table 4). For instance, for RNA1, σ_{app} amounts $\sim 13\%$ of $\langle R_{DA} \rangle_E$ for S_t linkers, whereas for other linkers σ_{app} does not exceed 3–4% of $\langle R_{DA} \rangle_E$. This fact indicates the presence of FRET heterogeneities in addition to the complex acceptor's photophysics, which is responsible for the minimal broadening of few percent of $\langle R_{DA} \rangle_E$.⁶¹

Maximum Entropy Deconvolution. To further support this conclusion, we reanalyzed all data with S_t linkers and used an unbiased model-free DA distance distribution instead of the assumed Gaussian distribution.^{47,48} We performed a maximum entropy (ME) deconvolution^{47,48} of R_{DA} distributions observed for S_t linkers on different time scales (Figure 7E and F): eTCSPC data for the nanosecond time scale (dashed lines) and PDA of smFRET intensity data for the millisecond time scale (solid lines). ME deconvolutions of both experiment types indicate the presence of multiple FRET states. eTCSPC data show that S_t linkers have two stable conformations (corresponding to two major “stable” peaks “s1” and “s2” in Figure 7E and F) possibly with some additional flexibility leading to broadening of these peaks. For RNA1, PDA data show a peak between the two DA distances found by eTCSPC (“mixing” peak “m” in Figure 7F), which suggests that additional averaging takes place on a time scale faster than the dwell time

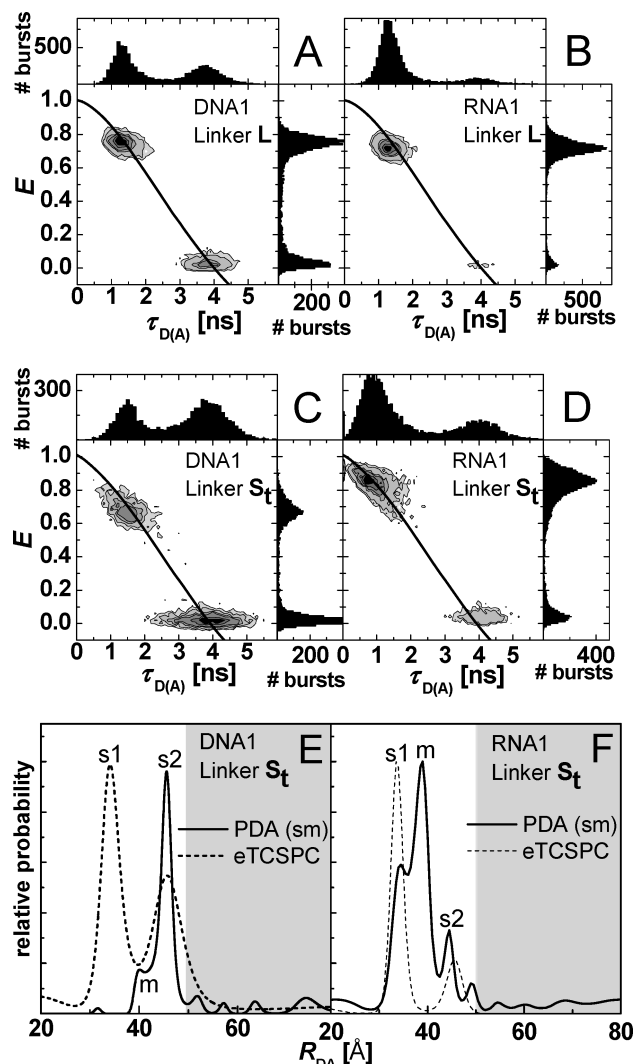


Figure 7. (A–D) 2D probability histograms of FRET efficiency E versus the donor lifetime $\tau_{D(A)}$ generated from smFRET data obtained for (A) DNA1, L linkers; (B) RNA1, L linkers; (C) DNA1, S_t linkers; and (D) RNA1, S_t linkers. The number of molecules (fluorescence bursts) in each bin is gray scale shaded from white (lowest) to black (highest). The corresponding 1D parameter histograms are given as projections. In all plots, solid lines indicate the E versus $\tau_{D(A)}$ dependence given by eqs S12 and S13 in section S2.6 of the Supporting Information. (E,F) Maximum entropy deconvolution of donor–acceptor distance distributions for (E) DNA1, S_t linkers and (F) RNA1, S_t linkers. The solid and dashed lines represent deconvolutions of single-molecule intensity data and donor decays measured by eTCSPC, respectively. The ME method allows one to extract distributions of fluorescence lifetimes from eTCSPC data,⁴⁸ which can be converted into distributions of R_{DA} . In PDA,⁶⁸ distance ME distributions are directly used to calculate FRET efficiency distributions. The shaded areas indicate the region where the observed peaks may represent multimolecular and photobleaching events. The peaks “s1” and “s2” correspond to stable dye conformations, whereas the peak “m” is likely due to averaging of the states s1 and s2 on the millisecond time scale.

(ms). Dynamic PDA^{78,80} provides no indication for “mixing” with millisecond characteristic times (see Figure S8 in the Supporting Information). Thus, transitions responsible for the “mixing” peak occur on a significantly faster time scale, whereas the “s1” and “s2” peaks are stable during the burst duration.

For DNA1, a significant discrepancy between PDA and eTCSPC data is observed (cf., Figure 7E and Tables 1 and 4). This effect could be due to a static quenching of the acceptor in the state corresponding to the shorter R_{DA} (peak “s1” in Figure 7E). If so, the s1 state would be nearly invisible in SMD measurements because both dyes are quenched in this state (D is quenched by high FRET), which also explains the disagreement between donor decay and acceptor rise data (Table 4). This explanation is supported by ensemble measurements of the acceptor fluorescence quantum yield Φ_{FA} , which is reduced for the S_t linker ($\Phi_{FA} = 0.21$) as compared to the L linker ($\Phi_{FA} = 0.38$, Table 1).

In summary, complex distributions of DA distances found for S_t linkers severely complicate their use in quantitative FRET studies. Thus, an optimal linker must always allow for free diffusion within the dye’s accessible volume. On the other hand, to minimize position uncertainties and to prevent dye–dye interactions, linkers should not be longer than necessary. For DNA and Alexa488, the I linker seems to fit best to these requirements. For Cy5, shorter S_d and S_t linkers work equally well because the dye has an extra 6-atom linker between the chromophore and the reactive coupling group (Figure 2). Hence, for undefined dye environments, where modeling of dye positions is impossible, I and S_d are well suitable linkers as they provide better absolute distance estimations.

3.5. Minimizing Uncertainties Due to the Orientational Factor κ^2 . For accurate FRET analysis, we must consider not only translational linker diffusion, which is usually slower than ns, but also orientational dynamics. Given subnanosecond local rotational correlation times in Table 2, dynamic orientational averaging can be assumed at least to some extent. Dye reorientation dynamics on the time scale of FRET (k_{FT} , see section 3.2.2)⁸¹ could, thus, be relevant only for very short distances, which is beyond the scope of this Article. The anisotropy measurements indicate, however, that the orientational distribution of both D and A is not strictly isotropic even for L-linkers (Table 2), contrary to the MD data from²⁵ (for details, see section S2.8 in the Supporting Information). Therefore, detailed analysis of related κ^2 effects for all linkers must be performed.

As mentioned above, the modeling of the accessible space of a dye is much less important for short linkers, making them more suitable for quantitative FRET measurements in the case of undefined environments. One potential problem when using short linkers, however, is the estimation of the orientation factor κ^2 . The assumption of $\kappa^2 \cong 2/3$ might not be justified for short linkers as their movement is more restricted. In this section, we estimate the range of possible values for κ^2 ^{19,82,83} for the different linkers of DNA1 and RNA1 and determine potential errors for FRET distance measurements if $\kappa^2 = 2/3$ is assumed in the calculation of R_0 .

To minimize the uncertainty in κ^2 , we use the residual anisotropies that result from the measurements of the donor-only, the acceptor-only (Figure 3B in Section 3.1.2.), and the FRET-sensitized acceptor (Figure 8A) anisotropy decays ($r_{\infty,D}$, $r_{\infty,A}$ and $r_{\infty,A(D)}$, respectively), to calculate the range of all possible values for κ^2 by eq 9:

$$\begin{aligned} \kappa^2 = & \frac{2}{3} + \frac{2}{3} S_D^{(2)} S^{(2)}(\beta_1) + \frac{2}{3} S_A^{(2)} S^{(2)}(\beta_2) + \frac{2}{3} S_D^{(2)} S_A^{(2)} S^{(2)}(\delta) \\ & + 6 S^{(2)}(\beta_1) S^{(2)}(\beta_2) + 1 + 2 S^{(2)}(\beta_1) + 2 S^{(2)}(\beta_2) \\ & - 9 \cos \beta_1 \cos \beta_2 \cos \delta \end{aligned} \quad (9)$$

In eq 9, β_1 and β_2 are the angles between the symmetry axes of the rotations of the dyes and the distance vector R_{DA} , while δ is the angle between the symmetry axes (Figure 8B).

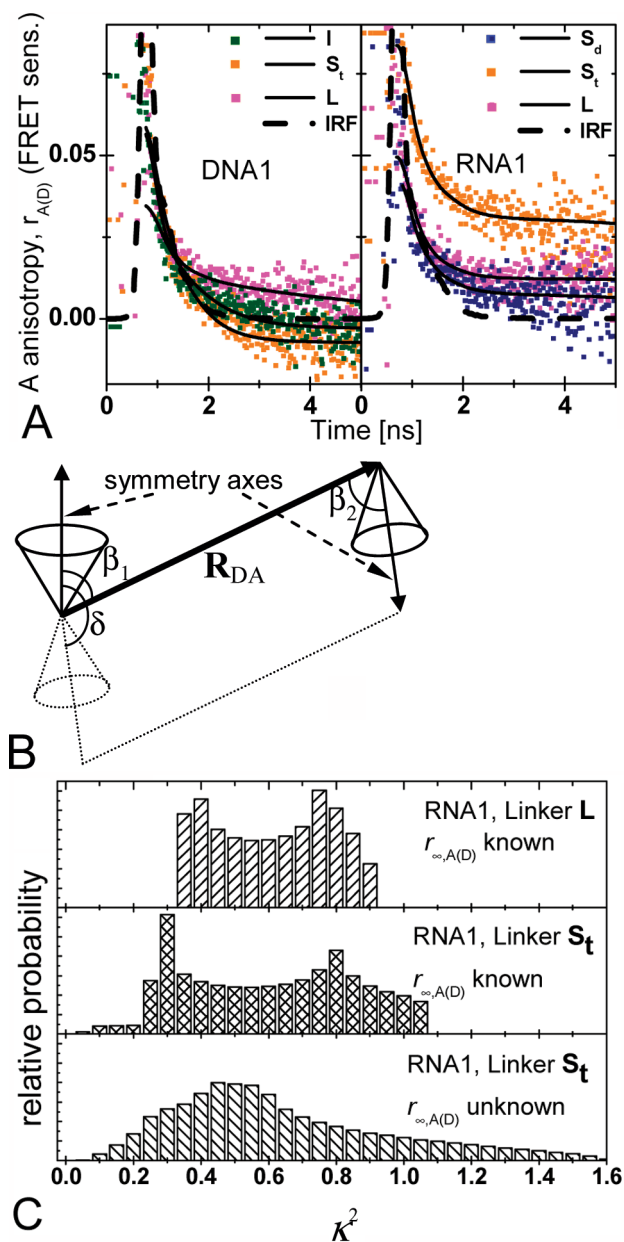


Figure 8. (A) FRET-sensitized acceptor anisotropy decays (dots), the rescaled instrument response functions (IRF, dashed line), and the anisotropy decays (black solid lines) of DNA (left) and RNA (right) (fit results, see Table 5). (B) Sketch showing angles that define the orientational factor κ^2 in eq 9: β_1 and β_2 are the angles between the symmetry axes of the rotations of the dyes and the distance vector R_{DA} , and δ is the angle between the symmetry axes. (C) Calculated probability distributions for possible values of κ^2 in RNA1: (top) for linker L, (middle) for linker S_t and in both cases taking into account $r_{\infty,A(D)}$; (bottom) for linker S_t and only using the offsets from the anisotropy decays of the donor and acceptor molecules $r_{\infty,D}$ and $r_{\infty,A}$.

The necessary second-rank order parameters $S^{(2)}$ are defined in eq 3 ($S_D^{(2)}$) and eq 4 ($S_A^{(2)}$) and formally by:

$$\begin{aligned} S^{(2)}(\delta) &= \frac{1}{2}(3 \cos^2 \delta - 1) = \frac{r_{\infty,A(D)}}{r_0 S_D^{(2)} S_A^{(2)}}, \quad S^{(2)}(\beta_1) \\ &= \frac{1}{2}(3 \cos^2 \beta_1 - 1), \quad \text{and } S^{(2)}(\beta_2) = \frac{1}{2}(3 \cos^2 \beta_2 - 1) \end{aligned} \quad (10)$$

Table 5. Fit Parameters for FRET-Sensitized Acceptor Anisotropy Decay ($r_0 = 0.38$) and Results for κ^2 Probability Distributions Taking into Account $r_{\infty,D}$, $r_{\infty,A}$ (Table 2), and $r_{\infty,A(D)}$

linker	sensitized anisotropy			border values: worst cases			mean value: typical case		
	r_s	ρ_1 , ns (b_1)	$\rho_{2(\text{global})}$, ns ($b_2 = r_{\infty,DA}$)	κ^2_{\min}	κ^2_{\max}	ΔR_{DA}	κ^2_{mean}	accuracy of R_{DA}	precision ΔR_{DA}
DNA1									
I	0.005	0.605 (0.042)	>60 (−0.003)	0.25	0.94	−6%...+15%	0.6	+3.1%	7.0%
S _t	0.002	0.400 (0.080)	>60 (−0.007)	0.22	0.92	−6%...+17%	0.59	+2.6%	5.9%
L	0.012	0.227 (0.056)	10 (0.010)	0.42	0.92	−6%...+ 7%	0.67	+1.7%	4.4%
RNA1									
S _d	0.012	0.260 (0.045)	>60 (0.007)	0.10	1.57	−13%...+37%	0.64	+0.8%	8.4%
S _t	0.038	0.258 (0.069)	>60 (0.031)	0.09	1.1	−9%...+28%	0.57	+3.9%	8.9%
L	0.016	0.246 (0.047)	>60 (0.012)	0.38	0.93	−6%...+ 9%	0.62	+2.2%	5.1%

κ^2 values (eq 9) cannot be calculated unambiguously because, in general, the angles β_1 and β_2 are not experimentally accessible. However, for an experimentally determined δ (eq 10), it is possible to define a range of possible values for β_1 and β_2 (eqs S16A and S16B in section S2.8 of the Supporting Information), which allows one to calculate the range of κ^2 values (eq 9) compatible with the experimental data ($r_{\infty,D}$, $r_{\infty,A}$, and $r_{\infty,A(D)}$).

Note that κ^2 does not explicitly depend on the cone opening half angles θ_D and θ_A (section 3.1.2) and the assumption of dye reorientation within a cone/disk (Figure 3C and D); that is, even if this approximation is considered as unrealistic, it is not needed to obtain eq 9. However, axially symmetric transition dipole orientation distributions are usually assumed for κ^2 estimation,¹⁹ which might be not exactly the case here (Figure 4). Therefore, it is absolutely reasonable to expect that additional averaging of mutual orientations of D and A by diffusion along the DNA or RNA groove would bring the effective κ^2 even closer to 2/3 than given by eq 9. This might be one of the reasons for the (unexpected) high accuracy of our FRET benchmark study using RNA2 and RNA3 (Figure 5C in section 3.2.2). Thus, the κ^2 estimations given below represent the worst case scenario. Moreover, in this study, we cannot investigate a possible correlation between R_{DA} and κ^2 as proposed in a few recent theoretical works^{20,75,76} for the case that the dye exhibits a slow exchange between different microenvironments. Our simplified (AV) simulations do not allow us to discuss this effect.

3.5.1. Estimation of κ^2 Using $r_{\infty,A(D)}$. The residual anisotropies $r_{\infty,D}$, $r_{\infty,A}$, and $r_{\infty,A(D)}$ can be used to calculate a probability distribution of κ^2 values. For the estimation of $r_{\infty,A(D)}$, the FRET-sensitized acceptor anisotropy decays were studied for all linkers of DNA1 and RNA1 (see section 2 for the measurement and analysis procedures). The data and the results of the analysis are shown in Figure 8A and Table 5, respectively. In all cases, the decays could be fitted by a biexponential decay consisting of a fast decay time $\rho_1 \approx 0.4$ ns (resulting from fast FRET (k_{FT}) and local reorientations (k_R)) and a slow decay time ρ_2 (global motion). As in section 3.1, the slow component is approximated by a time-independent offset $r_{\infty,A(D)}$ and considered to be the residual anisotropy. It is largest for the linker S_t in RNA and is negative for the linkers I and S_t in DNA, which indicates that the transition dipole moments of the dyes are preferentially orientated perpendicular to each other. Indeed, the differences between all anisotropy decays indicate distinct mean dye orientations. Moreover, smFRET measurements were analyzed by inspecting 2D probability histograms of steady state anisotropy r_s versus the

donor lifetime $\tau_{D(A)}$ to make sure that the resulting residual anisotropies $r_{\infty,A(D)}$ are due to restricted reorientation of the dyes and not due to a fraction of molecules where the dye is immobile. For linkers L, the FRET populations in the $r_s - \tau_{D(A)}$ 2D plots appear to be symmetric at $r_s \approx 0.1$ and $\tau_{D(A)} \approx 1.3$ ns. Thus, there is only one anisotropy population for each lifetime population. For shorter linkers, the anisotropy distributions become less symmetric. The existence of additional donor populations is most pronounced for S_t linkers with a fraction of a 2-fold increased steady-state anisotropy, which amounts to less than 29% (see Figure S9 in section S2.8 of the Supporting Information for all FRET pairs). However, in no case could a completely immobile dye species be detected.

Three distributions of possible κ^2 values are presented in Figure 8C with the corresponding parameters compiled in Table 5. The smallest and the largest possible values κ^2_{\min} and κ^2_{\max} represent the worst case scenario for a FRET distance measurement for which $\kappa^2 = 2/3$ is assumed. The resulting range of potential relative errors (worst case) for distance measurements ΔR_{DA} typically vary from −6% to +15% (Table 5). For a range of possible κ^2 values, one can define an accuracy (systematic error) and a precision (uncertainty; eqs S17A and S17B in section S2.8 of the Supporting Information), which would result in a typical uncertainty of κ^2 in the FRET distance measurements presented here. These parameters are shown in Table 5, and it is obvious that the errors and deviations are relatively small in all cases, although they become larger for short linkers (see also Figure 8C). Nevertheless, the distances measured using short linkers are reasonably accurate.

3.5.2. Estimation of κ^2 with Unknown $r_{\infty,A(D)}$. For the estimation of κ^2 , many groups only use $r_{\infty,D}$ and $r_{\infty,A}$,⁷ which result from the anisotropy decays of the D-only and the A-only molecules, respectively, and do not measure the FRET sensitized acceptor anisotropy decay. Hence, $r_{\infty,A(D)}$ is unknown, and δ must be allowed to be every value between 0° and 90°. For comparison, the probability distributions for κ^2 for all linkers of DNA1 and RNA1 were also calculated accordingly; that is, only $r_{\infty,D}$ and $r_{\infty,A}$ were taken into account. The results are shown in Table S4 in section S2.8 of the Supporting Information. There are only slight changes in accuracy and precision, but the range of possible κ^2 values nearly doubles. Thus, the worst-case ΔR_{DA} increases especially toward smaller distance values from typically $\sim -6\%$ for known $r_{\infty,A(D)}$ (section 3.5.1) to typically -15% for unknown $r_{\infty,A(D)}$. Figure 8C shows the probability distributions for the RNA molecules for the linkers L and S_t in case all three residual

anisotropies are taken into account (top and middle, respectively) and for linker S_t when only $r_{\infty,D}$ and $r_{\infty,A}$ are used (bottom).

It is worth mentioning that for the estimation of κ^2 the steady-state anisotropy r_s is commonly used instead of the residual anisotropy r_{∞} .^{7,29} This method, however, overestimates κ^2 -related errors even further as r_s is usually larger than r_{∞} .

4. CONCLUSION

In this work, we introduce new short dye linkers for labeling of DNA and RNA. FRET measurements on dsDNA and dsRNA model systems test their suitability for quantitative studies. For well-defined environments and if the dye diffuses freely in the sterically allowed, mean positions of the dyes and R_{DA} distributions can be accurately modeled using relatively simple and fast accessible volume simulations. The translational motions of the dyes appear to be slow on the time scale of the fluorescence lifetimes and depend on linker size and structure. For L, I, and S_d linkers, the motions are completely averaged out on the millisecond time scale, whereas S_t linkers exhibit a complex distribution of R_{DA} due to inhibited diffusion through the accessible volume. It became also clear that experimentally measured distances, especially for long linkers, cannot be directly compared to the structurally relevant mean position distance R_{mp} or the distance between C5 atoms of uracil. Because of broad distributions of DA distances, there is a large discrepancy between R_{mp} and $\langle R_{DA} \rangle_E$ (measured by smFRET) or $\langle R_{DA} \rangle$ (measured by eTCSPC), which must be always taken into account. When doing so, high precision distance measurements are possible. This correction and, therefore, modeling of dye positions is less important for short linkers. This makes them particularly suitable for undefined environments. However, due to the inhomogeneities of DA distances observable for S_t as a linker for Alexa488, we advise against its use and recommend the S_d or the I linker instead.

It became clear that, when calculating a probability distribution for possible values of κ^2 , its width can be further reduced when taking into account not only the residual anisotropies of the donor and acceptor $r_{\infty,D}$ and $r_{\infty,A}$ but also $r_{\infty,A(D)}$, which results from the FRET-sensitized acceptor anisotropy decay. Furthermore, for short and intermediate linkers, κ^2 -related errors are only slightly higher than for long linkers. When using them for quantitative FRET measurements of internally labeled nucleic acids with Alexa488 and Cy5 as a FRET pair, it is, therefore, safe to assume κ^2 to be 2/3. For unknown local environments, long linkers increase uncertainties significantly more.

■ ASSOCIATED CONTENT

S Supporting Information. Additional experimental procedures, DNA and RNA sample information, and additional results. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

stanislav.kalinin@uni-duesseldorf.de; sabine.mueller@uni-greifswald.de; cseidel@hhu.de

Author Contributions

^{||}These authors contributed equally.

■ ACKNOWLEDGMENT

C.A.M.S., S.M., H.N., and S.K. thank the German Science foundation (DFG) in the priority program SPP 1258 "Sensory and regulatory RNAs in prokaryotes" for funding this work. S.S. thanks the NRW Research School Biostruct for funding. We are grateful to Hayk Vardanyan for help in single-molecule measurements. We thank Evangelos Sisamakias, Hugo Sanabria, Ralf Kühnemuth, and Suren Felekyan for very fruitful discussions. We thank Anna K. Woźniak for providing the cartoon for the TOC graphic.

■ REFERENCES

- (1) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 719–726.
- (2) Clegg, R. M. *Methods Enzymol.* **1992**, *211*, 353–388.
- (3) van der Meer, B. W.; Cooker, G.; Chen, S. Y. *Resonance Energy Transfer: Theory and Data*; VCH Publishers: New York, 1994.
- (4) Sisamakias, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. *Methods Enzymol.* **2010**, *475*, 456–514.
- (5) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5*, 507–516.
- (6) Förster, T. *Ann. Phys.* **1948**, *437*, 55–75.
- (7) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Kluwer Academic/Plenum Publishers: New York, 1999.
- (8) Valeur, B. *Molecular Fluorescence: Principles and Applications*; Wiley-VCH Verlag: Weinheim, 2002.
- (9) Scholes, G. D. *Annu. Rev. Phys. Chem.* **2003**, *54*, 57–87.
- (10) Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6264–6268.
- (11) Deniz, A. A.; Dahan, M.; Grunwell, J. R.; Ha, T. J.; Faulhaber, A. E.; Chemla, D. S.; Weiss, S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3670–3675.
- (12) Rothwell, P. J.; Berger, S.; Kensch, O.; Felekyan, S.; Antonik, M.; Wöhrle, B. M.; Restle, T.; Goody, R. S.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1655–1660.
- (13) Margittai, M.; Widengren, J.; Schweinberger, E.; Schröder, G. F.; Felekyan, S.; Hausteiner, E.; König, M.; Fasshauer, D.; Grubmüller, H.; Jahn, R.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15516–15521.
- (14) Sanborn, M. E.; Connolly, B. K.; Gurunathan, K.; Levitus, M. *J. Phys. Chem. B* **2007**, *111*, 11064–11074.
- (15) Neubauer, H.; Gaiko, N.; Berger, S.; Schaffer, J.; Eggeling, C.; Tuma, J.; Verdier, L.; Seidel, C. A. M.; Griesinger, C.; Volkmer, A. *J. Am. Chem. Soc.* **2007**, *129*, 12746–12755.
- (16) Norman, D. G.; Grainger, R. J.; Uhrin, D.; Lilley, D. M. *J. Biochemistry* **2000**, *39*, 6317–6324.
- (17) Vámosi, G.; Gohlke, C.; Clegg, R. M. *Biophys. J.* **1996**, *71*, 972–994.
- (18) Ranjit, S.; Gurunathan, K.; Levitus, M. *J. Phys. Chem. B* **2009**, *113*, 7861–7866.
- (19) Dale, R. E.; Eisinger, J.; Blumberg, W. E. *Biophys. J.* **1979**, *26*, 161–193.
- (20) VanBeek, D. B.; Zwier, M. C.; Shorb, J. M.; Krueger, B. P. *Biophys. J.* **2007**, *92*, 4168–4178.
- (21) Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T. J.; Giraud, G.; Norman, D. G.; Ha, T.; Lilley, D. M. *J. Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11176–11181.
- (22) Clegg, R. M.; Murchie, A. I. H.; Zechel, A.; Carlberg, C.; Diekmann, S.; Lilley, D. M. *J. Biochemistry* **1992**, *31*, 4846–4856.
- (23) Clegg, R. M.; Murchie, A. I. H.; Zechel, A.; Lilley, D. M. *J. Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2994–2998.
- (24) Haas, E. *ChemPhysChem* **2005**, *6*, 858–870.
- (25) Woźniak, A. K.; Schröder, G.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18337–18342.
- (26) Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18964–18969.
- (27) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528–1533.

- (28) Cai, Q.; Kusnetzow, A. K.; Hideg, K.; Price, E. A.; Haworth, I. S.; Qin, P. Z. *Biophys. J.* **2007**, *93*, 2110–2117.
- (29) Muschielok, A.; Andrecka, J.; Jawhari, A.; Bruckner, F.; Cramer, P.; Michaelis, J. *Nat. Methods* **2008**, *5*, 965–971.
- (30) Xie, Y.; Maxson, T.; Tor, Y. J. *Am. Chem. Soc.* **2010**, *132*, 11896–11897.
- (31) Borjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. *J. Am. Chem. Soc.* **2009**, *131*, 4288–4293.
- (32) Lewis, F. D. *Pure Appl. Chem.* **2006**, *78*, 2287–2295.
- (33) Agrofoglio, L. A.; Gillaizeau, I.; Saito, Y. *Chem. Rev.* **2003**, *103*, 1875–1916.
- (34) Heck, R. F. *Org. React.* **1982**, *27*, 345–390.
- (35) Hobbs, F. W., Jr. *J. Org. Chem.* **1989**, *54*.
- (36) Dey, S.; Sheppard, T. L. *Org. Lett.* **2001**, *3*, 3983–3986.
- (37) Muschielok, A.; Michaelis, J. *FRETnps Tools, Version 2.6.1*; Ludwig Maximilians University: Munich, 2008.
- (38) Widengren, J.; Kudryavtsev, V.; Antonik, M.; Berger, S.; Gerken, M.; Seidel, C. A. M. *Anal. Chem.* **2006**, *78*, 2039–2050.
- (39) Cruickshank, K. A.; Stockwell, D. L. *Tetrahedron Lett.* **1988**, *29*, 5221–5224.
- (40) McKeen, C. M.; Brown, L. J.; Nicol, J. T. G.; Mellor, J. M.; Brown, T. *Org. Biomol. Chem.* **2003**, *1*, 2267–2275.
- (41) Schaller, H.; Weimann, G.; Khorana, H. G.; Lerch, B. J. *Am. Chem. Soc.* **1963**, *85*, 3821.
- (42) Shah, K.; Wu, H. Y.; Rana, T. M. *Bioconjugate Chem.* **1994**, *5*, 508–512.
- (43) Hakimelahi, G. H.; Proba, Z. A.; Ogilvie, K. K. *Can. J. Chem.* **1982**, *60*, 1106–1113.
- (44) Bannwarth, W. *Chimia* **1987**, *41*, 302–317.
- (45) Schmidt, C.; Welz, R.; Muller, S. *Nucleic Acids Res.* **2000**, *28*, 886–894.
- (46) O'Connor, D. V.; Phillips, D. *Time-correlated Single Photon Counting*; Academic Press: New York, 1984.
- (47) Livesey, A. K.; Skilling, J. *Acta Crystallogr., Sect. A* **1985**, *41*, 113–122.
- (48) Brochon, J. C. *Methods Enzymol.* **1994**, *240*, 262–311.
- (49) Kühnemuth, R.; Seidel, C. A. M. *Single Mol.* **2001**, *2*, 251–254.
- (50) Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. M. *J. Biotechnol.* **2001**, *86*, 163–180.
- (51) Arden-Jacob, J. Universität-Gesamthochschule, Siegen, 1992.
- (52) Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. *J. Phys. Chem.* **1996**, *100*, 5541–5553.
- (53) Eggeling, C.; Fries, J. R.; Brand, L.; Günther, R.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1556–1561.
- (54) Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2009**, *10*, 1389–1398.
- (55) Levitus, M.; Ranjit, S. Q. *Rev. Biophys.* **2010**, *44*, 123–151.
- (56) Chibisov, A. K.; Zakharova, G. V.; Gerner, H.; Sogulyaev, Y. A.; Mushkalo, I. L.; Tolmachev, A. I. *J. Phys. Chem.* **1995**, *99*, 886–893.
- (57) Widengren, J.; Schwill, P. J. *Phys. Chem. A* **2000**, *104*, 6416–6428.
- (58) Widengren, J.; Schweinberger, E.; Berger, S.; Seidel, C. A. M. *J. Phys. Chem. A* **2001**, *105*, 6851–6866.
- (59) Sauerwein, B.; Murphy, S.; Schuster, G. B. *J. Am. Chem. Soc.* **1992**, *114*, 7920–7922.
- (60) Buschmann, V.; Weston, K. D.; Sauer, M. *Bioconjugate Chem.* **2003**, *14*, 195–204.
- (61) Kalinin, S.; Sisamak, E.; Magennis, S. W.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 6197–6206.
- (62) Garcia de la Torre, J.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719–730.
- (63) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169–184.
- (64) Case, D. A.; et al. *AmberTools, Version 1.3*; University of California: San Francisco, CA, 2008.
- (65) DeLano, W. L. *The PyMOL Molecular Graphics System, Version 1.2r1*; DeLano Scientific: San Carlos, CA, 2002.
- (66) *Hyperchem (TM) Professional, Version 7.51*; Hypercube, Inc.: Gainesville, FL, 2003.
- (67) Antonik, M.; Felekyan, S.; Gaiduk, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2006**, *110*, 6970–6978.
- (68) Kalinin, S.; Felekyan, S.; Valeri, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2008**, *112*, 8361–8374.
- (69) Knox, R. S.; van Amerongen, H. J. *Phys. Chem. B* **2002**, *106*, 5289–5293.
- (70) Majumdar, Z. K.; Hickerson, R.; Noller, H. F.; Clegg, R. M. *J. Mol. Biol.* **2005**, *351*, 1123–1145.
- (71) Haas, E.; Wilchek, M.; Katchalski-Katzir, E.; Steinberg, I. Z. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1807–1811.
- (72) Olofsson, M.; Kalinin, S.; Zdunek, J.; Oliveberg, M.; Johansson, L. B. A. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3130–3140.
- (73) Kulinski, T.; Visser, A.; Okane, D. J.; Lee, J. *Biochemistry* **1987**, *26*, 540–549.
- (74) Borst, J. W.; Laptinok, S. P.; Westphal, A. H.; Kühnemuth, R.; Hornen, H.; Visser, N. V.; Kalinin, S.; Aker, J.; van Hoek, A.; Seidel, C. A. M.; Visser, A. J. W. G. *Biophys. J.* **2008**, *95*, 5399–5411.
- (75) Dolgikh, E.; Roitberg, A. E.; Krause, J. L. *J. Photochem. Photobiol., A* **2007**, *190*, 321–327.
- (76) Dolgikh, E.; Ortiz, W.; Kim, S.; Krueger, B. P.; Krause, J. L.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, *113*, 4639–4646.
- (77) Di Fiori, N.; Meller, A. *Biophys. J.* **2010**, *98*, 2265–2272.
- (78) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 7983–7995.
- (79) Maus, M.; Cotlet, M.; Hofkens, J.; Gensch, T.; De Schryver, F. C.; Schaffer, J.; Seidel, C. A. M. *Anal. Chem.* **2001**, *73*, 2078–2086.
- (80) Gopich, I. V.; Szabo, A. J. *Phys. Chem. B* **2007**, *111*, 12925–12932.
- (81) Isaksson, M.; Norlin, N.; Westlund, P. O.; Johansson, L. B. A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1941–1951.
- (82) Johansson, L. B. A.; Edman, P.; Westlund, P. O. *J. Chem. Phys.* **1996**, *105*, 10896–10904.
- (83) Johansson, L. B. A.; Bergstrom, F.; Edman, P.; Grechishnikova, I. V.; Molotkovsky, J. G. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 1563–1567.

Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity

*Simon Sindbert^{1, #}, Stanislav Kalinin,^{1, #, *} Hien Nguyen², Andrea Kienzler³, Lilia Clima³, Willi Bannwarth³, Bettina Appel², Sabine Müller^{*2}, Claus A. M. Seidel^{*1}*

Supporting Information

1. Supporting Materials and Methods

S1.1 Oligonucleotide synthesis

Deoxyoligonucleotide synthesis was carried out on an ExpediteTM 8909 Nucleic Acid Synthesis system on a 1 μ mol scale using phosphoramidite chemistry according to the general synthesis protocol ¹. Standard phosphoramidites as well as nucleosides coupled to CPG as solid support were obtained from Prologo/Sigma Aldrich. The modified building block **5** (67 μ mol/ml in CH₃CN) was incorporated during the automated synthesis. Cleavage of the synthesized deoxyoligonucleotides from the solid support and deprotection was performed.

Cleavage from solid support and deprotection was performed according to standard protocols ¹ by adding a 25% NH₃-solution. After exchanging the NH₄⁺ by K⁺ the oligonucleotide samples were desalted by size exclusion chromatography using NAP-10 columns. Purification of the modified deoxyoligonucleotides was either performed by preparative PAGE (20%, 40 x 20 x 0.2 cm) or by preparative HPLC. Pre-electrophoresis was performed over night at 400 V with tris-borate buffer. Deoxyoligonucleotide (5 μ L, 1 OD/ μ L in H₂O) and bromophenol-blue/xlenecyanol- solutions (5 μ L) were heated to 90°C for 2 min and rapidly cooled to 0 °C before being loaded onto the gel. Electrophoretic separation was performed for 18 h at 400 V. The gel was visualized at 366 nm and the bands containing the desired DNA were isolated and crunched. They were extracted over night with 2 times their volume of MilliQ water. After centrifugation the supernatant was collected and concentrated. Resulting oligonucleotide samples were desalted by size exclusion chromatography

on NAP-10 columns. Reversed phase HPLC runs were performed on a Merck/Hitachi system; reversed phase: EC-125/4-Nucleosil-100-5-C18 columns, solvent A = 0.1M Et₃NH(OAc) buffer at pH 7.0, B = CH₃CN. All deoxyoligonucleotides were analyzed by PAGE on polyacrylamide gels (20%) of 0.4 mm thickness. Pre-electrophoresis was performed for 2 h at 500 V with tris-borate buffer. Deoxyoligonucleotide (1 μL, 0.1 OD/μL in H₂O) and bromophenol-blue/xylene cyanol-solution (2 μL) were heated to 90 °C for 2 min and rapidly cooled to 0 °C before being loaded on the gel. Electrophoretic separation was performed for 2 h at 500 V and 4 mA. Deoxyoligonucleotide bands were stained with a solution of 3,3'-diethyl-9-methyl-4,5,4',5'-dibenzothiacarbocyanine bromide (Stains-All; Fluka)

Oligoribonucleotides were synthesized by the phosphoramidite method on a Pharmacia Gene Assembler Plus, at 1 μmol scale as described in ². Standard PAC-phosphoramidites as well as CPG-supports were obtained from ChemGenes. 1-(Benzylmercapto)-1*H*-tetrazole (emp Biotech) was used as activator ³. The linker modified nucleoside phosphoramidites were coevaporated three times with dry dichloromethane, kept under vacuum for 2 h to remove traces of solvents, stored in vacuum over P₂O₅ overnight and used in oligo synthesis as 0.1 M solution in acetonitrile. The coupling time for the linker building blocks was 5 min. The obtained RNA was cleaved from the support and deprotected using ammoniacal methanol and TEA x 3HF as described in ⁴, and purified by gel electrophoresis using 10% denaturing polyacrylamide gels. Elution was carried out using 0.5 M LiOAc followed by EtOH precipitation. Oligonucleotides were analyzed by PAGE and MALDI-MS.

S1.2 Labeling of Deoxyoligonucleotides with Cy5 and Alexa 488

For labelling with Cy5, DNA fragments (460.1 μg , 59.24 nmol) were dissolved in water (126.9 μL). To this a solution of Cy5 NHS ester (1 mg) dissolved in 10 μL DMF, DMF (33 μL) was added. The mixture was further diluted with sodium bicarbonate buffer pH 8.5 (21 μL) and water (21 μL). The mixture was vortexed at rt. for 24 h. Then 3M sodium acetate puffer, pH 5.2 (30 μL) and ethanol (825 μL) were added and the mixture was cooled on dry ice for 60 min. After centrifugation for 30 min at 11000 rpm the supernatant was removed and the obtained pellet was washed with 800 μL ice-cold ethanol. The crude DNA was purified by preparative reversed phase HPLC (EC-125/4- Nucleosil-100-5- C18 column, 0-40% B in 30 min, 40-100% B in 4 min, 100% B for 2 min, 100-0% B in 4 min ; t_R = 19.04 min) and desalted on NAP-10 columns. The purity was verified by analytical PAGE. For labelling with Alexa488, DNA fragments (535.0 μg , 55.63 nmol) were dissolved in DMF/dioxane/water 1:1:1 (166.9 μL) and $i\text{Pr}_2\text{EtN}$ (1.9 μL , 11.13 μmol) and Alexa Fluor 488 5-TFP (1 mg) was added and the mixture was vortexed at rt. for 24 h. Afterwards the solvent was removed under reduced pressure and the residue was washed with EtOH (3 x 500 μL). The product was purified by preparative gel electrophoresis (20%, 40 x 20 x 0.2 cm). The purity was verified by analytical PAGE.

S1.3 Labeling of oligoribonucleotides with Alexa 488 and Cy5

5 to 10 nmol amino linker modified RNA were dissolved in 15-25 μ l Borax buffer 0.1 M, at pH 9.4 for Alexa 488 coupling reactions, and at pH 8.7 for Cy5 coupling reactions. The mixture was added to a solution of 100 μ g dye in 5 μ l DMF (the solution of the dye in DMF was freshly prepared just before the coupling reaction). The coupling reaction was carried out at rt. overnight in the dark. Excess of dye was removed via gel filtration (Sephadex G25 fine, GE Healthcare). The labeled RNA was purified by denaturing PAGE. The sample was denatured at 90 °C for 3 min, immediately subjected onto a 15% 200 x 150 x 1.5 mm gel, and electrophoresis was run at 400 V for 5.5 h in the dark. The band corresponding to the labeled RNA was cleaved out, the RNA was eluted with 0.5 M lithium acetate, and precipitated from ethanol at -20 C (yield 15-37% for ALEXA 488, 26-54% for Cy5).

S1.4 Nucleic acid sequences and labeling positions

DNA sequences

DNA1 with different linkers L, I and S_t (10 bp separation, linkers pointing towards each other), labeling positions depicted in green and red for Alexa488 and Cy5, respectively.

5' -d (ACT GAT CGT AAG CTA CTG AAG CGT A) -3'
3' -d (TGA CTA GCA TTC GAT GAC TTC GCA T) -5'

RNA sequences

RNA1 with different linkers L, S_d and S_t (11 bp separation, linkers pointing towards each other), labeling positions depicted in green and red for Alexa488 and Cy5, respectively.

5' -GGC GGU GCC GAC UGC GAG CUU GCC A-3'
3' -CCG CCA CGG CUG ACG CU CAA CGG U-5'

RNA2 and RNA3 all with linkers L, labeling positions depicted in green and red for Alexa488 and Cy5, respectively. The numbers in parenthesis represent the basepair separation between the dyes whereas the + and – signs indicate the effect of the dye displacement towards the 3'-end on the DA distance (Figure 5C inset).

RNA2(16–)

5' -CCG GUG GUU AUA UUA CCU GGU ACG CCU UGA CGU GGG G-3'
3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA2(19–)

5' -CCG GUG GUU AUA UUA CCU GGU ACG CCU UGA CGU GGG G-3'
3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA2(22–)

5' -CCG GUG GUU AUA UUA CCU GGU ACG CCU UGA CGU GGG G-3'
3' -GGC CAC CAA UAU AAU GGA CCA UGC GGA ACU GCA CCC C-5'

RNA3(18+)

5' -CCC CAC GUC AAG GCG UGG UGG CCG AAG GUC GG-3'
3' -GGG GUG CAG UU CCG ACC ACC GGC UUC CAG CC-5'

RNA3(24+)

5' -CCC CAC GUC AAG GCG UGG UGG CCG AAG GUC GG-3'
3' -GGG GUC CAG UUC CGC ACC ACC GGC UUC CAG CC-5'

S1.5 Fluorescence spectroscopy

Time-resolved polarized fluorescence experiments and data analysis.

Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation sources were either a 470 nm diode laser (LDH-P-C 470, Picoquant, Berlin, Germany) operating at 8 MHz for donor excitation or a 635 nm diode laser (LDH-8-1 126, Picoquant, Berlin, Germany) operating at 10 MHz for direct acceptor excitation. The emission wavelength was set to 520 nm for donor emission and to 665 nm for acceptor emission, respectively. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path) resolution. Additional cut-off filters were used to reduce the contribution of the scattered light (>500 nm for donor and >640 nm for acceptor emission, respectively). All measurements were performed at room temperature. The concentrations of DNA/RNA molecules were kept below 1 μ M. The G -factor was calculated using steady state anisotropies of the solutions measured with a Fluorolog-3 (Horiba Jobin Yvon, Munich, Germany). Fluorescence intensity and anisotropy decay curves were fitted using the iterative re-convolution approach ⁵. The maximum number of counts was typically 25,000. The fits approximately range from the maximum of the instrument response functions (IRF) to the first time channel with less than 100 detected photons. The fluorescence intensity decays of FRET-labeled molecules (donor and acceptor emission) were fitted globally with the decays of the molecules only labeled with either the donor (donor only, D-only) or the acceptor (acceptor only, A-only) dye. The fluorescence decays $F(t)$ were modeled by single or double exponential decays (eq 1 and Section 3.1) or by assuming a Gaussian distribution of distances (eq 6 and 7 and see Section S2.6). The anisotropy decays $r(t)$ were formally described by double or triple exponential decays (rotational correlation times ρ_1 , ρ_2 and ρ_3) with free amplitudes (b_1 , b_2 and b_3) (eq S2B and see Section 3.1). Applying appropriate weighting ⁵ the anisotropy decays were recovered by global fitting of the sum and difference curves according to eqs S1,

$$F_{\text{sum}}(t) = F_{\parallel}(t) + 2GF_{\perp}(t) = F(t) \quad (\text{S1A})$$

$$F_{\text{diff}}(t) = F_{\parallel}(t) - GF_{\perp}(t) = F(t)r(t) \quad (\text{S1B})$$

All species irrespective of the fluorescence lifetime were assumed to exhibit the same anisotropy decay.

The fluorescence anisotropy decay $r(t)$ can be described as a product of the separate factors responsible for local dye reorientations and the overall rotation of DNA or RNA, respectively, as ⁶

$$r(t) = [(r_0 - r_{\infty}) \exp(-t / \rho_{\text{local}}) + r_{\infty}] \exp(-t / \rho_{\text{global}}) \quad (\text{S2A})$$

where r_0 is the fundamental anisotropy, and r_∞ is the residual anisotropy. In this study, two exponentials were often needed to describe the local fluorophore dynamics. This leads to the following more formal form of eq S2A with up to three rotational correlation times ρ_i and the corresponding anisotropy amplitudes b_i (eq S2B, also given as eq 2 in the main text)

$$r(t) = b_1 \exp(-t / \rho_1) + b_2 \exp(-t / \rho_2) + b_3 \exp(-t / \rho_3) \quad (\text{S2B})$$

In the case of timescale separation of local and overall motions ($\rho_{\text{overall}} \gg \rho_{\text{local}}$), the times ρ_1 and ρ_2 characterize only local dye reorientations (cf. eqs S2A and S2B).

Single-molecule fluorescence measurements

The fluorescent donor molecules (Alexa 488) are excited by a linearly polarized, active-mode-locked Argon-ion laser (Innova Saber, Coherent, Santa Clara, CA, USA, 496.5 nm, 73.5 MHz, ~ 300 ps) or by a 485 nm diode laser (LDH-D-C 485, Picoquant, Berlin, Germany) operating at 64 MHz. The laser light is focused into the dilute solution (< 50 pM) of labeled molecules by a 60x/1.2 water immersion objective. Each molecule generates a brief burst of fluorescence photons as it traverses the detection volume. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into a wavelength ranges below and above 595 nm by using a dichroic beamsplitter (595 DCXR, AHF, Tübingen, Germany). Additionally, red (HQ 720/150 nm for Cy5) and green (HQ 533/46 nm for Alexa 488 and Rh110) bandpass filters (both made by AHF, Tübingen, Germany) in front of the detectors ensure that only fluorescence photons coming from the acceptor and donor molecules are registered. An estimate of the focal geometry is acquired by determining the diffusion correlation time of 200 ± 13 μs for Rhodamine 110 and knowing its diffusion coefficient of 0.34 ± 0.03 $\mu\text{m}^2/\text{ms}$. Detection is performed using four avalanche photodiodes (SPCM-AQR-14, Laser Components, Germany or alternatively for the green channels PDM050CTC, PicoQuant, Berlin, Germany). The signals from all detectors are guided through a passive delay unit and two routers to two synchronized time-correlated single photon counting boards (SPC 132, Becker and Hickl, Berlin, Germany) connected to a PC. Bursts of fluorescence photons are distinguished from the background of 1-2 kHz by applying certain threshold intensity criteria⁷. Bursts during which bleaching of the acceptor occurs are excluded from further analysis by applying a criterion regarding the difference in macroscopic times, $|T_G - T_R| < 1$ ms, where T_G and T_R are the average macroscopic times in which all photons have been detected in the green and red channels respectively during one burst⁸.

S2. Supporting Results

S2.1 Fluorescence and anisotropy decays of Alexa488 and Cy5 attached to DNA and RNA

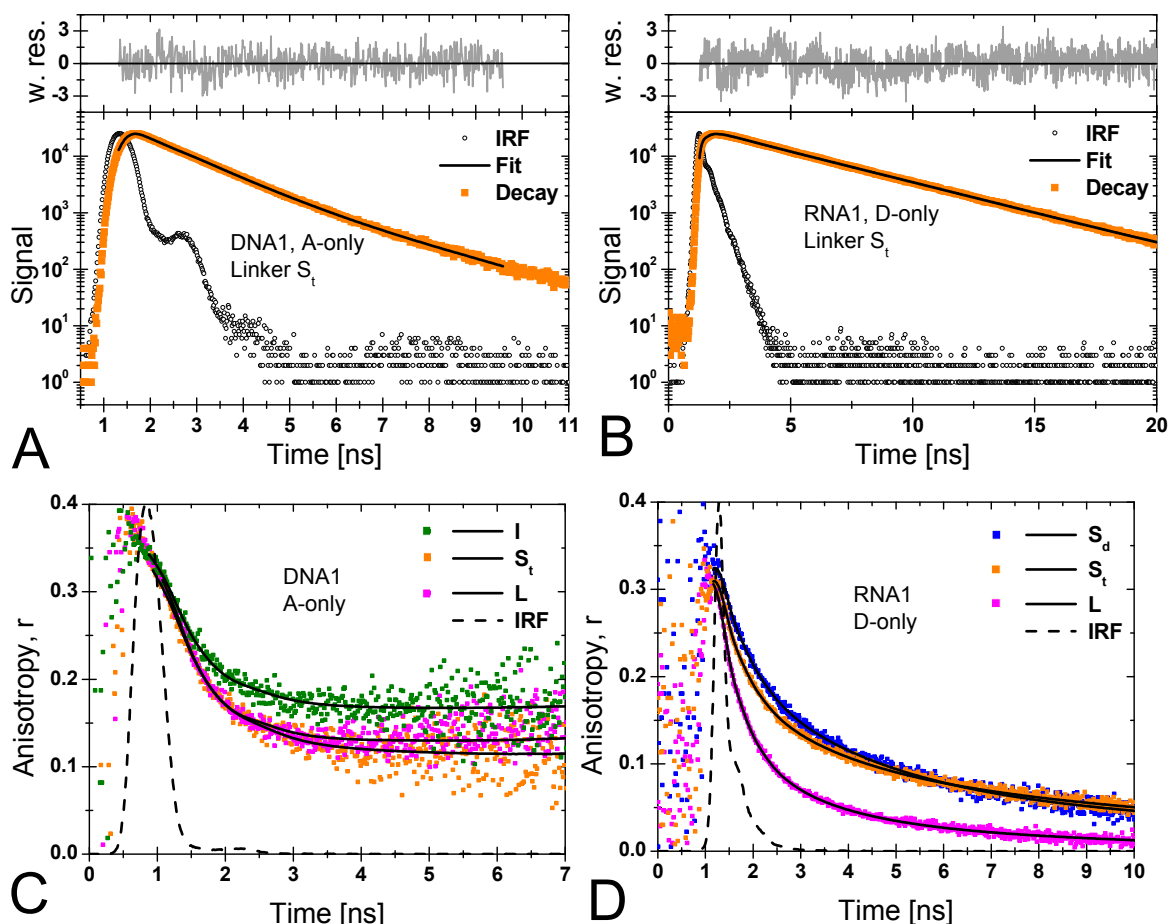


Figure S1. DNA (right side) and RNA (left side) eTCSPC measurements, (A, B) fluorescence lifetime and (C, D) fluorescence anisotropy measurements of single labeled nucleic acids.

(A): fluorescence decay of Cy5 attached to DNA1 via S_t linker, (B) fluorescence decay of Alexa488 attached to RNA1 via S_t linker. (B): Experimental data (orange filled squares), instrument response function (IRF, black open circles) and fits to the data (black solid lines) are shown. Weighted residuals are presented above each plot (gray solid lines).

(C): anisotropy decays of Cy5 attached to DNA1 via I (green), S_t (orange) and L (magenta) linkers.

(D): anisotropy decays of Alexa488 attached to RNA1 via S_d (blue), S_t (orange) and L (magenta) linkers. The anisotropy fits are shown as solid black lines (see S1.5 “Time-resolved polarized fluorescence experiments and data analysis” for more details) and the rescaled IRFs as dashed black lines. The presented fits approximately range from the maximum of the IRFs to the first time channel with less than 100 detected photons.

S2.2 Definitions of differently averaged DA distances

AV simulations provide sets of possible donor and acceptor positions, $\{\mathbf{R}_{D(i)}\}$ and $\{\mathbf{R}_{A(j)}\}$, respectively, with $i = 1 \dots n$ and $j = 1 \dots m$. All allowed positions are assumed to be equally probable. Thus, the distance between the mean dye positions (R_{mp}), the mean DA distance $\langle R_{\text{DA}} \rangle$ and the FRET-averaged distance $\langle R_{\text{DA}} \rangle_E$ can be defined as follows:

$$R_{\text{mp}} = \left| \langle \mathbf{R}_{D(i)} \rangle - \langle \mathbf{R}_{A(j)} \rangle \right| = \left| \frac{1}{n} \sum_{i=1}^n \mathbf{R}_{D(i)} - \frac{1}{m} \sum_{j=1}^m \mathbf{R}_{A(j)} \right| \quad (\text{S3})$$

$$\langle R_{\text{DA}} \rangle = \left\langle \left| \mathbf{R}_{D(i)} - \mathbf{R}_{A(j)} \right| \right\rangle_{i,j} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left| \mathbf{R}_{A(j)} - \mathbf{R}_{D(i)} \right| \quad (\text{S4})$$

$$\langle R_{\text{DA}} \rangle_E = R_0 \left(\langle E \rangle^{-1} - 1 \right)^{1/6} \quad \text{where} \quad \langle E \rangle = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{1}{1 + \left| \mathbf{R}_{A(j)} - \mathbf{R}_{D(i)} \right|^6 / R_0^6} \right) \quad (\text{S5})$$

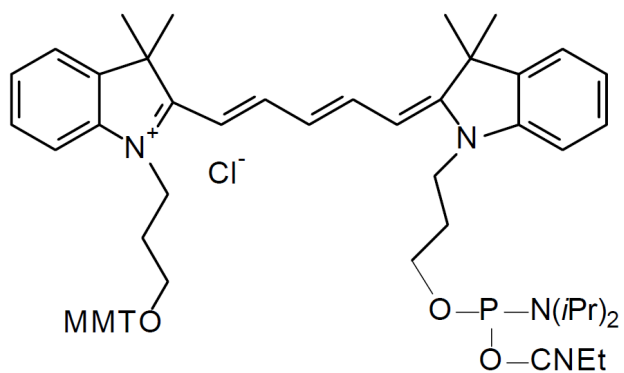


Figure S2. Cy5 phosphoramidite ⁹

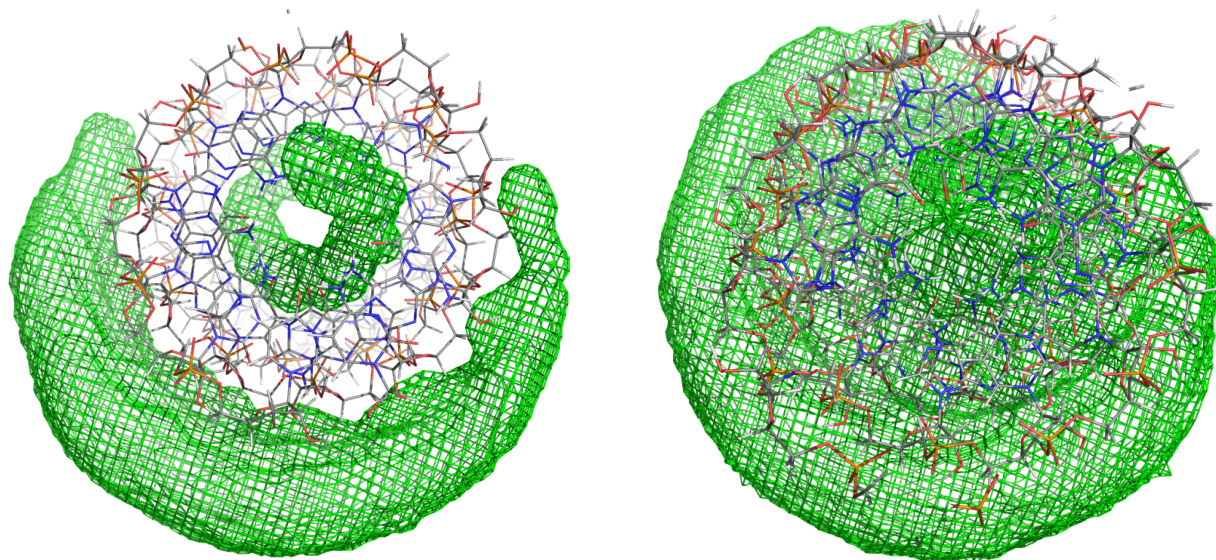


Figure S3. Top view of Alexa488 AVs simulated using (left) $R_{\text{dye}} = 3.5 \text{ \AA}$ (right) superposition of clouds obtained for $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5$. Other parameters are given in the caption of Figure 4 in the main text.

S2.3 R_{mp} to $\langle R_{DA} \rangle_E$ conversion function for L linkers

We generated a series of AVs for D and A dyes attached to A-RNA (Table S2), separated by -7 to 30 dsRNA basepairs. For each pair of AVs we calculated R_{mp} and $\langle R_{DA} \rangle_E$ (eqs S3 and S5).

Corresponding R_{mp} and $\langle R_{DA} \rangle_E$ values are plotted in Figure S2. The solid red line represents a polynomial approximation to the $\langle R_{DA} \rangle_E (R_{mp})$ dependence given by $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$.

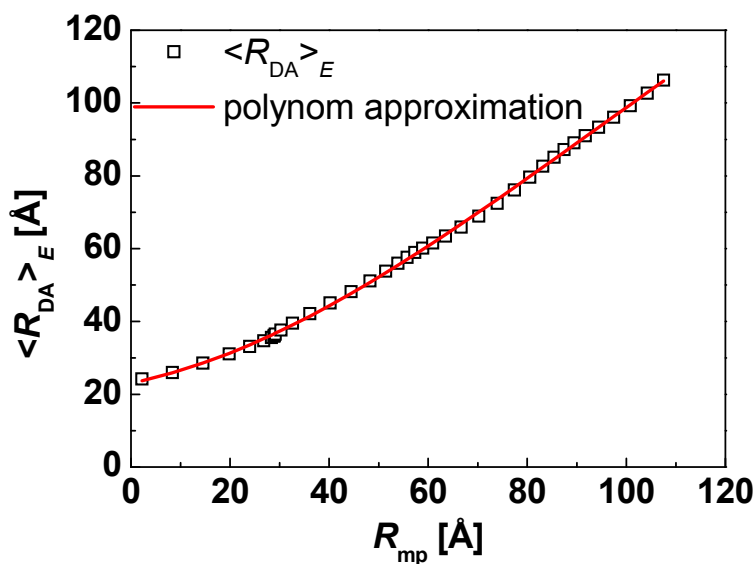


Figure S4. R_{mp} to $\langle R_{DA} \rangle_E$ conversion function and its polynomial approximation $\langle R_{DA} \rangle_E = -2.68 \times 10^{-5} R_{mp}^3 + 7.53 \times 10^{-3} R_{mp}^2 + 0.272 R_{mp} + 23.1$.

S2.4 Photon distribution analysis (PDA)

PDA accurately predicts the shape of single-molecule FRET efficiency histograms in the presence of FRET, explicitly taking into account shot noise and background contributions¹⁰⁻¹². PDA applies to smFRET data measured with two or more detection channels, e.g. “green” (G) and “red” (R), which collect fluorescence of the donor and acceptor, respectively. PDA calculates the probability of observing a certain combination of photon counts $P(S_G, S_R)$

$$P(S_G, S_R) = \sum_{F_G+B_G=S_G; F_R+B_R=S_R} P(F)P(F_G, F_R | F)P(B_G)P(B_R) \quad (S6)$$

The intensity distribution of the fluorescence only contribution to the signal, $P(F)$, is obtained from the total measured signal intensity distribution $P(S)$ by deconvolution assuming that the background signals B_G and B_R obey Poisson distributions, $P(B_G)$ and $P(B_R)$, with known mean intensities $\langle B_G \rangle$ and $\langle B_R \rangle$. $P(F_G, F_R | F)$ represents the conditional probability of observing a particular combination of green and red fluorescence photons, F_G and F_R , provided the total number of registered fluorescence photons is F , and can be expressed as a binomial distribution¹². Subsequently, $P(S_G, S_R)$ may be further manipulated to generate a theoretical histogram of any FRET-related parameter as discussed elsewhere¹¹.

In practice, a fixed DA distance is usually not sufficient to describe FRET species, and a Gaussian distance distribution with a mean of $\langle \tilde{R} \rangle$ and an apparent distribution half width (σ_{app}) has to be used instead. As shown in¹³ $\langle \tilde{R} \rangle$ is slightly biased towards longer distances as compared to $\langle R_{DA} \rangle_E$ (eq S7)

$$\langle \tilde{R} \rangle = \langle R_{DA} \rangle_E \langle \Phi_{FA} \rangle^{1/6} \langle \Phi_{FA}^{-1/6} \rangle \quad (S7)$$

where Φ_{FA} is the acceptor fluorescence quantum yield (see Table S1). In this work, the correction factors $\langle \Phi_{FA} \rangle^{1/6} \langle \Phi_{FA}^{-1/6} \rangle$ are very close to unity (Table S1) and can be disregarded.

In this work, the following parameters have been used for PDA: $\langle B_G \rangle = 1.23$ kHz; $\langle B_R \rangle = 0.49$ kHz; crosstalk: 1.7 %; Donor fluorescence quantum yield: 0.8; Alexa488-Cy5 Förster radius: 52 Å; green/red detection efficiency ratio: 0.78.

Depending on the labeling position, the acceptor fluorescence quantum yield varies considerably and has to be estimated for each sample. It is assumed that only dynamic acceptor quenching takes place and Φ_{FA} is thus proportional to the species-averaged fluorescence lifetime of A, $\langle \tau_A \rangle_x$. As

reference sample we use Cy5-labeled DNA with $\langle \tau_A \rangle_x = 1.17$ ns and $\Phi_{FA} = 0.32$ ¹⁴ to correct for the presence of $\sim 20\%$ of dark cis-state at single-molecule conditions. The obtained acceptor quantum yields and the corresponding correction factors $\langle \Phi_{FA} \rangle^{1/6} \langle \Phi_{FA}^{-1/6} \rangle$ are presented in Table S1.

Table S1A. Fluorescence quantum yields of the acceptor (Cy5) calculated from $\langle \tau_A \rangle_x$.

Samples	$\langle \tau_A \rangle_x$, ns	Φ_{FA}	$\langle \Phi_{FA} \rangle^{1/6} \langle \Phi_{FA}^{-1/6} \rangle$
DNA1 L	1.16	0.32	1.008
DNA1 I	1.29	0.35	1.010
DNA1 S _t	1.17	0.32	1.011
RNA1 L	1.10	0.30	1.004
RNA1 S _d	1.20	0.33	1.005
RNA1 S _t	1.10	0.30	1.004
RNA2 (16+), (19+), (22+)	1.07	0.29	1.007
RNA3(18-)	1.08	0.30	1.006
RNA3(22-)	1.72	0.47	1.008

Table S1B. Comparison of fluorescence quantum yields measured using reference fluorophores with values estimated from the species-averaged lifetimes $\langle \tau \rangle_x$.

Samples	Measured Φ_{FD}	Estimated Φ_{FD}	Measured Φ_{FA}	Estimated Φ_{FA}
DNA1 L	0.92	0.96	0.38	0.39
DNA1 I	1.03	0.96	0.40	0.43
DNA1 S _t	0.98	0.97	0.21	0.39
RNA1 L	0.83	0.86	0.36	0.37
RNA1 S _d	0.70	0.76	0.38	0.40
RNA1 S _t	0.84	0.83	0.36	0.37

S2.5 Benchmark study for a FRET-based A-RNA structural model

Experimental $\langle R_{DA} \rangle_E$ and calculated R_{mp} and $\langle R_{DA} \rangle_E$ values for two internally labeled dsRNAs (RNA2 and RNA3, see section 1.4 for sequences and labeling positions)

Table S2. DA distances in RNA measured by smFRET; data are fitted using PDA.

	SMD experiment (PDA)		AV simulation^b	
Sample	$\langle R_{DA} \rangle_E$, Å	σ_{app} , Å	R_{mp} , Å	$\langle R_{DA} \rangle_E$, Å
RNA2(16–)	41.4	1.5	35.5	41.0
RNA2(19–)	48.8	1.9	46.8	49.6
RNA2(22–)	55.2	2.1	54.7	56.1
RNA3(18+)	60.5	2.9	60.5	61.1
RNA3(24+)	76.4	3.8 ^a	79.6	78.9

^a unstable; fixed at 5% of $\langle R_{DA} \rangle_E$, justified in ¹³.

^b structural features of A-RNA are given in Table S3

Table S3A. Structural and FRET modeling features of B-DNA and A-RNA

Parameter	B-DNA	A-RNA
Helix handedness	Right	Right
bp/turn	10	11
Rise/bp, (Å)	3.4	2.81
C5 displacement, (Å)	0	+/- 0.8
angular shift for Cy5 along xy-plane, correction for AV model	20°	20°

Table S3B. Mean positions of D and A attached via L linkers, expressed in terms of rotations about the DNA or RNA helical axis.

Parameter	D: B-DNA	A: B-DNA	D: A-RNA	A: A-RNA
Distance to the helical axis (Å)	12.0	11.5	9.9	9.7
Rotation angle (deg) ^a	-20	-20	-58	-62

^a positive values correspond to 5' to 3' helix direction

S2.6 Fluorescence decays in the presence of FRET

Donor fluorescence decay

Due to local quenching the fluorescence decay of the donor in the absence of FRET is often multi-exponential (eq S8)

$$F_{D(0)}(t) = \sum_i x_D^{(i)} \exp(-t / \tau_{D(0)}^{(i)}) \quad (S8)$$

To extend eqs 6 and 7 (main text) to this case, it is usually assumed that for given DA distance and orientation, FRET rate (k_{FRET}) is independent of the donor lifetime. This is strictly true if quenching does not change the donor radiative lifetime, which is generally reasonable. eq 6 can be then written as

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} P(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (S9)$$

Sensitized acceptor fluorescence decay

Extensive discussion of FRET-excited acceptor fluorescence relaxation $F_{A(D)}(t)$ can be found, for example, in ¹⁵⁻¹⁷. Briefly, for monoexponential decays of both fluorophores and a single FRET rate $F_{A(D)}(t)$ is given by eq S10.

$$F_{A(D)}(t) \propto -\exp(-t / \tau_{D(0)} - k_{\text{FRET}} t) + \exp(-t / \tau_A) \quad (S10)$$

In general, by taking into account multiexponential decays of D and A and a probability of direct acceptor excitation p_{ex}^A one obtains ¹⁷:

$$F_{A(D)}(t) = (1 - p_{\text{ex}}^A) \sum_{i,j} \left(\frac{x_D^{(i)} x_A^{(j)} k_{\text{FRET}}}{k_{\text{FRET}} + 1 / \tau_{D(0)}^{(i)} - 1 / \tau_A^{(j)}} \left[-\exp(-t / \tau_{D(0)}^{(i)} - k_{\text{FRET}} t) + \exp(-t / \tau_A^{(j)}) \right] \right) + p_{\text{ex}}^A F_A(t) \quad (S11)$$

In eq S11, $F_A(t)$ stands for the fluorescence decay of the directly excited acceptor.

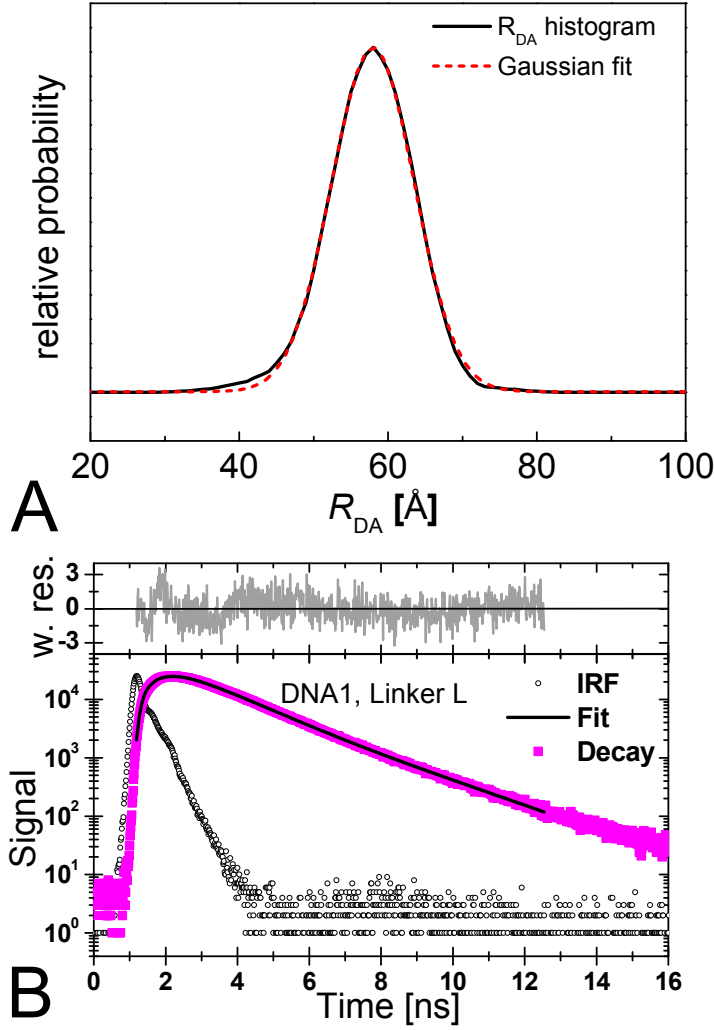


Figure S5. (A) R_{DA} distribution calculated for the RNA2(22–) sample by using the AV approach (black solid line) and a fitted Gaussian distribution (dashed red line). (B) eTCSPC measurement of FRET mediated fluorescence decay of Cy5 obtained for the DNA1 FRET sample with L linker. Experimental data (magenta filled squares), instrument response function (IRF, black open circles) and the fit (black solid line) are shown. Weighted residuals are presented in the upper plot (gray solid line). A Gaussian distribution of distances is assumed (for parameters see Table 5 in the main text). $\chi_r^2 = 1.34$. The fit approximately ranges from the maximum of the IRFs to the first time channel with less than 100 detected photons.

S2.7 Analysis of FRET broadening

Theoretical E - $\tau_{D(A)}$ dependence in the presence of fast linker dynamics

For a single DA distance, a simple E - $\tau_{D(A)}$ relation is expected ⁶.

$$E_{\text{static}} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad (\text{S12})$$

However, because of different averaging of intensity parameters (E , F_D/F_A , or S_G/S_R) and the apparent fluorescence lifetime, deviations from eq S12 are expected for the case of R_{DA} -distributions ¹⁸. In the dynamic averaging, which results in a multi-exponential fluorescence decay of the donor the FRET efficiency is related to the species-average lifetime $\langle \tau \rangle_x$ according to eq S13,

$$E_{\text{static}} = 1 - \frac{\langle \tau \rangle_x}{\tau_{D(0)}} \quad (\text{S13})$$

Clearly for a fluorescence burst having at most few 100 photons, it is impossible to resolve multiple donor lifetimes and determine $\langle \tau \rangle_x$. Thus, the maximum likelihood estimator (MLE) ref ¹⁹ is commonly used to compute a single lifetime for single-molecule fluorescence bursts provides approximately the fluorescence-weighted average lifetime, $\langle \tau \rangle_f$. Because it is generally impossible to describe analytically the relation between the differently averaged lifetimes, we use an empirical dependence with an empirical polynomial with the coefficients c_i to obtain $\langle \tau \rangle_x$

$$\langle \tau \rangle_x = \sum_{i=0}^n c_i (\langle \tau \rangle_f)^i \quad (\text{S14})$$

The set of $\{c_i\}$ used depends on σ_{DA} i.e. on linker length. In this work we used:

$c_0 = -0.0192489$, $c_1 = 0.230201$, $c_2 = 0.47411$, $c_3 = -0.0906313$, $c_4 = 0.00512842$ ($n = 4$) for L linkers, and

$c_0 = -0.0220651$, $c_1 = 0.373858$, $c_2 = 0.413748$, $c_3 = -0.0880843$, $c_4 = 0.00603218$ ($n = 4$) for S_t , S_d and I linkers.

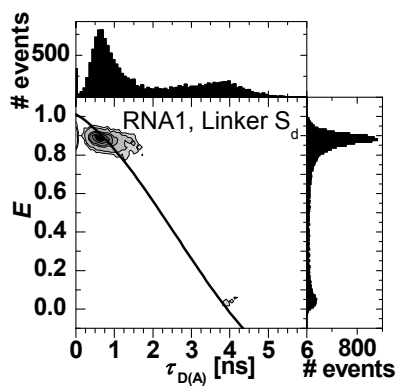


Figure S6. 2D probability histograms of FRET efficiency E versus the donor lifetime $\tau_{D(A)}$ generated from smFRET data obtained for RNA1, S_d linkers. The number of molecules (fluorescent bursts) in each bin is gray scale shaded from white (lowest) to black (highest). The corresponding 1D parameter histograms are given as projections. In all plots, solid lines indicate the E vs. $\tau_{D(A)}$ dependence given by eqs S13 and S14.

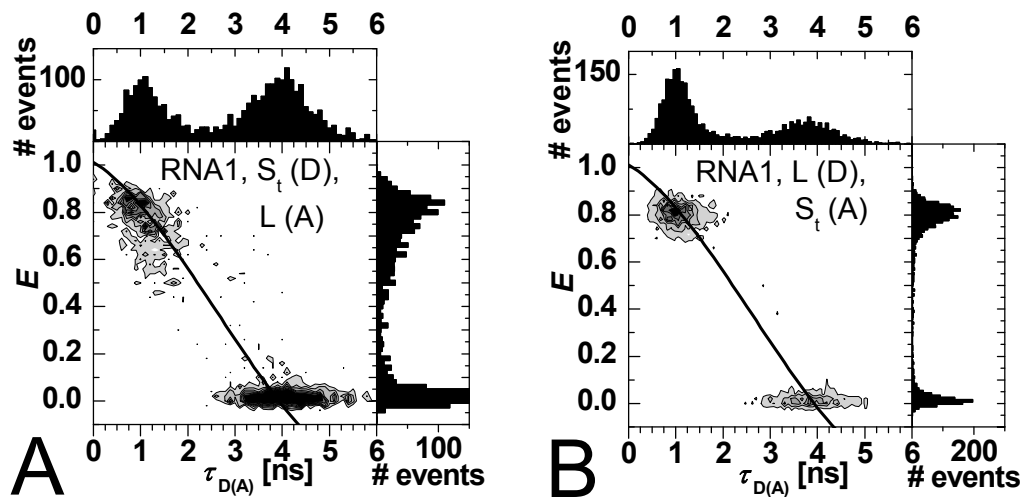


Figure S7. 2D probability histograms of FRET efficiency E versus the donor lifetime $\tau_{D(A)}$ generated from smFRET data obtained for (A) RNA1, S_t linker for the donor dye and L linker for the acceptor dye; (B) RNA1, L linker for the donor dye and S_t linker for the acceptor dye. The number of molecules (fluorescent bursts) in each bin is gray scale shaded from white (lowest) to black (highest). The corresponding 1D parameter histograms are given as projections. In all plots, solid lines indicate the E vs. $\tau_{D(A)}$ dependence given by eqs S13 and S14.

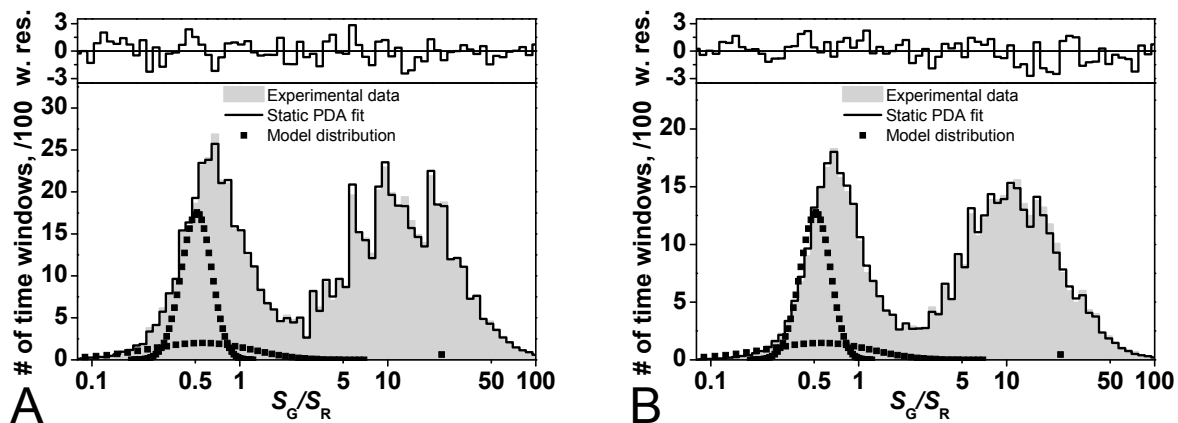


Figure S8. Dynamic PDA of FRET data obtained for the DNA1 sample with dyes attached via S_t linkers. The data are divided into (A) 1 ms and (B) 2 ms time windows. The data are fitted globally using a static model function formally described by: 2 Gaussian distance distributions with $\langle R_{DA} \rangle_1 = 45.1 \text{ \AA}$; $\sigma_{DA1} = 1.8 \text{ \AA}$ (29.2%); $\langle R_{DA} \rangle_2 = 45.7 \text{ \AA}$; $\sigma_{DA2} = 6.5 \text{ \AA}$ (11.8%); Donor-only (57.6%) and an impurity (apparent $R_{DA} = 97.8 \text{ \AA}$; 1.0%). Good fits to the data indicated by $\chi^2_r = 1.33$ and $\chi^2_r = 1.43$ for cases (A) and (B), respectively, suggest that FRET dynamics occurs on much faster or/and slower timescale.

S2.8 κ^2 -calculations

Calculation of order parameters from MD data

Assuming that a set of transition dipole orientations described by corresponding unit vectors $\{\boldsymbol{\mu}\}$ is computed by using MD simulations, the second rank order parameter is easily calculated (eq S15),

$$S^{(2)} = \left\langle \frac{3(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j)^2 - 1}{2} \right\rangle \quad (\text{S15})$$

In eq S15, $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ represent two randomly chosen transition dipole moments, and the average $\langle \rangle$ is calculated over a sufficient number ($>10^6$) of (i, j) pairs.

For the MD data set from ref¹⁴ we obtained $S^{(2)} = 0.013$ for both D and A, which considerably disagrees with the experimental values of -0.23 and 0.58, respectively.

Possible β angles for known δ -angle

The range of β angles (see Figure 10A main text) compatible with a given δ -angle is given by eq S16,

$$|\delta - \beta_1| < \beta_2 < \min(\delta + \beta_1, \pi/2) \quad (\text{S16A})$$

$$0 < \beta_1 < \pi/2 \quad (\text{S16B})$$

For symmetry reasons it is unnecessary to consider the case of $\beta > \pi/2$. A figure schematically showing a range of possible values of β_1 and β_2 is presented in²⁰ (Figure 4).

Definitions of accuracy and precision of distance measurements in the context of the κ^2 -problem

Let us assume a DA pair is characterized by a single “true” DA distance R_{DA} and $\kappa^2 \neq 2/3$. Using a Förster radius calculated for $\kappa^2 = 2/3$ obviously results in an apparent DA distance $R_{\text{DA}}^{(2/3)} = (3/2 \kappa^2)^{-1/6} R_{\text{DA}}$. For a range of possible values of κ^2 one can naturally define the accuracy and precision of distance measurements (eq S17):

$$\text{Accuracy} = \langle (3/2 \kappa^2)^{-1/6} \rangle - 1 \quad (\text{S17A})$$

$$\text{Precision} = [\text{var}((3/2 \kappa^2)^{-1/6})]^{1/2} \quad (\text{S17B})$$

Possible κ^2 -values for unknown $r_{\infty, A(D)}$

Table S4. Results for κ^2 probability distributions taking into account, $r_{\infty, D}$ and $r_{\infty, A}$ (see Table 2, main text).

DNA1						
Linker	κ^2_{\min}	κ^2_{\max}	ΔR_{DA}	κ^2_{mean}	accuracy	precision
I	0.23	1.54	-15 % ... +16 %	0.67	+2.6%	7.6%
S_t	0.22	1.45	-14 % ... +17 %	0.67	+2.3%	6.9%
L	0.4	1.27	-11 % ... +8 %	0.67	+1.5%	5.0%
RNA1						
S_d	0.23	1.59	-16 % ... +16 %	0.67	+2.8%	8.0%
S_t	0.09	1.63	-16 % ... +28 %	0.67	+3.3%	9.3%
L	0.32	1.39	-13 % ... +12 %	0.67	+1.9%	6.0%

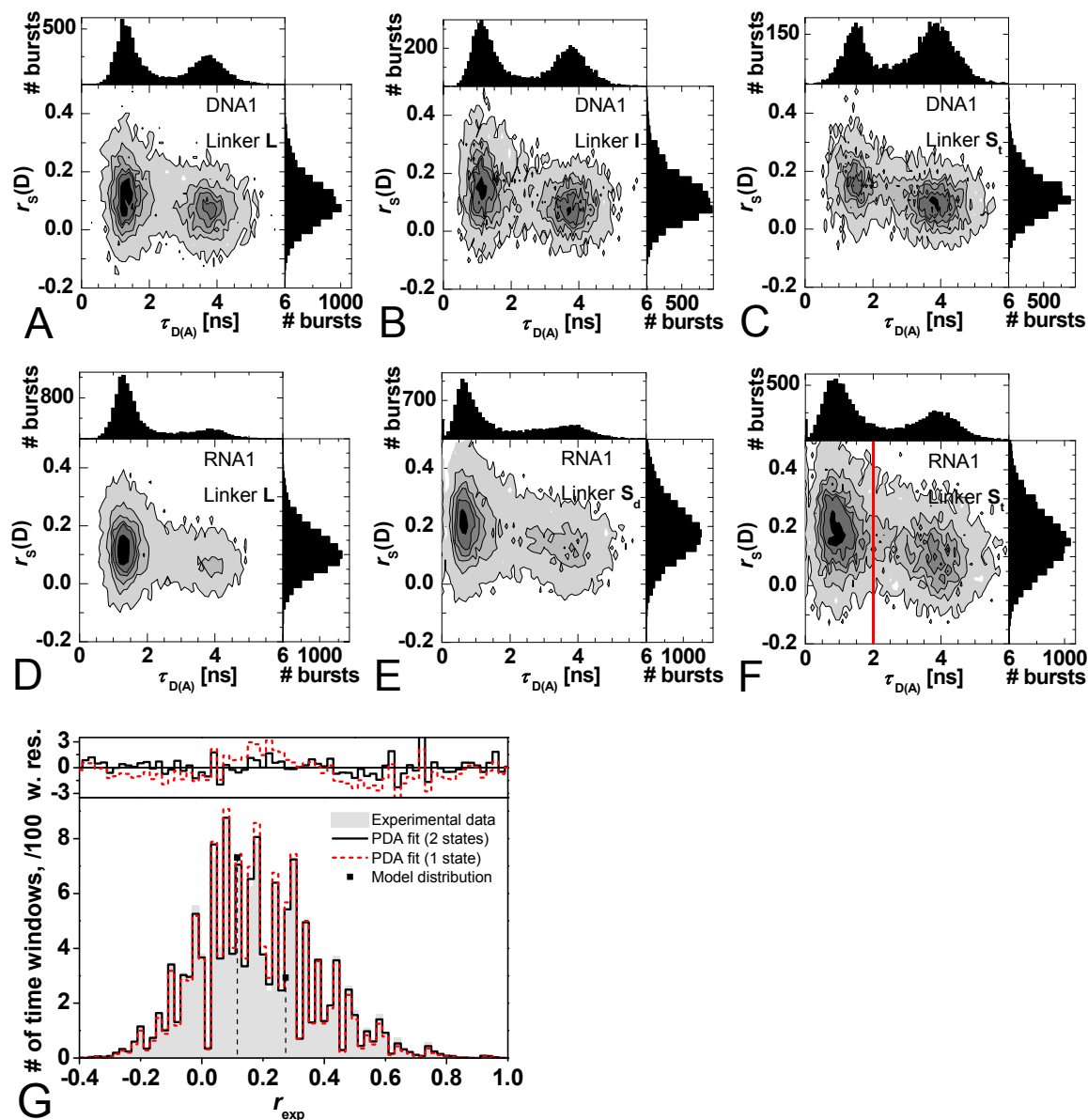


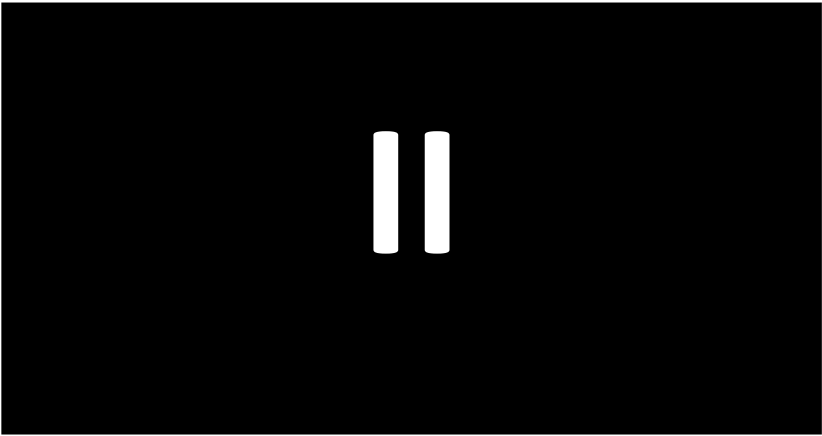
Figure S9. (A-F) 2D histograms of fluorescence steady state donor anisotropy versus the donor lifetime in the presence of FRET. The corresponding 1D parameter histograms are given as projections. (A) DNA1, linker L, (B) DNA1, linker I, (C) DNA1 linker S_t, (D) RNA1 linker L, (E) RNA1 linker S_d, (F) RNA1 linker S_t, and (G) Anisotropy PDA for the FRET subensemble of RNA1 with linker S_t as indicated in Fig. S9F (left of the red line, i.e. $\tau_{D(A)} < 2$ ns). The anisotropy histogram of experimental data (gray area) is fitted with one (red dashed line, $r = 0.157$, $\chi_r^2 = 2.06$) and two states (black solid line, $r_1 = 0.114$ (71.4%) and $r_2 = 0.276$ (28.6%), $\chi_r^2 = 0.95$). Weighted residuals are shown in the upper plot.

References

- (1) Bannwarth, W. *Chimia* **1987**, *41*, 302-317.
- (2) Schmidt, C.; Welz, R.; Muller, S. *Nucleic Acids Res.* **2000**, *28*, 886-894.
- (3) Welz, R.; Muller, S. *Tetrahedron Lett.* **2002**, *43*, 795-797.
- (4) Welz, R.; Bossmann, K.; Klug, C.; Schmidt, C.; Fritz, H. J.; Muller, S. *Angew. Chem. Int. Ed.* **2003**, *42*, 2424-2427.
- (5) O'Connor, D. V.; Phillips, D. *Time-correlated Single Photon Counting*; Academic Press: New York, 1984.
- (6) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Kluwer Academic/ Plenum Publishers: New York, 1999.
- (7) Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. M. *J. Biotechnol.* **2001**, *86*, 163-180.
- (8) Eggeling, C.; Widengren, J.; Brand, L.; Schaffer, J.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. A* **2006**, *110*, 2979-2995.
- (9) Glen Research, www.glenresearch.com/ProductFiles/10-5915.html.
- (10) Kalinin, S.; Felekyan, S.; Valeri, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2008**, *112*, 8361-8374.
- (11) Kalinin, S.; Felekyan, S.; Antonik, M.; Seidel, C. A. M. *J. Phys. Chem. B* **2007**, *111*, 10253-10262.
- (12) Antonik, M.; Felekyan, S.; Gaiduk, A.; Seidel, C. A. M. *J. Phys. Chem. B* **2006**, *110*, 6970-6978.
- (13) Kalinin, S.; Sisamakakis, E.; Magennis, S. W.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 6197-6206.
- (14) Woźniak, A. K.; Schröder, G.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. *Proc.Natl.Acad.Sci.USA.* **2008**, *105*, 18337-18342.
- (15) Kulinski, T.; Visser, A.; Okane, D. J.; Lee, J. *Biochemistry* **1987**, *26*, 540-549.
- (16) Borst, J. W.; Laptinok, S. P.; Westphal, A. H.; Kühnemuth, R.; Hornen, H.; Visser, N. V.; Kalinin, S.; Aker, J.; van Hoek, A.; Seidel, C. A. M.; Visser, A. J. W. G. *Biophys. J.* **2008**, *95*, 5399-5411.
- (17) Olofsson, M.; Kalinin, S.; Zdunek, J.; Oliveberg, M.; Johansson, L. B. A. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3130-3140.
- (18) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 7983-7995.
- (19) Maus, M.; Cotlet, M.; Hofkens, J.; Gensch, T.; De Schryver, F. C.; Schaffer, J.; Seidel, C. A. M. *Anal. Chem.* **2001**, *73*, 2078-2086.
- (20) Ivanov, V.; Li, M.; Mizuuchi, K. *Biophys. J.* **2009**, *97*, 922-929.

Reference with 16 or more authors (main text)

- (61) Case, D.A.; Darden, T. A.; Cheatham, T.E.; III, Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. *AmberTools*, Version 1.3; University of California: San Francisco, 2008.



FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang

Stanislav Kalinin^{1#}, Thomas Peulen^{1,*}, Simon Sindbert^{1,*}, Paul J. Rothwell^{1/2}, Sylvia Berger¹, Tobias Restle^{2/4}, Roger S. Goody², Holger Gohlke³, Claus A. M. Seidel^{1#}

¹Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany, ²Max-Planck-Institut für molekulare Physiologie, Otto-Hahn-Str. 11, 44227 Dortmund, Germany, ³Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany, ⁴UK S-H - Campus Lübeck, Institut für Molekulare Medizin, Ratzeburger Allee 160, 23538 Lübeck, Germany

* contributed equally.

corresponding authors: stanilav.kalinin@uni-duesseldorf.de, cseidel@hhu.de.

Abstract

A comprehensive toolbox for FRET-restrained modeling of biomolecules and their complexes is presented for quantitative applications in structural biology. A dramatic improvement in the precision of FRET-derived structures is achieved through explicitly considering spatial distributions of dye positions, which greatly reduces uncertainties due to flexible dye linkers. The precision and confidence levels of the models are calculated by rigorous error estimation. The accuracy of this approach is demonstrated by docking a DNA/DNA 19/35 primer/template to HIV-1 reverse transcriptase. The derived model agrees with the known X-ray structure with an RMSD of 0.5 Å. Moreover, the formerly unknown configuration of the flexible single strand template overhang was determined by FRET-guided “screening” of a large structural ensemble created by molecular dynamics simulations. We found a preferential structure with the 5'-end of the overhang bound to the fingers domain of RT, which might have important implications concerning

proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis.

Introduction

In recent years, single-molecule (sm) fluorescence spectroscopy¹ has come of age, and sm Förster Resonance energy transfer (FRET), functioning as a “spectroscopic ruler”², is providing important insights into structural heterogeneity and function of biomolecules under *in vitro*^{3,4} and *in vivo* conditions⁵. Currently, we derive our knowledge of biomolecular structure to a large extent from traditional methods such as X-ray crystallography, which determine highly resolved but static models. However, biomolecules are dynamic and undergo intrinsic motions^{6,7}. FRET has the key advantage that it allows for the observation of several biomolecular conformations in solution^{8,9} with high time resolution determined by fluorescence lifetime of the dyes in the order of a few nanoseconds.

One common misconception about FRET is that it only provides low accuracy information for structural modeling. The uncertain fluorophore positions with respect to their attachment points and the orientation dependence of the FRET efficiency (“ κ^2 -problem”) are considered fundamental limitations of FRET. Explicitly modeling the dye behavior^{10,11} by considering the structure of the biomolecule and calculating the distribution of dye positions is the key to increasing the spatial resolution of FRET with flexibly linked dyes. In contrast, the conversion of FRET data into distances between the labeling sites results in an unnecessary loss of accuracy. Several approaches to derive FRET-restrained structures of biomolecules and of their complexes have been published^{8,12-19}. However, many questions have remained unsolved yet. These include the uncertainty of dye positions due to flexible dye linkers, averaging of the FRET efficiency over distributions of donor-acceptor distances, and potential effects of the spatial arrangement of structural units on dye distributions. Another issue is the missing information on the quality of resulting structural models and how it is influenced by uncertainties of “input” FRET data. There is also little evidence as to the accuracy of FRET-restrained 3D modeling, which could be gained by comparison to known

structures. Finally, a productive combination of FRET with state-of-the-art *in silico* modeling approaches for the generation of candidate model structures is needed.

To illustrate the problem of flexibly linked dyes, we consider Alexa488 and Cy5 dyes attached to DNA using standard C6 linkers (Figure 1A). There is ample theoretical^{11,20-24} and experimental²⁵ evidence for the existence of donor-acceptor distance distributions due to flexible dye linkers. In Figure 1A, the green and red surfaces show the space accessible to the donor (D) and acceptor (A) fluorophores, respectively, as determined by a geometric accessible volume (AV) algorithm^{14,26}. It is obvious that (1) the mean dye positions (colored spheres) are far from dye attachments points (black crosses) and that (2) AVs are large, requiring an averaging of FRET observables over a distribution of donor-acceptor distances. Thus, although the relationship between the distance R_{DA} , the Förster radius R_0 , and the FRET efficiency E is well known (equation (1))^{27,28},

$$E = 1 / (1 + R_{DA}^6 / R_0^6) \quad (1),$$

its value is very limited if R_{DA} is (mis)interpreted as a mean position distance $R_{mp} = \left| \langle \vec{R}_D \rangle - \langle \vec{R}_A \rangle \right|$ between the position vectors \vec{R} or, even worse, as a distance between the dyes' attachment points. In fact, due to different averaging of E and R_{DA} , $\langle E \rangle$ does not correspond to R_{mp} ^{10,25}. A distance formally calculated using equation (1) will be referred to as donor-acceptor FRET-averaged distance, $\langle R_{DA} \rangle_E$. Figure 1B reveals large differences between R_{mp} and $\langle R_{DA} \rangle_E$ of up to 10 Å (~30% for $\langle R_{DA} \rangle_E = 35$ Å) as calculated for the system shown in Figure 1A. This confirms that considering distributions of DA distances is essential for quantitative FRET.

However, assuming that the spatial distributions of the donor and acceptor fluorophores (and, thus, the DA distance distributions) are accurately predicted, the expected FRET efficiency can be calculated with very high precision. Realistic modeling of the dye behavior is critical for this task. Ideally this is done by means of MD simulations^{11,16,20}. However, in many cases the AV approach, which requires a few seconds of computational time, appears to be a reasonable approximation²⁵ that we use here.

The assumptions of the AV algorithm can be experimentally verified by analyzing complementary fluorescence parameters (e.g. the fluorescence lifetime distribution and anisotropy to rule out immobile dyes) and/or by calibrating a FRET pair using molecules

of known structure such as dsDNA. Moreover, systematic errors related to the AV approach are likely to average out if many DA distances are measured. In this way uncertainties due to long dye linkers are largely eliminated. Ångström precision can be achieved when positioning labeled macromolecules, even if a single FRET measurement does not provide the distance between two labeling sites with high accuracy.

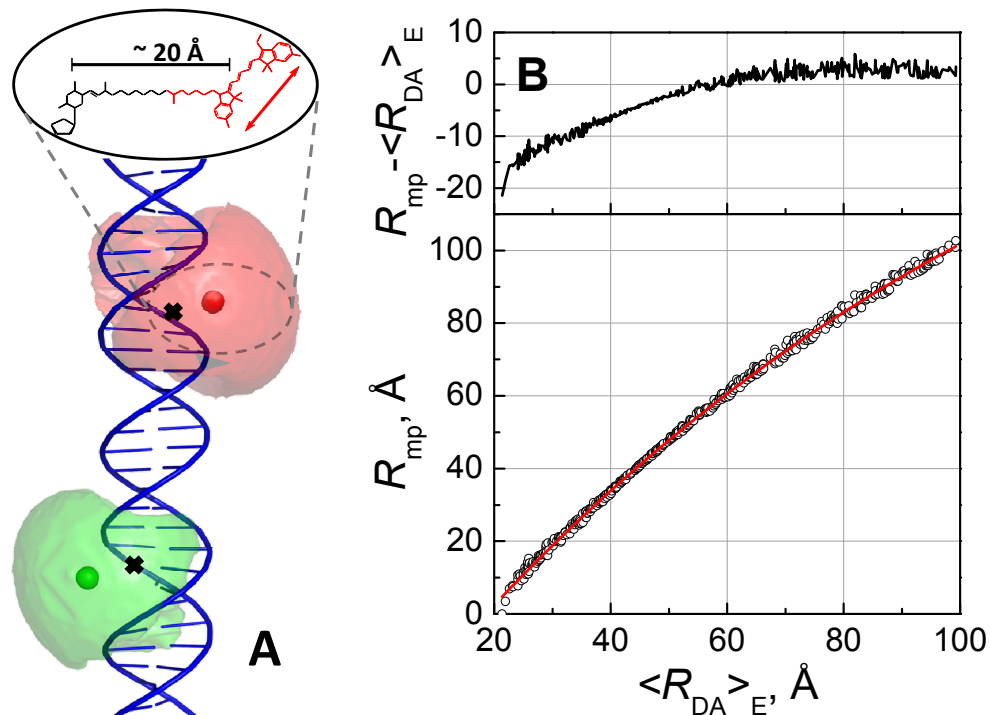


Figure 1. Effect of dye position distributions on FRET. (A) Accessible volumes of Alexa488 (green) and Cy5 (red) attached to a dsRNA via a C6 linker (ellipse). The mean positions of the dyes are depicted as spheres (green for Alexa488, red for Cy5) and the attachment atoms of the linkers as black crosses. The distance between the dyes' attachment points (C5 atoms) is 43.6 Å, whereas the distance between the dyes' mean positions (R_{mp}) is 52.6 Å. (B) Bottom: An $\langle R_{DA} \rangle_E$ to R_{mp} conversion function (red line) generated by fitting a 3rd order polynomial to $\langle R_{DA} \rangle_E/R_{mp}$ value pairs (open circles) calculated for a set of randomly oriented AVs of Alexa488 and Cy5 ($R_0 = 52$ Å) for dsRNA. The rms deviation between the data and the polynomial approximation is 0.9 Å over the whole range of $\langle R_{DA} \rangle_E$ and 0.6 Å for $30 \text{ Å} < \langle R_{DA} \rangle_E < 70 \text{ Å}$. Top: the difference between R_{mp} and $\langle R_{DA} \rangle_E$ can reach 10 Å in the range accessible to FRET.

For utilizing the advantage of single-molecule fluorescence detection we established a toolbox for FRET-restrained high-precision structural modeling of biomolecules, which also considers their mobility and structural heterogeneity.

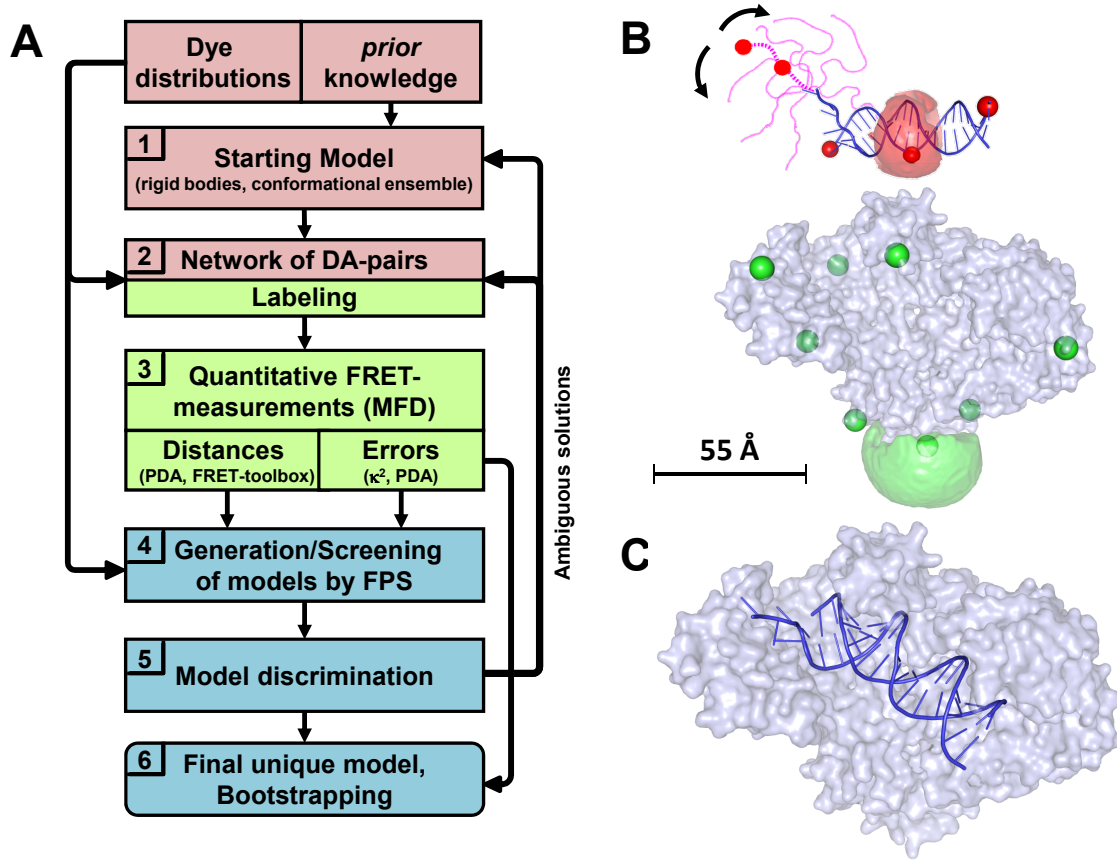


Figure 2. FRET-restrained positioning and screening. (A) Flow diagram of FRET restrained high-precision structural modeling. It comprises three main parts: design of the experiment (shaded in red), measurement and analysis of a set of samples (green), and generation and validation of structural models (blue); (B) crystal structures of the protein and the (ds)DNA (both from PDB-ID 1R0A). The donor labeling positions (green spheres) are named according to the RT subunit (p51 or p66) and the position of the introduced cysteine mutation to which they were coupled: p51Q6C, p51K173C, p51E194C, p51K281C, p66Q6C, p66T27C, p66E194C, and p66K287C. Five acceptor labeling positions on the primer/template DNA (red spheres) are named according to strand (dp or dt) and the position of the labeled nucleotide. The last paired nucleotides are referred to as position 1. For labeled positions on the template overhang the position is negative with regard to position 1, and for the primer coupled fluorophores the number is positive. Thus, the labeled dp/dt complexes are named dp(1)/dt, dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15) (Supplementary Section S1.2). As an example, FRET-pairs are named RT(p51E194C):dp/dt(-15). AV clouds are shown for labeling position p51K173C (green) and dp(10) (red). Parameters used for generation of AV clouds: for donor positions: $L_{\text{link}} = 20 \text{ Å}$, $w_{\text{link}} = 4.5 \text{ Å}$, $R_{\text{dye}} = 3.5 \text{ Å}$, for acceptor positions: $L_{\text{link}} = 23 \text{ Å}$ for dp(10), dp(19), dt(-6) and dt(-15), $L_{\text{link}} = 8.5 \text{ Å}$ for dp(1), $w_{\text{link}} = 4.5 \text{ Å}$, $R_{\text{dye}(1)} = 11 \text{ Å}$, $R_{\text{dye}(2)} = 3 \text{ Å}$, $R_{\text{dye}(1)} = 1.5 \text{ Å}$; (C) Structural model of the RT:dp/dt complex obtained by rigid body docking.

The workflow of our approach is depicted in Figure 2A. Overall, six steps are needed to generate a FRET-restrained structural model: (1) A starting model is generated utilizing prior knowledge from known structures, homology modeling or *ab initio* modeling. (2) Taking the positional distributions of the coupled dyes into account, we use the start model to design a network of dye positions most useful for FRET positioning and screening (FPS, step 4). (3) FRET is quantitatively measured by single-molecule

multiparameter fluorescence detection (smMFD). We perform a rigorous data analysis and error estimation of FRET-derived donor-acceptor distances by analyzing photon distributions and time-resolved anisotropies of the dyes. (4) Possible structural models are searched for and evaluated with respect to their agreement with the FRET data by FPS. Here, two complementary approaches were used to generate structural models: possible arrangements were discovered by repeated rigid-body docking of known substructures or, alternatively, a model-free search was performed by screening suitable models in a large structure library. (5) The possible models are ranked according to their violation of FRET restraints and are assigned to clusters of related structure organization to judge the uniqueness of the structure models. (6) In the final step, the precision (RMSD) of the structure models is determined by bootstrapping.

For validation of this approach, we studied the human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT). HIV-1 RT is a heterodimer composed of two subunits: a 66-kD chain (p66) and a 51-kD chain (p51). RT is responsible for transcription of viral RNA into double stranded DNA^{29,30}. We performed a smFRET study of RT complexed with a 19/35 DNA/DNA primer/template (dp/dt) (Figure 2B and 2C) and characterized in detail the rigid double strand (ds) and flexible single strand (ss) DNA parts of the complex. Several crystal structures were determined for the productive complex in the open educt state (P-E, e.g., refs. ³¹⁻³⁴), which allows us to determine the accuracy of FPS for dsDNA (Figure 2C). So far, the ssDNA template overhang was not resolved by X-ray crystallography. However, there is biochemical evidence that the properly bound template overhang plays an important role in translocation of nucleic acid during processive DNA synthesis ²⁹, e.g., by helping to resolve secondary structures within the substrate, and the overall dp/dt binding affinity to RT increases with the overhang length by a factor of seven³⁵. Moreover, interactions of the ss region of the template with the enzyme were shown to affect the geometry of the ds region within the active site ³⁶. This is thought to affect fidelity of DNA synthesis and could be responsible for altered sensitivity towards certain nucleoside RT inhibitors. Combining FPS and molecular dynamics (MD) simulations we found a well-defined configurational space of the overhang preferentially interacting with one region of the protein.

Results

In the following we describe FRET-restrained high-precision structural modeling (Figure 2A) for the RT:dp/dt complex.

Step 1: Starting models. The complex partners and the labeling positions (colored spheres) are illustrated in Figure 2B. As prior knowledge we used the crystal structure with a 2.9 Å resolution from ³⁴ (PDB-ID: 1R0A), where the RT:dp/pt complex is in the educt state (P-E) (i.e., in the state immediately before incorporation of the next nucleotide³⁷) with RT in an open conformation.

To test the accuracy of FRET-restrained modeling of the HIV-1 RT dp/dt interaction (see step 4 below), we separated the dp/dt from the protein and applied our FPS procedure. For determining the conformation of the template overhang missing in the crystal structure, we generated a starting model for the MD simulations by attaching the single strand to the crystalized DNA such that it sticks out straight from the protein.

Step 2: Network of DA pairs. Eight donor labeling positions were selected on the enzyme and five acceptor labeling positions were chosen on the primer/template DNA (Figure 2B). Overall, 36 independent single-molecule FRET measurements were planned for the RT:dp/dt complex.

Step 3.1: Quantitative smFRET measurements by MFD. smMFD experiments avoid ensemble averaging by analyzing single-molecule events. The distance information is usually deduced from the FRET efficiency (E)^{27,28}, which can be calculated either from the donor and acceptor fluorescence F_D and F_A or from donor fluorescence lifetimes $\tau_{D(A)}$ and $\tau_{D(0)}$ with and without acceptor, respectively³⁸.

$$E = \frac{F_A}{(\gamma' F_D + F_A)} = 1 - \frac{\tau_{D(A)}}{\tau_D} \quad \text{with } \gamma' = \Phi_{FA}/\Phi_{FD(0)} \quad (2)$$

In equation (2), the correction factor γ' accounts for fluorescence quantum yields Φ_F of donor and acceptor. In MFD, all fluorescence parameters are acquired simultaneously³⁸, which enables a multi-dimensional analysis. The correlated FRET analysis by equation (2) helps avoiding most of the pitfalls of ensemble FRET measurements, such as incomplete labeling, fluorophore quenching, and the inability to resolve multiple FRET states^{8,37,38}.

In Figure 3A, 2D burst frequency histograms of the F_D/F_A signal ratio and donor anisotropy r_D versus $\tau_{D(A)}$ are presented for the complex RT(p51E194C):dp(1)/dt. In agreement with ref. ³⁷ three complex types are found: Dead-end (D-E, olive), productive complex in the product state (P-P, orange), and productive complex in the educt state (P-E, red). The observed populations follow the theoretically expected dependencies (blue lines) between $\tau_{D(A)}$ and F_D/F_A or r_D , respectively. This indicates that (1) no significant dye quenching takes place, which could result in errors in the recovered R_{DA} , and (2) no long-lived immobile dye population is observed. 2D FRET analysis of other data sets can be found in the Supplementary Section S4. Here we calculate E from intensities (equation (2)) because F_D and F_A obey well-defined statistics³⁹⁻⁴¹, while the lifetime information ensures that the observed effects are due to FRET. Using both 2-D analysis and photon distribution analysis (PDA, see step 3.2, analysis in the 1D F_D/F_A histogram in Figure 3A)^{39,42} gives unsurpassed sensitivity for characterization of FRET populations derived from single-molecule FRET experiments.

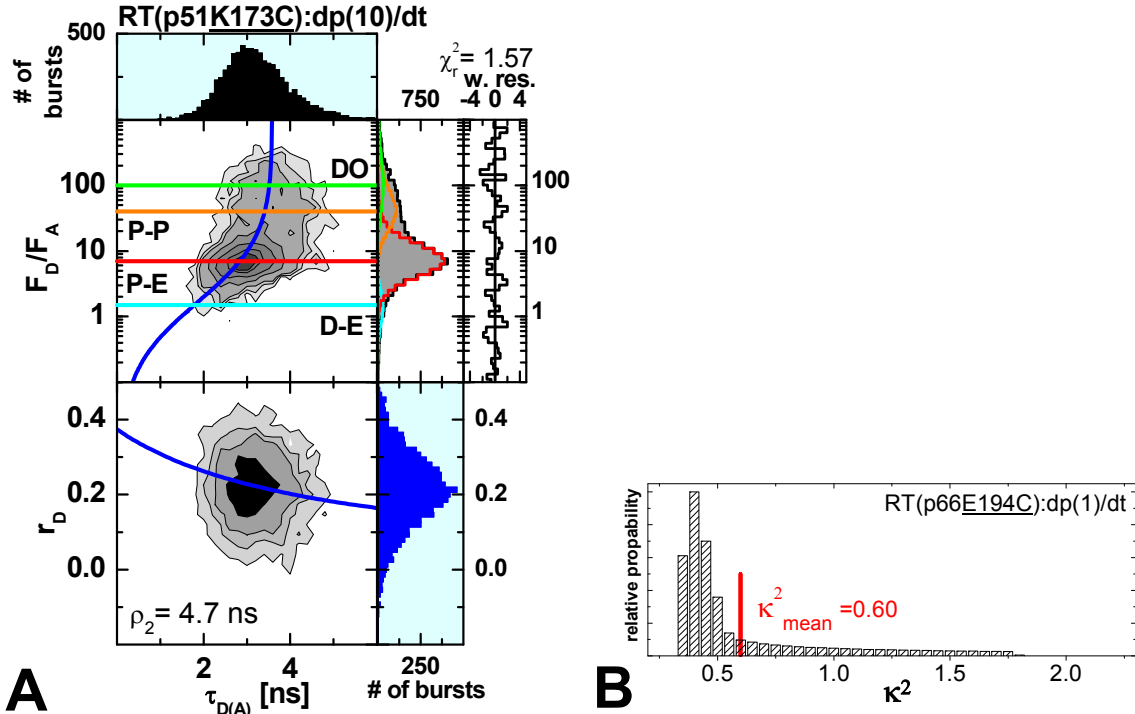


Figure 3. Distance measurements and error estimation. (A) 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for sample RT(p51K173C):dp(10)/dt. The number of molecules (fluorescence bursts) in each bin is gray scale colored from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs. $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; blue) is overlaid (see Supplementary Section S3.3.2 for details). The solid line in the r_D - $\tau_{D(A)}$ diagram are given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $\rho = 3.5$ ns (blue), $r_0 = 0.374$. The 1D F_D/F_A histogram is fitted by PDA: experimental data (grey area) is fitted (black dotted line) using the following parameters: 55% of FRET state 1 (red solid line): $\langle R_{DA} \rangle_{E(1)} = 62$ Å, 23% of FRET state 2 (orange solid line): $\langle R_{DA} \rangle_{E(2)} = 88$ Å, and 4% of FRET-state 3 (cyan solid line): $\langle R_{DA} \rangle_{E(3)} = 47$ Å, σ_{app} for all FRET states: 9 %; 18 % of D-only (light green solid line). The fit yields $\chi_r^2 = 1.57$ (weighted residuals are plotted to the right of the 1D F_D/F_A histogram). The FRET-states 1, 2, and 3 correspond to the previously described protein in educt state P-E (red), protein in product state (P-P, orange), and the dead-end complex (D-E, cyan), respectively. (B) The derived κ^2 values result in an error of 10.0% in $\langle R_{DA} \rangle_E$ (see Supplementary Section S3.3.4).

Step 3.2: Input data for FPS-distances and uncertainties. PDA and similar techniques^{11,41,43,44} simultaneously provide mean distances $\langle R_{DA} \rangle_E$ and uncertainties (ΔR_{DA}) by explicitly taking photon statistics into account. PDA needs fewer free parameters compared to the traditional approach of fitting multiple Gaussian peaks. It provides meaningful fit quality parameters that allow one to justify the chosen fit model. To estimate the errors of fitted parameters, $\Delta R_{DA}(E)$, we explore the parameter space for sets of variables providing acceptable fits (Supplementary Section S3.3.3). Error estimation is usually very difficult if E -distributions are fitted by Gaussians using

general-purpose fitting software, because the proper model function and the standard deviations of data points are unavailable.

Uncertainties in the mutual orientation of donor and acceptor (κ^2 errors, $\Delta R_{\text{DA}}(\kappa^2)$) can be estimated by analyzing anisotropy decays⁴⁵ accessible in MFD (Supplementary Section S3.3.4). Figure 3B shows a typical distribution of possible κ^2 values compatible with experimental anisotropy decays. Due to a weak dependence of R_0 on κ^2 ^{28,46} this broad distribution results in only ~10 % uncertainty in the distance. More advanced κ^2 estimation procedures^{47,48} can be conveniently incorporated into our set of tools.

The overall errors ΔR_{DA} are determined following error propagation rules (equation (3))

$$\Delta R_{\text{DA}}^2 = \Delta R_{\text{DA}}^2(E) + \Delta R_{\text{DA}}^2(\kappa^2) + \dots \quad (3)$$

This procedure was applied for 20 distances in the previously crystallized part of the complex (Supplementary Table S2) for the dsDNA and for 16 distances in the template overhang (Supplementary Table S3).

For generating or judging FRET-based structure models, the knowledge of $\langle E \rangle$ and, thus, $\langle R_{\text{DA}} \rangle_E$ is sufficient. These parameters can be calculated for any putative structure using an AV or MD model and are directly comparable with experimental data. In practice, however, it is useful to convert $\langle R_{\text{DA}} \rangle_E$ into R_{mp} (see Figure 1B bottom) because this avoids repeated calculations (or transformations) of AVs during iterative structure optimization (see Supplementary Section S3.4.1 for further details).

To calculate $\langle R_{\text{DA}} \rangle_E$ from the AV model we assumed static averaging of distances and dynamic reorientation (not necessarily resulting in $\langle \kappa^2 \rangle = 2/3$) on the timescale of FRET. We have previously shown^{10,25} that this approximation works well for dyes attached to DNA and RNA via long methylene linkers. Although the reorientation timescale is not as fast for D dyes attached to RT (Figure 3A lower plot), fast rotations of A bound to DNA justify the assumption of dynamic κ^2 -averaging.

Step 4: Rigid body docking of dsDNA with FRET restraints. To find the position and orientation of dsDNA with respect to RT in best agreement with FRET data, the weighted data-model deviation (χ_E^2) has to be minimized for the set of n distances,

$$\chi_E^2 = \sum_{i=1}^n \frac{(R_{\text{DA}(i)} - R_{\text{model}(i)})^2}{\Delta R_{\text{DA}(i)}^2} \quad (4)$$

This optimization problem can be defined for $\langle R_{\text{DA}} \rangle_E$ values or (more conveniently) for converted mean position distances R_{mp} . To solve this problem, we assumed the complex partners to be rigid bodies and applied a rigid body dynamics approach to dock the partners using FRET restraints. Although a large number of more sophisticated approaches exist (e.g. refs. ^{49,50}), to our knowledge none of these methods allows for the explicit modeling of fluorophores and averaging of measured quantities over distributions of DA distances. Here, we estimate the coordinates of mean dye positions by AV simulations and then fix them with respect to the labeled substructure. If a distance is measured between certain D and A dyes, this is accounted for by adding a “spring” connecting the dyes’ mean positions, having an equilibrium length of R_{mp} and a strength derived from the corresponding ΔR_{DA} (see Supplementary Section S3.4.2 for implementation details). Relaxing this system is equivalent to minimizing the χ_E^2 parameter given by equation (4). Two steps are distinguished in the rigid body dynamics approach:

“Search”. In the first step we generate a large number of complex starting from random configurations of the binding partners excluding those which clash. To prevent clashes between RT and DNA we introduce strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way the positioning is guided by an overall quality parameter (reduced chi-squared parameter, χ_r^2 see methods).

“Refinement”. In the second step, AVs are recalculated accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure the dyes are *not* attached to. The resulting mean dye positions are used to re-optimize the structure.

Step 5: Analysis of docking/screening results. The obtained structures are presented in a cluster plot together with corresponding best structure models (Figure 4). Clusters represent groups of solutions with similar χ_r^2 values and low RMSD within the group (Supplementary Section S3.5). In Figure 4, the symbol sizes reflect the group size. Blue lines show the thresholds with respect to the best possible fit with $\chi_{r,\min}^2$ (see Methods). After the coarse "search" step (Figure 4A; 6 Å clash tolerance) three groups of solutions are found below the threshold, with a few Å RMSD (all dsDNA atoms) between these solutions. Compared to the X-ray structure, the RMSD values for these groups are 6.3 Å, 5.0 Å, and 9.7 Å, respectively. After one "refinement" iteration, when AVs are recalculated and the clash tolerance is reduced to 2 Å, only one solution remains below the threshold (1.3 Å RMSD compared to the X-ray structure; Figure 4B). There are also a few very similar solutions with 5 – 6 Å RMSD versus the best one that could be possibly considered, but all solutions with RMSD > 10 Å already have unreasonably high χ_r^2 -values. After the second round of refinement with a 1 Å clash tolerance (Figure 4C) the separation between the best structure and other solutions is further increased. The final best structure shows an RMSD value of only 0.5 Å with respect to the X-ray structure (Figure 4C). The agreement is remarkably good considering the dye linker lengths of ~20 Å. Even if we require higher statistical significance level of >99% (dashed cyan line in Figure 4C) and accept the solutions with RMSD of ~4 to ~8 Å (cyan), the deviation from the X-ray structure is still much smaller than the sum of dye linker lengths, which justifies the term "high-precision FRET". Compared to typical κ^2 uncertainties (~9 %), the resulting errors are also small, which is most likely due to averaging towards $\langle \kappa^2 \rangle = 2/3$ when multiple DA distances are considered; the same applies to possible systematic errors of the AV approach. Accounting for clashes has a clear effect on the selectivity, but even when using unrealistically mild restrictions ("search" step, Figure 4A) FRET models already agree well with the known structure.

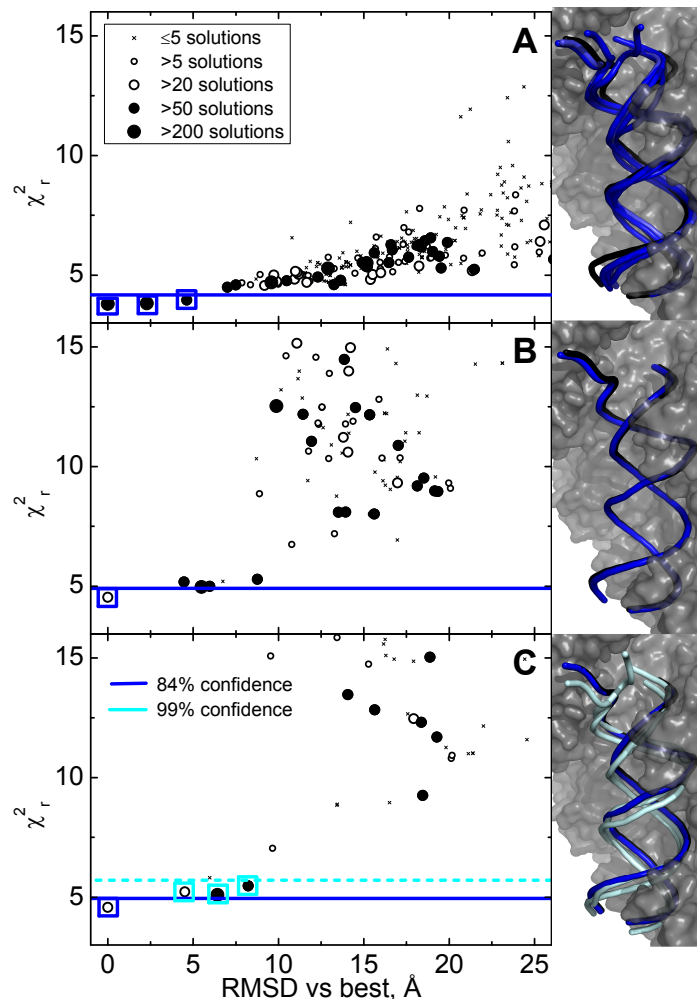


Figure 4. Cluster analysis of docking results. Rigid body docking of RT and dp/dt (both from crystal structure PDB-ID 1R0A). Left: cluster analysis of obtained solutions is shown for (A) initial search phase with 6 Å clash tolerance, (B) first refinement iteration with 2 Å clash tolerance, and (C) second refinement iteration with 1 Å clash tolerance. The symbol size reflects the number of solutions in the shown cluster. Horizontal lines represent confidence thresholds as defined by $\chi_r^2 < \chi_{r,\min}^2 + 0.378$ (blue solid line, ~84% confidence, $\sim 1.4 \sigma$) and $\chi_r^2 < \chi_{r,\min}^2 + 1.09$ (cyan dashed line in panel C, ~99% confidence, $\sim 2.6\sigma$). The thresholds are derived from the chi-squared distribution with 14 degrees of freedom ($\chi^2 < 19.3$ and $\chi^2 < 29.2$, respectively). Right: DNA from X-ray structure (black) overlaid with FPS solutions (blue, cyan) indicated in cluster plots by colored squares.

Step 6: Estimation of precision and accuracy (see also Supplementary Section S3.5).

The uncertainties of FRET-restrained modeling include possible alternative solutions

(local minima of χ_r^2) with similar data-model deviations and uncertainties of the best solution itself. Assuming that a unique solution had been found at this point, we applied a procedure similar to bootstrapping⁵¹ to estimate the precision for the best structure.

The model distance set calculated for the optimized structure (red structure in Figure 4C) was perturbed by simultaneously adding normally distributed random numbers with a zero mean and a standard deviation given by the experimental errors $\{\Delta R_{DA}\}$ to all model distances. This procedure would be impossible without meaningful “input” errors $\{\Delta R_{DA}\}$ from step 3.2. The structure was then re-optimized with the perturbed distance restraints. The procedure was repeated 100 times yielding 100 perturbed structures of the ds dp/dt part, where the phosphate atoms are shown as colored spheres (Figure 5A).

This set of structures represents a distribution of possible positions of all backbone atoms consistent with experimental data. For each phosphate atom the uncertainty was calculated (Figure 5B) using equation (5):

$$\text{RMSD} = \left\langle \left| \vec{R}_{\text{best model}} - \vec{R}_{\text{perturbed model}} \right|^2 \right\rangle^{1/2} \quad (5)$$

For the best FRET model, we estimated an average precision of 2.4 Å RMSD for all P atoms of the dsDNA (Figure 5B). The X-ray structure (0.5 Å RMSD to the FRET structure over all dsDNA atoms) is well within these uncertainty limits. The solutions obtained at earlier stages of rigid body docking are clearly less precise and less accurate showing that the refinement steps are essential.

Alternatively, cross-validation or similar tests⁵² can be performed (Supplementary Section S3.5.2). However, in contrast to X-ray or NMR data, the redundancy of FRET data is usually low. In practice one often chooses new labeling positions with the objective of distinguishing between ambiguous solutions (Figure 2A). Thus, discarding a few FRET restraints might make the position of a unit completely undefined. For this reason, bootstrapping is preferred over procedures where some data points are completely discarded. For RT, both ways of error estimation work well (Supplementary Section S3.5.2).

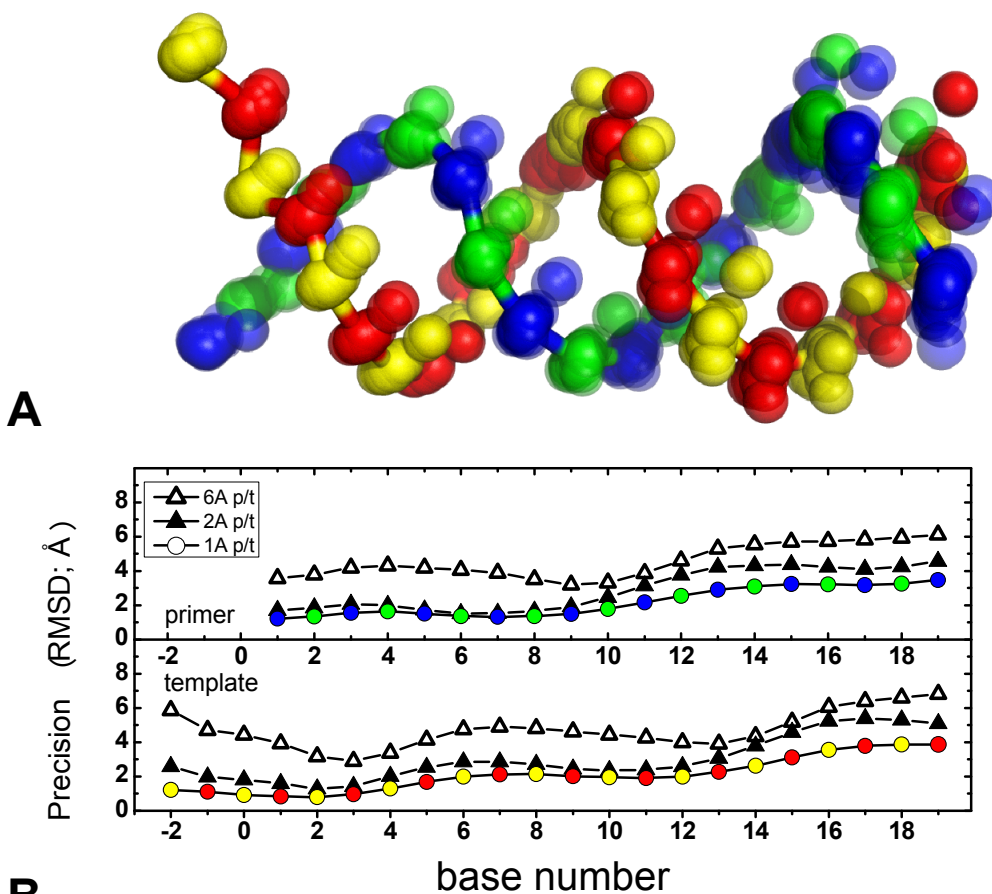


Figure 5. Bootstrapping of docked dsDNA dp/dt. (A): Ensemble of structures generated with perturbed distances and analyzed by equation (5) and a 1 Å clash tolerance. For better comparison the phosphate atoms of the DNA-backbone are colored alternately yellow/red or blue/green for the primer and template strands, respectively. The DNA is oriented as in the figure below. (B): Uncertainty of phosphate atom positions calculated for each P atom using equation (5) after the initial search with a 6 Å clash tolerance (open triangles) and after two refinement steps with 2 Å (full triangles) and 1 Å (circles) clash tolerances. The average RMSD values for all phosphate atoms are 5.0 Å, 3.3 Å, and 2.4 Å, respectively.

Extension to flexible parts of the complex: Conformer selection (screening) using experimental FRET data. An alternative strategy of finding a structure compatible with FRET data is to generate a large number of putative structures (e.g., by extracting conformations from an MD trajectory) and to "filter" the results with respect to agreement with FRET data by calculating χ_r^2 for each structure. Structures with a low χ_r^2 and of good configurational quality are then selected. An obvious advantage of this approach is that state-of-the-art algorithms for conformational sampling can be used in case of flexible molecules instead of rigid body docking. As a proof of concept we generated a conformational ensemble of the template overhang (Figure 2B) applying all-

atom MD simulations in explicit solvent⁵³. Ten trajectories (2855 conformations, orange dots for the N1 atom of the last nucleobase in Figure 6A) were filtered using 16 distances where the acceptor dye was attached to the template overhang dp/dt(-6) and dp/dt(-15) (Supplementary Table S3). The structure of the overhang, that fits best to the FRET data, is depicted in magenta together with the approximate 1σ confidence interval indicated by green dots (150 conformations) representing the N1 atom of the nucleobase of nucleotide dt(15). The conformational ensemble satisfying FRET data is represented by three major configurations (bold) in Figure 6B preferentially located in positively charged regions. The green isopleths for the N1 atom determined by MD simulations (Figure 6C) illustrate good agreement between these regions and the structures satisfying FRET restraints. The structures wind around the fingers domain with the 5'-end binding to RT in a loop structure within positively charged protein regions. Although we cannot exclude the existence of other minor conformer populations for template overhang, sub-ensemble TCSPC data shows no excessive broadening of DA distance distributions measured for the template overhang as compared to the dsDNA part (Supplementary Section S3.3.5, Tables S4 and S5). Thus, a significant population of free unbound overhang can be excluded. While interactions of the ss template region in close proximity to the primer terminus directly affect active site geometry and, thus, fidelity of the enzyme, the effects of template/RT interactions beyond positions -6 are not immediately obvious. However, during reverse transcription the enzyme has to pass regions with extended secondary and even short double stranded structures. It is conceivable that extensive interactions between RT and the template relatively far upstream of the site of nucleotide incorporation help resolving such obstacles.

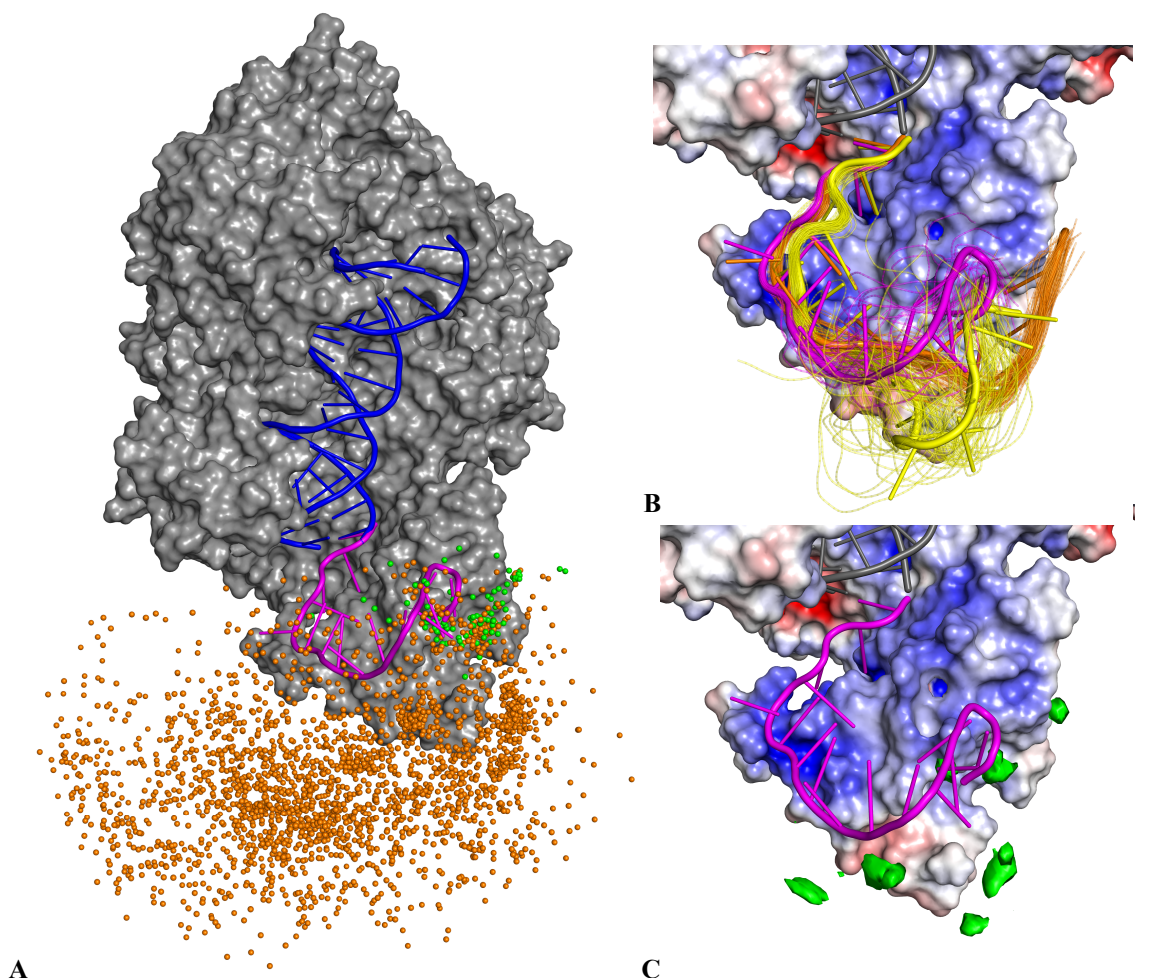


Figure 6. Structure of ssDNA dp/dt overhang obtained by screening of MD trajectories. (A) The ensemble of ssDNA overhang structures generated by all-atom MD simulations in explicit solvent (orange – 2855 conformations) and conformations filtered by smTCSPC using a confidence interval of 1σ (green – 150 conformations). Dots represent the N1 atom of the nucleobase of the nucleotide dt(-15). The structure that best fits to the FRET data is shown as a magenta cartoon ($\chi^2_r=0.88$). (B) Conformations within a confidence interval of 1σ of the smTCSPC data. The electrostatic potential of the protein as determined by APBS has been mapped onto the protein surface using a color scale (dark blue: $6 \text{ k}_B\text{T}/e_c$; dark red: $-6 \text{ k}_B\text{T}/e_c$). (C) Green isopleths show regions of preferred residence of the N1 atom of the nucleobase of nucleotide dt(-15) as determined from MD simulations. The isopleths encompass regions with at least 2% of the maximal residence likelihood. Note the overall good agreement between these regions and the location of the 1σ confidence interval determined by smTCSPC (panel B).

In conclusion, FPS improves the precision of FRET-derived structures dramatically. The combination with advanced computer simulations allows for a detailed molecular description of the proposed structure models. We developed a procedure to characterize the uniqueness and precision of FRET-restrained models, based on a precise spectroscopic estimation of “input” uncertainties. Utilizing the single-molecule advantage of FRET, heterogeneous systems can be easily investigated. Together with filtered

fluorescence correlation spectroscopy⁵⁴, FRET harbors the potential to study conformational control of biomolecular function in complex systems within a nanosecond to minute time range and associate it to detailed dynamic structures without spatial averaging.

Methods and Materials

All samples, applied methods and supporting data are described in detail in the Supplement. All used data used for FPS are compiled in the Supplementary Tables S2-S5.

Materials. Sample preparation and labeling was done according to ⁵⁵. We used RT mutants containing single accessible cysteine positions which were labeled with Alexa 488 as donor dye. Cy5 was used as acceptor dye and was attached to different positions on a 19/35 DNA/DNA primer/template. This dye pair has a Förster radius $R_0 = 52$ Å. Further details are given Supplementary Section S1.

Experimental conditions. The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris•HCl, pH 7.8, 6 mM MgCl₂, and 400 µM ascorbic acid. See Supplementary Section S2.1 for further details.

Experimental methods. The multiparameter fluorescence detection (MFD) is performed using an inverted confocal microscope with excitation by a pulsed laser at 496 nm. Fluorescence detection is performed with an additional pinhole defining a detection volume of 2 fl. Sample molecules diffusing through the detection volume cause a brief (~1ms) burst of fluorescence. Dilute solutions of molecules (~50 pM) ensure that only single molecules are detected. The fluorescence signal is divided into parallel and perpendicular components and in wavelength ranges below and above 620 nm (green and red, respectively) and single photons are detected by 4 avalanche photodiodes. For each photon the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength is recorded. Fluorescence bursts are distinguished from the background signal of 1 to 2 kHz by applying certain threshold intensity criteria ⁵⁶. See Supplementary Section S2.2 for further details.

Analysis methods. Distances (Supplementary Section S3.3.2) and their uncertainties (Supplementary Sections S3.3.3 and S3.3.4) were determined by PDA and time-resolved sub-ensemble anisotropies. The positional distribution of certain species was analyzed by sub-ensemble time correlated single-photon counting (Supplementary Section S3.3.5). The distances were computed according to equation (1).

FPS. Details on the generation and screening of models are given in Supplementary Section S3.4.

Method for rigid body docking of dsDNA with FRET restraints in Step 4 (for more details see Supplementary Section S3.4.2) “*Search*” In the first step, we generate a

large number of rigid body models that correspond to local minima of χ_r^2 (see Supplementary equation (S15)). For this, the optimization is started from a random configuration of the binding partners excluding those with clashes between them. The Verlet algorithm⁵⁷ (see Supplementary Section S3.4.2.2) is applied to model translational and rotational motions of units until the system is relaxed and certain convergence criteria are fulfilled. Although probably not the most efficient optimization algorithm, rigid body dynamics is straightforward to use and allows exploration of the local minima. In addition, clashes between molecules are prevented by introducing strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way, rigid body dynamics effectively minimizes the reduced chi-squared parameter that accounts for violations of FRET restraints and of van der Waals radii, $\chi_r^2 = (\chi_E^2 + \chi_{clash}^2)/(n - p)$ (p is the number of degrees of freedom, here 6).

Initially clashes are allowed to some extent to ensure penetration of DNA into the nucleic acid binding cleft, and structures showing considerable clashes can be filtered out later. To ensure exhaustive sampling of the configurational space of the binding partners, docking with random initial conditions is repeated many (10^2 - 10^4) times.

“*Refinement*”. The second step is to re-model the AVs accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure the dyes are *not* attached to. These modified AVs are then used to calculate new mean dye positions, followed by a re-optimization of the structure. At the same time, we simultaneously reduce the clash tolerance to make clashes between complex partners practically forbidden. This procedure can be repeated several times for each structure

until new iterations do not further improve the agreement with experimental data or change the solution significantly.

Analysis of docking/screening results. Solutions are considered ambiguous if the respective χ_r^2 values do not differ significantly. We typically apply a threshold $\chi_r^2 < \chi_{r,\min}^2 + [2/(n-p)]^{1/2}$, which roughly corresponds to the variance of the chi-squared distribution of $2 \times (\text{degrees of freedom})^{58}$ (blue lines in Figure 4). The fact that $\chi_{r,\min}^2$ is often larger than one is attributed to systematic experimental errors and to possible limitations of the AV and/or rigid body models. In this work this problem is compensated by oversampling. Other criteria defining different levels of significance can be applied here in a straightforward way.

MD simulations are described in Supplementary Section S3.4.3. The discrimination between models and the determination of quality parameters is described in Supplementary Section S3.5.

Acknowledgments

CAMS and SK thank the German Science foundation (DFG) in the priority program SPP 1258 "Sensory and regulatory RNAs in prokaryotes" for funding this research (SE 1195/12-2). The authors acknowledge the Volkswagen-Foundation for funding to C.A.M.S. (I/74 470) and R.S.G. The authors acknowledge the International Helmholtz Research School on Biophysics and Soft Matter (BioSoft) for funding to T.P. and the NRW Research School Biostruct for funding to S.S.

References

1. Orrit, M. & Bernard, J. Single Pentacene Molecules Detected by Fluorescence Excitation in a p-Terphenyl Crystal. *Phys. Rev. Lett.* **65**, 2716-2719 (1990).
2. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819-846 (1978).
3. Ha, T. *et al.* Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc.Natl.Acad.Sci.USA*. **93**, 6264-6268 (1996).
4. Weiss, S. Fluorescence spectroscopy of single biomolecules. *Science* **283**, 1676-1683 (1999).
5. Sakon, J. J. & Weninger, K. R. Detecting the conformation of individual proteins in live cells. *Nat. Meth.* **7**, 203-U256 (2010).
6. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).
7. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **324**, 203-207 (2009).
8. Margittai, M. *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc.Natl.Acad.Sci.USA*. **100**, 15516-15521 (2003).
9. Borgia, M. B. *et al.* Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **474**, 662-U142 (2011).
10. Woźniak, A. K., Schröder, G., Grubmüller, H., Seidel, C. A. M. & Oesterhelt, F. Single molecule FRET measures bends and kinks in DNA. *Proc.Natl.Acad.Sci.USA*. **105**, 18337-18342 (2008).
11. Hoefling, M. *et al.* Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. *Plos One* **6**, e19791 (2011).
12. Mekler, V. *et al.* Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599-614 (2002).
13. Andrecka, J. *et al.* Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **105**, 135-140 (2008).
14. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Meth.* **5**, 965-971 (2008).
15. Choi, U. B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.* **17**, 318-U384 (2010).
16. Brunger, A. T., Strop, P., Vrljic, M., Chu, S. & Weninger, K. R. Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.* **173**, 497-505 (2011).
17. Sabir, T., Schroder, G. F., Toulmin, A., McGlynn, P. & Magennis, S. W. Global Structure of Forked DNA in Solution Revealed by High-Resolution Single-Molecule FRET. *J. Am. Chem. Soc.* **133**, 1188-1191 (2011).
18. McCann, J. J., Zheng, L. Q., Chiantia, S. & Bowen, M. E. Domain Orientation in the N-Terminal PDZ Tandem from PSD-95 Is Maintained in the Full-Length Protein. *Structure* **19**, 810-820 (2011).
19. Balci, H., Arslan, S., Myong, S., Lohman, T. M. & Ha, T. Single-Molecule Nanopositioning: Structural Transitions of a Helicase-DNA Complex during ATP Hydrolysis. *Biophys. J.* **101**, 976-984 (2011).
20. Best, R. B. *et al.* Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. USA* **104**, 18964-18969 (2007).
21. Dolgih, E. *et al.* Theoretical Studies of Short Polyproline Systems: Recalibration of a Molecular Ruler. *J. Phys. Chem. A* **113**, 4639-4646 (2009).
22. Dolgih, E., Roitberg, A. E. & Krause, J. L. Fluorescence resonance energy transfer in dye-labeled DNA. *J. Photochem. Photobiol. A* **190**, 321-327 (2007).
23. VanBeek, D. B., Zwier, M. C., Shorb, J. M. & Krueger, B. P. Fretting about FRET: Correlation between kappa and R. *Biophys. J.* **92**, 4168-4178 (2007).
24. Olofsson, M., Kalinin, S., Zdunek, J., Oliveberg, M. & Johansson, L. B. A. Tryptophan-BODIPY: A versatile donor-acceptor pair for probing generic changes of intraprotein distances. *Phys. Chem. Chem. Phys.* **8**, 3130-3140 (2006).

25. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* **133**, 2463-2480 (2011).
26. Cai, Q. *et al.* Nanometer distance measurements in RNA using site-directed spin Labeling. *Biophys. J.* **93**, 2110-2117 (2007).
27. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Third edn, (Springer, 2006).
28. van der Meer, B. W., Cooker, G. & Chen, S. Y. *Resonance Energy Transfer: Theory and Data*. (VCH Publishers, 1994).
29. Goette, M., Rausch, J. W., Marchand, B., Sarafianos, S. & Le Grice, S. F. J. Reverse transcriptase in motion: Conformational dynamics of enzyme-substrate interactions. *Biochim. Biophys. Acta - Proteins and Proteomics* **1804**, 1202-1212 (2010).
30. Liu, S. X., Abbondanzieri, E. A., Rausch, J. W., Le Grice, S. F. J. & Zhuang, X. W. Slide into action: Dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* **322**, 1092-1097 (2008).
31. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790 (1992).
32. Jacobo-Molina, A. *et al.* Crystal Structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **90**, 6320-6324 (1993).
33. Esnouf, R. *et al.* Mechanism of inhibition of HIV-1 reverse transcriptase by non nucleoside inhibitors. *Nat. Struct. Biol.* **2**, 303-308 (1995).
34. Peletskaya, E. N., Kogon, A. A., Tuske, S., Arnold, E. & Hughes, S. H. Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. *J. Virol.* **78**, 3387-3397 (2004).
35. Patel, P. H. *et al.* Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351-5363 (1995).
36. Upadhyay, A. K., Talele, T. T. & Pandey, V. N. Impact of template overhang-binding region of HIV-1 RT on the binding and orientation of the duplex region of the template-primer. *Mol. Cell. Biochem.* **338**, 19-33 (2010).
37. Rothwell, P. J. *et al.* Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655-1660 (2003).
38. Sisamak, E., Valeri, A., Kalinin, S., Rothwell, P. J. & Seidel, C. A. M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **475**, 456-514 (2010).
39. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970-6978 (2006).
40. Gopich, I. V. & Szabo, A. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *J. Phys. Chem. B* **107**, 5058-5063 (2003).
41. Gopich, I. V. & Szabo, A. Single-molecule FRET with diffusion and conformational dynamics. *J. Phys. Chem. B* **111**, 12925-12932 (2007).
42. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C. A. M. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B* **112**, 8361-8374 (2008).
43. Nir, E. *et al.* Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103-22124 (2006).
44. Santoso, Y., Torella, J. P. & Kapanidis, A. N. Characterizing Single-Molecule FRET Dynamics with Probability Distribution Analysis. *ChemPhysChem* **11**, 2209-2219 (2010).
45. Dale, R. E., Eisinger, J. & Blumberg, W. E. Orientational freedom of molecular probes - Orientation factor in intra-molecular energy transfer. *Biophys. J.* **26**, 161-193 (1979).
46. van der Meer, B. W. Kappa-squared: from nuisance to new sense. *Rev. in Mol. Biotechnol.* **82**, 181-196 (2002).
47. Muschielok, A. & Michaelis, J. Application of the Nano-Positioning System to the Analysis of Fluorescence Resonance Energy Transfer Networks. *J. Phys. Chem. B* **115**, 11927-11937 (2011).

48. Isaksson, M., Norlin, N., Westlund, P. O. & Johansson, L. B. A. On the quantitative molecular analysis of electronic energy transfer within donor - acceptor pairs. *Phys. Chem. Chem. Phys.* **9**, 1941-1951 (2007).
49. Lyskov, S. & Gray, J. J. The RosettaDock server for local proteinprotein docking. *Nucleic Acids Res.* **36**, W233-W238 (2008).
50. van Dijk, M. & Bonvin, A. Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res.* **38**, 5634-5647 (2010).
51. Efron, B. Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion. *Annals of Statistics* **14**, 1301-1304 (1986).
52. Brunger, A. T. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475 (1992).
53. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688 (2005).
54. Felekyan, S., Kalinin, S., Sanabria, H., Valeri, A. & Seidel, C. A. M. Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *ChemPhysChem* **13**, 1036-1053 (2012).
55. Rothwell, J. P. *Structural Investigation of the HIV-1 RT using single pair Fluorescence Energy Transfer*, University of Dortmund, (2002).
56. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163-180 (2001).
57. Martyna, G. J., Tobias, D.J., Kleinm M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177-4189 (1994).
58. Soong, T. T. *Fundamentals of probability and statistics for engineers.* (John Wiley & Sons, 2004).

FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang

Stanislav Kalinin^{1#}, Thomas Peulen^{1,*}, Simon Sindbert^{1,*}, Paul J. Rothwell^{1/2}, Sylvia Berger¹, Tobias Restle^{2/4}, Roger S. Goody², Holger Gohlke³, Claus A. M. Seidel^{1#}

¹Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany, ²Max-Planck-Institut für molekulare Physiologie, Otto-Hahn-Str. 11, 44227 Dortmund, Germany, ³Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany, ⁴UK S-H - Campus Lübeck, Institut für Molekulare Medizin, Ratzeburger Allee 160, 23538 Lübeck, Germany

* contributed equally.

corresponding authors: stanilav.kalinin@uni-duesseldorf.de, cseidel@hhu.de.

Supporting Information

S1.	MATERIALS.....	3
S1.1.	HIV-RT:dp/pt complex	3
S1.2.	dt/dp sequences.....	3
S1.3.	Structures of modified nucleotides	4
S1.4.	Secondary structure of dp/dt.....	4

S2.	EXPERIMENTAL CONDITIONS	5
S2.1.	Measurement conditions	5
S2.2.	Multiparameter Fluorescence Detection (MFD).....	5
S3.	FPS METHOD	6
S3.1.	Starting Model.....	6
S3.2.	Network of DA-pairs	7
S3.3.	Quantitative FRET-measurements.....	7
S3.3.1.	MFD-measurements	7
S3.3.2.	Distances through PDA	7
S3.3.3.	Statistical uncertainties in PDA	10
S3.3.4.	Estimation of the κ^2 related distance uncertainties	12
S3.3.5.	Sub-ensemble TCSPC (seTCSPC) of FRET data	13
S3.4.	Generation/Screening of models	14
S3.4.1.	Obtaining mean position distances: additional details.....	14
S3.4.2.	Rigid body docking: implementation details	16
S3.4.3.	Structure ensemble generation via MD simulation	19
S3.5.	Model discrimination.....	20
S3.5.1.	Cluster analysis on dsDNA	20
S3.5.2.	Precision estimation: overfitting tests (dsDNA).....	21
S3.5.3.	FRET-guided screening for the template overhang	21
S4.	SUMMARY OF EXPERIMENTAL RESULTS	22
S4.1.	Tables.....	22
S4.2.	MFD plots.....	27
S4.3.	seTCSPC analysis of fluorescence lifetime and anisotropy.....	36
S4.4.	Obtained κ^2 distributions	38

S1. Materials

S1.1. HIV-RT:dp/pt complex

Sample preparation and labeling was done according to ¹. We used RT mutants containing a single accessible cysteine at positions 6, 27, 194 and 287 on the p66 subunit (RT(p66Q6C), RT(p66T27C), RT(p66E194C) and RT(p66K287C), respectively) and at positions 6, 173, 194 and 281 of the p51 subunit (RT(p51Q6C), RT(p51K173C), RT(p66E194C) and RT(p66K281C), respectively) ² (see Figure 2B in the main text). These cysteines were labeled with the green donor fluorophore Alexa488-C5 maleimide ³. The red acceptor dye Cy5 was attached to the primer strand at positions 1, 10 and 19 (dp(1)/dt, dp(10)/dt and dp(19)/dt, respectively) or the template strand at positions -15 and -6 (dp/dt(-15) and dp/dt(-6), respectively) of a 19/35 DNA/DNA primer/template (see S1.2 and S1.4 for sequences, labeling positions, secondary structure and numbering of dp/dt) either by internal labeling with a C6-aminolink with the NHS-ester of Cy5 (for dp(10)/dt, dp(19)/dt, dp/dt(-6) and dp/dt(-15)) or to the 3' end of the primer with the phosphoramidite derivative of Cy5 (for dp(1)/dt) (see S1.3). The sequence of the primer/template is based on the HIV-1 viral primer binding site.

S1.2. dt/dp sequences

Table S1. Sequences and labeling positions (red) of primer (dp) and template (dt) strands.

sample	sequence and labeling position
dp	5'-d(TTGTCCCTGTTCTGGGCGCC)-3'
dp(1)	5'-d(TTGTCCCTGTTCTGGGCGC C)-3'
dp(10)	5'-d(TTGTCCCTG T CTGGGCGCC)-3'
dp(19)	5'-d(T TGTCCCTGTTCTGGGCGCC)-3'
dt	5'-d(TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA)-3'
dt(-6)	5'-d(GGGTTAATC T CTGCATGGCGCCCGAACAGGGACAA)-3'
dt(-15)	5'-d(T GGTTAATCTCTGCATGGCGCCCGAACAGGGACAA)-3'

S1.3. Structures of modified nucleotides

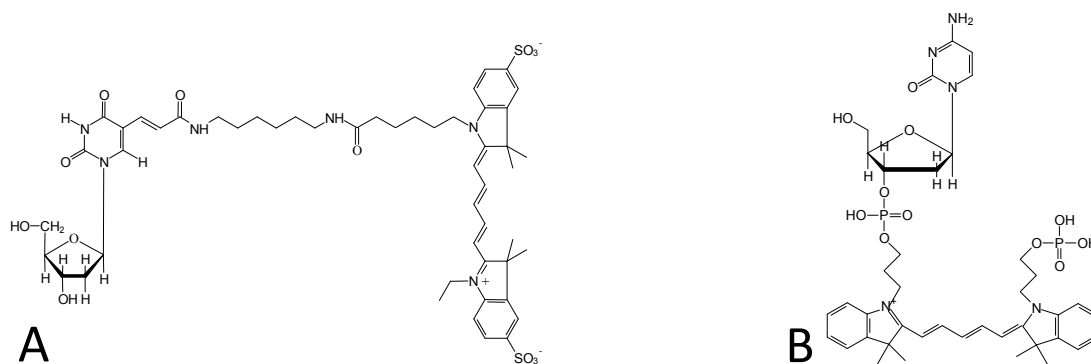


Figure S1. Structures of the fluorescent modified nucleotides used. (A) dThymine labeled with C6-aminolink with the NHS-ester of Cy5 used at positions dp(10), dp(19), dt(-6) and dt(-15). (B) 3' labeled dCytosine with the phosphoramidite derivative of Cy5 used for position dp(1).

S1.4. Secondary structure of dp/dt

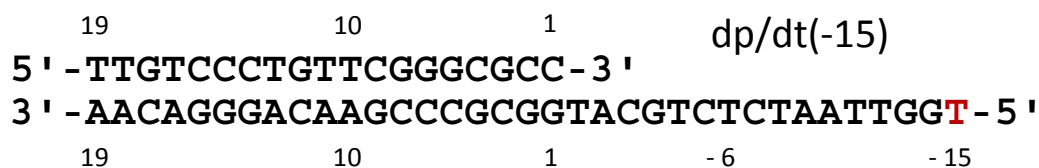


Figure S2. Secondary structure of DNA dp/dt(-15) with labeling position (red).

S2. Experimental conditions

Experiments were performed according to ⁴:

S2.1. Measurement conditions

The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris·HCl, pH 7.8, 6 mM MgCl₂, and 400 μ M ascorbic acid. Because of the extremely low concentrations used in single-molecule studies (50 pM RT and 200 pM dp/dt), high binding affinities are required to ensure stable complex formation. By reducing the KCl concentration from the 50 mM used in previous studies on this system ⁵ to 10 mM, we were able to reduce the dissociation constant, K_d , dramatically. Data from classical ensemble titrations indicated an upper limit for the K_d value at this salt concentration of <100 pM. The actual single-molecule measurements confirmed that the affinity was high enough to ensure complex formation, because most RT molecules sampled contained bound (labeled) substrate. Ensemble measurements on single nucleotide incorporation under these salt conditions showed similar kinetic behavior to that seen at higher salt, with three kinetic phases being observed ^{1,5}. Before measurements, a “stock” complex solution was made at a concentration of 250 nM protein and 1 μ M dp/dt. This solution was equilibrated on ice for at least 30 min. Immediately before a measurement, the stock solution was diluted rapidly into the measurement buffer 1:10, followed by a further 1:10 dilution before a final 1:50 dilution to a final concentration of 50 pM protein and 200 pM dp/dt. A droplet of the sample solution (50 μ l) was applied to a coverslip which formed the bottom of a closed chamber with a water-saturated atmosphere. Coating the coverslip with κ -casein led to complete suppression of adsorption of the sample molecules during the measurement time of up to an hour.

S2.2. Multiparameter Fluorescence Detection (MFD)

MFD is performed using a confocal (pinhole $\varnothing = 100\mu\text{m}$) epi-illuminated microscope with excitation by a pulsed linearly polarized argon-ion laser at 496 nm. The laser is focused into the solution with a 60x 1.2 water immersion objective. Fluorescence detection is performed with the same objective, but with an additional pinhole that results in a detection volume of 2 fl, as estimated from fluorescence correlation spectroscopy (FCS) measurements. Further set-up parameters obtained via FCS are the ratio ω_0/z_0 of 4, and the characteristic diffusion times t_D for

Rhodamine 110 and RT:dp/dt complex single-labeled at position p66K287C of 0.3 and 1.3 ms, respectively. Here, ω_0 and z_0 are the distances from the center of the laser beam focus in the radial and axial directions, respectively, at which the collected fluorescence intensity has dropped by a factor of e^2 , compared with its peak value. Sample molecules diffusing freely through the solution occasionally pass through the detection volume, resulting in a brief (~ 1 ms) burst of fluorescence. Dilute solutions of molecules (~ 50 pM) ensure that only single molecules are detected, each molecule producing a single burst. The fluorescence signal is divided into parallel and perpendicular components in wavelength ranges below and above 620 nm (green and red, respectively; filters: HQ535/50, HQ730/140). The photons are detected by 4 avalanche photodiodes (APD) coupled to a counting board (SPC 431, Becker & Hickel, Berlin) and a personal computer. For each photon the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength is recorded. Fluorescence bursts are distinguished from the background signal of 1 to 2 kHz by applying certain threshold intensity criteria ⁶.

S3. FPS Method

S3.1. Starting Model

Dye distributions by the AV approach. We model dye distributions by the AV approach ^{7,8} according to ⁹. The dyes are approximated by a sphere with an empirical radius of R_{dye} , where the central atom of the fluorophore is connected by a flexible linkage of a certain effective length $\underline{L}_{\text{link}}$ and width w_{link} to the nucleobase. The overall length of the linkage is given by the actual length of the linker and the internal chemical structure of the dye. A geometric search algorithm finds all dye positions within the linkage length from the attachment point which do not cause steric clashes with the macromolecular surface. All allowed positions are considered as equally probable which allows one to define an accessible volume for the dye (AV). To take the three quite different spatial dimensions of Cy5 (see Figure S1) into account, we used its real physical dimensions for each calculation of a position distribution and performed three independent AV simulations with three different radii $R_{\text{dye}(i)}$ and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. See caption of Figure 2 in the main text for values used for $R_{\text{dye}(i)}$, $\underline{L}_{\text{link}}$ and w_{link} .

Prior Knowledge: States expected to be observed for the HIV-1 RT:dp/dt complex. Under the given measurement conditions the RT:dp/dt complexes adopts different conformational states designated as product complex in educt state (P-E), product-complex in product state (P-P) and a dead-end complex (D-E). The P-P and P-E are two species generally close in transfer efficiency representing different steps in the polymerization reaction ⁴. The dead-end complex (D-E) is most likely not directly contributing to the enzymatic activity of the protein ⁴. The measurements were performed in absence of dNTPs and pyrophosphate. Thus, all three states are present under the given measurement conditions. However, as the protein is preferentially in the educt-state (P-E), the state with the highest amplitude was assigned to the P-E the state and the less populated state to the P-P state. Depending on the relative orientation and distance of the donor and acceptor fluorophores the two states could not always be resolved as the change in transfer efficiency upon transition from P-E to P-P is not big enough. Hence, the states P-E and P-P were sometimes fitted with one state. Only a small fraction of the protein/DNA complex is in the enzymatically not contributing D-E complex. Thus an assignment of the P-E state is in most measurements unambiguous.

S3.2. Network of DA-pairs

Details on the labeling positions can be found in Section S1.

S3.3. Quantitative FRET-measurements

S3.3.1.MFD-measurements

For experimental setup for MFD see Section S2.2

S3.3.2.Distances through PDA

As specified in the main text, we calculate distances by measuring the fluorescence intensities of D and A (F_D and F_A , respectively). R_{DA} is then given by combining equation (1) and (2) in the main text which yields equation (S1):

$$R_{DA} = \left(\frac{\Phi_{FA}}{\Phi_{FD(0)}} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_0 \quad (S1)$$

or equivalently by equation (S2):

$$R_{DA} = \left(\Phi_{FA} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_{0r} \quad (S2)$$

Here, we use the reduced Förster radius R_{0r} ¹, which, in contrast to R_0 , does not depend on $\Phi_{FD(0)}$. Throughout this work we used $\Phi_{FA} = 0.32$ ($\Phi_{FA} = 0.32$ due to the presence of $\sim 20\%$ cis-trans isomerization instead of $\Phi_{FA} = 0.4$ which is expected for ensemble measurements¹⁰) and $R_{0r} = 53.97 \text{ \AA}$ (using $R_{0r} = 53.97 \text{ \AA}$ is equivalent to $R_0 = 52 \text{ \AA}$ and $\Phi_{FD(0)} = 0.8$). F_D and F_A can be calculated from the signals measured in the green and red detection channels S_G and S_R , respectively, via equations (S3) and (S4):

$$F_D = \frac{F_G}{g_G} = \frac{S_G - \langle B_G \rangle}{g_G} \quad (S3)$$

$$F_A = \frac{F_R}{g_R} = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \quad (S4)$$

where F_G and F_R are the fluorescence signals in the green and the red signal channels, respectively, α is the crosstalk factor which is determined as the ratio between donor photons detected in the red channels and those detected in the green channels ($\alpha = F_{R(D)} / F_{G(D)}$) for the D only labeled sample, g_G and g_R are the detection efficiencies in the green and red channels, respectively, and $\langle B_G \rangle$ and $\langle B_R \rangle$ are the mean background intensities in the green and red channels, respectively. For the determination of R_{DA} we only need to know the ratio between F_D and F_A , therefore, we only need to know the ratio between the detection efficiencies g_G/g_R (see below for the determination of g_G/g_R).

To accurately predict the shape of S_G/S_R (or equivalently F_D/F_A) histograms in the presence of FRET we use PDA, which explicitly takes into account shot noise, background contributions and additional broadening due to complex acceptor photophysics^{11,12,13, Kalinin, 2010 #2744}. PDA calculates the probability of observing a certain combination of photon counts $P(S_G, S_R)$

$$P(S_G, S_R) = \sum_{F_G + B_G = S_G; F_R + B_R = S_R} P(F) P(F_G, F_R | F) P(B_G) P(B_R) \quad (S5)$$

The intensity distribution of the fluorescence only contribution to the signal, $P(F)$, is obtained from the total measured signal intensity distribution $P(S)$ by deconvolution assuming that the background signals B_G and B_R obey Poisson distributions, $P(B_G)$ and $P(B_R)$, with known mean intensities $\langle B_G \rangle$ and $\langle B_R \rangle$. $P(F_G, F_R | F)$ represents the conditional probability of observing a particular combination of green and red fluorescence photons, F_G and F_R , provided the total number of registered fluorescence photons is F , and can be expressed as a binomial distribution¹³. Subsequently, $P(S_G, S_R)$ may be further manipulated to generate a theoretical histogram of any FRET-related parameter as discussed elsewhere¹².

Species needed in PDA. Typically a model was used which accounts for two FRET-populations representing the educt-state (P-E) and the product-state (P-P), a D-only population, and, if necessary, populations accounting for impurities and for a dead-end complex (see above Section S3.1) were added as well (see section S4.2 for the assignments of the states in MFD plots for all experiments). The FRET states were fitted using the same value for the global relative additional (beyond the shot noise) width σ_{app} (see above Section S3.1). Of the P-E and P-P states the distance of the one with the larger amplitude was assumed to be P-E and, thus, was chosen to be used for structural modeling. For some datasets only one FRET state is visible. In those cases we assume that the product and educt state are overlapping. If the P-P and P-E-state have similar amplitudes (ratio smaller than 4:3) an assignment of the distances to the corresponding states is not possible and the amplitude weighted distance $\langle R \rangle_x$ with the uncertainties err_+ , err_- has been used for modeling (see Table S2). In general, for a model with n FRET states and a D-only fraction, $2n+1$ fit parameters are needed: n mean DA distances, n fractions, and an additional (beyond the shot noise) distribution width σ_{app} expressed as a fraction of the corresponding mean distance. This additional distribution width can be attributed mainly to complex acceptor photophysics and thus can be fitted globally as justified in¹⁴. As a result, PDA needs much fewer free parameters than the classical approach of fitting multiple Gaussian peaks, which requires up to $3n+1$ parameters (n mean DA distances, n fractions, $n+1$ peak widths).

To be able to convert model distances into probabilities of observing green photons, the detection efficiency ratio g_G/g_R is needed (see equations (S3) and (S4)). The smFRET measurements were performed over a period of more than one year. For the individual measurement sessions the g_G/g_R -ratio of the setup was calibrated for each measurement session by requiring that the linker-corrected static FRET line¹⁵ goes through the observed FRET

populations in the 2D histograms of F_D/F_A vs $\tau_{D(A)}$ (see e.g. Figure 3A main text, see Tables S2 and S3 for the determined values for g_G/g_R). The linker-corrected static FRET lines were calculated with the following equation:

$$F_D/F_A = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left(\frac{\langle \tau_{D(0)} \rangle_x}{c_3 \langle \tau_{D(A)} \rangle_f^3 + c_2 \langle \tau_{D(A)} \rangle_f^2 + c_1 \langle \tau_{D(A)} \rangle_f + c_0} - 1 \right) \right. \quad (S6)$$

whereas $\langle \tau_i \rangle_x$ and $\langle \tau_i \rangle_f$ are species and fluorescence averaged lifetimes, respectively. The polynomial coefficients in equation (S6) are compiled in Table S7. They are obtained as described in ¹⁵ with one modification. In this work we also consider the fact that in most of cases the fluorescence decay of D dyes (Alexa488) bound to RT is multi-exponential. Thus, the formalism described in ¹⁵ has to be extended as follows. First, the donor fluorescence decay is formally fitted using a multi-exponential relaxation model (see Table S6). For each donor sub-species, we assume that its quantum yield is proportional to the corresponding fluorescence lifetime, and calculate the donor and acceptor fluorescence (equation (2) in the main text) as well as the lifetime distribution in the presence of FRET (equation (2) in the main text) as described ¹⁵. The total donor and acceptor fluorescence intensities are weighed sums of those calculated for all donor sub-species, and the overall lifetime distribution is a superposition of species-specific lifetime distributions, weighted by the species fractions. From these data we can calculate $\langle \tau \rangle_f$ and $\langle E \rangle$ ($\langle E \rangle = \langle F_A \rangle / [\gamma' \langle F_D \rangle + \langle F_A \rangle]$) for any given $\langle R_{DA} \rangle$. This procedure is then repeated for a series of $\langle R_{DA} \rangle$ values, yielding corresponding sets of $\langle \tau \rangle_f$ and $\langle E \rangle$. This dependence is approximated with a polynomial function (equation (S6)). A program for computing various FRET lines is available from the authors (see <http://www.mpc.uni-duesseldorf.de/seidel/software.htm>).

S3.3.3. Statistical uncertainties in PDA

Confidence intervals estimation for multiple fit parameters is performed as follows. All free fit parameters are varied simultaneously in a random manner. The χ_r^2 -value is calculated at 100000 random points yielding 100-1000 points with χ_r^2 -values below $\chi_{r,\min}^2 + (2/N_{\text{bins}})^{1/2}$ (here N_{bins} is the number of histogram bins, and $\chi_{r,\min}^2$ is the reduced chi-squared of the best fit). The range where such fits are possible is assigned as 1σ confidence interval (Figure S3). Whereas one

could calculate χ_r^2 thresholds more strictly from the chi-squared distribution¹⁶, in practice $\chi_{r,\min}^2$ is often affected by experimental imperfections and can be considerably larger than one. For this reason, we prefer the simple test mentioned above which relates χ_r^2 values to that of the best fit.

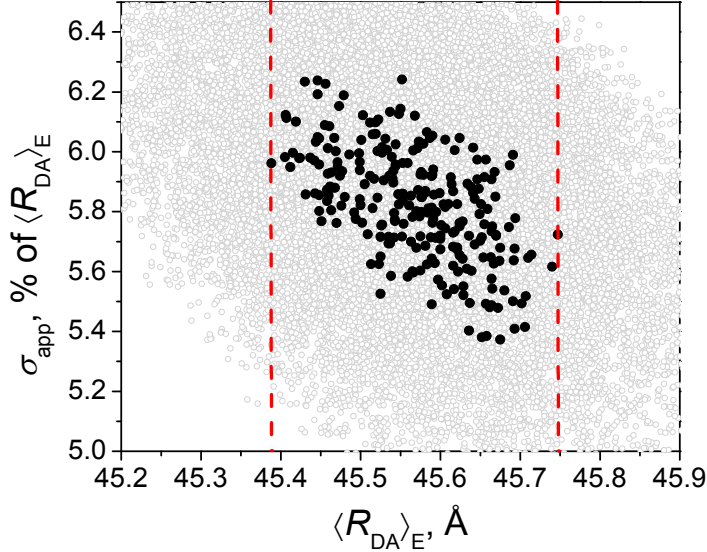


Figure S3. Confidence interval estimation in PDA. Here $\langle R_{DA} \rangle_E$ of the P-E state is plotted against its apparent width (σ_{app}) for RT(p66Q6C):dp(10)/dt. Black points represent sets of parameters which provide acceptable fits with $\chi_r^2 < \chi_{r,\min}^2 + (2/N_{bins})^{1/2}$, where $\chi_{r,\min}^2$ is the reduced chi-squared for the best fit, and N_{bins} is the number of FRET histogram bins. Gray dots correspond to $\chi_r^2 < \chi_{r,\min}^2 + 1$, making it clear that pre-defined search intervals are sufficiently large. Red lines show resulting 1σ confidence interval for $\langle R_{DA} \rangle_E$.

It is clear from Figure S3 that in our case the statistical uncertainties ($< \pm 0.2\text{\AA}$) are very small compared to those due to κ^2 (see further). However, photon statistics becomes more relevant when minor and/or overlapping FRET states are considered. If an assignment of the P-P and P-E state is unclear the errors were calculated as described above. In our experience, for minor FRET states with $< 20\%$ equilibrium population, statistical errors can easily exceed those due to κ^2 uncertainties (unpublished results).

S3.3.4. Estimation of the κ^2 related distance uncertainties

The procedure is described in details in ^{9,17,18}. Briefly, the knowledge of residual anisotropies (r_∞) of the directly excited D and A, and that of the A excited via FRET ($r_{\infty,D}$, $r_{\infty,A}$ and $r_{\infty,A(D)}$, respectively), allows one to reduce the range of possible κ^2 -values. In our case a sub-ensemble analysis of single-molecule data was performed to estimate $r_{\infty,D}$. Figure S4 shows typical decays of polarized components $F_{\parallel}(t)$ and $F_{\perp}(t)$ which are globally fitted with the following model:

$$\begin{aligned} F_{\parallel}(t) &= F(t) \cdot (1 + (2 - 3l_1) \cdot r(t)) / 3 + B_{\parallel} \\ F_{\perp}(t) &= GF(t) \cdot (1 - (1 - 3l_2) \cdot r(t)) / 3 + B_{\perp} \end{aligned} \quad (S7)$$

In equation (S7) $F(t)$ is the fluorescence decay typically modeled by a bi-exponential relaxation, G is the ratio of detection efficiencies of parallel and perpendicular channels, factors l_1 and l_2 describe polarization mixing in high-NA objectives ¹⁹, and B_{\parallel} and B_{\perp} represent background contributions in parallel and perpendicular detection channels, respectively. The anisotropy decay $r(t)$ reflects local motions of the dye and global rotations of the macromolecule according to equation (S8)

$$r(t) = \left[(r_0 - r_\infty) e^{-\frac{t}{\rho_{local}}} + r_\infty \right] e^{-\frac{t}{\rho_{global}}} \cong (r_0 - r_\infty) e^{-\frac{t}{\rho}} + r_\infty \quad (S8)$$

In equation (S8), r_0 is the fundamental anisotropy (fixed to $r_0 = 0.38$). A fitted anisotropy decay is presented in Figure S4. For $r_{\infty,A}$ we use 0.13 from ⁹ and for $r_{\infty,A(D)}$ we used steady state anisotropies $r_{s,A(D)}$ which were determined via PDA according to ¹². Using the formalism described in ¹⁷, we obtain κ^2 uncertainties ($\Delta R_{DA}(\kappa^2)$) corresponding to each FRET distance. A typical κ^2 -distribution is shown in Figure 2 (main text). All values determined for ΔR_{DA} are listed in Tables S2 and S3.

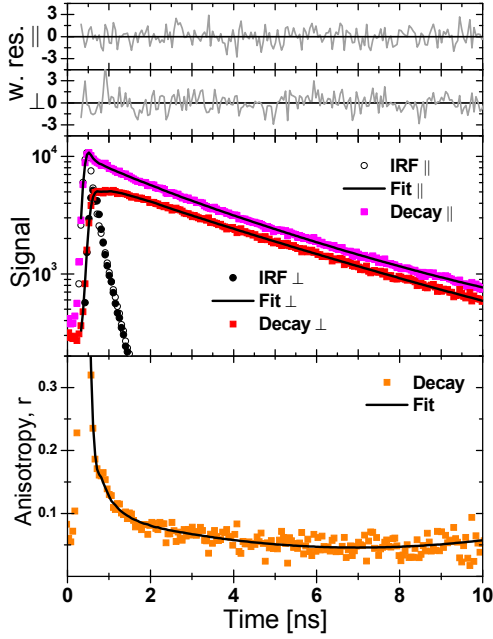


Figure S4. (Middle) $F_{\parallel}(t)$ and $F_{\perp}(t)$ decays obtained for the D-only population of sample RT(p51E194C):dp/dt. The instrument response functions (IRF) for parallel and perpendicular detection channels are shown as open and filled circles, respectively. The fits to $F_{\parallel}(t)$ and $F_{\perp}(t)$ (equation (S7)) are displayed as black solid lines. Weighed residuals are shown in the upper panel. (Bottom) The experimental anisotropy decay calculated as $r(t) = [F_{\parallel}(t) - GF_{\perp}(t)]/[F_{\parallel}(t) + 2GF_{\perp}(t)]$ overlaid with that obtained from the fitted model functions (equations. (S7) and (S8)). The resulting r_{∞} is 0.131 (the apparent anisotropy plateau is lower because of background contributions).

S3.3.5. Sub-ensemble TCSPC (seTCSPC) of FRET data

Due to local quenching the fluorescence decay of the donor in the absence of FRET is often multi-exponential (equation (S9))

$$F_{D(0)}(t) = \sum_i x_D^{(i)} \exp(-t / \tau_{D(0)}^{(i)}) \quad (\text{S9})$$

Furthermore, it is usually assumed that for given DA distance and orientation, FRET rate (k_{FRET}) is independent of the donor lifetime. This is strictly true if quenching does not change the donor radiative lifetime, which is generally reasonable. In the presence of FRET, the donor fluorescence decay can be then expressed as

$$F_D(t) = \sum_i x_D^{(i)} \int_{R_{DA}} P(R_{DA}) \exp\left(-\frac{t}{\tau_{D(0)}^{(i)}} \left[1 + (R_0 / R_{DA})^6\right]\right) dR_{DA} \quad (\text{S10})$$

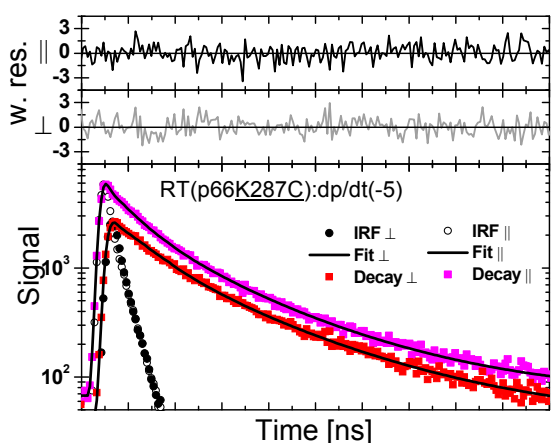
In this work we assumed Gaussian distribution of donor-acceptor distances ($P(R_{DA})$) with a mean of $\langle R_{DA} \rangle$ and a half-width of σ_{DA} . In addition, a fraction of Donor-only molecules (x_{DOnly}) was considered (equation (S11)):

$$F(t) = (1 - x_{\text{DOnly}})F_D(t) + x_{\text{DOnly}}F_{D(0)}(t) \quad (\text{S11})$$

Equation (S11) in combination with equation (S7) was used to fit the polarized components of the donor fluorescence relaxation recovered from smMFD data.

As shown in Figure S5 and Tables S4 and S5 there are no differences in width between the samples with an acceptor located on the double stranded part of the DNA (dp(1)/dt, dp(10)/dt, dp(19)dt) and the samples with the acceptor fluorophore attached to the template overhang (dp/dt(-6), dp/dt(-15)).

(A)



(B)

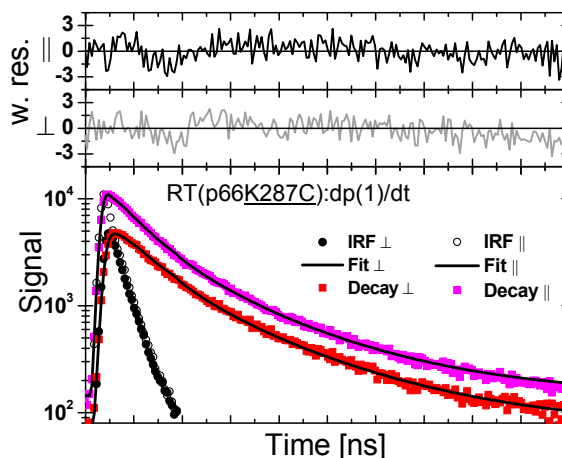


Figure S5: Decay histograms and fits of a Gaussian model function distributed in R_{DA} . The fluorescence properties (lifetimes) and anisotropies of the donor were fixed to the values of the donor sample p66K287C:dp/dt, the background fluorescence was fixed to values determined by independent background measurements. Free fitting parameters are the mean DA distance $\langle R_{DA} \rangle$ and the width of the distance distribution σ_{DA} . The obtained parameters are: (A) $\langle R_{DA} \rangle = 43 \text{ \AA}$, $\sigma_{DA} = 13 \text{ \AA}$, $\chi^2_r = 1.0$, for RT(p66K287C):dp/dt(-5), and (B), $\langle R_{DA} \rangle = 34 \text{ \AA}$, $\sigma_{DA} = 15 \text{ \AA}$, $\chi^2_r = 1.4$ for RT(p66K287C):dp(10)/dt.

S3.4. Generation/Screening of models

S3.4.1. Obtaining mean position distances: additional details

As in this work we focus on FRET between dyes attached to a macromolecule via flexible linkers, we must consider the contribution of linker length to the DA distance. In most of cases it is not negligible and can be even comparable with absolute donor-acceptor distances (Figure 1A main text). As discussed in detail in ⁹, dye linkers are responsible for two major effects: (1) obviously, a considerable displacement of the mean dye position with respect to the attachment

point is expected, and (2) averaging of FRET efficiencies over possible DA distances and orientations takes place. The second consideration makes the well-known equation $E = 1/(1 + R_{DA}^6 / R_0^6)$ inapplicable even to calculate the distance between mean dye positions (R_{mp}). To illustrate this effect and hence the difference between R_{mp} and $\langle R_{DA} \rangle_E$ AVs of Alexa488 and Cy5 dyes attached to DNA using "standard" C6 linkers were simulated. Figure 1A (main text) schematically shows simulated dye clouds, making it also clear that the C6 linker length (~ 20 Å between the attachment point and the center of the dye) is comparable with DA distances accessible to FRET (~ 35 - 70 Å for the Alexa488-Cy5 pair). By varying the labeling position of the acceptor dye, we obtained a set of mean position distances and corresponding mean FRET efficiencies $\langle E \rangle$. In Figure S6, the values of $\langle E \rangle$ are plotted against R_{mp} values. Significant deviations between the Förster dependence (equation (1) in the main text) and the obtained relationship between $\langle E \rangle$ and R_{mp} are apparent. Thus, if the average FRET efficiency $\langle E \rangle$ is directly used to calculate R_{mp} , errors of up to 10 Å ($\sim 30\%$) are expected (Figure 1B main text).

By interpolation of an $\langle R_{DA} \rangle_E$ - R_{mp} dependence using a known set of dye clouds, a $\langle R_{DA} \rangle_E$ - R_{mp} conversion function^{9,10} can be generated. We should note that $\langle E \rangle$ depends not only on R_{mp} , but also on the mutual orientation of the dye clouds (of given size and shape). Fortunately, this effect is relatively weak even for asymmetric dye position distributions such as presented in Figure 1A. Figure 1B (main text) shows the conversion function generated for random orientation of the dyes' AVs. The average deviation between the data and the fitted 3rd order polynomial function is only 0.6 Å (30 Å $< \langle R_{DA} \rangle_E < 70$ Å), which is much smaller than typical uncertainties of FRET-derived distances. Obviously, for more spherical AVs this deviation would be even less pronounced. This justifies the use of a conversion function at least during the initial optimization steps. With some effort the $\langle E \rangle$ - R_{mp} conversion function can be also obtained empirically²⁰ by studying a set of molecules with known structure (e.g. dsDNA).

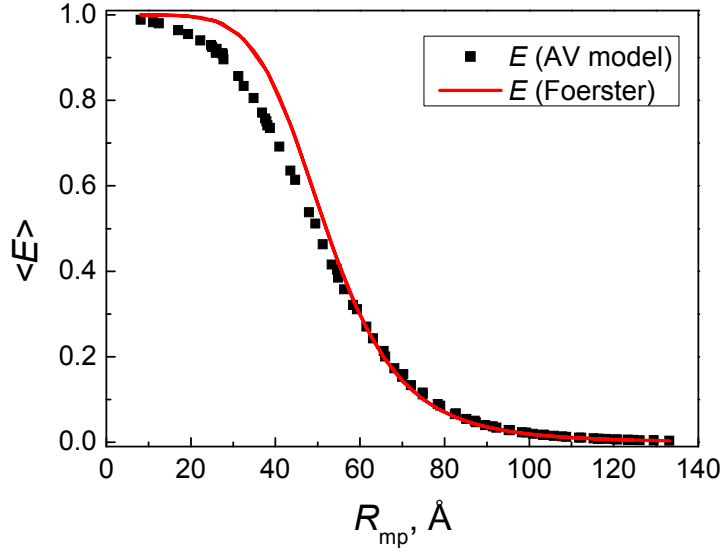


Figure S6. The relationship between R_{mp} and $\langle E \rangle$ averaged over spatial distributions of donor and acceptor estimated using the AV approach as shown in Figure 1A in the main text (for simplicity, $\kappa^2 = 2/3$ is assumed for every DA pair). The red solid line represents the dependence given by equation (1) in the main text with $R_{DA} = R_{mp}$.

S3.4.2. Rigid body docking: implementation details

S3.4.2.1. Mechanical model.

As discussed in the main text, the purpose of rigid body docking is to minimize the weighed deviation between n experimentally obtained distances $\{R_{DA}\}$ and corresponding model distances $\{R_{model}\}$, given the uncertainties $\{\Delta R_{DA}\}$:

$$\chi_E^2 = \sum_n \frac{(R_{DA} - R_{model})^2}{(\Delta R_{DA})^2} \quad (S12)$$

In equation (S12), $\{R_{DA}\}$ can represent a set of R_{mp} or $\langle R_{DA} \rangle_E$ values. The first option is easier to implement in combination with rigid body dynamics, whereas directly calculating deviations between experimental and model $\langle R_{DA} \rangle_E$ is more appropriate for structure screening. Now we notice that the right-hand part of equation (S12) is equivalent to the energy of a network of mean dye positions (points in space), connected with n springs with relaxed lengths of $\{R_{DA}\}$ and corresponding spring constants $k = 2/(\Delta R_{DA})^2$. The coordinates of mean dye positions are

obtained by MD or AV simulations and then fixed with respect to corresponding labeled macromolecules, which are treated as rigid bodies. Minimizing χ^2_E in equation (S7) is equivalent to relaxing this rigid body system. Here and further in this section we use reduced energy which results in unusual units for k [$1/\text{\AA}^2$] and other familiar quantities. If not stated otherwise, all distances are expressed in \AA , and masses are expressed in Da.

In addition to FRET restraints, clashes between different subunits are prevented by considering clash contributions to the total “energy”, equivalently to equation (S12), by equation (S13):

$$\chi^2_{\text{clash}} = \sum_{i,j} \begin{cases} 0 & , r_{ij} \geq r_{wi} + r_{wj} \\ (r_{wi} + r_{wj} - r_{ij})^2 / r_{\text{ctol}}^2 & , r_{ij} < r_{wi} + r_{wj} \end{cases} \quad (\text{S13})$$

where r_{ij} is the distance between atoms i and j which belong to different subunits, r_{wi} and r_{wj} are their van der Waals radii, and r_{ctol} is the pre-defined clash tolerance. We typically used 6 \AA during initial search and 2 to 1 \AA during refinement (see “step 4” in the main text). More realistic potentials (e.g. the Lennard-Jones potential ²¹) can be also used here instead of the harmonic potential in equation (S13). The simplified approach (equation (S13)) is justified by low accuracy of FRET data (a few \AA) as compared to possible violations of van der Waals radii ($r_{wi} + r_{wj} - r_{ij} < 0.1 \text{\AA}$). Anyhow, by choosing a sufficiently small r_{ctol} the atoms can be made as “hard” as desired, in which case the contribution of χ^2_{clash} becomes negligible. The reduced chi-squared parameter to be minimized is then given by equation (S14)

$$\chi_r^2 = (\chi_E^2 + \chi_{\text{clash}}^2) / (n - p) \rightarrow \min \quad (\text{S14})$$

where n is the number of distance restraints (in our case $n = 20$) and p is the number of degrees of freedom, which is equal to $6 \times (\text{number of bodies} - 1)$, in our case $p = 6$.

S3.4.2.2. Time evolution of the rigid body system

The position and orientation of each subunit at any time t is described by a coordinate vector of its center of mass $\mathbf{x}(t)$ and a rotational matrix $\mathbf{Q}(t)$. To model translational and rotational motions we used the Verlet algorithm ²² with damping,

$$\mathbf{x}(t + \Delta t) = (2 - \nu)\mathbf{x}(t) - (1 - \nu)\mathbf{x}(t - \Delta t) + \mathbf{F}\Delta t^2/m \quad (\text{S15})$$

In equation (S15), Δt is the simulation time step, the factor ν accounts for viscosity (see equation (S20)), \mathbf{F} is the total force acting on a subunit, and m stands for its mass. The “forces” are derived from violations of FRET distances, clashes between subunits, and optionally other constraints (e.g. flexible chemical linkages between subunits). Since here we are only interested in finding energy minima (rather than in investigating trajectories), we use a simplified representation of the moment of inertia (I) by assuming the same (mean) value of I for all axes. In analogy with equation (S15), we obtain for the rotations

$$\mathbf{Q}(t + \Delta t) = \mathbf{Q}\left(\frac{\mathbf{T}}{|\mathbf{T}|}, \theta\right) \mathbf{Q}(\mathbf{w}, -\nu\theta_w) \mathbf{Q}(t) \mathbf{Q}^{-1}(t - \Delta t) \mathbf{Q}(t) \quad (\text{S16})$$

In equation (S16), \mathbf{T} is the total torque vector; the angle θ is given by

$$\theta = \Delta t^2 |\mathbf{T}| / I \quad (\text{S17})$$

the rotational matrices with parameters represent a rotation by angle θ about an axis \mathbf{u} ,

$$\mathbf{Q}(\mathbf{u}, \theta) = \begin{pmatrix} c + u_x^2 d & u_x u_y d - u_z s & u_x u_z d + u_y s \\ u_x u_y d + u_z s & c + u_y^2 d & u_y u_z d - u_x s \\ u_x u_z d - u_y s & u_y u_z d + u_x s & c + u_z^2 d \end{pmatrix} \quad (\text{S18})$$

with: $\mathbf{u} = (u_x, u_y, u_z)$; $c = \cos\theta$; $d = 1 - \cos\theta$; $s = \sin\theta$

and \mathbf{w} and θ_w fulfil equation (S19)

$$\mathbf{Q}(\mathbf{w}, \theta_w) = \mathbf{Q}(t) \mathbf{Q}^{-1}(t - \Delta t) \quad (\text{S19})$$

To minimize oscillations and to improve convergence, viscosity factor ν must be chosen so that the system is close to being critically damped. ν is initially estimated by equation (S20)

$$\nu = 2\Delta t \sqrt{K/M} \quad (\text{S20})$$

with an option of additional fine-tuning to improve convergence. In equation (S20) K is the sum of all spring constants and M is the total mass of all subunits.

Each rigid body simulation is run until the following criteria are met: kinetic energy $E_K < 0.001$; $|\mathbf{F}| < 0.001$; $|\mathbf{T}| < 0.02$ (in the units described above). After the initial search procedure, typically

more than 99% of structures reach convergence as defined by these criteria. We should also mention that more advanced optimization procedures ²³ can be also applied at this step. However, in our experience, even with rigid body dynamics >95% convergence probability is achieved in most of cases.

S3.4.3. Structure ensemble generation via MD simulation

In the complex determined by X-ray crystallography (PDB code: 1R0A) four unpaired nucleotides of the template overhang are present. To generate a starting structure for the template overhang 12 nucleotides were added to the existing strand (sequence: 5'-GGGTTAATCTCT-3') such that the unpaired bases point straight away from the protein.

MD simulations were performed with the AMBER 11 suite of programs ²⁴ together with the force field as described by Cornell et al. ²⁵ using modifications suggested by Simmerling et al. ²⁶ for the protein, and the “bsc0” parameterization for the DNA ²⁷. The complex structure was placed into an octahedral periodic box of TIP3P water molecules ²⁸ with Na⁺ ions added to reach electroneutrality of the system. The distance between the edges of the water box and the closest atom of the complex was at least 10 Å, resulting in a system of ~147000 atoms. The system was minimized by 50 steps of steepest descent minimization followed by 450 steps of conjugate gradient minimization. The particle mesh Ewald (PME) method ²⁹ was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using SHAKE ³⁰. The time-step for all MD simulations was 2 fs, with a direct-space, non-bonded cutoff of 8 Å. Applying harmonic restraints with force constants of 5 kcal mol⁻¹ Å⁻² to all solute atoms but the added nucleotides, canonical ensemble (NVT)-MD was carried out for 50 ps, during which the system was heated from 100 K to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD was used for 50 ps to adjust the solvent density. From there, with the harmonic restraints applied to all solute atoms but the first 15 nucleotides of the template overhang, ten different trajectories were spawned by adjusting the simulation temperature to 300.0 K, 300.1K, ..., 300.9 K, using a time constant of 10 ps for heat-bath coupling. After additional 20 ns of equilibration time for each trajectory, the following 50 to 65 ns of NVT-MD for each trajectory were used for analysis with the program ptraj of the AMBER suite, with conformations saved every 20 ps. This resulted in 571 ns of total simulation time for production.

For filtering against FRET data (see Figure 6A, main text), conformations were extracted from each trajectory at intervals of 200 ps and pooled together. Likewise, the trajectories were pooled for generating a 3D histogram of the preferred location of the N1 atoms in either t(-15); the spacing of the cubic grid is 1 Å. Isopleths show preferred regions of occupancy by the N1 atoms at a contour level of 90% with respect to the most highly occupied cube. Figures were generated with PyMOL.

S3.5. Model discrimination

S3.5.1. Cluster analysis on dsDNA

After the search phase, the obtained solutions are sorted by “energy” (or equivalently by χ_r^2) and plotted as shown in Figure S7A. Obviously, there are groups of very similar solutions (clusters), which are separated by steps in the χ_r^2 plot and corresponding peaks in the RMSD plot (Figure S7C).

To generate a cluster plot as shown in Figure 4 (main text), we applied thresholds to $\Delta\chi_r^2$ and RMSD plots as shown by red lines in Figure S7. Solutions for which both $\Delta\chi_r^2$ and RMSD fell below these thresholds were grouped with previous structures. As a result, several clusters of similar structures were obtained as shown in Figure 4 using different symbol sizes.

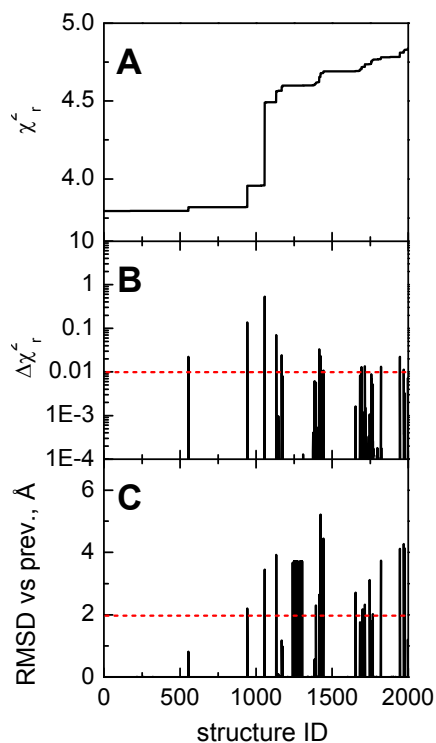


Figure S7. (A) χ_r^2 for 2000 best solutions. (B) Increase of χ_r^2 compared to the previous structure. (C) RMSD between current and previous structures. Best 2000 of 10000 structures generated during the search phase with $r_{\text{ctol}} = 6$ Å are shown. Red lines indicate clustering thresholds. Resulting clusters are shown in Figure 4 (main text).

S3.5.2. Precision estimation: overfitting tests (dsDNA)

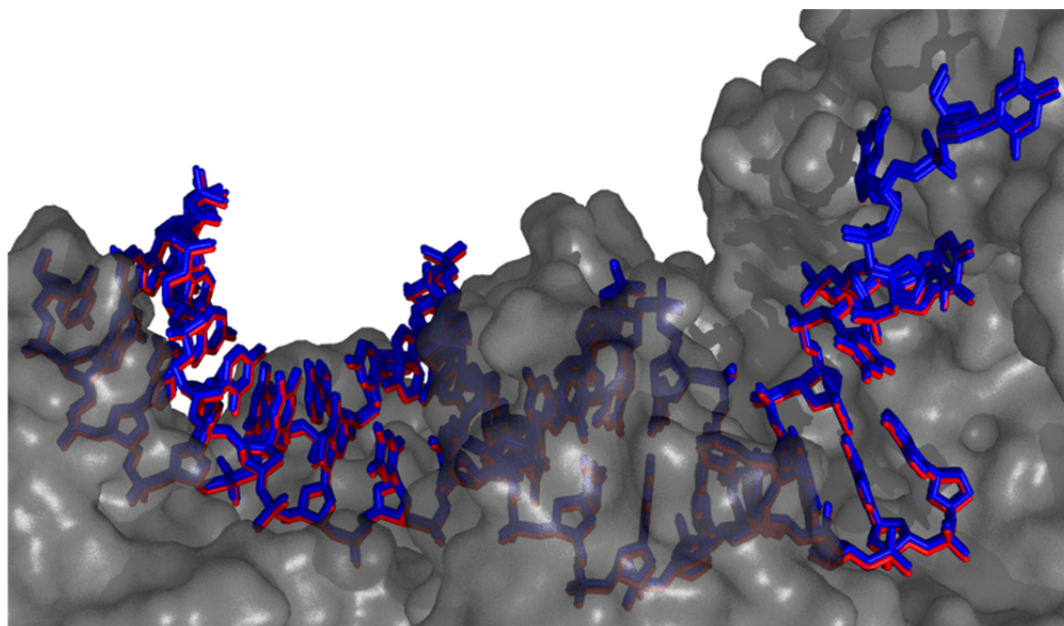


Figure S8. Overlay of 5 structures obtained using 80% of FRET restraints each (blue) and the best solution described in the main text (red).

For these tests we randomly discarded 20% of distances in each case and re-optimized the structure using the remaining 80% of FRET data. Figure S8 shows the overlay of 5 resulting structures with the best solution (red). Compared to the optimal solution, the mean RMSD of DNA phosphate atoms calculated for these 5 structures is 0.5 Å, the worst-case RMSD is 1.1 Å. This test is similar to commonly used cross-validation tests³¹ and clearly demonstrates that our results are not critically dependent on any single distance restraint.

S3.5.3. FRET-guided screening for the template overhang

The width of the states identified by seTCSPC and PDA on the single-stranded DNA is comparable to the donor-acceptor distribution on the double stranded part of the DNA (see Figure S5 and Tables S4 and S5). Moreover, structures that do not wind around the fingers domain are clearly inconsistent with the FRET data.

S4. Summary of experimental results

S4.1. Tables

Table S2. Major distances $\langle R_{DA} \rangle_E$ (amplitudes in brackets) resulting from PDA used for structural modeling of dsDNA, corresponding κ^2 -errors $\Delta R_{DA}(\kappa^2)$, fitted apparent widths of $\langle R_{DA} \rangle_E$ -distributions σ_{app} , and average amplitude weighted distances $\langle R \rangle_x$, with corresponding errors $err_{+/-}$, and the used experimental setup.

sample	$\langle R_{DA} \rangle_E$, Å (fraction)	$\Delta R_{DA}(\kappa^2)$	σ_{app}/R	$\langle R \rangle_x$, Å	err_- , Å	err_+ , Å	setup
RT(p66Q6C):dp(1)/dt	46(70%)	9.8%	7.8%	46	4.5	4.5	b
RT(p66Q6C):dp(10)/dt	46(58%)	9.9%	5.9%	46	4.5	4.5	b
RT(p66Q6C):dp(19)/dt	73(69%)	9.0%	8.5%	73	6.4	6.4	a
RT(p66T27C):dp(1)/dt	30(20%)	8.8%	9.0%	30	2.8	2.8	b
RT(p66T27C):dp(19)/dt	73(59%)	8.8%	8.0%	73	6.4	6.4	a
RT(p66E194C):dp(1)/dt	41(52%)	7.4%	7.2%	41	3.1	3.1	c
RT(p66E194C):dp(10)/dt	44(35%)	7.3%	7.9%	44	3.2	3.2	c
RT(p66E194C):dp(19)/dt	83(58%)	7.3%	9.2%	83	6.1	6.1	c
RT(p66K287C):dp(1)/dt	44(26%)	10.3%	9.0%	44	4.5	4.5	c
RT(p66K287C):dp(19)/dt	45(50%)	9.6%	9.1%	45	4.4	4.4	a
RT(p51Q6C):dp(1)/dt	68(45%) 76(36%)	9.3%	9.0%	71	10.5	9.8	b
RT(p51Q6C):dp(19)/dt	65(80%)	9.4%	7.3%	65	6.2	6.2	b
RT(p51K173C):dp(1)/dt	60(61%)	10.1%	7.1%	60	6.0	6	b
RT(p51K173C):dp(10)/dt	62(55%)	10.0%	8.5%	62	6.2	6.2	b
RT(p51K173C):dp(19)/dt	67(71%)	10.0%	9.0%	67	6.6	6.6	b
RT(p51E194C):dp(1)/dt	63(37%) 79(32%)	7.3%	9.0%	70.6	13.9	12.9	c
RT(p51E194C):dp(10)/dt	69(53%)	7.3%	9.0%	68	5.1	5.1	c
RT(p51E194C):dp(19)/dt	55(40%)	7.4%	5.0%	55	4.1	4.1	c
RT(p51K281C):dp(10)/dt	81(54%)	9.6%	9.0%	81	7.8	7.8	b
RT(p51K281C):dp(19)/dt	35(37%)	10.0%	6.2%	35	3.5	3.5	b

In two cases the major peak could not be clearly assigned. In these cases the weighted mean was used for modeling. The setups have the following green/red-detection efficiency ratios: setup a: $g_G/g_R=0.80$; setup b: $g_G/g_R=0.54$; setup c: $g_G/g_R=0.70$

Table S3. Distances $\langle R_{DA} \rangle_E$ resulting from PDA (major-state) used for structural modeling of ssDNA, corresponding uncertainties in ΔR_{DA} and fitted apparent widths of $\langle R_{DA} \rangle_E$ -distributions σ_{app} .

sample	$\langle R_{DA} \rangle_E$, Å	ΔR_{DA}	σ_{app}	setup
RT(p66Q6C):dp/dt(-6)	48(53%)	9.1%	10%	a
RT(p66Q6C):dp/dt(-15)	49(61%)	9.1%	10%	a
RT(p66T27C):dp/dt(-6)	34(7%) 42(4%)	8.8%	13%	a
RT(p66T27C):dp/dt(-15)	29(30%) 45(23%)	9.0%	7%	b
RT(p66E194C):dp/dt(-6)	44(34%)	7.2%	6%	c
RT(p66E194C):dp/dt(-15)	46(28%)	7.2%	7%	c
RT(p66K287C):dp/dt(-6)	36(34%)	10.0%	7%	a
RT(p66K287C):dp/dt(-15)	43(32%)	10.1%	13%	b
RT(p51Q6C):dp/dt(-6)	76(82%)	9.4%	11%	b
RT(p51Q6C):dp/dt(-15)	74(83%)	9.5%	7%	b
RT(p51K173C):dp/dt(-6)	62(63%)	10.0%	8%	b
RT(p51K173C):dp/dt(-15)	61(69%)	10%	9%	b
RT(p51E194C):dp/dt(-6)	68(53%)	7.2%	8%	c
RT(p51E194C):dp/dt(-15)	68(38%)	7.4%	7%	c
RT(p51K281C):dp/dt(-6)	76(76%)	10.0%	10%	a
RT(p51K281C):dp/dt(-15)	77(59%)	9.9%	9%	a

The setups have the following green/red-detection efficiency ratios: setup a: $g_G/g_R=0.80$; setup b: $g_G/g_R=0.54$; setup c: $g_G/g_R=0.70$

Table S4: seTCSPC distances of FRET-populations selected by the F_D/F_A -ratio and PDA-distances of the major state in the single stranded overhang

	R [Å]	σ_{DA} [Å]	x_D	F_D/F_A	s_v	s_h	χ^2_r
RT(p51Q6C):dp/dt(-6)	74.7	n.r.		all	0.1	0.1	2.4
RT(p51Q6C):p/t(-15)	74.3	n.r.	0	all	0.1	0.1	2.3
RT(p51K173C):dp/dt(-6)	62.3	n.r.		all	0.2	0.1	1.2
RT(p51K173C):p/t(-15)	62.5	n.r.		0.1-35	0.3	0.1	1.5
RT(p51E194C):dp/dt(-6)	69.6	n.r.	20	all	0.3	0.3	2.6
RT(p51E194C):p/t(-15)	75.2	n.r.		0.1-6	0.3	0.2	1.2
RT(p51K281C):dp/dt(-6)	70.2	n.r.	20	all	0.2	0.1	2
RT(p51K281C):p/t(-15)	72.1	n.r.		all	0.2	0.1	1.8
RT(p66Q6C):dp/dt(-6)	55.8	15		0.1-10	0.3	0.2	1.3
RT(p66Q6C):p/t(-15)	49.4	19.7		0.1-10	0.2	0.1	1.2
RT(p66T27C):dp/dt(-6)	43.3	18		0.1-6	0.5	0.3	1
RT(p66T27C):p/t(-15)	45.5	17		0.1-8	0.3	0.2	1.2
RT(p66E194C):dp/dt(-6)	58	11		0.1-12	0.7	0.3	1
RT(p66E194C):p/t(-15)	53.8	12		0.6-3.5	0.5	0.2	1.2
RT(p66K287C):dp/dt(-6)	37.2	11		0.1-6	0.2	0.2	1
RT(p66K287C):p/t(-15)	42.1	11	0.1	0.1-11	0.2	0.1	1.2

F_D/F_A , R , σ , D , F_D/F_A , s_v , s_h correspond to the donor/acceptor fluorescence intensity ratio, the donor/acceptor distance, the width of the fitted Gaussian donor/acceptor distance distribution, the donor fraction and the scatter amplitude in the vertical and horizontal channel, n.r. is short for not resolved

Table S5: seTCSPC of FRET-population selected by F_D/F_A -ratio located on the double-stranded DNA

	R [Å]	σ_{DA} [Å]	x_D [%]	F_D/F_A	s_v	s_h	χ^2_r
RT(p51 <u>E194C</u>):p(1)/t	75.3	n.r.		0.1-6	0.5	0.3	1.2
RT(p51 <u>E194C</u>):p(10)/t	73.8	n.r.		9-55	0.4	0.2	1.1
RT(p51 <u>E194C</u>):p(19)/t	62.2	6.6		9-74	0.3	0.2	1.3
RT(p51 <u>K173C</u>):p(1)/t	64.9	n.r.		0.1-100	0.2	0.1	1.6
RT(p51 <u>K173C</u>):p(10)/t	64.9	n.r.		0.1-40	0.3	0.1	1.4
RT(p66 <u>E194</u>):p(1)/t	50.2	11.7		0.1-6	0.6	0.3	1.7
RT(p66 <u>E194</u>):p(10)/t	52.5	19.9		0.1-16	0.4	0.2	2.3
RT(p66 <u>E194</u>):p(19)/t	75.2	n.r.	20.0	5-1200	0.5	0.2	3.4
RT(p66 <u>K287C</u>):p(1)/t	40.0	14.5		0.05-5	0.2	0.1	1.2
RT(p66 <u>K287C</u>):p(19)/t	31.4	15.1		0.05-7	0.1	0.1	1.4
RT(p66 <u>Q6C</u>):p(1)/t	45.9	11.2		0.1-5	0.2	0.1	1.3
RT(p66 <u>Q6C</u>):p(10)/t	45.8	10.3		0.1-10	0.3	0.1	1.1
RT(p66 <u>Q6C</u>):p(19)/t	77.1	15.3		0.2	0.3	0.1	2.2
RT(p66 <u>T27C</u>):p(1)/t	45.8	12.5		0.05-5	0.6	0.4	1.1
RT(p66 <u>T27C</u>):p(19)/t	83.0	n.r.	2.0	5-1000	0.3	0.1	1.6
RT(p51 <u>E194C</u>):p(1)/t	75.3	n.r.		0.1-6	0.5	0.3	1.2

F_D/F_A , R , σ , D , F_D/F_A , s_v , s_h , b_v , b_h correspond to the donor/acceptor fluorescence intensity ratio, the donor/acceptor distance, the width of the fitted Gaussian donor/acceptor distance distribution, the donor fraction, the scatter amplitude and background in the vertical and horizontal channel. n.r. is short for not resolved

Table S6: Fitting parameters of the time resolved donor-decays - lifetimes $\tau_{l,2}$ species amplitudes $x_{l,2}$ calculated quantum yield $\Phi(\tau)$, anisotropies (r_l , r_∞), rotational correlation times ρ_l , ρ_2

	x_l	τ_l [ns]	x_2	τ_2 [ns]	r_l	ρ_l [ns]	r_∞	ρ_2 [ns]
RT(p66Q6C):dp/dt	0.82	3.72	0.18	1.13	0.14	0.36	0.24	17.9
RT(p66T27C):dp/dt	0.82	3.63	0.18	0.68	0.18	0.21	0.20	11.4
RT(p66E194C):dp/dt	1.00	3.87			0.25	0.14	0.13	5.5
RT(p66K287C):dp/dt	0.80	3.75	0.20	0.97	0.11	0.30	0.27	19.6
RT(p51Q6C):dp/dt	0.84	3.71	0.16	0.95	0.14	0.34	0.24	14.2
RT(p51K173C):dp/dt	0.83	3.73	0.17	1.03	0.12	0.31	0.26	15.3
RT(p51E194C):dp/dt	0.88	3.68	0.12	0.92	0.25	0.10	0.13	5.4
RT(p51K281C):dp/dt	0.84	3.91	0.16	1.32	0.13	0.28	0.25	16.5

The fundamental anisotropy r_0 was fixed to 0.38.

Table S7: Average fluorescence lifetimes $\langle \tau_{D(0)} \rangle_f$, quantum yields $\Phi_{FD(0)}$, polynomial coefficients c_i , and estimated widths of donor/acceptor distribution σ used for the calculation of the static FRET-lines through Figure S10.1-8. In all measurements $\Phi_{FA} = 0.32$.

	$\langle \tau_{D(0)} \rangle_f$	$\Phi_{FD(0)}$	c_0	c_1	c_2	c_3	σ [Å]
RT(p66Q6C):dp/dt	3.558	0.635	-0.0560	0.6002	0.3064	-0.0534	6
RT(p66T27C):dp/dt	3.513	0.605	-0.0599	0.6471	0.2925	-0.0534	6
RT(p66E194C):dp/dt	3.870	0.755	-0.0425	0.4862	0.2927	-0.0406	12
RT(p66K287C):dp/dt	3.581	0.623	-0.0606	0.6392	0.2927	-0.0524	6
RT(p51Q6C):dp/dt	3.582	0.638	-0.0560	0.6028	0.3006	-0.0518	6
RT(p51K173C):dp/dt	3.585	0.638	-0.0564	0.6042	0.3012	-0.0521	6
RT(p51E194C):dp/dt	3.589	0.653	-0.0518	0.5695	0.3064	-0.0509	12
RT(p51K281C):dp/dt	3.754	0.682	-0.0545	0.5791	0.2932	-0.0473	6

S4.2. MFD plots

Assignment of states and color scheme in all MFD plots. The single molecule data was fitted with PDA as explained in section 3.3.2. The different expected states (P-P, P-E, D-E) are colored as follows. The major FRET-population is assigned to the P-P state and colored red. The P-E state is the peak closer in distance to the P-P state. If the P-P state and the D-E state could not be assigned by the proximity to the P-E state, the higher populated state is assumed to be the P-P state. The P-P state is colored orange. The remaining narrow FRET-peak is assigned to the D-E and colored blue. In the PDA-analysis donor only is colored green. In some cases an additional usually broad peak is necessary to fit the data which is most likely present due to impurities and bleaches molecules.

General description of all MFD figures. Measuring 36 FRET pairs, we present 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms of the lifetime- and anisotropy-distribution are displayed as projections. The theoretical relationship between F_D/F_A and $\tau_{D(A)}$ is given by the static FRET line (red line) using equation S6 with all parameters compiled in Table S7. The solid red and orange lines in the r_D - $\tau_{D(A)}$ diagram are given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$ $r_0 = 0.374$ whereas ρ is the rotational correlation time and r_0 the fundamental anisotropy. The PDA analysis of selected bursts, χ_r^2 and the weighted residuals and of the fits are shown in the upper right panels (The PDA-analysis is not the simple projection of the 2D-histogram. In particular for high-FRET states they species fractions may differ.). The gray area in the PDA- F_D/F_A histogram corresponds to the experimental data and the colored lines to the fitted states. The protein may be in three distinct states (P-P, P-E, and D-E), whereas P-E is generally the major state. Acceptor bleaching and impurities were fitted with Gaussians of free width; the relevant states were fitted with Gaussians with global relative widths of the states σ/R_i . The donor fraction is denoted by x_D . The relative amplitudes are given in brackets.

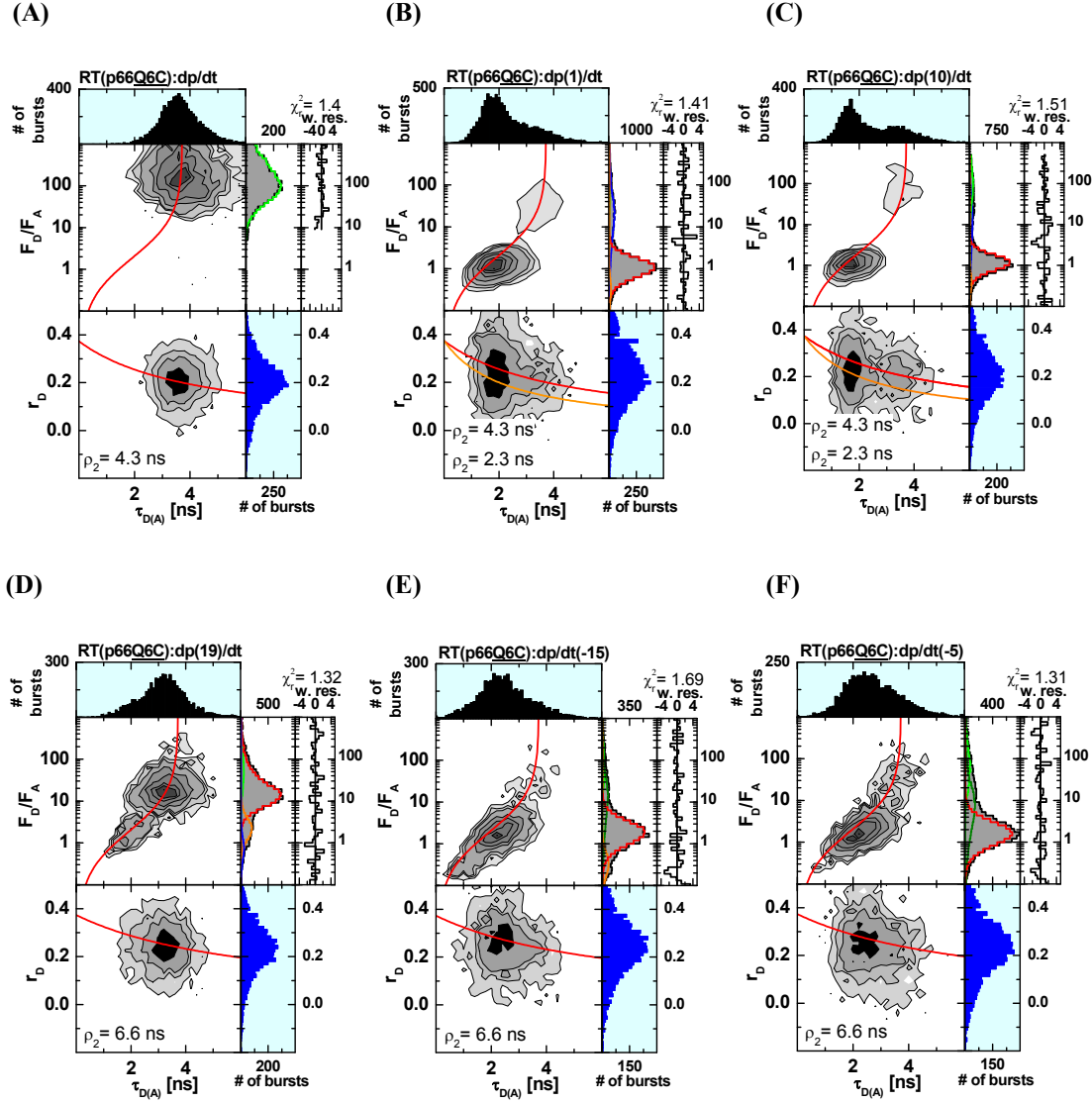


Figure S10.1: (A) RT(p66Q6C):p/t donor only sample (B) RT(p66Q6C):dp(1)/dt 46 Å(70%) (red) 37 Å(3%) (orange) $\sigma_{app}/R=7.8\%$, 76 Å(23%) $\sigma=20$ Å (olive), $x_D=9\%$ (green) (C) RT(p66Q6C):dp(10)/dt 46 Å(58%) (red) 38 Å(3%) (orange) $\sigma_{app}/R=5.9\%$, 66 Å(15%) $\sigma=19$ Å (olive), $x_D=23\%$ (green) (D) RT(p66Q6C):dp(19)/dt 73 Å(69%) (red) 51 Å(14%) (orange) 42 Å(4%) (blue) $\sigma_{app}/R=8.5\%$, $x_D=14\%$ (green) (E) RT(p66Q6C):dp/dt(-15) 49 Å(61%) (red) 36 Å(9%) (orange) $\sigma_{app}/R=10\%$, 68 Å(19%) (olive), $x_D=11\%$ (green) (F) RT(p66Q6C):dp/dt(-6) 48 Å(53%) (red) $\sigma_{app}/R=10\%$, $x_D=20\%$ (green).

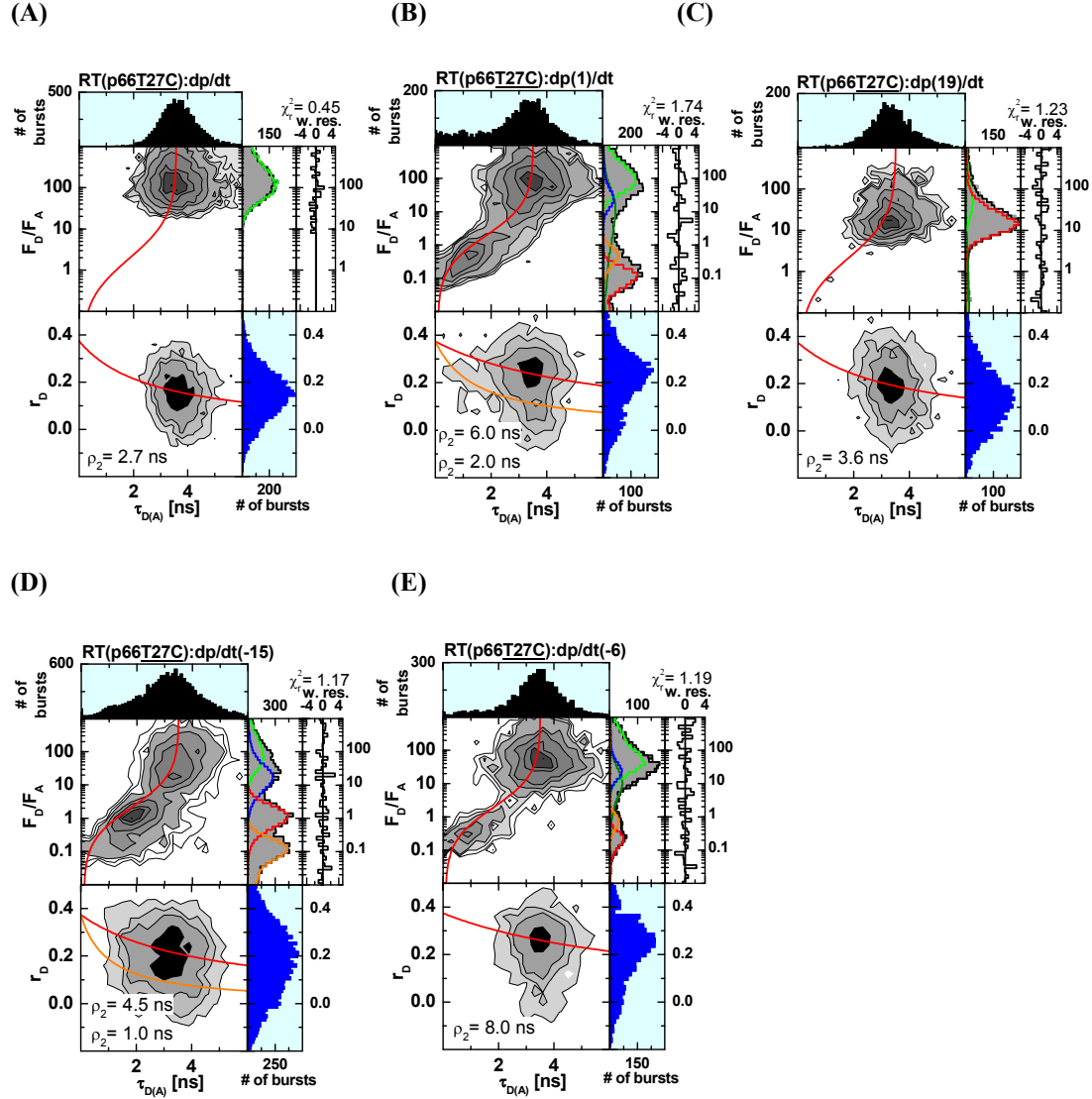


Figure S10.2: (A) RT(p66T27C):dp/dt donor only sample (B) RT(p66T27C):dp(1)/dt 30 Å (20%) (red) 39 Å (7%) (orange) 79 Å (6%) (blue) $\sigma_{app}/R=9.0\%$, 57 Å (20%) $\sigma=25$ Å (olive), $x_D=48\%$ (green) (C) RT(p66T27C):dp(19)/dt 73 Å (59%) (red) $\sigma_{app}/R=8.0\%$, 27 Å (16%) $\sigma=20$ Å $x_D=25\%$ (green) (D) RT(p66T27C):dp/dt(-15) 29 Å (30%) (red) 45 Å (23%) (orange) $\sigma_{app}/R=13\%$, 74 Å (20%) $\sigma=12$ Å (olive), $x_D=27\%$ (green) (E) RT(p66T27C):dp/dt(-6) 34 Å (7%) (red) 42 Å (4%) (orange) $\sigma_{app}/R=7\%$, 67 Å (14%) $\sigma=13$ Å (olive), donor only fraction 75% (green).

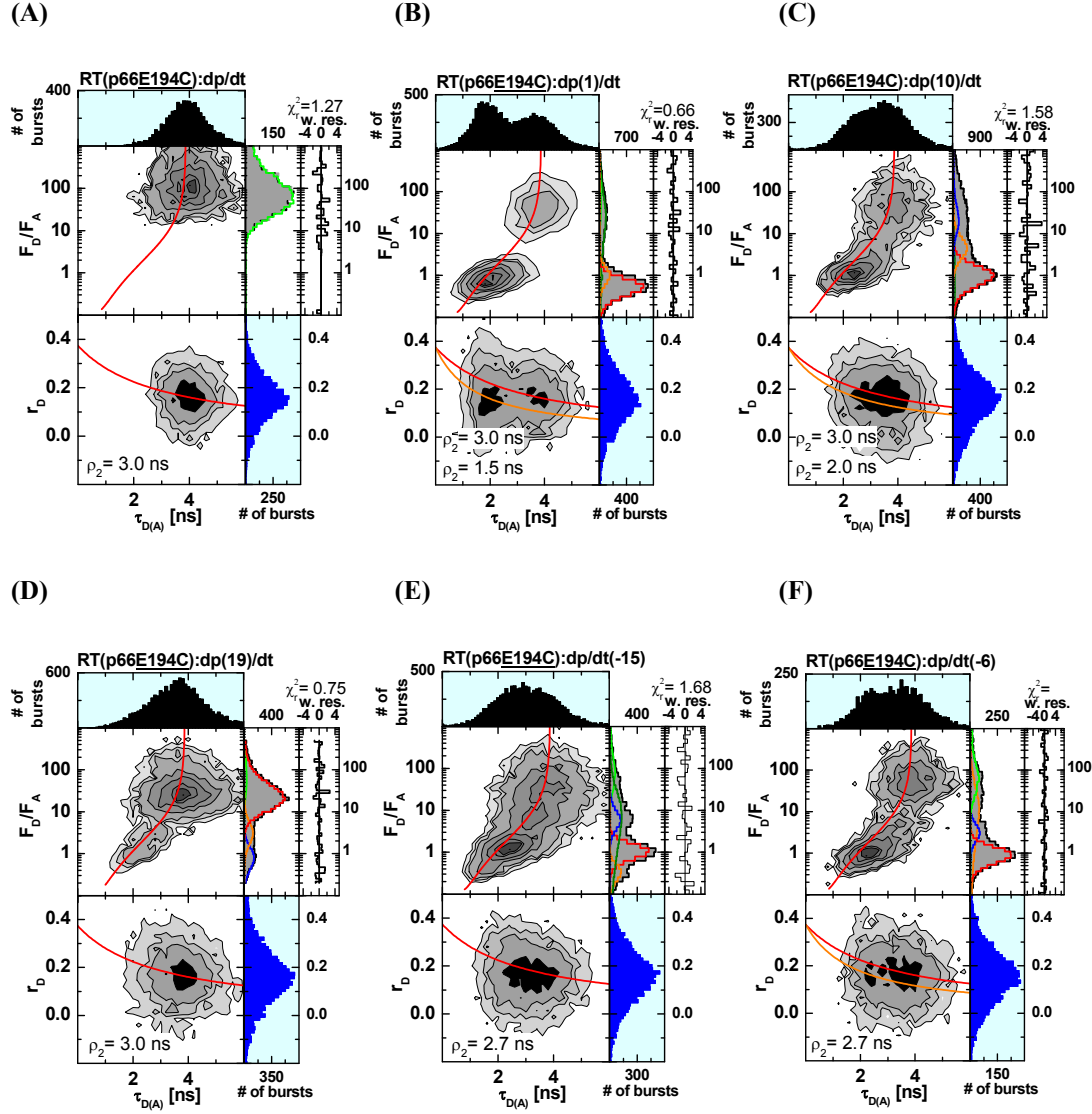


Figure S10.3: (A) RT(p66E194C):dp/dt – donor only sample, (B) RT(p66E194C):dp(1)/dt 41 Å(52%) (red) 46 Å(6%) (orange) $\sigma_{app}/R=7.2\%$, 84 Å $\sigma=39$ Å (olive), $x_D=7.7\%$ (green), red dirt 30 Å (3%) (C) RT(p66E194C):dp(10)/dt 44 Å(35%) (red) 58 Å(13%) (orange) 79 Å (9%) $\sigma_{app}/R=7.9\%$, 35 Å(7%) $\sigma=22$ Å (olive), $x_D=36\%$ (D) RT(p66E194C):dp(19)/dt 83 Å(58%) (red) 55 Å(11%) (orange) 43 Å(12%) (blue) $\sigma_{app}/R=9.2\%$, $x_D=20\%$ (E) RT(p66E194C):dp/dt(-15) 46 Å(28%) (red) 61 Å(12%) (blue) 38 Å(8%) (orange) $\sigma/R=6\%$, 60 Å(35%) $\sigma=24$ Å, $x_D=17\%$ (F) RT(p66E194C):dp/dt(-6) 44 Å(34%) (red) 55 Å(8%) (orange) $\sigma_{app}/R=7\%$, 64 Å(28%) $\sigma=21$ Å, $x_D=30\%$

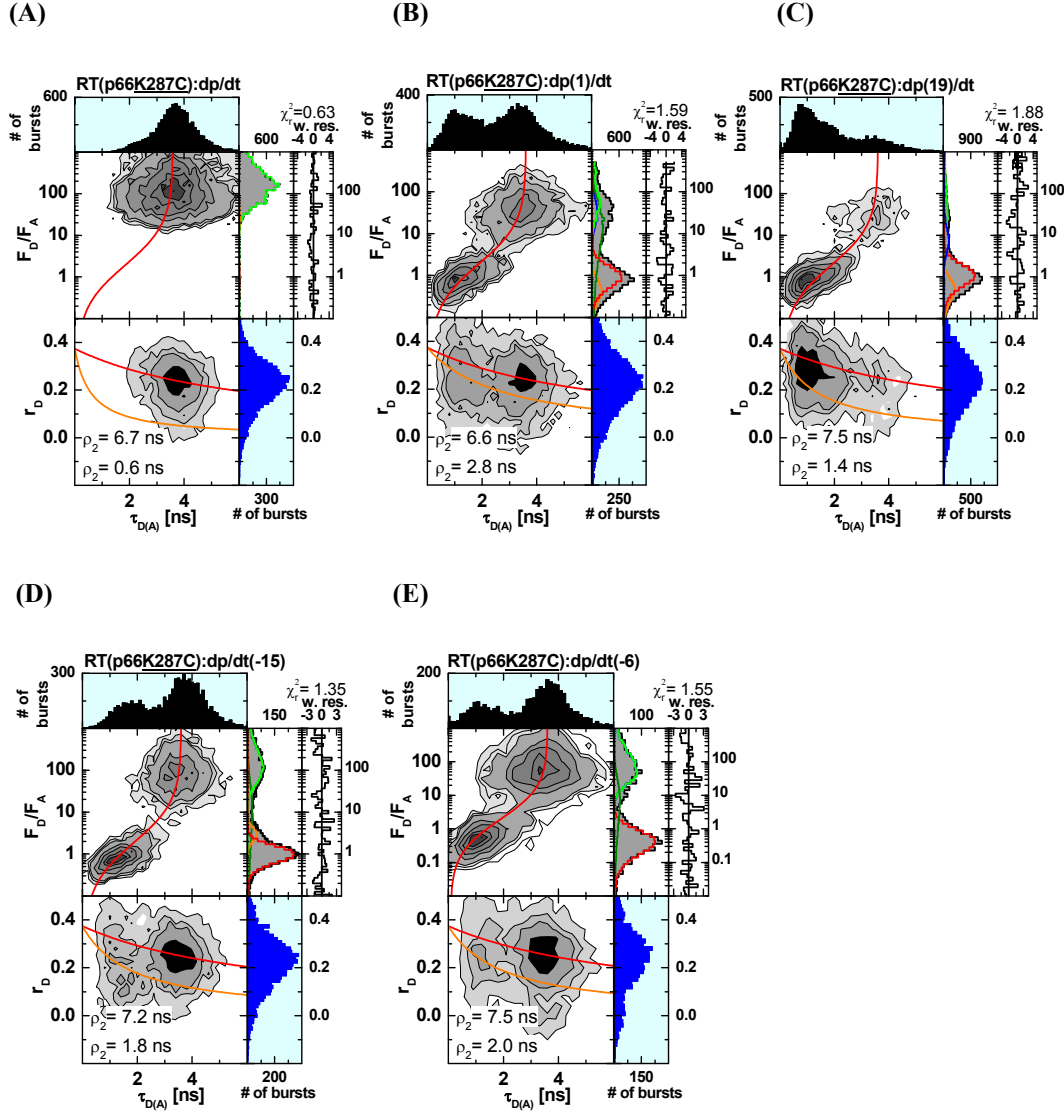


Figure S10.4: (A) RT(p66K287C):dp/dt donor only sample (B) RT(p66K287C):dp(1)/dt 44 Å(26%) (red) 38 Å(9%) (orange) 87 Å(8%) $\sigma_{app}/R=9.0\%$, 67 Å(32%) $\sigma=25$ Å (olive), $x_D=24\%$ (green) (C) RT(p66K287C):dp(19)/dt 45 Å(50%) (red) 39 Å(18%) (orange) 64 Å(9%) (blue) $\sigma_{app}/R=9.1\%$, $x_D=23\%$ (green) (D) RT(p66K287C):dp/dt(-15) 43 Å(32%) (red) 51 Å(5%) (orange) $\sigma_{app}/R=7\%$, 57 Å(15%) $\sigma=23$ Å (olive), $x_D=48\%$ (green) (E) RT(p66K287C):dp/dt(-6) 36 Å(34%) (red) $\sigma_{app}/R=13\%$, 51 Å(15%) $\sigma=23$ Å (olive), $x_D=51\%$ (green)

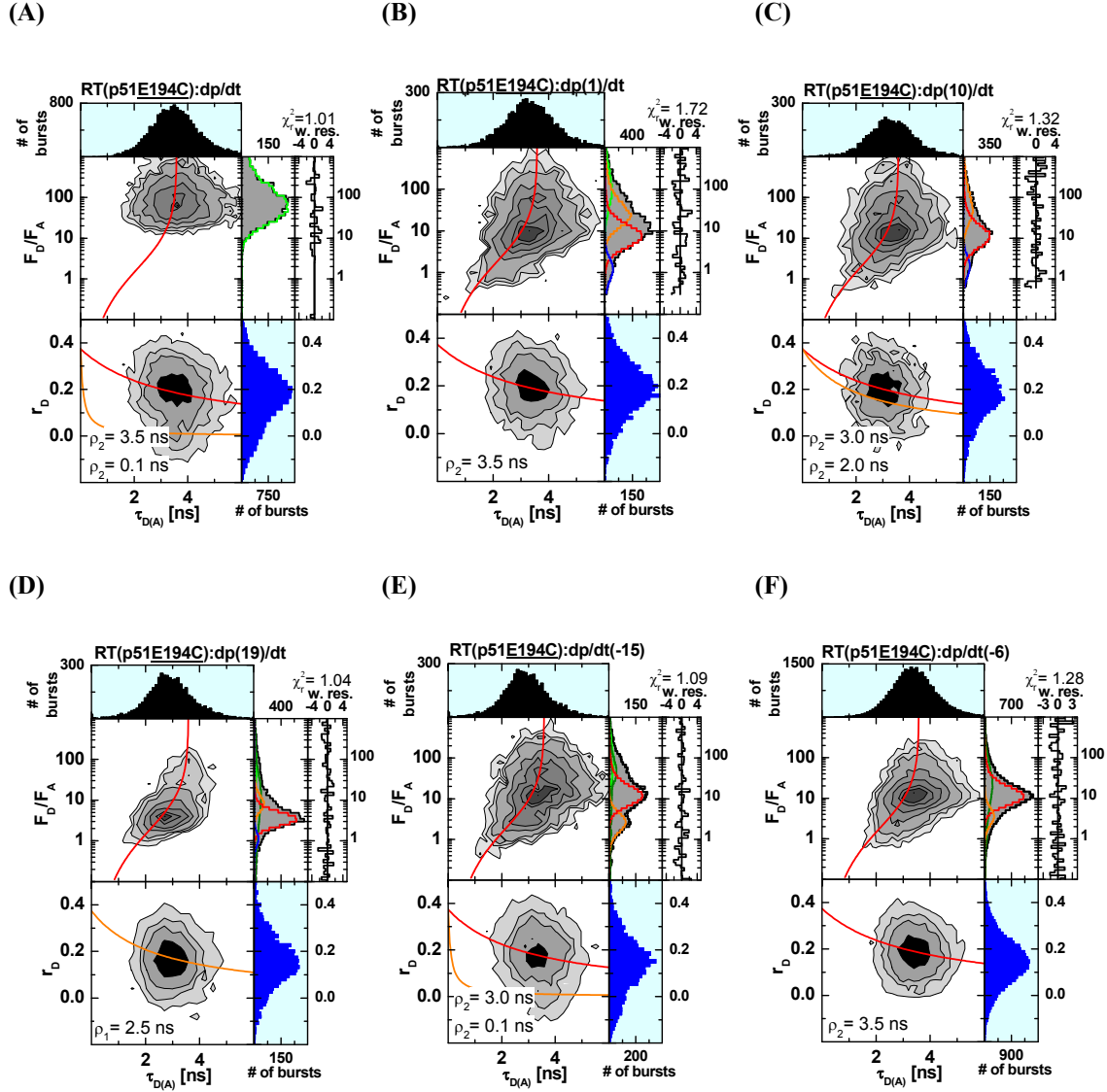


Figure S10.5: (A) RT(p51E194C):dp/dt donor only sample (B) RT(p51E194C):dp(1)/dt 63 Å(37%) (red) 79 Å(32%) (orange) 46 Å(8%) (blue) $\sigma_{app}/R=9.0\%$, $x_D=24\%$ (green) (C) RT(p51E194C):dp(10)/dt 68 Å(53%) (red) 98 Å(36%) (orange) 50 Å (11%) (blue) $\sigma_{app}/R=9.0\%$ (D) RT(p51E194C):dp(19)/dt 55 Å(40%) (red) 64 Å(11%) (orange) 45 Å(4%) (blue) $\sigma/R=5.0\%$, 64 Å(32%) $\sigma=24$ Å (olive), $x_D=12\%$ (green) (E) RT(p51E194C):dp/dt(-15) 68 Å(38%) 52 Å (16%) $\sigma_{app}/R=8\%$, 65 Å(15%), $x_D=31\%$ (F) RT(p51E194C):dp/dt(-6) 68 Å(53%) (red) 55 Å(11%) (orange) 45 Å(2%) (blue) $\sigma_{app}/R=7\%$, 75 Å (29%) $\sigma=21$ Å (olive), $x_D=4\%$

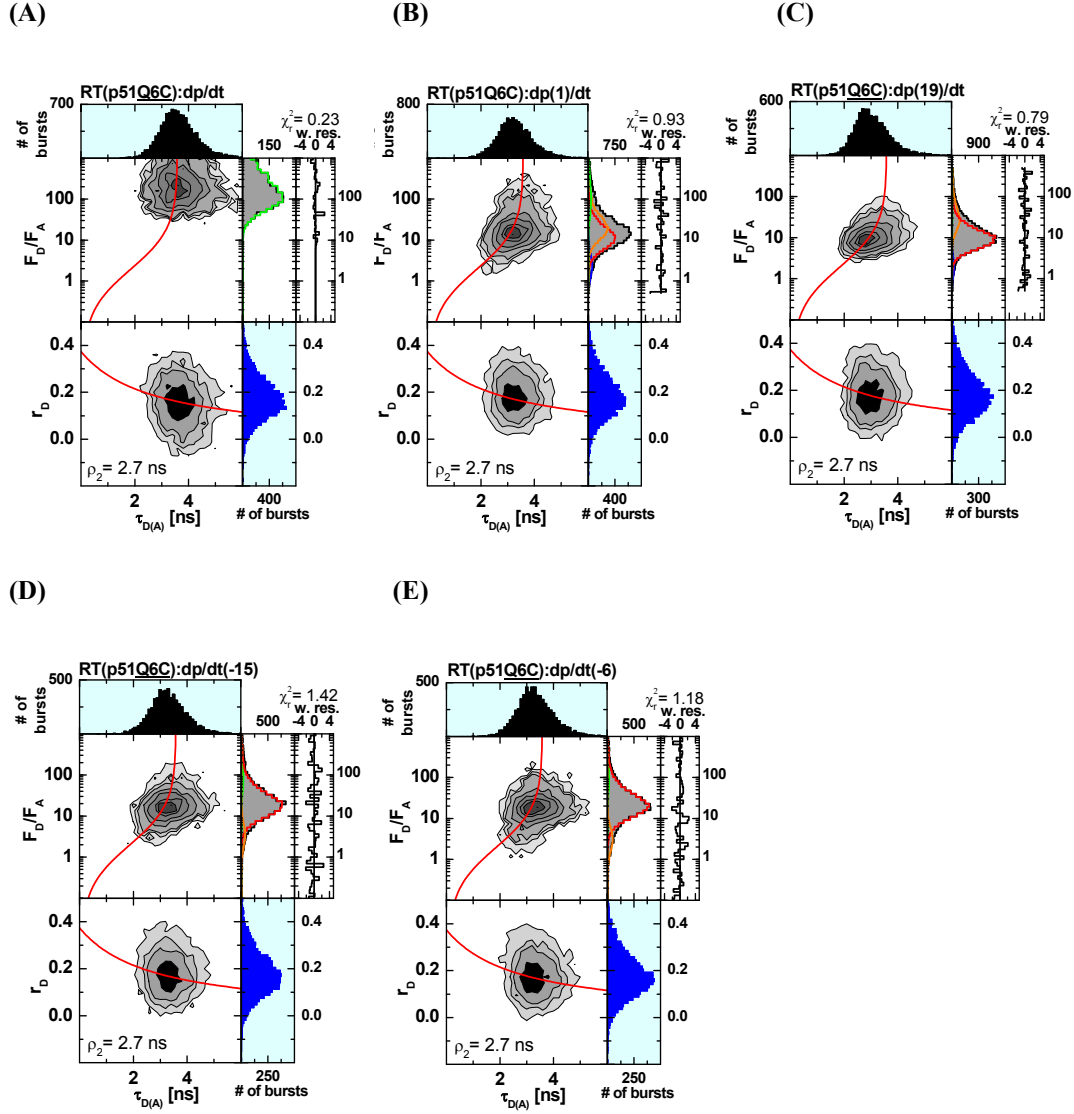


Figure S10.6: (A) RT(p51Q6C):dp/dt donor only sample (B) RT(p51Q6C):dp(1)/dt 68 Å(45%) (red) 76 Å(36%) (orange) 51 Å(6%) (blue) $\sigma_{app}/R=9.0\%$ 65 Å(26%) $\sigma=11$ Å (olive), $x_D=15\%$ (C) RT(p51Q6C):dp(19)/dt 65 Å(80%) (red) 83 Å(16%) (orange) 49 Å(3%) (blue) $\sigma_{app}/R=7.3\%$, $x_D=3\%$ (green) (D) RT(p51Q6C):dp/dt(-15) 74A(83%) (red) 53 Å(5%) (orange) $\sigma_{app}/R=11\%$, $x_D=12\%$ (E) RT(p51Q6C):dp/dt(-6) 76A(82%) 54 Å(9%) $\sigma_{app}/R=11\%$, $x_D=9\%$

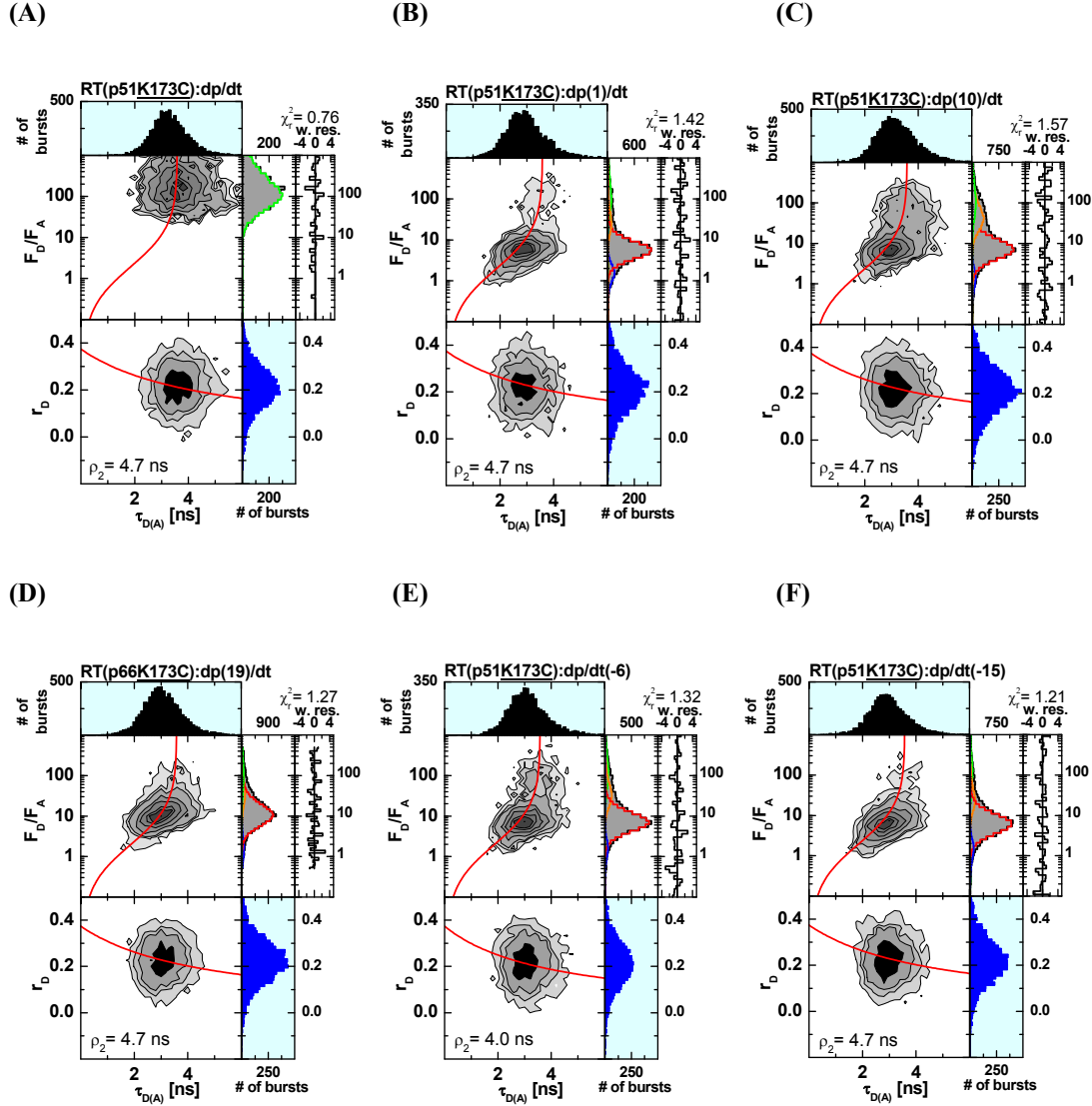


Figure S10.7: (A) RT(p51K173C):dp/dt donor only sample (B) RT(p51K173C):dp(1)/dt 60 Å(61%) (red) 79 Å (7%) (orange) 48 Å(9%) (blue) $\sigma_{app}/R=7.1\%$, $x_D=23\%$ (C) RT(p51K173C):dp(10)/dt 62 Å(55%) (red) 88 Å(23%) (orange) 47 Å(4%) (blue) $\sigma_{app}/R=8.5\%$, $x_D=18\%$ (D) RT(p51K173C):dp(19)/dt 67 Å(71%) (red) 86 Å(10%) (orange) 48 Å(5%) (blue) $\sigma_{app}/R=9\%$, $x_D=14\%$ (E) RT(p51K173C):dp/dt(-6) 62 Å(63%) (red) 83 Å(14%) (orange) 48 Å(4%) (blue) $\sigma_{app}/R=8\%$, $x_D=19\%$ (green) (F) RT(p51K173C):dp/dt(-15) 61 Å(69%) 78 Å(11%) 47 Å(6%) $\sigma_{app}/R=9\%$, $x_D=15\%$

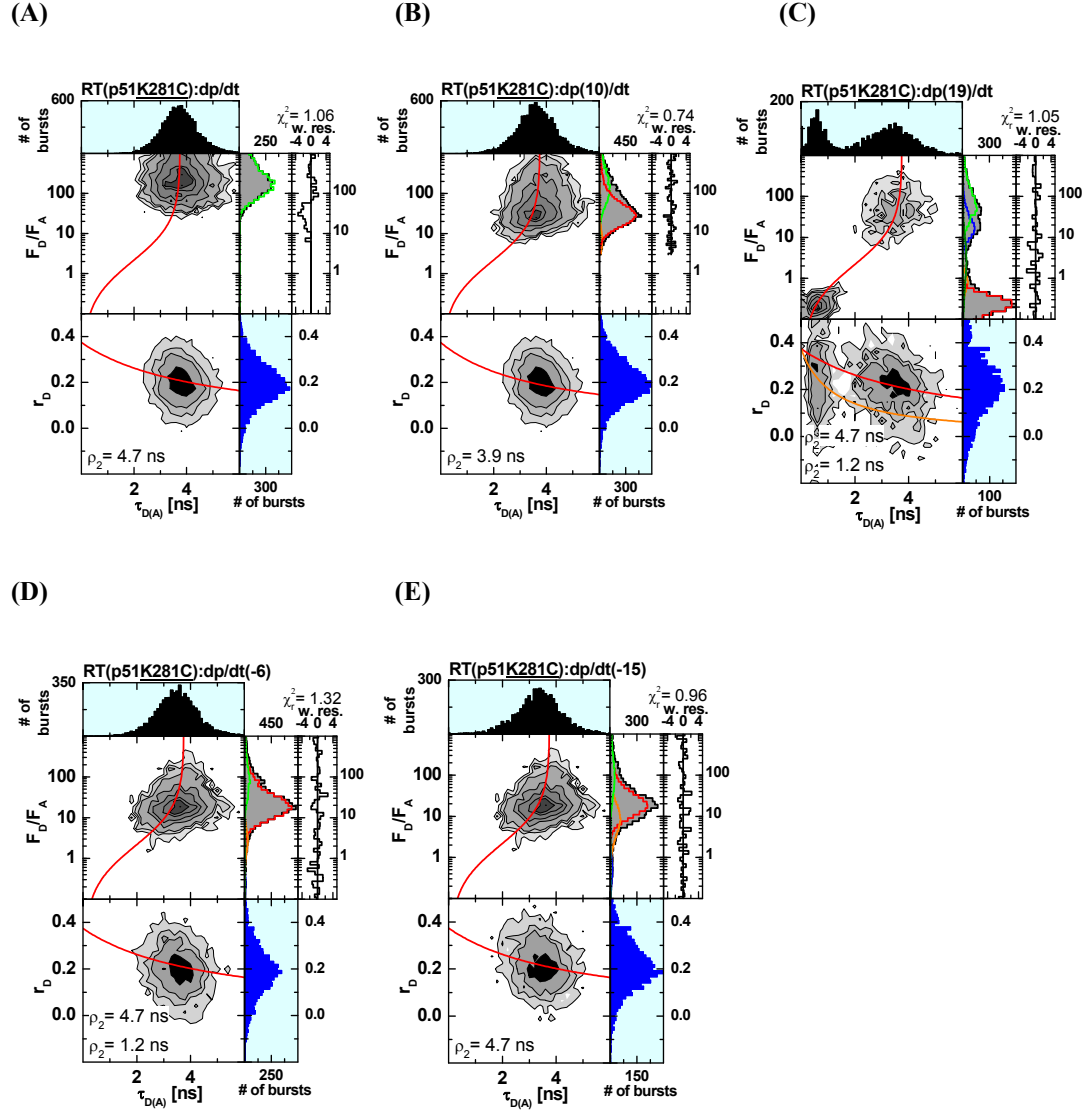


Figure S10.8: (A) RT(p51K281C):dp/dt donor only sample (B) RT(p51K281C):dp(10)/dt 81 Å(54%) (red) 62 Å(3%) (orange) $\sigma_{app}/R=9.0\%$, $x_D=43\%$ (green) (C) RT(p51K281C):dp(19)/dt 19bp 35 Å(37%) (red) 43 Å (3%) (orange) 74 Å (12%) (blue) $\sigma_{app}/R=6\%$, 51 Å(36%) $\sigma=13$ Å (olive) , $x_D=22\%$ (green) (D) RT(p51K281C):dp/dt(-6) 76 Å (76%) (red) 52 Å (4%) (orange) 34 Å (1%) (blue) $\sigma_{app}/R=10\%$, $x_D=20\%$ (green) (E) RT(p51K281C):dp/dt(-15) 77 Å(59%) (red) 64 Å(11%) (orange) 39 Å(3%) (blue) $\sigma_{app}/R=9\%$, $x_D=28\%$ (green)

S4.3. seTCSPC analysis of fluorescence lifetime and anisotropy

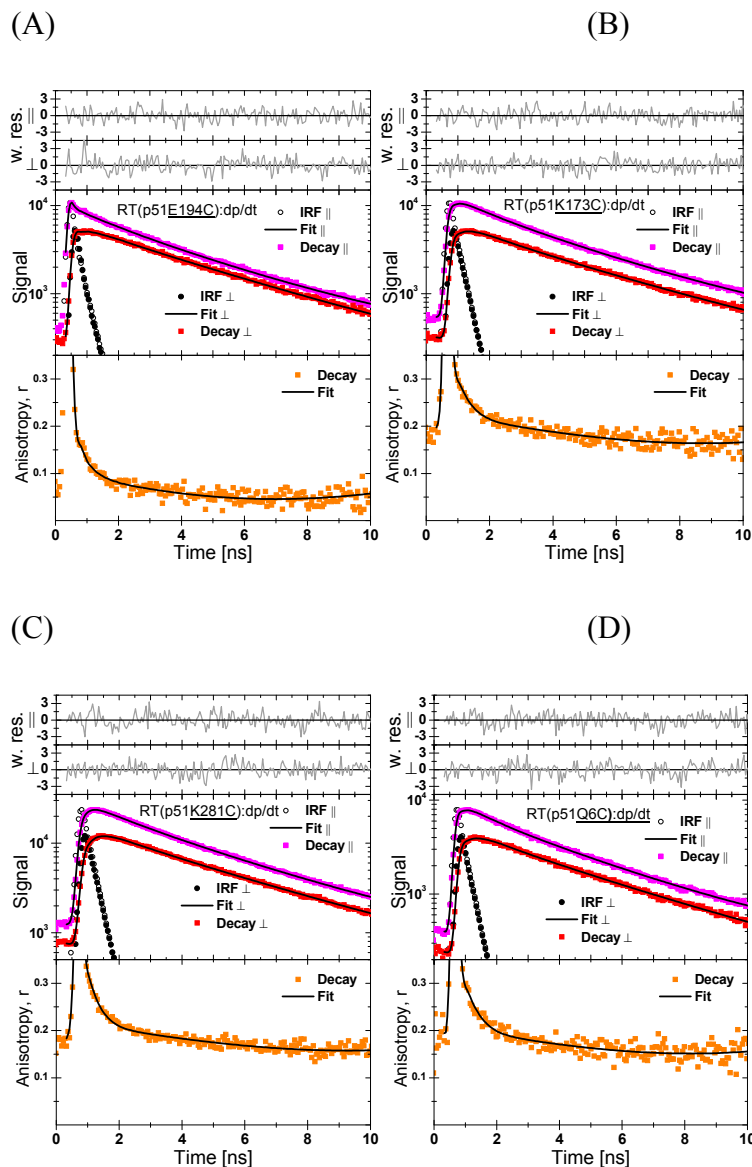
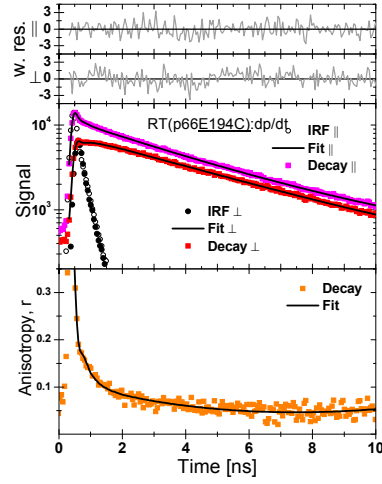
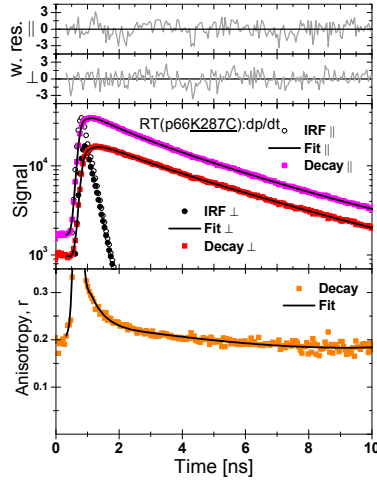


Figure S11.1: Donor only decays, $r_0=0.38$ (fixed): **(A)** RT(p51E194C):dp/dt - vv scatter 7%, bg=248, vp scatter 9% bg=126, lifetimes=3.7ns(88%), 0.9ns(12%), anisotropy $r_1=0.25$ $r_\infty=0.13$ $\rho_1=0.10$ $\rho_2=5.45$ **(B)** RT(p51K173C):dp/dt vv scatter 1%, bg=307, vp scatter 1% bg=130, lifetimes=3.7ns(83%), 1.0ns(17%), anisotropy $r_1=0.12$ $r_\infty=0.26$ $\rho_1=0.31$ $\rho_2=15.3$ **(C)** RT(p51K281C):dp/dt vv scatter 1%, bg=567, vp scatter 1.3% bg=232, lifetimes=3.9ns(84%), 1.3ns(16%), anisotropy $r_1=0.13$ $r_\infty=0.25$ $\rho_1=0.28$ $\rho_2=16.5$ **(D)** RT(p51Q6C):dp/dt vv scatter 0.5%, bg=229, vp scatter 0.8% bg=98, lifetimes=3.7ns(84%), 1.0ns(16%), anisotropy $r_1=0.14$ $r_\infty=0.24$ $\theta_1=0.34$ $\theta_2=14.2$

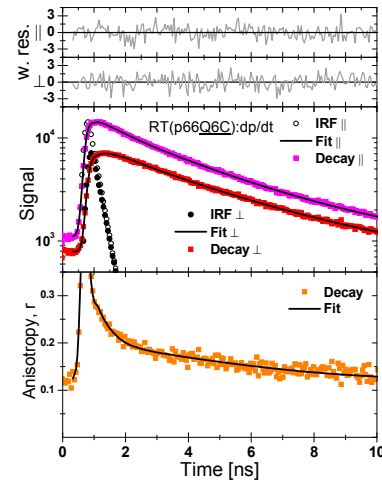
(A)



(B)



(C)



(D)

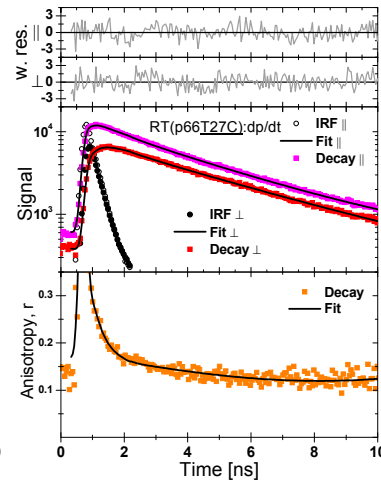


Figure S11.2: Donor only decays, $r_0=0.38$ (fixed): **(A)** RT(p66E194C):dp/dt - vv scatter 4.6%, bg=352, vp scatter 6.2% bg=197, lifetime=3.9ns(100%), anisotropy $r_1=0.25$ $r_\infty=0.13$ $\rho_1=0.14$ $\rho_2=5.53$; **(B)** RT(p66K287C):dp/dt vv scatter 2.3%, bg=888, vp scatter 3.2% bg=384, lifetimes=3.75ns(80%), 0.97ns(20%), anisotropy $r_1=0.11$ $r_\infty=0.27$ $\rho_1=0.30$ $\rho_2=19.6$; **(C)** RT(p66Q6C):dp/dt vv scatter 1.0%, bg=799, vp scatter 1.4% bg=557, lifetimes=3.72ns(82%), 1.13ns(18%), anisotropy $r_1=0.14$ $r_\infty=0.24$ $\rho_1=0.36$ $\rho_2=17.9$; **(D)** RT(p66T27C):dp/dt vv scatter 0.7%, bg=349, vp scatter 1.0% bg=164, lifetimes=3.63ns(82%), 0.68ns(18%), anisotropy $r_1=0.18$ $r_\infty=0.20$ $\rho_1=0.21$ $\rho_2=11.4$;

S4.4. Obtained κ^2 distributions

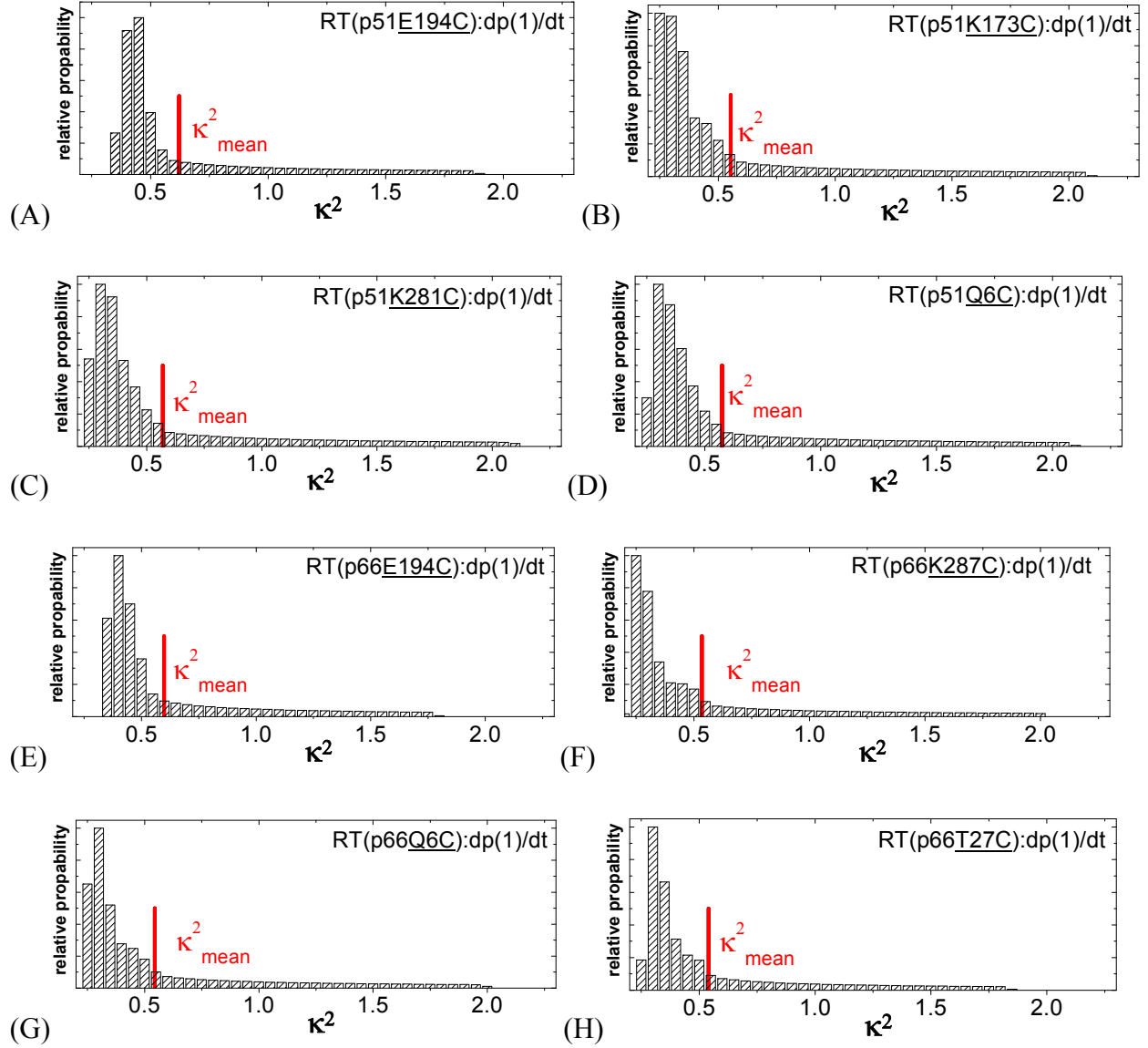


Figure S12.1: Possible κ^2 values for sample: (A) RT(p51E194C):dp(1)/dt $\langle \kappa^2 \rangle = 0.62$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% (B) RT(p51K173C):dp(1)/dt $\langle \kappa^2 \rangle = 0.55$ uncertainty in $\langle R_{DA} \rangle_E$ 9.7% (C) RT(p51K281C):dp(1)/dt $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 9.7% (D) RT(p51Q6C):dp(1)/dt $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 9.4% (E) RT(p66E194C):dp(1)/dt $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 7.3% (F) RT(p66K287C):dp(1)/dt $\langle \kappa^2 \rangle = 0.53$ uncertainty in $\langle R_{DA} \rangle_E$ 10.3% (G) RT(p66Q6C):dp(1)/dt $\langle \kappa^2 \rangle = 0.55$ uncertainty in $\langle R_{DA} \rangle_E$ 9.8% (H) RT(p66T27C):dp(1)/dt $\langle \kappa^2 \rangle = 0.54$ uncertainty in $\langle R_{DA} \rangle_E$ 9.0%

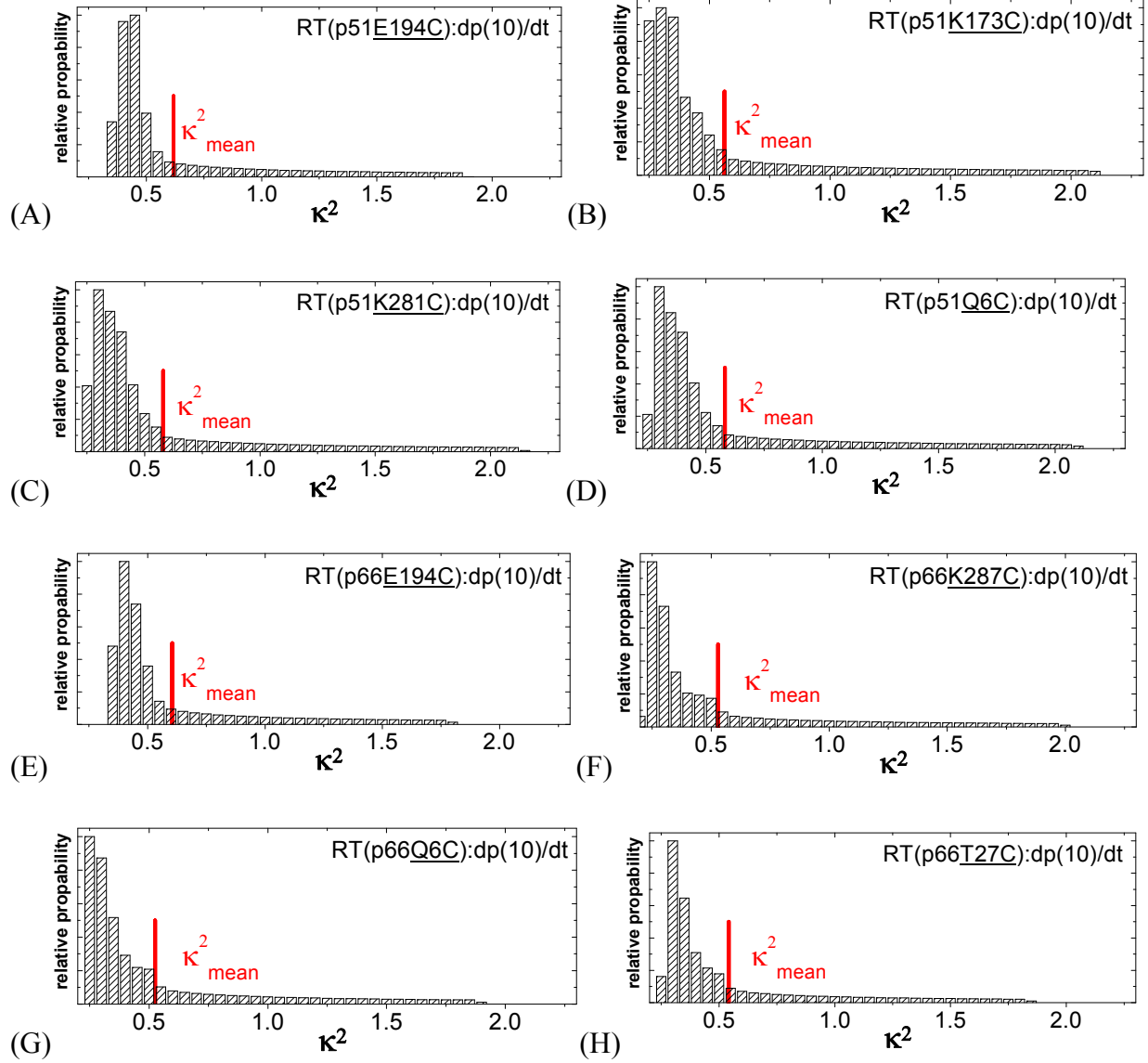


Figure S12.2: Possible κ^2 values for sample: (A) RT(p51E194C):dp(10)/dt $\langle \kappa^2 \rangle = 0.62$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% (B) RT(p51K173C):dp(10)/dt $\langle \kappa^2 \rangle = 0.56$ uncertainty in $\langle R_{DA} \rangle_E$ 9.9% (C) RT(p51K281C):dp(10)/dt $\langle \kappa^2 \rangle = 0.58$ uncertainty in $\langle R_{DA} \rangle_E$ 9.5% (D) RT(p51Q6C):dp(10)/dt $\langle \kappa^2 \rangle = 0.58$ uncertainty in $\langle R_{DA} \rangle_E$ 9.3% (E) RT(p66E194C):dp(10)/dt $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 7.3% (F) RT(p66K287C):dp(10)/dt $\langle \kappa^2 \rangle = 0.53$ uncertainty in $\langle R_{DA} \rangle_E$ 10.3% (G) RT(p66Q6C):dp(10)/dt $\langle \kappa^2 \rangle = 0.53$ uncertainty in $\langle R_{DA} \rangle_E$ 9.8% (H) RT(p66T27C):dp(10)/dt $\langle \kappa^2 \rangle = 0.54$ uncertainty in $\langle R_{DA} \rangle_E$ 9.0%

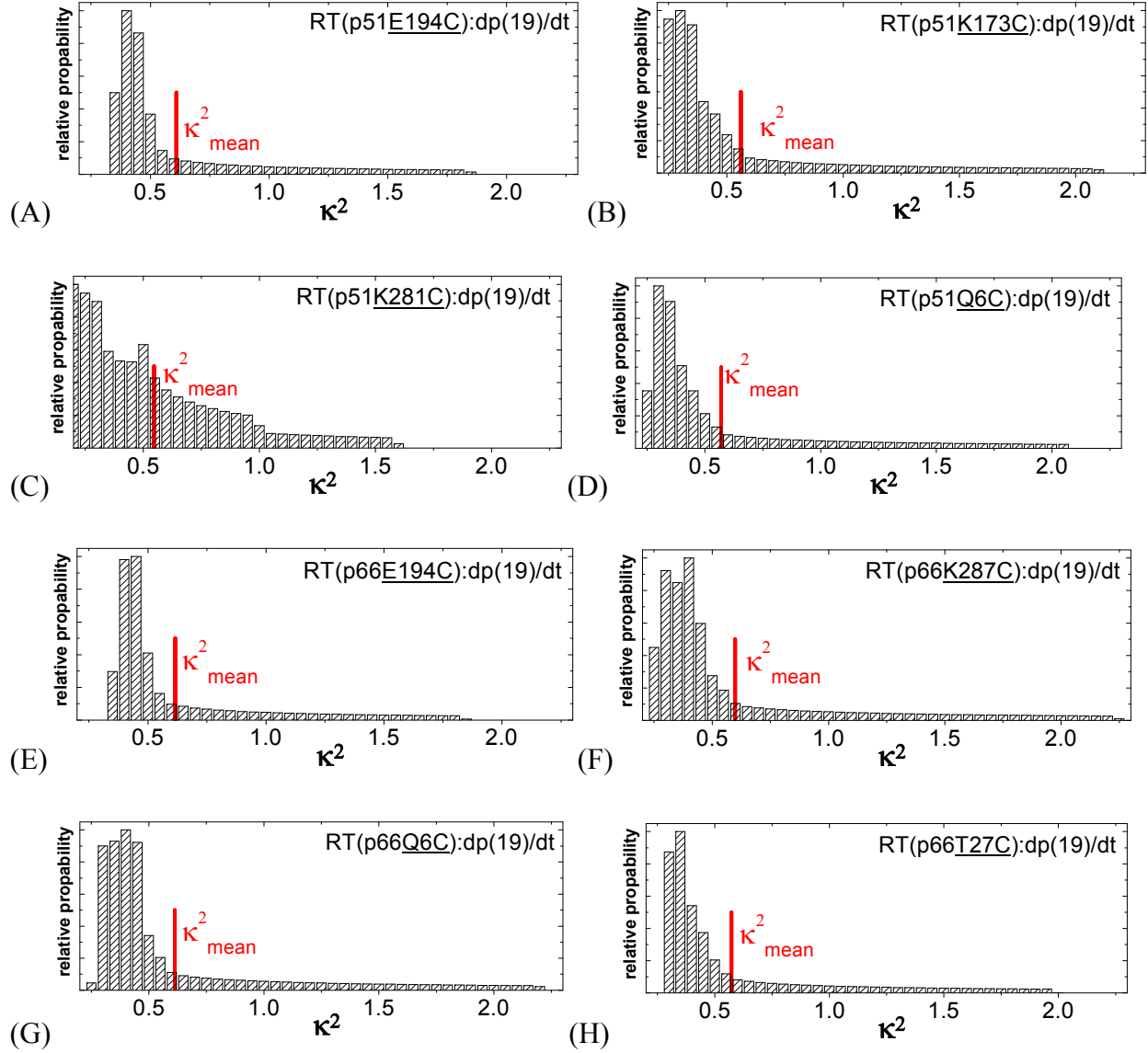


Figure S12.3: Possible κ^2 values for sample **(A)** RT(p51E194C):dp(19)/dt $\langle \kappa^2 \rangle = 0.61$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% **(B)** RT(p51K173C):dp(19)/dt $\langle \kappa^2 \rangle = 0.56$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% **(C)** RT(p51K281C):dp(19)/dt $\langle \kappa^2 \rangle = 0.55$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% **(D)** RT(p51Q6C):dp(19)/dt $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 9.4% **(E)** RT(p66E194C):dp(19)/dt $\langle \kappa^2 \rangle = 0.62$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% **(F)** RT(p66K287C):dp(19)/dt $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 9.6% **(G)** RT(p66Q6C):dp(19)/dt $\langle \kappa^2 \rangle = 0.61$ uncertainty in $\langle R_{DA} \rangle_E$ 8.9% **(H)** RT(p66T27C):dp(19)/dt $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 8.8%

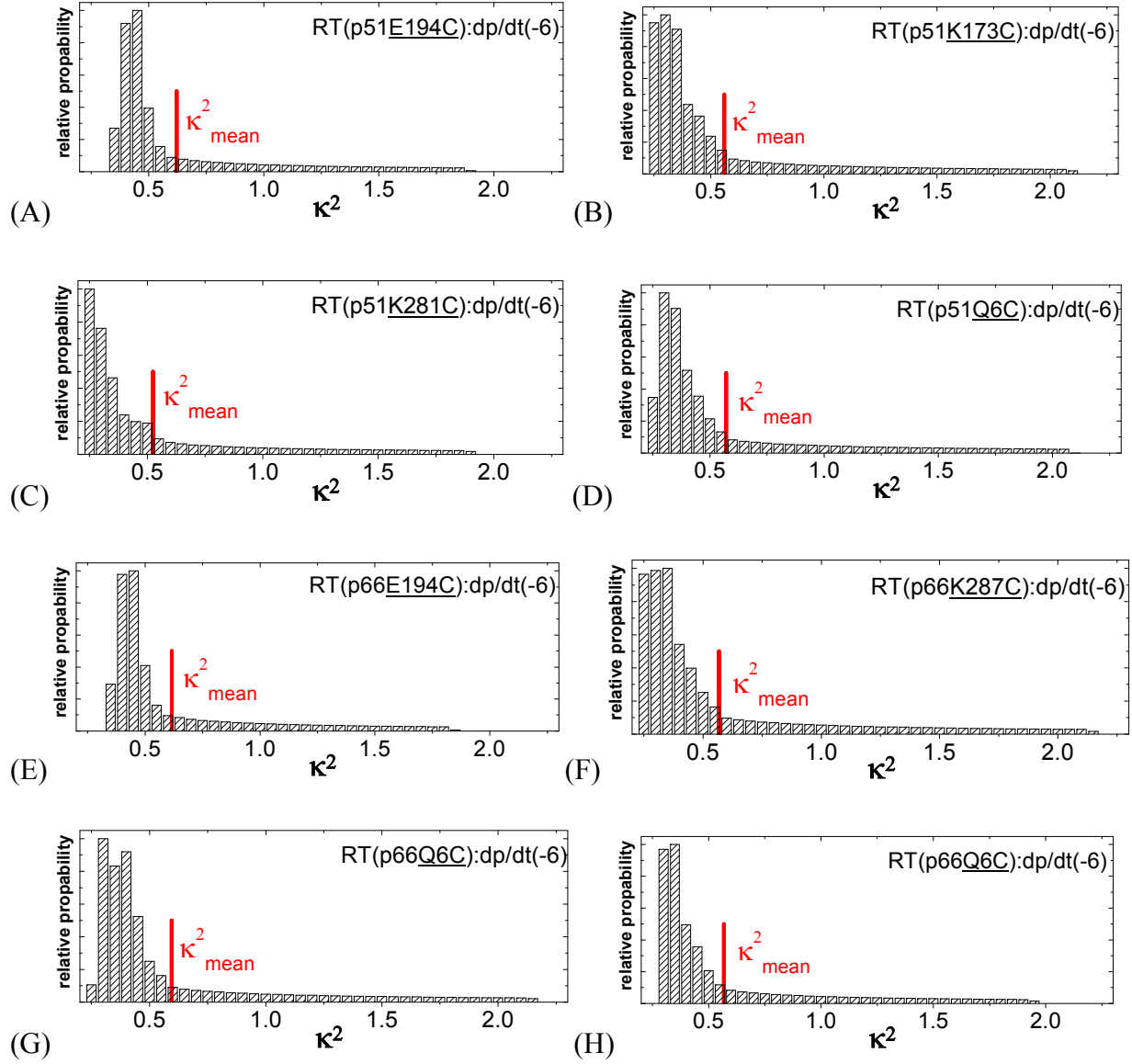


Figure S12.4: Possible κ^2 values of sample **(A)** RT(p51E194C):p/t(-6) $\langle \kappa^2 \rangle = 0.62$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% **(B)** RT(p51K173C):dp/dt(-6) $\langle \kappa^2 \rangle = 0.56$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% **(C)** RT(p51K281C):dp/dt(-6) $\langle \kappa^2 \rangle = 0.52$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% **(D)** RT(p51Q6C):p/t(-6) $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 9.4% **(E)** RT(p66E194C):p/t(-6) $\langle \kappa^2 \rangle = 0.62$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% **(F)** RT(p66K287C):p/t(-6) $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% **(G)** RT(p66Q6C):dp/dt(-6) $\langle \kappa^2 \rangle = 0.59$ uncertainty in $\langle R_{DA} \rangle_E$ 9.1% **(H)** RT(p66T27C):dp(19)/dt $\langle \kappa^2 \rangle = 0.57$ uncertainty in $\langle R_{DA} \rangle_E$ 8.8%

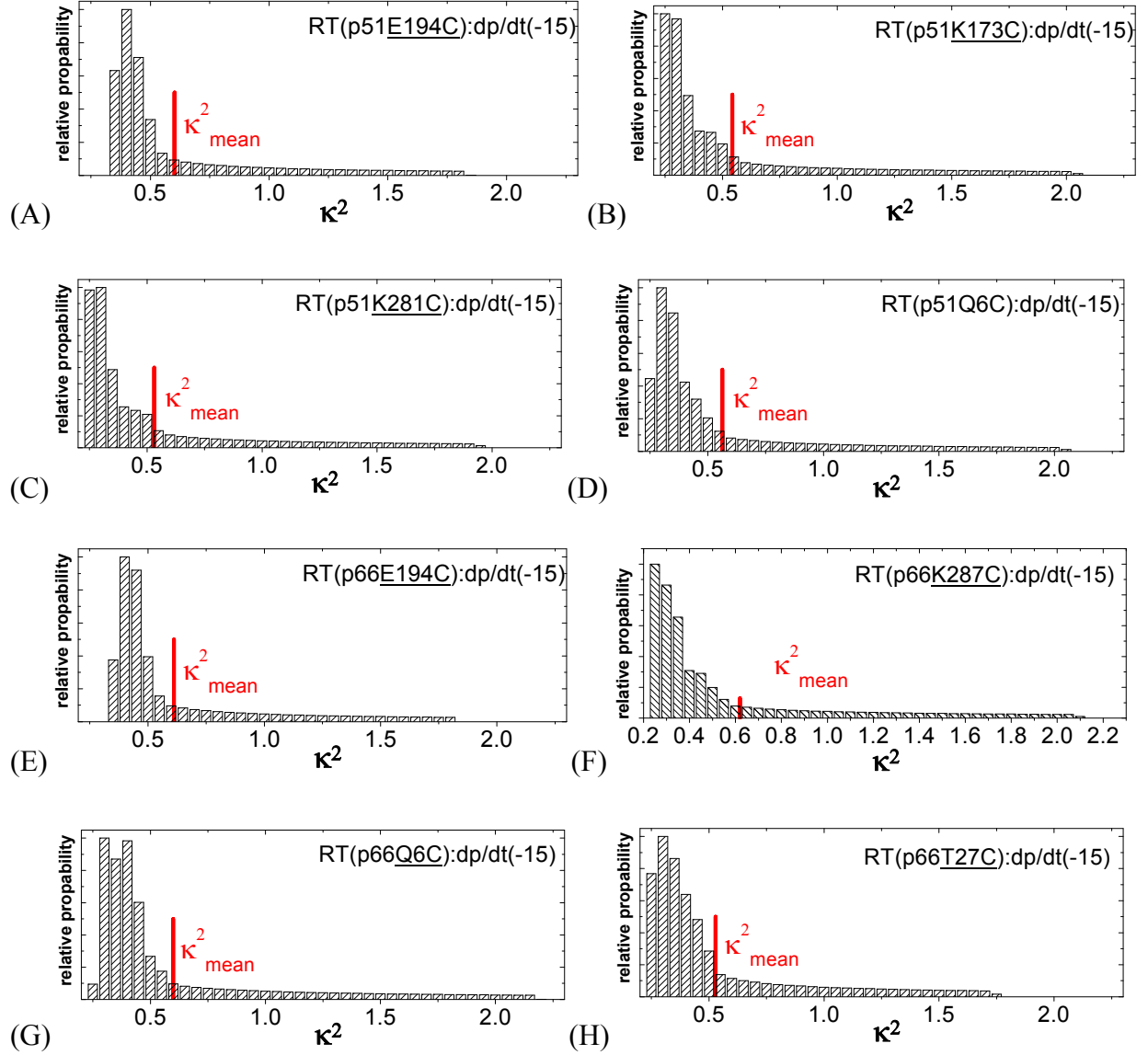
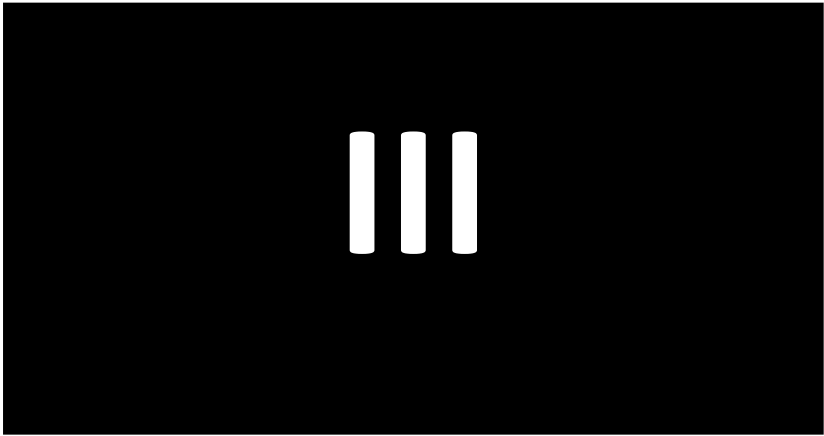


Figure S12.5: Possible κ^2 values of sample (A) RT(p51E194C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 7.4% (B) RT(p51K173C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.54$ uncertainty in $\langle R_{DA} \rangle_E$ 10.0% (C) RT(p51K281C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.53$ uncertainty in $\langle R_{DA} \rangle_E$ 9.9% (D) RT(p51Q6C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.56$ uncertainty in $\langle R_{DA} \rangle_E$ 9.5% (E) RT(p66E194C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.61$ uncertainty in $\langle R_{DA} \rangle_E$ 7.2% (F) RT(p66K287C):p/t(-15) $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 10.1% (G) RT(p66Q6C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.60$ uncertainty in $\langle R_{DA} \rangle_E$ 9.1% (H) RT(p66T27C):dp/dt(-15) $\langle \kappa^2 \rangle = 0.52$ uncertainty in $\langle R_{DA} \rangle_E$ 9.0%

References

1. Rothwell, J. P. *Structural Investigation of the HIV-1 RT using single pair Fluorescence Energy Transfer*, University of Dortmund, (2002).
2. Kensch, O., Restle, T., Wohrl, B. M., Goody, R. S. & Steinhoff, H. J. Temperature-dependent Equilibrium between the Open and Closed Conformation of the p66-Subunit of HIV-1 Reverse Transcriptase Revealed by Site-directed Spin Labelling. *J. Mol. Biol.* **301**, 1029-1039 (2000).
3. Kensch, O. *et al.* HIV-1 Reverse Transcriptase-Pseudoknot RNA Aptamer Interaction Has a Binding Affinity in the Low Picomolar Range Coupled with High Specificity. *J. Biol. Chem.* **275**, 18271-18278 (2000).
4. Rothwell, P. J. *et al.* Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655-1660 (2003).
5. Wohrl, B. M., Krebs, R., Goody, R. S. & Restle, T. Refined model for primer/template binding by HIV-1 reverse transcriptase: Pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J. Mol. Biol.*, 333-344 (1999).
6. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163-180 (2001).
7. Cai, Q. *et al.* Nanometer distance measurements in RNA using site-directed spin Labeling. *Biophys. J.* **93**, 2110-2117 (2007).
8. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Meth.* **5**, 965-971 (2008).
9. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* **133**, 2463-2480 (2011).
10. Woźniak, A. K., Schröder, G., Grubmüller, H., Seidel, C. A. M. & Oesterhelt, F. Single molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA*. **105**, 18337-18342 (2008).
11. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C. A. M. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B* **112**, 8361-8374 (2008).
12. Kalinin, S., Felekyan, S., Antonik, M. & Seidel, C. A. M. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *J. Phys. Chem. B* **111**, 10253-10262 (2007).
13. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970-6978 (2006).
14. Kalinin, S., Sisamak, E., Magennis, S. W., Felekyan, S. & Seidel, C. A. M. On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *J. Phys. Chem. B* **114**, 6197-6206 (2010).

15. Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C. A. M. Detection of structural dynamics by FRET: A photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B* **114**, 7983-7995 (2010).
16. Soong, T. T. *Fundamentals of probability and statistics for engineers*. (John Wiley & Sons, 2004).
17. Dale, R. E., Eisinger, J. & Blumberg, W. E. Orientational freedom of molecular probes - Orientation factor in intra-molecular energy transfer. *Biophys. J.* **26**, 161-193 (1979).
18. Ivanov, V., Li, M. & Mizuuchi, K. Impact of Emission Anisotropy on Fluorescence Spectroscopy and FRET Distance Measurements. *Biophys. J.* **97**, 922-929 (2009).
19. Koshioka, M., Saski, K. & Masuhara, H. Time-Dependent Fluorescence Depolarization Analysis in Three Dimensional Microscopy. *Appl. Spectrosc.* **49**, 224-228 (1995).
20. Choi, U. B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.* **17**, 318-U384 (2010).
21. Atkins, P. W. & de Paula, J. *Atkins' Physical Chemistry*. (Oxford University Press, 2001).
22. Martyna, G. J., Tobias, D.J., Klein, M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177-4189 (1994).
23. Marushchak, D., Grenklo, S., Johansson, T., Karlsson, R. & Johansson, L. B. A. Fluorescence depolarization studies of filamentous actin analyzed with a genetic algorithm. *Biophys. J.* **93**, 3291-3299 (2007).
24. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688 (2005).
25. Cornell, W. D. *et al.* A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
26. Simmerling, C., Strockbine, B. & Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258-11259 (2002).
27. Perez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817-3829 (2007).
28. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926-935 (1983).
29. Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).
30. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).
31. Brunger, A. T. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475 (1992).



Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling

Simon Sindbert^{1,*}, Stanislav Kalinin^{1,*}, Hayk Vardanyan¹, Christian Hanke⁴, Tomasz Soltysinski², Grzegorz Lach², Sabine Müller⁵, Janusz Bujnicki³, Holger Gohlke⁴, Claus A. M. Seidel¹

¹Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, Geb 26.32, 40225 Düsseldorf, Germany, ²International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland, ³Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznan, Poland, ⁴Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University, Universitätsstr. 1, 40225 Düsseldorf, Germany, ⁵Ernst-Moritz-Arndt-Universität Greifswald, Institut für Biochemie, Bioorganische Chemie, Felix-Hausdorff-Strasse 4, 17487, Greifswald, Germany

* contributed equally

Abstract

Like for many other non-coding RNAs, the helical four-way junction (RNA4WJ) is an essential structural motif of the hairpin ribozyme. Using FRET restrained high-precision structural modeling in combination with MD and coarse-grained simulations, we resolve the structures of three coexisting conformers of a fully Watson-Crick base paired RNA4WJ based on the hairpin ribozyme. 51 different donor-acceptor-pairs were measured using single-molecule multi-parameter fluorescence detection (smMFD). For each dataset, the single-molecule approach allowed for the simultaneous extraction of three distances (and their corresponding errors) belonging to one major FRET state and two minor states. Distinct Mg^{2+} -affinities were used for the assignment of the two minor states to the corresponding conformers. Rigid body models for the major and both minor conformers were obtained by docking rigid ds A-RNA helices explicitly taking into account dye position distributions. The three rigid body models were refined by all-atom MD simulations and coarse-grained RNA folding using FRET-restraints. A cluster analysis gives confidence levels for the proposed ensemble of models, and the models' quality was assessed via rigorous error estimation. The achieved precisions are significantly better than the uncertainty of the dye position with respect to the macromolecule.

Introduction

Non-protein coding RNAs perform essential functions in living organisms (Amaral, Dinger et al. 2008; Mercer, Dinger et al. 2009; Moazed 2009; Walter, Woodson et al. 2009). They commonly exhibit helical junctions as exemplified by the hairpin (Murchie, Thomson et al. 1998; Fedor 1999; Walter, Burke et al. 1999; Zhao, Wilson et al. 2000; Tan, Wilson et al. 2003), hammerhead (Pley, Flaherty et al. 1994; Scott, Finch et al. 1995) or VS ribozyme (Lafontaine, Norman et al. 2002), the U1 snRNA (Duckett, Murchie et al. 1995), bacteriophage RNA-based genomes (Beekwilder, Nieuwenhuizen et al. 1995; Groeneveld, Thimon et al. 1995), riboswitches (Chowdhury, Maris et al. 2006) and ribosomal RNA (Shen and Hagerman 1994; Orr, Hagerman et al. 1998). In ribozymes, the junction accelerates the folding into the biologically active tertiary structure by up to three orders of magnitude (Lilley 1998; Walter, Burke et al. 1999; Doudna and Cech 2002; Tan, Wilson et al. 2003; Pljevaljic, Millar et al. 2004; Al-Hashimi and Walter 2008). So far, there are no structural models of RNA4WJs that are not influenced by interactions with proteins or between stems, e.g. via bases of internal loops. The effects caused by sole stacking interactions are unknown.

In this work we examine an RNA4WJ that was derived from the hairpin ribozyme (Rupert, Massey et al. 2002) (PDB-ID: 1M5K, Fig. 1.2 in the SI). Its sequence was modified at the outer parts of the arms to ensure Watson-Crick base pairing throughout the whole molecule with no additional unpaired bases and no pairing between the arms (4H junction (Lilley, Clegg et al. 1996), see Fig. 1 and Fig. 1.2 in the SI). Single-molecule FRET studies showed that the 4WJ of the hairpin ribozyme exhibits Mg^{2+} dependent dynamic heterogeneities (Walter, Murchie et al. 1998; Hohng, Wilson et al. 2004; Wilson, Nahas et al. 2005), furthermore, the presence of at least three coexisting conformers was shown (Hohng, Wilson et al. 2004). This has prevented structural determination via classical methods such as X-ray crystallography, NMR spectroscopy or SAXS. Moreover, transient minor populations have never been structurally characterized by FRET before.

To resolve the structures of the three coexisting conformers of the RNA4WJ we apply FRET restrained high-precision structural modeling, which has been shown to improve the assessment of quality and precision of FRET-derived structures dramatically (Kalinin 2012). Furthermore, the single-molecule advantage of FRET is most suitable to investigate heterogeneous systems. Additionally, we use state of the art bioinformatics and computational biophysics to refine the FRET derived models with the aim to obtain meaningful all-atom structures.

Different positions on the RNA were chosen for internal labeling for FRET with Alexa488 as a donor (D) and Cy5 as an acceptor (A) dye. In total 51 combinations of different DA pairs were sequentially analyzed (see Fig. 1).

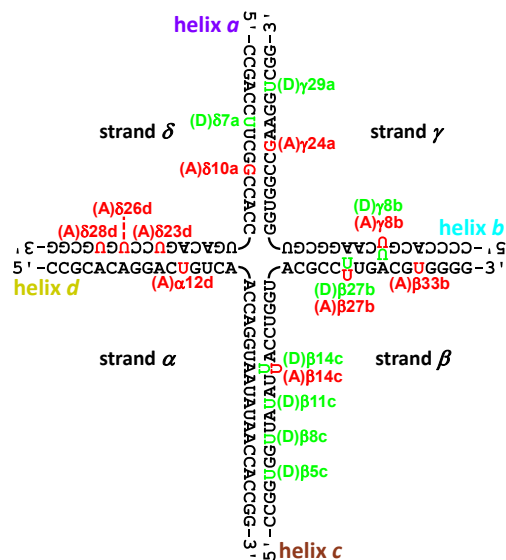


Figure 1. Secondary structure of the RNA4WJ with labeling positions for donor (green) and acceptor (red) dyes. The names of the labeling positions indicate the type of dye (D for Alexa488 and A for Cy5) followed by the name of single strand (α , β , γ or δ), the number of the labeled base starting from the 5'-end, and the arm of the junction (a, b, c or d).

Quantitative single-molecule FRET measurements by smMFD

Single-molecule multi-parameter fluorescence detection (smMFD, see Section 3.3.2 in the SI) simultaneously acquires all fluorescence parameters (Sisamak, Valeri et al. 2010), which allows to calculate the FRET efficiency based on both fluorescence intensity and fluorescence lifetime. Thus, unlike ensemble FRET measurements, it detects incomplete labeling, characterizes fluorophore quenching and mobility, and resolves multiple FRET states (Margittai, Widengren et al. 2003; Rothwell, Berger et al. 2003; Sisamak, Valeri et al. 2010). DA distances (R_{DA}) are obtained from the FRET indicator, F_D/F_A (ratio of D and A fluorescence) (Sisamak, Valeri et al. 2010) via Eq. 1:

$$R_{DA} = \left(\Phi_{FA} \frac{F_D}{F_A} \right)^{\frac{1}{6}} R_{0r} \quad (1)$$

where Φ_{FA} is the fluorescence quantum yield of A (Section 4.1 in the SI for the determination of Φ_{FA} for every A labeling position), and R_{0r} is the reduced Förster radius (Rothwell, Berger et al. 2003). See Section 3.4 in the SI for further details.

All 51 DA-pairs were analyzed in free diffusion by confocal smMFD. A high concentration of Magnesium ions. (20 mM $MgCl_2$) was chosen to avoid structural transitions within the observation time (~ 1 ms) (Hohng, Wilson et al. 2004). Fig. 2A shows typical 2D MFD-histograms of the F_D/F_A fluorescence ratio and the donor anisotropy r_D versus $\tau_{D(A)}$ for sample $(D)\beta 11c_{-}(A)\delta 23d$. The three FRET states (indicated by horizontal lines) and the D-only species follow the theoretically expected dependence between $\tau_{D(A)}$ and F_D/F_A for a static molecule (see Eq. 3.5 in the SI). The low donor anisotropies r_D are a prove for its fast rotational diffusion time ρ , which minimizes dye orientation problems in the FRET analysis.

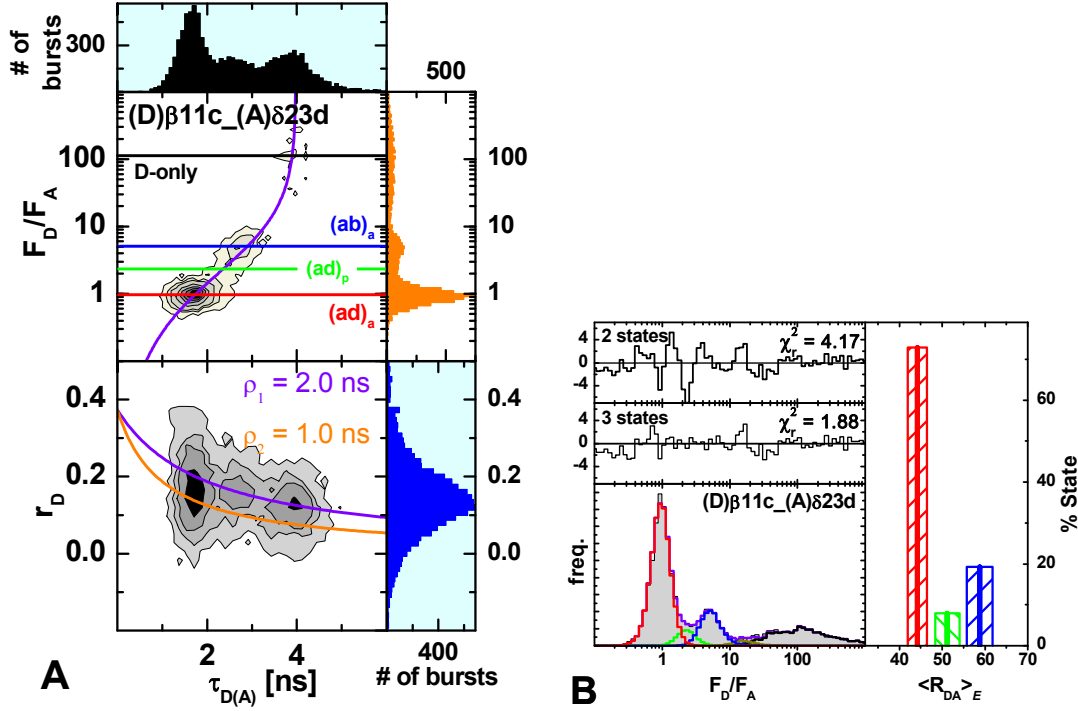


Figure 2. (A) 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D (lower panel) versus $\tau_{D(A)}$ for sample $(D)\beta11c_ (A)\delta23d$. The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. The theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; purple, see Section 3.5 in the SI for details) and horizontal lines indicating the D-only (black) and the three FRET states belonging to conformers $(ad)_a$ (red), $(ad)_p$ (green) and $(ab)_a$ (blue) are overlaid in the F_D/F_A vs $\tau_{D(A)}$ plot. The solid purple and orange lines in the r_D - $\tau_{D(A)}$ diagram are given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with rotation correlation times $\rho_1 = 1.0$ ns (orange) and $\rho_2 = 2.0$ ns (purple), $r_0 = 0.374$. (B) PDA for sample $(D)\beta11c_ (A)23d$ (selected bursts). The F_D/F_A histogram of experimental data (gray area) is fitted (purple solid line) using the following parameters: 45.1 % of $\langle R_{DA} \rangle_{E(1)} = 44.2$ Å (red, $(ad)_a$); 4.9 % of $\langle R_{DA} \rangle_{E(2)} = 51.2$ Å (green, $(ad)_p$); 11.9 % of $\langle R_{DA} \rangle_{E(3)} = 58.8$ Å (blue, $(ab)_a$); $\sigma_{app} = 3.8$ % of $\langle R_{DA} \rangle_E$; 36.2 % of D-only; 1.9 % of impurities with apparent $R_{DA} = 71.5$ Å (also present in D-only samples); $\chi_r^2 = 1.88$. Weighted residuals are shown in the middle plot. The upper plot shows the weighted residuals for PDA with only two FRET states ($\chi_r^2 = 4.17$) (see Section 4.2 in the SI). The right panel shows the distances and relative amplitudes (solid lines) and the confidence intervals for the distances (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$) resulting from Eq. 2.

Distances and errors. To extract values of $\langle R_{DA} \rangle_E$ from noisy single molecule data we use photon distribution analysis (PDA) (Antonik, Felekyan et al. 2006; Kalinin, Felekyan et al. 2008) (see Fig. 2B and Section 3.4 in the SI) as it takes into account shot noise and provides a meaningful reduced chi-squared

value (χ_r^2) directly derived from photon statistics. For approximately half of the 51 datasets three FRET states were necessary to reach a satisfying fit quality (χ_r^2 and weighted residuals; see Fig. 2B) by PDA with one major state (red) with ~ 60 to 85 % relative amplitude and two minors (blue, green) with usually similar amplitudes of each ~ 4 to 20 %, respectively (see Section 4.2 in the SI for PDA of $(D)\beta 11c_A\delta 23d$ with two FRET states). This strongly suggests the presence of at least 3 quasi-static conformers in equilibrium, which has been found in previous studies (Hohng, Wilson et al. 2004). For the other half, two states were sufficient (one major and one minor). In these cases, the second “hidden” minor state is overlapped by one of the two visible ones. If the amplitude of the visible minor state was smaller than 25 %, the other minor was usually assumed to be overlapped by the major, if not by the minor. See Section 4.3 in the SI for a table with all distances, amplitudes and errors for each of the three conformers.

Additionally, PDA allows us to estimate errors for the fitted parameters due to photon statistics, in particular for $\langle R_{DA} \rangle_E \{ \Delta R_{DA}(E) \}$ (see Section 3.7 in the SI for more details). Another contribution to ΔR_{DA} originates from the uncertainty of the mutual orientation of the donor and acceptor dyes (κ^2 errors, $\Delta R_{DA}(\kappa^2)$). For ds A-RNA, labeled with Alexa488 and Cy5 as a DA-pair using C6 (hexamethylene) linkers it is 5 % (Sindbert, Kalinin et al. 2011), hence, we use this value for $\Delta R_{DA}(\kappa^2)$ throughout this work. The overall error ΔR_{DA} is then calculated according to error propagation rules (Eq. 2):

$$\Delta R_{DA}^2 = \Delta R_{DA}^2(E) + \Delta R_{DA}^2(\kappa^2) \quad (2)$$

Fig. 2B shows an exemplary PDA of smMFD data, the resulting distances and relative amplitudes of the fitted FRET states and their resulting overall uncertainties. For the major peaks $\Delta R_{DA}(E)$ contributes only weakly to the overall error (between 0.5 and 2 %), hence, ΔR_{DA} is mostly between 5 and 6 %. For the minor peaks $\Delta R_{DA}(E)$ can be considerably higher, especially if they are overlapped by others (mostly between 3 and 10 %, up to ~ 20 % for a few cases, see Section 4.3 in the SI for all resulting values for ΔR_{DA}).

Assignment of FRET peaks

In four-way junctions of nucleic acids, in general, up to 4 different conformer patterns for stacked helix pairs are possible (Hohng, Wilson et al. 2004). The conformers are distinguished by coaxial stacking pattern of helices (e.g. helix *a* can stack either on helix *b* or *d*, thus, *ad* or *ab*) and by parallel (p) or antiparallel (a) direction of the two single strands which continuously run through the stacked helix pairs (see Fig. 3A).

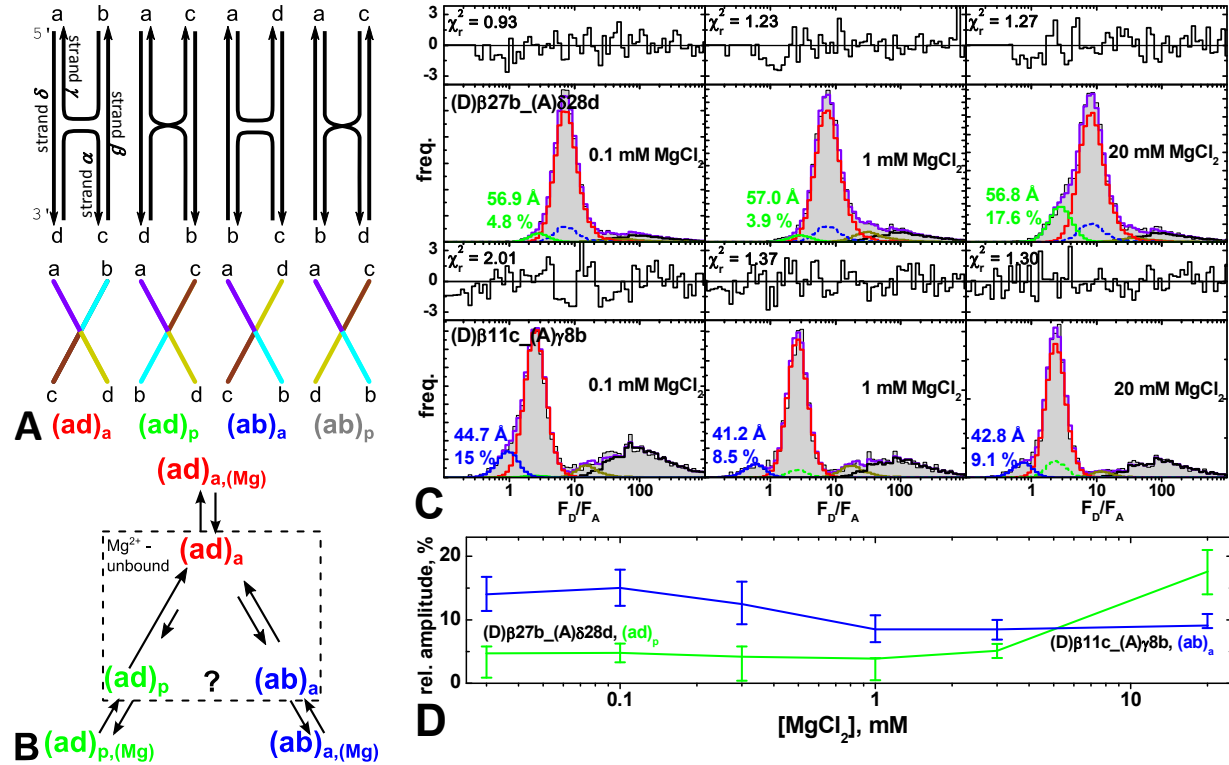


Figure 3. (A) Four possible conformers with stacked helix pairs of a four-way junction of nucleic acids. (B) Possible transitions between the three conformers and their respective Mg²⁺-bound and unbound states. (C) PDAs (selected bursts) of Mg²⁺-titrations (0.1 mM, left, 1 mM, middle and 20 mM, right) for samples (D)β27b_(A)δ28d (upper panel) and (D)β11c_(A)γ8b (lower panel). F_D/F_A histograms of experimental datasets (gray areas) are fitted (purple solid lines) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, green for (ad)_a, $\langle R_{DA} \rangle_{E(2)}$, red for (ad)_p and $\langle R_{DA} \rangle_{E(3)}$, blue for (ab)_a) one D-only (black) and one impurity state (dark yellow) (see Table 4-5 in the SI for all PDA fit parameters). Weighted residuals are shown in the respective upper plots. (D) Relative FRET amplitudes of the state assigned to conformer (ad)_p for (D)β27b_(A)δ28d and the state assigned to (ab)_a for (D)β11c_(A)γ8b measured at various concentrations of MgCl₂.

Using all 51 distances from the major FRET populations to model a structure yields conformer $(ad)_a$ (see Fig. 4A below). The two other conformers visible in our data we assume to be $(ad)_p$ and $(ab)_a$, as justified below. The presence of conformer $(ab)_p$ can neither be confirmed nor excluded as for every combination of labeled helix pair, its FRET populations could overlap with those of the other three (see Fig. 3A). For each dataset, the minor FRET peaks have to be assigned to either $(ad)_p$ or $(ab)_a$, in order to be able to use them for structure modeling. However, because of similar equilibrium fractions, they cannot be reliably assigned by their amplitudes. The assignment was, therefore, based on chemical properties of these states as well as the geometric model of the RNA4WJ (Fig. 3A). First, we noticed that at low Mg^{2+} concentrations one minor peak significantly decreases in amplitude, whereas the other one remains stable. This observation can be attributed to fast species interconversion at low Mg^{2+} concentrations (Hohng, Wilson et al. 2004; Buck, Noeske et al. 2010; Buck, Wacker et al. 2011). A possible kinetic scheme is shown in Fig. 3B. Considering the time resolution of our experiment, FRET peaks due to species with sub-ms lifetimes cannot be resolved and apparently disappear. Thus, minor peaks can be unambiguously assigned according to their behavior at low Mg^{2+} , which is in principle sufficient for structural modeling. Two exemplary Mg^{2+} titrations are presented in Fig. 3C and D. At low Mg -concentrations, the minor high FRET peak slightly increases in amplitude for $(D)\beta11c_{-}(A)\gamma8b$ and significantly decreases for $(D)\beta27b_{-}(A)\delta28d$. Referring to the geometric model of RNA4WJ (Fig. 3A), the “stable” and “unstable” species resemble $(ab)_a$ and $(ad)_p$ conformers, respectively, which can be shown as follows. The model suggests that, in most of cases, the DA distances should be determined mainly by the angle between the helical axes. For instance, for labeling on helices b and d the minor state yielding the smaller distance can be assigned to $(ad)_p$ (Fig. 3A). Thus, in the case of $(D)\beta27b_{-}(A)\delta28d$ (Fig. 3C) the minor high FRET peak is due to the $(ad)_p$ state, and for $(D)\beta11c_{-}(A)\gamma8b$ it is due to $(ab)_a$. We performed 14 titrations which all confirmed the above assignment, that is, the minor FRET peak that disappears at low Mg^{2+} is consistent with the assumed geometry of $(ad)_p$. In addition, Mg^{2+} titrations were performed for all samples with $(A)\alpha12d$ which is close to the junction, and all samples with labeled Guanines $((A)\gamma24a$ and $(A)\delta10a)$ where the dye’s mean position is far away from the RNA’s helical axis (see Section 3.2 in the SI). In other 19 cases the minor peaks were assigned based on the geometric model (Fig. 3A). All PDA results are summarized in Tables 4-4 (20 mM $MgCl_2$) and 4-5 (Mg^{2+} - titrations) in the SI. We finally end up with a set of 51 distances $\langle R_{DA} \rangle_E$ and corresponding errors ΔR_{DA} for each of the three conformers (see Table 4-3 in the SI).

Rigid body docking

To generate 3D structural models accurately describing the orientations and positions the junction's four helices, we, initially, treaded the individual helices as rigid bodies having perfect A-RNA form. This simplification is justified due to Watson-Crick base pairing throughout the whole molecule. Thus, significant deviations from A-RNA structure are expected only within close proximity to the junction. The four ds A-RNAs were generated with the nucleic acid builder (NAB) which is a part of AmberTools (Case 2008). Additionally to the 51 FRET distance restraints, four strong constraints between the 3O atoms of the 3' ends and the P atoms of the 5' ends of single strands were introduced representing chemical bonds between neighboring helices at the junction.

To find structures which agree best with the measured distances, we applied a rigid body dynamics approach while explicitly modeling the accessible volumes (AVs, see Section 3.1 in the SI) of fluorophores (Kalinin 2012). This makes it possible to estimate mean dye positions with respect to RNA and average over distributions of DA distances (for further details see Sections 3.8 and 3.9 in the SI and (Kalinin 2012)). The weighted data-model deviation for a set of 51 + 4 distances (see Eq. 3) is minimized by docking the four dsRNAs. This is done sufficiently often (1000 iterations) yielding all local minima.

$$\chi_E^2 = \sum_{i=1}^n \frac{(R_{DA(i)} - R_{\text{model}(i)})^2}{\Delta R_{DA(i)}^2} \quad (3)$$

Resulting structures and quality validation

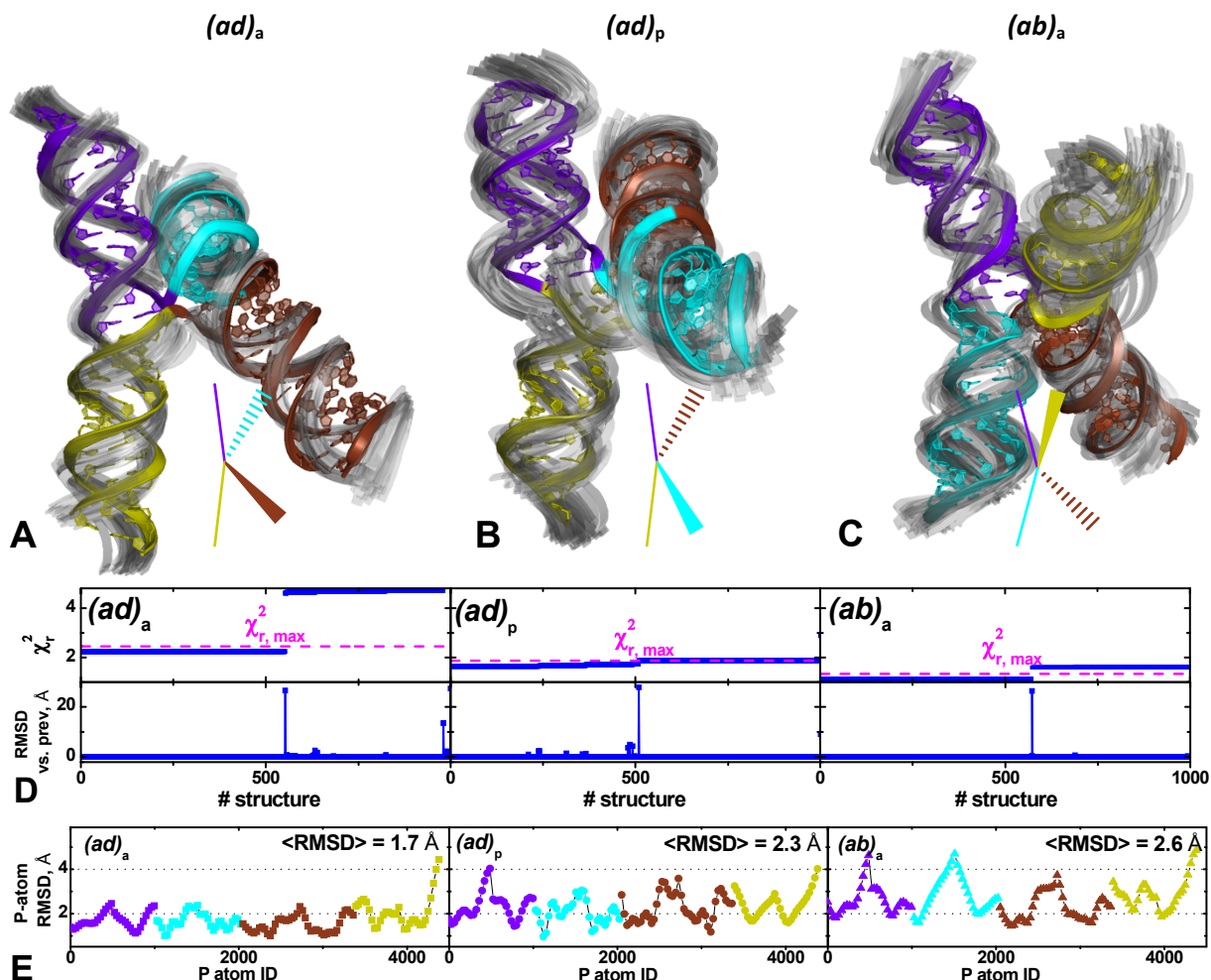


Figure 4. (A - C) Structures (cartoon representation) with the lowest χ_r^2 after refinement of the rigid body models by MD simulations (colored) for conformers $(ad)_a$ (A), $(ad)_p$ (B) and $(ab)_a$ (C) overlaid with 100 structures (grey transparent) indicating the precision (uncertainties) of the helix positions and orientations resulting from bootstrapping of the rigid body model. Bottom, respectively: Sketches depicting the mutual orientation of helices a (purple), b (cyan), c (brown) and d (dark yellow). (D) χ_r^2 (upper panels) and RMSD vs. the previous structure (lower panels) plotted against the structure ID found after docking in ascending order with respect to χ_r^2 for $(ad)_a$ (left), $(ad)_p$ (middle) and $(ab)_a$ (right). The dashed magenta lines represent 84 % confidence thresholds (see Section 3.10 in the SI): $\chi_{r,max}^2 = 2.46$, 1.86 and 1.34 for $(ad)_a$, $(ad)_p$ and $(ab)_a$, respectively. (E) uncertainty of phosphate atom positions in helices a (purple), b (cyan), c (brown) and d (dark yellow) calculated by bootstrapping for the respective solutions with the lowest χ_r^2 after docking for $(ad)_a$ (left, squares), $(ad)_p$ (middle, circles) and $(ab)_a$ (right, triangles). The average RMSD values over all P atoms are 1.7 Å, 2.3 Å, and 2.6 Å, respectively.

Analysis of docking results. The respective “best” structures (lowest χ_r^2) after docking (see Section 4.6 in the SI) and further refinement by MD simulations (see below) are shown in Fig. 4A - C. It is obvious that all three correspond to their respective conformers in terms of helix orientations (see Fig. 3A). However, stacking between the helix pairs (*ad* and *bc* for (*ad*)_a and (*ad*)_p or *ab* and *cd* for (*ab*)_a) is never perfect. Instead, clear kinks are visible. This is in accordance with recent studies, which, compared to the distances measured in the 4WJ, yielded significantly different distances along a dsRNA with the same sequence as helix pair *bc* between corresponding labeling positions (Sindbert, Kalinin et al. 2011). In the resulting structure for (*ad*)_p the stacking of the two helix pairs is not as pronounced as for the other two conformers which is due to the parallel orientation of its strands. The two strands (γ and α) crossing between the two stacked helix pairs interchange between helices on opposite sides the junction instead of between two neighboring helices describing a U-turn (see Fig. 3A). χ_r^2 and “RMSD vs. the previous structure” for all docked structures are plotted in Fig. 4D for the three conformers. Steps in the χ_r^2 plot and corresponding peaks in the RMSD plot (Fig. 4D) separate groups of very similar solutions (clusters). For (*ad*)_a and (*ab*)_a unique solutions are found below the 84 % confidence threshold ($\chi_{r,\max}^2$, see Section 3.10 in the SI for details). Hence, they can be assumed to be unique. For (*ad*)_p there are several solutions below χ_{\max}^2 . However, they deviate from the best one by only up to 2.9 Å RMSD which is comparable to the model’s precision of 2.3 Å (see below).

For all three conformers, approximately half of the found solutions after docking deviate strongly from the respective best ones (RMSD = 26.6 Å for (*ad*)_a, 28.1 Å for (*ad*)_p and 26.6 Å for (*ab*)_a) and can be excluded with a confidence of > 99.99 % ((*ad*)_a), 84 % ((*ad*)_p) and 97 % ((*ab*)_a), respectively. Comparing these “second best” structures (see Figure 4.5 in the SI) with the according best ones shows that they are “rough” quasi-mirrored images. This is due the fact that for most labeling positions (labeled uracils) the dye linker points into the major groove of dsRNA. Thus, the mean positions of the dyes are close to the helix axis and are, therefore, invariant to mirror image transformation of the structure. For labeled guanines, dye linkers point into the minor groove resulting in AVs that are significantly displaced from the helical axis (see Figure 3.1 in the SI). Labeling positions (*A*) γ 24*a* and (*A*) δ 10*a* were specifically chosen to distinguish between the two quasi - mirrored solutions for conformer (*ad*)_a resulting in a significantly increased χ_r^2 for the “second best” solution (see Fig. 4D).

For a few cases, the major FRET peak $((ad)_a)$ yielded a significant shift in distance at low Mg^{2+} -concentrations compared to the corresponding measurements at 20 mM (see Section 4.3 in the SI). We attribute these changes to a slightly changed conformation of $(ad)_a$ in its Mg^{2+} -unbound state which has been predicted to resemble a 90° cross (Walter, Murchie et al. 1998). A structure was generated with the 30 distances measured at 0.1 mM $MgCl_2$ for $(ad)_a$ (see Section 4.5 in the SI). It exhibits a clear $(ad)_a$ conformation only with angles between the stacked helix pairs closer to 90° compared to the structure at 20 mM $MgCl_2$.

Precision estimation. To estimate the uncertainties of the respective best solutions after docking we apply a bootstrapping procedure (Efron 1986) according to (Kalinin 2012) (see Section 3.11 in the SI). In Fig. 4A - C the respective final solutions after refinement via MD simulations (see below) are overlaid with the corresponding sets of perturbed structures and in Figure 4E the uncertainties of all P-atoms are plotted for the three conformers. The average RMSD values over all P atoms are 1.7 Å, 2.3 Å, and 2.6 Å, for $(ad)_a$, $(ad)_p$, and $(ab)_a$ states, respectively. Due to the smaller measurement errors, the uncertainties are smaller for $(ad)_a$ than for the minor conformers.

Model refinement

MD simulations. The three final rigid body models are used as starting structures for further refinement through all-atom molecular dynamics (MD) simulations in explicit solvent (Case, Cheatham et al. 2005) (see Section 3.12 in the SI). Throughout the simulation the positions of all P-atoms except those within 6 bps distance from the junction were restrained by harmonic potentials such that their fluctuations are equal to positional uncertainties (RMSD values) obtained by bootstrapping for the rigid body models (see Figure 4E). The resulting trajectories were then filtered with respect to agreement with FRET data. For this, AVs were modeled to the MD-derived structures, resulting DA distances were calculated and compared to measured ones. Figures 4A-C show the resulting structures with the lowest value of χ_r^2 , i.e., structures with best agreement with FRET distance restraints. Comparing χ_r^2 between the best MD and rigid body models shows that for conformers $(ad)_a$ and $(ad)_p$, χ_r^2 is significantly reduced from 1.93 and 1.57 to 1.75 and 1.40, respectively, while for $(ab)_a$ it increased from 1.06 to 1.32 (the values of χ_r^2 shown here include only violations of FRET restraints).

The models of the RNA4WJ from FRET-filtered molecular dynamics simulations have both the global geometry consistent with the FRET restraints and the local stereochemistry encoded in the MD force field (see Section 3.12 in the SI). In particular, all bases at the junction are properly stacked after MD-refinement. The optimization of the local structure is achieved without significant violations of the global geometry as judged by comparison to respective rigid body models (see Section 4.6 in the SI). The RMSDs over all P atoms, excluding those within 6 bps distance from the junction, are 3.17 Å, 2.96 Å and 2.41 Å for $(ad)_a$, $(ad)_p$ and $(ab)_a$, respectively.

SimRNA. As an alternative approach, a coarse grained modeling program SimRNA (Rother, Rother et al. 2012) was used for the *de novo* folding of the RNA4WJ (see Section XXX in the SI). SimRNA uses a reduced representation of RNA (only five centers of interaction per residue), a Monte Carlo sampling scheme, and a statistical (knowledge-based) potential to estimate the energy of interactions, and as a result of these simplifications, compared to all-atom MD simulations, it is faster by a factor of ~1000. The formation of helices was enforced by including distance restraints for residues expected to form base pairs (excluding the 6 pairs in each helix closest to the junction). Furthermore, 51 distance restraints were used between the P atoms of the 18 labeled nucleotides with uncertainties according to the RMSD values calculated by bootstrapping for the respective Ps (see Figure 4E). A series of

simulations were performed to minimize both the energy of interactions, and to satisfy the restraints. The resulting models were converted to full atom representations and energy-minimized. They are shown in Figure 4.6 in the SI. The RMSD over all P atoms compared to the MD results are 3.52 , XXX and XXX Å for $(ad)_a$, $(ad)_p$ and $(ab)_a$, respectively.

Conclusions and outlook

FRET restrained high-precision structural modeling allowed establishing highly accurate structural models of three coexisting conformers of the RNA4WJ. Even though distance uncertainties for the minor populations were often significantly larger compared to major ones, the precisions of the minor models are only slightly worse. In comparison to, e.g., NMR spectroscopy, only a few (51) distance restraints were used. However, both FRET-restrained molecular dynamics simulations and coarse-grained structure predictions provided meaningful and consistent all-atom models. Due to its dynamic behavior and the resulting structural heterogeneity, structure determination of the RNA4WJ by traditional structural biology approaches is very difficult. Making use of the single-molecule advantage, we were able to structurally resolve three conformers in parallel by FRET including the coexisting transient minor conformers. Although the sequence of the RNA4WJ was significantly altered compared to the natural hairpin ribozyme, we believe that the present results are nevertheless structurally relevant because the sequence in the junction region was left unchanged; also, because of the absence of interactions between the arms, the overall conformations are entirely determined by the junction itself. Thus, for the first time, the structure of an RNA4WJ was determined in a state uninfluenced by, e.g., interactions between the arms or with other macromolecules.

As demonstrated in this work, experimental restraints from FRET can be used in computational modeling of structures by post-filtering of, e.g., MD-derived ensembles. However, as a mean to improve the method, AV modeling could be implemented into a computational method. Using simple starting structures based on secondary structure elements (e.g. nucleic acid sequence) FRET restraints could then be used as a guiding potential which leads to relaxed solutions unbiased by the rigid body assumptions. Moreover, FRET-restrained high-precision structural modeling is also applicable to structurally heterogeneous and flexible proteins, whose overall structures are notoriously difficult to determine. Together with filtered fluorescence correlation spectroscopy (Felekyan, Kalinin et al. 2012), FRET harbors the potential to study conformational control of biomolecular function in complex systems within a nanosecond to minute time range and associate it to detailed dynamic structures without spatial averaging.

References

- Al-Hashimi, H. M. and N. G. Walter (2008). "RNA dynamics: it is about time." Current Opinion in Structural Biology **18**(3): 321-329.
- Amaral, P. P., M. E. Dinger, et al. (2008). "The eukaryotic genome as an RNA machine." Science **319**(5871): 1787-1789.
- Antonik, M., S. Felekyan, et al. (2006). "Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis." Journal of Physical Chemistry B **110**(13): 6970-6978.
- Beekwilder, M. J., R. Nieuwenhuizen, et al. (1995). "Secondary Structure Model for the Last 2 Domains of Single-Stranded Rna Phage Q-Beta." Journal of Molecular Biology **247**(5): 903-917.
- Buck, J., J. Noeske, et al. (2010). "Dissecting the influence of Mg(2+) on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain." Nucleic Acids Research **38**(12): 4143-4153.
- Buck, J., A. Wacker, et al. (2011). "Influence of ground-state structure and Mg(2+) binding on folding kinetics of the guanine-sensing riboswitch aptamer domain." Nucleic Acids Research **39**(22): 9768-9778.
- Case, D. A., T. E. Cheatham, et al. (2005). "The Amber biomolecular simulation programs." Journal of Computational Chemistry **26**(16): 1668-1688.
- Case, D. A. T. A. D. T. E. C. I. C. L. S. J. W. R. E. D. R. L. M. C. R. (2008). Amber San Francisco, University of California.
- Chowdhury, S., C. Maris, et al. (2006). "Molecular basis for temperature sensing by an RNA thermometer." EMBO Journal **25**(11): 2487-2497.
- Doudna, J. A. and T. R. Cech (2002). "The chemical repertoire of natural ribozymes." Nature **418**(6894): 222-228.
- Duckett, D. R., A. I. H. Murchie, et al. (1995). "The Global Folding of 4-Way Helical Junctions in Rna, Including That in U1 Snrna." Cell **83**(6): 1027-1036.
- Efron, B. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion." Annals of Statistics **14**(4): 1301-1304.
- Fedor, M. J. (1999). "Tertiary structure stabilization promotes hairpin ribozyme ligation." Biochemistry **38**(34): 11040-11050.
- Felekyan, S., S. Kalinin, et al. (2012). "Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules." ChemPhysChem **13**: 1036-1053.
- Groeneveld, H., K. Thimon, et al. (1995). "Translational Control of Maturation-Protein Synthesis in Phage Ms2 - a Role for the Kinetics of Rna Folding." Rna-a Publication of the Rna Society **1**(1): 79-88.
- Hohng, S., T. J. Wilson, et al. (2004). "Conformational flexibility of four-way junctions in RNA." Journal of Molecular Biology **336**(1): 69-79.
- Kalinin, S., S. Felekyan, et al. (2008). "Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis." Journal of Physical Chemistry B **112**(28): 8361-8374.
- Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. (2012). "FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang."
- Lafontaine, D. A., D. G. Norman, et al. (2002). "The global structure of the VS ribozyme." EMBO Journal **21**(10): 2461-2471.
- Lilley, D. M. J. (1998). "Folding of branched RNA species." Biopolymers **48**(2-3): 101-112.
- Lilley, D. M. J., R. M. Clegg, et al. (1996). "A nomenclature of junctions and branchpoints in nucleic acids. Recommendations 1994." Journal of Molecular Biology **255**(3): 554-555.

- Margittai, M., J. Widengren, et al. (2003). "Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1." Proc.Natl.Acad.Sci.USA **100**: 15516-15521.
- Mercer, T. R., M. E. Dinger, et al. (2009). "Long non-coding RNAs: insights into functions." Nature Reviews Genetics **10**(3): 155-159.
- Moazed, D. (2009). "Small RNAs in transcriptional gene silencing and genome defence." Nature **457**(7228): 413-420.
- Murchie, A. I. H., J. B. Thomson, et al. (1998). "Folding of the Hairpin Ribozyme in Its Natural Conformation Achieves Close Physical Proximity of the Loops." Molecular Cell **1**: 873-881.
- Orr, J. W., P. J. Hagerman, et al. (1998). "Protein and Mg²⁺-induced conformational changes in the S15 binding site of 16 S ribosomal RNA." Journal of Molecular Biology **275**(3): 453-464.
- Pley, H. W., K. M. Flaherty, et al. (1994). "3-Dimensional Structure of a Hammerhead Ribozyme." Nature **372**(6501): 68-74.
- Pljevaljcic, G., D. P. Millar, et al. (2004). "Freely diffusing single hairpin ribozymes provide insights into the role of secondary structure and partially folded states in RNA folding." Biophysical Journal **87**(1): 457-467.
- Rother, K., M. Rother, et al. (2012). Template-based and template-free modeling of RNA 3D structure: Inspirations from protein structure modeling. In RNA 3D structure analysis and prediction. N. B. Leontis and E. Westhof. Berlin, Springer-Verlag.
- Rothwell, P. J., S. Berger, et al. (2003). "Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 Reverse Transcriptase:primer/template complexes." Proceedings of the National Academy of Sciences of the United States of America **100**: 1655-1660.
- Rupert, P. B., A. P. Massey, et al. (2002). "Transition state stabilization by a catalytic RNA." Science **298**(5597): 1421-1424.
- Scott, W. G., J. T. Finch, et al. (1995). "The Crystal-Structure of an All-Rna Hammerhead Ribozyme - a Proposed Mechanism for Rna Catalytic Cleavage." Cell **81**(7): 991-1002.
- Shen, Z. and P. J. Hagerman (1994). "Conformation of the central, 3-helix junction of the 5 S ribosomal-RNA of *Sulfolobus acidocaldarius*." Journal of Molecular Biology **241**(3): 415-430.
- Sindbert, S., S. Kalinin, et al. (2011). "Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity." Journal of the American Chemical Society **133**(8): 2463-2480.
- Sisamakos, E., A. Valeri, et al. (2010). "Accurate single-molecule FRET studies using multiparameter fluorescence detection." Methods in Enzymology **475**: 456-514.
- Tan, E., T. J. Wilson, et al. (2003). "A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate." Proceedings of the National Academy of Sciences of the United States of America **100**(16): 9308-9313.
- Walter, F., A. I. H. Murchie, et al. (1998). "Global structure of four-way RNA junctions studied using fluorescence resonance energy transfer." Rna-a Publication of the Rna Society **4**(6): 719-728.
- Walter, F., A. I. H. Murchie, et al. (1998). "Folding of the four-way RNA junction of the hairpin ribozyme." Biochemistry **37**(50): 17629-17636.
- Walter, N. G., J. M. Burke, et al. (1999). "Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction." Nature Structural Biology **6**(6): 544-549.
- Walter, N. G., S. A. Woodson, et al. (2009). Non-protein coding RNAs. Heidelberg, Springer.
- Wilson, T. J., M. Nahas, et al. (2005). "Folding and catalysis of the hairpin ribozyme." Biochemical Society Transactions **33**: 461-465.
- Zhao, Z. Y., T. J. Wilson, et al. (2000). "The folding of the hairpin ribozyme: Dependence on the loops and the junction." Rna-a Publication of the Rna Society **6**(12): 1833-1846.

Structures for three coexisting conformers of an RNA four-way junction solved by FRET restrained high-precision structural modeling

Simon Sindbert^{1,*}, Stanislav Kalinin^{1,*}, Hayk Vardanyan¹, Christian Hanke⁴, Tomasz Soltysinski², Grzegorz Lach², Sabine Müller⁵, Janusz Bujnicki³, Holger Gohlke⁴, Claus A. M. Seidel¹

¹Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, Universitätsstraße 1, Geb 26.32, 40225 Düsseldorf, Germany, ²International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland, ³Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznan, Poland, ⁴Institute of Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University, Universitätsstr. 1, 40225 Düsseldorf, Germany, ⁵Ernst-Moritz-Arndt-Universität Greifswald, Institut für Biochemie, Bioorganische Chemie, Felix-Hausdorff-Strasse 4, 17487, Greifswald, Germany

* contributed equally

Supporting Information

Table of Contents

1	Materials	3
1.1	Oligonucleotides	3
1.2	DNA sequences	3
1.3	Modified U and dG building blocks	3
1.4	Secondary structures of hairpin ribozyme and 4H RNA four-way junction	4
2	Procedures	5
2.1	Hybridization of the RNA four-way junction	5
2.2	Measurement buffer	5
3	Methods	7
3.1	Calculation of dye distributions via the AV approach	7
3.2	AVs for U and dG labeling positions	7

3.3	Fluorescence spectroscopy	8
3.3.1	Time-resolved polarized fluorescence experiments and data analysis.	8
3.3.2	Single-molecule multi-parameter fluorescence detection (smMFD)	9
3.4	Distance determination via Photon Distribution Analysis (PDA)	10
3.5	Static FRET line and distribution of possible σ_{DA} -values	11
3.6	Determination of detection efficiency ratio g_G/g_R	13
3.7	Confidence intervals for fit parameters in PDA	13
3.8	Rigid body docking	13
3.9	$R_{mp}-\langle R_{DA} \rangle_E$ conversion function.....	14
3.10	Model discrimination of docking results.....	14
3.11	Estimation of model precision by bootstrapping.....	15
3.12	Molecular Dynamics simulations	15
4	Results	17
4.1	Measurement of D only fluorescence decays quantum yields Φ_{FA}	17
4.2	Fitting data set of $(D)\beta 11c_-(A)\delta 23d$ with only 2 FRET states	19
4.3	Distances and errors for measurements at 20 mM $MgCl_2$ and model distances for the rigid body models.....	20
4.4	Fit parameters for all PDA fits	21
4.5	Conformer $(ad)_a$ at 0.1 mM $MgCl_2$	25
4.6	Rigid body models.....	28
4.7	“2 nd best” docking results	29
4.8	SimRNA models.....	30
5	References	31

1 Materials

1.1 Oligonucleotides

Ultrapure labeled RNA and DNA with hexamethylen linkers (PAGE Grade), and all unlabeled sequences were purchased from Purimex (Grebenstein, Germany). See Figure 1.1 for chemical structures of modified U and dG nucleotides with linker and dye and Figure 1 in the main text for the sequences and labeling positions of the RNA strands.

1.2 DNA sequences

Sequences and labeling positions (green for Alexa488 and red for Cy5) for the dsDNA used for calibration of the detection efficiency ratio (see Section 3.6).

5' -d(GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT **T****A** CAA CTC CGA AAT AGG CCG) -3'

5' -d(CGG CCT ATT TCG GAG TTG TAA ACA GAG ATC GCC **T****A** AAC GTT CGC CTA GAC TAG TCC AAG TAT TGC) -3'

1.3 Modified U and dG building blocks

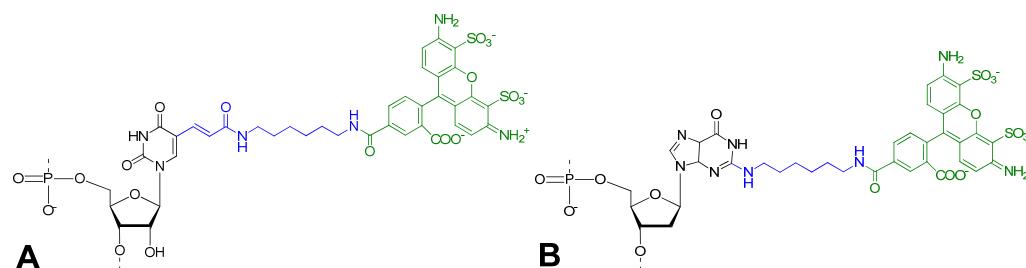


Figure 1.1. Structures of modified U (A) and dG (B) nucleotides with hexamethylen (C6) linkers (blue) and fluorescent dyes Alexa488 (green).

1.4 Secondary structures of hairpin ribozyme and 4H RNA four-way junction

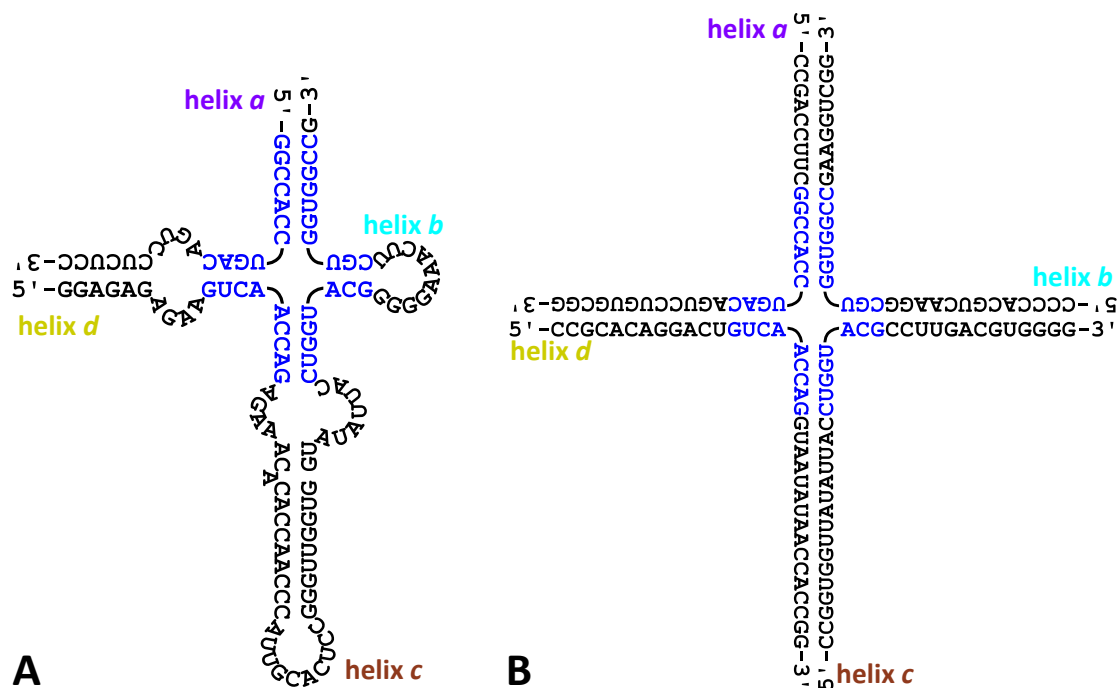


Figure 1.2. Secondary structures of (A) the hairpin ribozyme (Rupert, Massey et al. 2002) (PDB-ID: 1M5K) and (B) the RNA four-way junction used in this work. The matching base pairs are depicted in blue.

2 Procedures

2.1 Hybridization of the RNA four-way junction

The hybridization buffer contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl, and 20 mM MgCl_2 , pH 6.5. The concentration of donor-labeled (or acceptor labeled for AO samples) RNA single strand in the buffer ranged between 0.5 and 2 μM . For FRET molecules the ratio between the amount of donor, acceptor and the two unlabeled strands was 1 : 3 : 4 : 4. For donor and acceptor only molecules the ratio of labeled to the three unlabeled strands was 1 : 4 : 4 : 4. The solution was heated up to a temperature of 85°C inside a thermo-cycler (primus 96 advanced, peqLab, Erlangen, Germany) with 0.1°C/s and was then cooled down to 27°C with 2°C/h. It was then quickly cooled down to 4°C.

2.2 Measurement buffer

The measurement buffer contained 20 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, 100 mM KCl and between 0.03 and 20 mM MgCl_2 , pH 6.5. Additionally, approximately 0.5 mM of Trolox (Rasnik, McKinney et al. 2006) was added.

3 Methods

3.1 Calculation of dye distributions via the AV approach

Dye distributions were modeled using the AV approach (Cai, Kusnetzow et al. 2007; Muschielok, Andrecka et al. 2008) according to (Sindbert, Kalinin et al. 2011). The dyes are approximated by a sphere with an empirical radius of R_{dye} , where the central atom of the fluorophore is connected by a flexible linkage of a certain effective length L_{link} and width w_{link} to the nucleobase. The overall length of the linkage is given by the actual length of the linker and the internal chemical structure of the dye. A geometric search algorithm finds all dye positions within the linkage length from the attachment point which do not cause steric clashes with the macromolecular surface. All allowed positions are considered as equally probable which allows one to define an accessible volume for the dye (AV). To take the three quite different dimensions of a fluorophore into account, we used the real physical dimensions for each calculation of a position distribution and performed three independent AV simulations with three different radii $R_{\text{dye}(i)}$ and superimposed them. Thus, the obtained position distribution represents an average weighted by the number of allowed positions. Throughout this work the following values for $R_{\text{dye}(i)}$, L_{link} and w_{link} were used: $R_{\text{dye}(1)} = 5 \text{ \AA}$, $R_{\text{dye}(2)} = 4.5 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for Alexa488 (D) and $R_{\text{dye}(1)} = 11 \text{ \AA}$, $R_{\text{dye}(2)} = 3 \text{ \AA}$ and $R_{\text{dye}(3)} = 1.5 \text{ \AA}$ for Cy5 (A), $L_{\text{link}} = 20 \text{ \AA}$ for Alexa488 and $L_{\text{link}} = 22 \text{ \AA}$ for Cy5, $w_{\text{link}} = 4.5 \text{ \AA}$. As attachment atoms, C5s and C2s were chosen for labeled Us and dGs (see Figure 1.1), respectively.

3.2 AVs for U and dG labeling positions

In Figure 3.1, accessible volumes calculated for dyes attached to modified U and dG bases are compared. It is apparent that mean dye position is much farther from the helical axis ($\sim 17 \text{ \AA}$) in the case of dG labeling. This helps us to distinguish “quasi-mirrored” structures. In contrast, in the case of U labeling, the distance from the helical axis is much smaller ($\sim 10 \text{ \AA}$), and “mirroring” of a structure has less effect on the mean dye position. As a result, without FRET distances involving (A)Y24a and (A) $\delta 10a$ positions, rigid body docking finds two solutions with similar χ^2_r for the (ad)_a conformer.

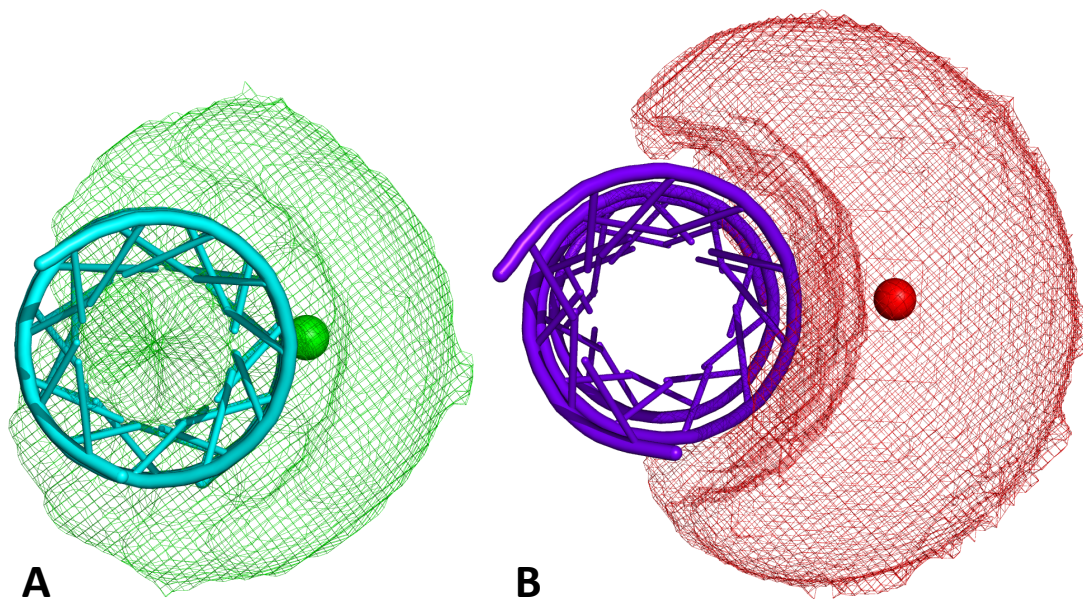


Figure 3.1: Accessible volumes of Alexa488 (A, green) and Cy5 (B, red) attached at positions (D)β8c (A, modified U) and (A)δ10a (B, modified G), respectively. The mean dye positions are displayed as green and red spheres. The following parameters were used for the AV simulation: Alexa488: $L_{\text{linker}} = 20 \text{ Å}$; $w_{\text{linker}} = 4.5 \text{ Å}$; $R_{\text{dye}(1)} = 5 \text{ Å}$; $R_{\text{dye}(2)} = 4.5 \text{ Å}$; $R_{\text{dye}(3)} = 1.5 \text{ Å}$; Cy5: $L_{\text{linker}} = 22 \text{ Å}$; $w_{\text{linker}} = 4.5 \text{ Å}$; $R_{\text{dye}(1)} = 11 \text{ Å}$; $R_{\text{dye}(2)} = 3 \text{ Å}$; $R_{\text{dye}(3)} = 1.5 \text{ Å}$.

3.3 Fluorescence spectroscopy

3.3.1 Time-resolved polarized fluorescence experiments and data analysis.

Ensemble time-correlated single-photon-counting (eTCSPC) measurements were performed using an IBH-5000U (IBH, Scotland) system. The excitation sources were either a 470 nm diode laser (LDH-P-C 470, Picoquant, Berlin, Germany) operating at 8 MHz for donor excitation or a 635 nm diode laser (LDH-8-1 126, Picoquant, Berlin, Germany) operating at 10 MHz for direct acceptor excitation. The emission wavelength was set to 520 nm for donor emission and to 665 nm for acceptor emission, respectively. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path) resolution. Additional cut-off filters were used to reduce the contribution of the scattered light (>500 nm for donor and >640 nm for acceptor emission, respectively). All measurements were performed at room temperature. The concentrations of RNA molecules were kept below 1 μM .

Fluorescence intensity decay curves were fitted using the iterative re-convolution approach (O'Connor and Phillips 1984). The maximum number of counts was typically 50,000. The fits approximately range from the maximum of the instrument response functions (IRF) to the first time channel with less than 100 detected photons. The fluorescence decays $F(t)$ were modeled by double exponential decays (Eq. 3.1):

$$F(t) = x_1 \exp(-t/\tau_1) + x_2 \exp(-t/\tau_2)$$

Eq. 3.1

$$\text{with } \langle \tau \rangle_x = x_1 \tau_1 + x_2 \tau_2$$

3.3.2 Single-molecule multi-parameter fluorescence detection (smMFD)

The fluorescent donor molecules (Alexa 488) are excited by a linearly polarized, active-mode-locked Argon-ion laser (Innova Saber, Coherent, Santa Clara, CA, USA, 496.5 nm, 73.5 MHz, ~ 300 ps) or by a 485 nm diode laser (LDH-D-C 485, Picoquant, Berlin, Germany) operating at 64 MHz. The laser light is focused into the dilute solution (< 50 pM) of labeled molecules by a 60x/1.2 water immersion objective. Each molecule generates a brief burst of fluorescence photons as it traverses the detection volume. This photon-train is divided initially into its parallel and perpendicular components via a polarizing beamsplitter and then into a wavelength ranges below and above 595 nm by using a dichroic beamsplitter (595 DCXR, AHF, Tübingen, Germany). Additionally, red (HQ 720/150 nm for Cy5) and green (HQ 533/46 nm for Alexa 488 and Rh110) bandpass filters (both made by AHF, Tübingen, Germany) in front of the detectors ensure that only fluorescence photons coming from the acceptor and donor molecules are registered. An estimate of the focal geometry is acquired by determining the diffusion correlation time of 200 ± 13 μ s for Rhodamine 110 and knowing its diffusion coefficient of 0.34 ± 0.03 μ m²/ms. Detection is performed using four avalanche photodiodes (SPCM-AQR-14, Laser Components, Germany or alternatively for the green channels PDM050CTC, or τ -SPAD-100, both PicoQuant, Berlin, Germany). The signals from all detectors are guided through a passive delay unit and two routers to two synchronized time-correlated single photon counting boards (SPC 132 or SPC 832, Becker and Hickl, Berlin, Germany) connected to a PC. Bursts of fluorescence photons are distinguished from the background of 1-2 kHz by applying certain threshold intensity criteria [Eggeling2001]. Bursts during which bleaching of the acceptor occurs are excluded from further analysis by applying a criterion regarding the difference in macroscopic times, $|T_G - T_R| < 0.5$ ms, where T_G and T_R are the average

macroscopic times in which all photons have been detected in the green and red channels respectively during one burst (Eggeling, Widengren et al. 2006).

3.4 Distance determination via Photon Distribution Analysis (PDA)

As specified in the main text, we calculate DA distances (R_{DA}) by measuring the fluorescence intensities of D and A (F_D and F_A , respectively). In Eq. 1 in the main text we use the reduced Förster radius R_{0r} (Rothwell 2002), which, in contrast to the Förster radius R_0 (Sisamakris, Valeri et al. 2010), does not depend on the quantum yield of D in absence of FRET $\Phi_{FD(0)}$. Throughout this $R_{0r} = 53.97 \text{ \AA}$ (using $R_{0r} = 53.97 \text{ \AA}$ is equivalent to using $R_0 = 52 \text{ \AA}$ and $\Phi_{FD(0)} = 0.8$). F_D and F_A can be calculated from the signals measured in the green and red detection channels S_G and S_R , respectively, via Eq. 3.2 and Eq. 3.3:

$$F_D = \frac{F_G}{g_G} = \frac{S_G - \langle B_G \rangle}{g_G} \quad \text{Eq. 3.2}$$

$$F_A = \frac{F_R}{g_R} = \frac{S_R - \alpha F_G - \langle B_R \rangle}{g_R} \quad \text{Eq. 3.3}$$

where F_G and F_R are the fluorescence signals in the green and the red signal channels, respectively, α is the crosstalk factor which is determined as the ratio between donor photons detected in the red channels and those detected in the green channels ($\alpha = F_{R(D)} / F_{G(D)}$) for the D only labeled sample, g_G and g_R are the detection efficiencies in the green and red channels, respectively (see Section 3.6 for the determination of g_G/g_R), and $\langle B_G \rangle$ and $\langle B_R \rangle$ are the mean background intensities in the green and red channels, respectively.

To accurately predict the shape of S_G/S_R (or equivalently F_D/F_A) histograms in the presence of FRET we use PDA, which explicitly takes into account shot noise, background contributions and additional broadening due to complex acceptor photophysics (Antonik, Felekian et al. 2006, Kalinin, 2010 #2744; Kalinin, Felekian et al. 2007; Kalinin, Felekian et al. 2008). PDA calculates the probability of observing a certain combination of photon counts $P(S_G, S_R)$

$$P(S_G, S_R) = \sum_{F_G+B_G=S_G; F_R+B_R=S_R} P(F)P(F_G, F_R|F)P(B_G)P(B_R) \quad \text{Eq. 3.4}$$

The intensity distribution of the fluorescence only contribution to the signal, $P(F)$, is obtained from the total measured signal intensity distribution $P(S)$ by deconvolution assuming that the background signals B_G and B_R obey Poisson distributions, $P(B_G)$ and $P(B_R)$, with known mean intensities $\langle B_G \rangle$ and $\langle B_R \rangle$. $P(F_G, F_R | F)$ represents the conditional probability of observing a particular combination of green and red fluorescence photons, F_G and F_R , provided the total number of registered fluorescence photons is F , and can be expressed as a binomial distribution (Antonik, Felekyan et al. 2006). Subsequently, $P(S_G, S_R)$ may be further manipulated to generate a theoretical histogram of any FRET-related parameter as discussed elsewhere (Kalinin, Felekyan et al. 2007).

In this work we used a model accounting for up to three FRET states and a D-only contribution. Additional broadening of FRET states was accounted for by a global parameter σ_{app} as justified in (Kalinin, Sisamakos et al. 2010). In some cases an impurity (mostly 1-3%) with an apparent distance of typically 70 - 90 Å had to be taken into account. This state was in most of cases present in respective D-only samples with amplitude of a few percent. Thus, for n FRET states $2n+1$ to $2n+3$ fit parameters were required depending on whether the impurity state was considered. The fit quality was judged by the reduced chi-squared (χ^2_r) parameter and by visually inspecting weighted residuals plots.

In those cases when only 2 FRET states (one major and one minor) were visible the second “invisible” minor state was assumed to be overlapped by either the visible minor or the visible major depending on their relative amplitudes. If the amplitude of the visible minor state was larger than 25 %, it was assumed that this FRET state contained both minor states. Thus, two states with the same distance and amplitude were used. Else, the “invisible” minor was assumed to have a relative amplitude of 10 % and the same distance as the major FRET state.

3.5 Static FRET line and distribution of possible σ_{DA} -values

The static FRET line represents the expected dependence between FRET indicators derived from intensities (e.g. F_D/F_A) and the fluorescence lifetime of the donor. In the simplest case it is given by the well-known equation $E = 1 - \tau_{\text{DA}}/\tau_{\text{D}(0)}$. In reality this relationship does not hold because the distributions of

donor-acceptor distances due to flexible dye linkers are not accounted for. In addition, non-exponential fluorescence decay of the donor dye itself must be considered (see Section 4.1 for the fluorescence decay fits for all D positions). These effects can be corrected for as described in (Kalinin, Valeri et al. 2010; Kalinin 2012). As there is no analytical expression for the $E(\tau_{DA})$ dependence that considers the above effects, a polynomial approximation is used. In this work we used the following approximation:

$$F_D/F_A = \frac{\Phi_{FD(0)}}{\Phi_{FA}} \left/ \left(\frac{\langle \tau_{D(0)} \rangle_x}{c_3 \langle \tau_{D(A)} \rangle_f^3 + c_2 \langle \tau_{D(A)} \rangle_f^2 + c_1 \langle \tau_{D(A)} \rangle_f + c_0} - 1 \right) \right. \quad \text{Eq. 3.5}$$

Where $\langle \tau_i \rangle_x$ and $\langle \tau_i \rangle_f$ are species and fluorescence averaged mean lifetimes, respectively. The polynomial coefficients $c_{(i)}$ are calculated assuming 6 Å half-width of the DA distance distribution (σ_{DA}) (Sindbert, Kalinin et al. 2011). See Table 4-1 for the resulting coefficients. For highly asymmetric AVs (see Section 3.2) σ_{DA} depends also on the mutual orientation of D and A clouds, which implies that individual σ_{DA} -values should be used for different samples and/or even FRET states. Considering various possible orientations of calculated dyes' AVs, we estimated that σ_{DA} can vary between ca. 5.5 and 12 Å being somewhat correlated with $\langle R_{DA} \rangle$ (Figure 3.2). To fit static FRET lines to the observed FRET states, values of σ_{DA} between 6 and 9 Å were required, which is within the expected range.

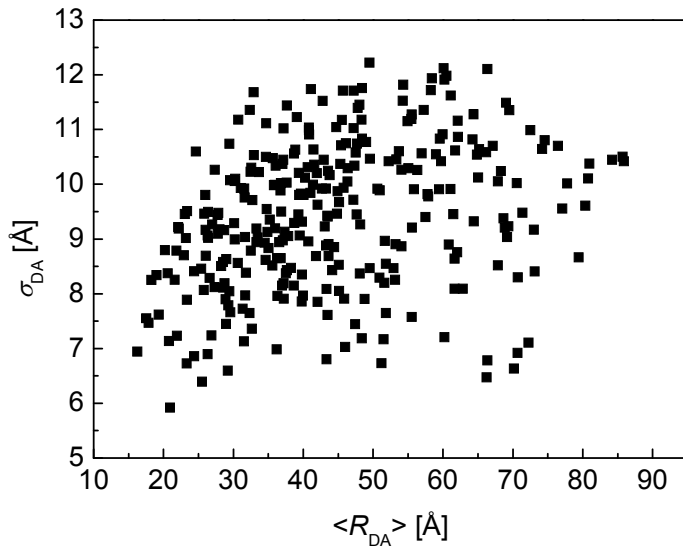


Figure 3.2. Possible values of σ_{DA} calculated for randomly positioned and oriented AVs of Alexa488 and Cy5 initially calculated for positions (D) β 8c and (A) δ 23d, respectively.

3.6 Determination of detection efficiency ratio g_G/g_R

To be able to convert model distances into probabilities of observing green photons, the detection efficiency ratio g_G/g_R is needed (g_G and g_R stand for the detection efficiencies of “green” and “red” channels, respectively). These values are calculated for each measurement session by requiring that the linker-corrected static FRET line (Kalinin, Valeri et al. 2010) (see Eq. 3.5) goes through the observed FRET population in a 2D histogram of F_D/F_A vs $\tau_{D(A)}$ for a measurement of a FRET labeled dsDNA (see Section 1.2). For the FRET line it is assumed that $\sigma_{DA} = 6 \text{ \AA}$, $\Phi_{FD(0)} = 0.8$ and $\tau_{D(0)} = 4.1 \text{ ns}$ (mono-exponential decay).

3.7 Confidence intervals for fit parameters in PDA

To estimate the errors of fitted parameters due to photon statistics, $\{\Delta R_{DA}(E)\}$, we explore the parameter space for sets of variables providing acceptable fits. All free fit parameters are varied simultaneously in a random manner. The χ^2_r -value is calculated at 100000 random points yielding 100-1000 points with χ^2_r -values below $\chi^2_{r,min} + (2/N_{bins})^{1/2}$ (here N_{bins} is the number of histogram bins, and $\chi^2_{r,min}$ is the reduced chi-squared of the best fit). The range where such fits are possible is assigned as 1σ confidence interval. Whereas one could calculate χ^2_r thresholds more strictly from the chi-squared distribution (Soong 2004), in practice $\chi^2_{r,min}$ is often affected by experimental imperfections and can be considerably larger than one. For this reason, we prefer the simple test mentioned above which relates χ^2_r values to that of the best fit. For the cases of overlapped minor states, their amplitudes were allowed to vary by up to $\pm 50 \%$ while exploring the parameter space.

3.8 Rigid body docking

Rigid body docking was performed as described in ref. (Kalinin 2012). Briefly, we used a mechanical model where the four RNA helices were assumed to be rigid bodies. Mean dye positions were rigidly fixed to the labeled RNA helix and connected with “springs” with relaxed lengths given by the corresponding values of R_{mp} . The strengths of the “springs” were derived from experimental errors ΔR_{DA} .

The system was then relaxed to the nearest minimum of its “energy”, which corresponds to a local χ_r^2 minimum. The procedure was repeated many times for random initial conditions, which ensures exhaustive sampling of the configuration space. In a second run (“refinement”) the AVs are re-modeled for all solutions found in the first run. This takes into account possible sterical clashes of fluorophores with RNA4WJ arms they are *not* attached to. All structures are then optimized using the new AVs. The solutions were sorted by χ_r^2 and clustered as shown in Figure 4D in the main text.

3.9 R_{mp} - $\langle R_{DA} \rangle_E$ conversion function

Although the optimization problem (Eq 3 in the main text) can be defined for $\langle R_{DA} \rangle_E$ values, this would require re-calculation of the dyes’ AVs after each iteration. To avoid this, we generated a conversion function between R_{mp} and $\langle R_{DA} \rangle_E$. This was done by generating a large number of random orientations and positions of dyes’ AVs. For each pair of dye clouds, R_{mp} and $\langle R_{DA} \rangle_E$ were calculated. The resulting $R_{mp}(\langle R_{DA} \rangle_E)$ dependence was approximated with a 3rd order polynomial and used as a conversion function. The RMS deviation between the polynomial approximation and the $R_{mp}(\langle R_{DA} \rangle_E)$ dependence was typically 0.5-0.7 Å.

3.10 Model discrimination of docking results

Model discrimination is done according to (Kalinin 2012). Solutions are considered ambiguous if the respective χ_r^2 values do not differ significantly. We typically apply a threshold given by Eq. 3.6:

$$\chi_r^2 < \chi_{r,\min}^2 + [2/(n-p)]^{1/2} = \chi_{r,\max}^2 \quad \text{Eq. 3.6}$$

where n is the number of distance restraint (51, the four restraints between the helix ends at the junction are not considered here as they cause a reduced number of degrees of freedom) and p is the number of degrees of freedom (9, not 18 as justified before) and which roughly corresponds to the variance of the chi-squared distribution of $2 \times (\text{degrees of freedom})$ (Soong 2004) (magenta lines in Figure 4D in the main text). The fact that $\chi_{r,\min}^2$ is often larger than one is attributed to systematic

experimental errors and to possible violations of the AV and/or rigid body models. Other criteria defining different levels of significance can be applied here in a straightforward way.

3.11 Estimation of model precision by bootstrapping

The precision of rigid body models was estimated by bootstrapping (Efron 1986) as described in (Kalinin 2012). Briefly, all model distances found for the optimized structures were simultaneously perturbed by adding normally distributed random numbers with a mean of zero and the standard deviation given by the experimental errors $\{\Delta R_{DA}\}$. Afterwards the structures were re-optimized using the perturbed distances. This procedure was repeated 100 times yielding a set of structures representing the distribution of possible positions of the helices. For this set, RMSD values were calculated for each phosphate atom with respect to the original model. The average RMSD over all P atoms of a structure is used to characterize the overall precision of the model.

3.12 Molecular Dynamics simulations

Molecular dynamics (MD) simulations were performed with the Amber 11 suite of programs (Case, Cheatham et al. 2005), together with the force field as described by Hornak et al. (Hornak, Abel et al. 2006), using modifications suggested by Pérez et al. (Perez, Marchan et al. 2007), Banás et al. (Banas, Hollas et al. 2010), and Joung et al. (Joung and Cheatham 2008).

The starting structure for each conformer, which was obtained from rigid body docking, was placed in an octahedral periodic box of TIP3P water molecules (Jorgensen, Chandrasekhar et al. 1983). The distance between the edges of the water box and the closest atom of the RNA was at least 11 Å, resulting in a system of ~250,000 atoms.

The system was minimized by 50 steps of steepest descent minimization followed by 450 steps of conjugate gradient minimization. The particle mesh Ewald method (Darden, York et al. 1993) was used to treat long range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using SHAKE (Ryckaert, Ciccotti et al. 1977). The time step for all MD simulations was 2 fs,

with a direct-space nonbonded cutoff of 8 Å. Applying harmonic restraints with force constants of 5 kcal mol⁻¹ Å⁻² to all solute atoms, we carried out canonical ensemble (NVT)-MD for 50 ps, during which the system was heated from 100 to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD was used for 150 ps to adjust the solvent density. Finally, the force constants of the harmonic restraints on solute atom positions were gradually reduced to 1 kcal mol⁻¹ Å⁻² during 50 ps of NVT-MD.

From the following 40 ns of NVT-MD at 300 K, conformations were extracted every 20 ps. Harmonic restraints were applied on phosphorus atoms further away than six bases from the junction region, using force constants that have been chosen such that the positional uncertainties of the phosphorous atoms as calculated by bootstrapping for the rigid body docking models are reproduced.

4 Results

4.1 Measurement of D only fluorescence decays quantum yields Φ_{FA}

According to Eq. 1 in the main text, for the determination of R_{DA} the acceptor quantum yield Φ_{FA} needs to be known (see also Section 3.4). Furthermore, to determine the static FRET line the fluorescence decay of the donor in absence of FRET needs to be determined (see Section 3.5). It has been shown that quenching of Alexa488 and Cy5 attached to dsRNAs using long C6 dye linkers is mostly dynamic (Sindbert, Kalinin et al. 2011). Thus, to determine Φ_{FA} and the D fluorescence decays for each A and D labeling position, respectively, eTCSPC measurements for the single labeled molecules were performed (see Section 3.3.1). Values for Φ_{FA} were assumed to be proportional to the species averaged lifetime $\langle \tau \rangle_x = x_1 \tau_1 + x_2 \tau_2$ and calculated by calibrating against a sample labeled with Cy5 with a known fluorescence quantum yield. For calibration we used single labeled dsDNA with a C6-hexamethylen linker and Cy5 with $\langle \tau \rangle_x = 1.16$ ns and $\Phi_{FA} = 0.32$ due to the presence of $\sim 20\%$ cis-trans isomerization under single-molecule conditions (instead of $\Phi_{FA} = 0.4$ expected for ensemble measurements) (Woźniak, Schröder et al. 2008; Sindbert, Kalinin et al. 2011). For the calibration of g_G/g_R with the dsDNA (see Section 3.6) we also use $\Phi_{FA} = 0.32$ for the static FRET line. Therefore, even if the determined values for Φ_{FA} are wrong, errors due to wrong calibration cancel out and Eq. 1 in the main text will yield the correct distance R_{DA} . For D samples, we assumed that free Alexa488 has $\tau_{D(0)} = 4.1$ ns and $\Phi_{FD(0)} = 0.8$ (Woźniak, Schröder et al. 2008; Sindbert, Kalinin et al. 2011). The results of the fits of the fluorescence decays and resulting values for the fluorescence and the species averaged lifetimes ($\langle \tau \rangle_x$ and $\langle \tau \rangle_f$, respectively) and for Φ_{FA} are compiled in Table 4-1. The fitted fluorescence decays of the D-only molecules were used to calculate static FRET lines according to Section 3.5 while assuming $\sigma_{DA} = 6$ Å. The resulting polynomial coefficients are listed in Table 4-1.

Table 4-1. Fluorescence lifetimes (τ) and quantum yields (Φ_{FA}) of single labeled RNA4WJ samples

D-only					
	τ_1 , ns (x_1)	τ_2 , ns (x_2)	$\langle \tau \rangle_x$, ns	$\langle \tau \rangle_f$, ns	
(D) $\beta 5c$	4.10(92%)	1.34(8%)	3.87	4.02	
(D) $\beta 8c$	4.03(90%)	1.23(11%)	3.74	3.93	
(D) $\beta 11c$	4.03(95%)	2.28(5%)	3.94	3.98	
(D) $\beta 14c$	4.12(93%)	0.61(7%)	3.86	4.08	
(D) $\beta 27b$	3.87(79%)	0.40(21%)	3.15	3.77	
(D) $\gamma 8b$	3.85(81%)	1.18(19%)	3.34	3.67	
(D) $\gamma 29a$	3.62(75%)	0.79(25%)	2.92	3.43	
(D) $\delta 7a$	3.99(82%)	0.51(18%)	3.35	3.89	
A-only					
	τ_1 , ns (x_1)	τ_2 , ns (x_2)	$\langle \tau \rangle_x$, ns	$\langle \tau \rangle_f$, ns	Φ_{FA}
(A) $\alpha 12d$	1.90(27%)	1.03(74%)	1.26	1.38	0.35
(A) $\beta 14c$	1.82(23%)	1.03(77%)	1.21	1.30	0.34
(A) $\beta 27b$	1.63(27%)	0.99(73%)	1.16	1.23	0.32
(A) $\beta 33b$	1.91(72%)	1.14(28%)	1.69	1.76	0.47
(A) $\gamma 8b$	1.81(25%)	1.03(75%)	1.23	1.32	0.34
(A) $\gamma 24a$	1.53(41%)	0.93(59%)	1.17	1.25	0.32
(A) $\delta 10a$	1.68(39%)	0.99(61%)	1.26	1.35	0.35
(A) $\delta 23d$	1.65(22%)	1.00(78%)	1.15	1.21	0.32
(A) $\delta 26d$	1.76(62%)	1.20(38%)	1.54	1.59	0.43
(A) $\delta 28d$	1.90(79%)	1.13(21%)	1.73	1.79	0.48

Table 4-2. Polynomial coefficients c_i used for the calculation of the static FRET-lines (see Section 3.5).

D position	c_0	c_1	c_2	c_3
(D) $\beta 5c$	-0.0501	0.5383	0.2748	-0.0391
(D) $\beta 8c$	-0.0516	0.5528	0.2795	-0.0413
(D) $\beta 11c$	-0.0449	0.5003	0.2837	-0.0391
(D) $\beta 14c$	-0.0518	0.5563	0.2638	-0.0373
(D) $\beta 27b$	-0.0672	0.7212	0.2455	-0.0443
(D) $\gamma 8b$	-0.0580	0.6107	0.2937	-0.0500
(D) $\gamma 29a$	-0.0675	0.7118	0.2845	-0.0569
(D) $\delta 7a$	-0.0644	0.6786	0.2494	-0.0418

4.2 Fitting data set of $(D)\beta 11c_{-}(A)\delta 23d$ with only 2 FRET states

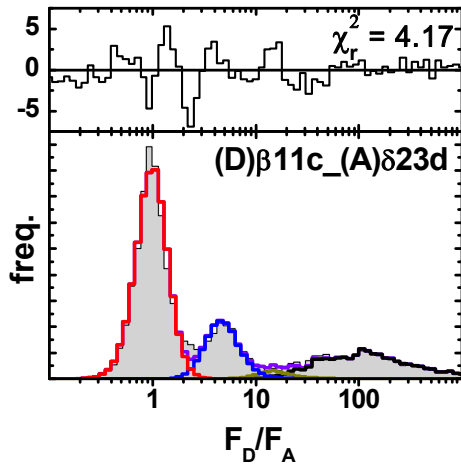


Figure 4.1. PDA for sample $(D)\beta 11c_{-}(A)\delta 23d$ (selected bursts). F_D/F_A histogram of experimental data (gray area) is fitted (purple solid line) using the following parameters: 47.5 % of $\langle R_{DA} \rangle_{E(1)} = 44.5 \text{ \AA}$ (red); 13.9 % of $\langle R_{DA} \rangle_{E(2)} = 57.8 \text{ \AA}$ (blue); $\sigma_{app} = 4.5 \text{ %}$ of $\langle R_{DA} \rangle_{E(1)}$; 36.5 % of D-only; 2.0 % of impurities with apparent $R_{DA} = 70.5 \text{ \AA}$; $\chi_r^2 = 4.17$. Weighted residuals are shown in the upper plot.

4.3 Distances and errors for measurements at 20 mM MgCl₂ and model distances for the rigid body models

Table 4-3. Values for measured distances $\langle R_{DA} \rangle_E$ and their relative amplitudes x , measurement errors ΔR_{DA} and model distances R_{model} resulting from rigid body docking.

		$(ad)_a$			$(ad)_p$			$(ab)_a$		
DA-pair	σ_{app}	$\langle R_{DA} \rangle_E, \text{\AA}$ (x_1)	ΔR_{DA}	R_{model}	$\langle R_{DA} \rangle_E, \text{\AA}$ (x_2)	ΔR_{DA}	R_{model}	$\langle R_{DA} \rangle_E, \text{\AA}$ (x_3)	ΔR_{DA}	R_{model}
(D) β 5c_(A) α 12d	4.1%	53.4(78.9%)	5.1%	51.6	53.4(10.2%)	15.5%	50.5	60.2(10.9%)	13.6%	55.3
(D) β 8c_(A) α 12d	3.9%	58.7(73.8%)	5.0%	52.8	58.7(11.8%)	10.9%	52.1	48.3(14.5%)	5.5%	50
(D) β 11c_(A) α 12d	3.8%	56.4(85.6%)	5.0%	50.6	46.6(9.7%)	5.7%	47.1	41(4.7%)	7.0%	43.2
(D) β 14c_(A) α 12d	5.9%	44.2(78.3%)	5.0%	38.8	44.2(10.8%)	10.4%	36.2	44.2(10.8%)	10.4%	35.6
(D)Y8b_(A) α 12d	4.9%	46.8(82.2%)	5.0%	43.9	36.9(8.9%)	11.6%	41	36.9(8.9%)	11.6%	38.1
(D)Y29a_(A) α 12d	3.6%	55.1(85.7%)	5.2%	52.1	47.7(4%)	9.7%	51.1	55.1(10.3%)	11.6%	50.9
(D) δ 7a_(A) α 12d	4.6%	39.8(70.5%)	5.1%	36.6	49.1(14.8%)	10.6%	36.9	49.1(14.8%)	10.6%	34.3
(D)Y8b_(A) β 14c	4.3%	44.3(80.8%)	5.2%	43.6	50.5(7.5%)	7.2%	43.6	44.3(11.7%)	13.5%	41
(D)Y29a_(A) β 14c	4.5%	52.2(83.8%)	5.1%	54.9	42.2(3.1%)	6.7%	39.8	56.1(13.1%)	10.7%	69.5
(D) δ 7a_(A) β 14c	4.3%	43.9(88%)	5.0%	48.6	38.9(6.2%)	6.7%	30.8	51(5.8%)	6.3%	52.6
(D)Y29a_(A) β 27b	3.7%	42.1(69.5%)	5.0%	40.9	50.8(7.6%)	7.9%	51.2	60.5(22.9%)	6.7%	58.9
(D) δ 7a_(A) β 27b	3.7%	39.3(66.4%)	5.1%	37.1	46.3(13.5%)	7.8%	49.9	51.8(20.1%)	5.8%	50.8
(D)Y29a_(A) β 33b	4.3%	49.5(80.7%)	5.0%	51.5	59.8(8.1%)	6.5%	63.3	75.1(11.2%)	10.4%	79.6
(D) δ 7a_(A) β 33b	4.4%	52.7(67.9%)	5.0%	54.6	59.7(12.6%)	5.7%	65.3	73.4(19.4%)	5.1%	66.1
(D) β 11c_(A)Y24a	3.5%	51.7(82.3%)	5.1%	47.4	44.9(6.8%)	5.8%	46.7	51.7(10.9%)	9.3%	53.8
(D) β 8c_(A)Y24a	4.0%	48.3(66.3%)	5.1%	46.8	54.6(11.9%)	7.7%	58.8	63.2(21.8%)	5.6%	57.4
(D) β 14c_(A)Y24a	3.5%	41.8(57.2%)	5.1%	42.4	49.2(24.6%)	6.7%	41.2	54.7(18.2%)	7.0%	54.3
(D) β 27b_(A)Y24a	3.5%	45.1(54.8%)	5.4%	44.5	41.4(20.7%)	6.3%	40.4	50.3(24.5%)	5.5%	47.4
(D) β 5c_(A)Y24a	4.5%	49.9(57.3%)	5.2%	54.1	45.4(19.5%)	6.5%	57.3	62.6(23.2%)	5.6%	69.9
(D) β 5c_(A)Y8b	4.8%	64.1(73.5%)	5.3%	65.9	64.1(12.2%)	11.4%	65.4	54.3(14.3%)	7.2%	58.4
(D) β 8c_(A)Y8b	4.8%	59.3(72.9%)	5.1%	60.3	59.3(12.7%)	13.3%	60.8	48.9(14.4%)	5.4%	51.8
(D) β 11c_(A)Y8b	4.0%	50.7(78.4%)	5.1%	51.0	50.7(10.4%)	14.4%	49.3	41.2(11.2%)	6.6%	40
(D) β 14c_(A)Y8b	4.7%	46.1(82.5%)	5.3%	43.6	53.1(6.7%)	7.2%	43.6	46.1(10.9%)	9.8%	41.1
(D) δ 7a_(A)Y8b	4.3%	41.2(67%)	5.0%	38.3	48.9(16.5%)	7.2%	48.2	48.9(16.5%)	7.2%	49.2
(D) β 11c_(A) δ 10a	3.5%	52.8(81.6%)	5.1%	48.5	46.5(8.4%)	8.3%	49.2	52.8(9.9%)	12.3%	51.4
(D) β 8c_(A) δ 10a	5.8%	51.6(75.7%)	5.2%	47.0	63(24.3%)	15.9%	60.6	63(0%)	15.9%	53.9
(D) β 27b_(A) δ 10a	3.5%	45.8(58%)	6.2%	46.5	42.2(16.2%)	7.3%	42.3	50.5(25.8%)	7.2%	48.2
(D) β 14c_(A) δ 10a	3.5%	42(59.5%)	5.1%	42.5	48.8(24.4%)	6.4%	42.5	54.3(16.1%)	6.0%	51.5
(D) β 5c_(A) δ 10a	4.5%	49.8(62.6%)	5.1%	53.1	56.4(21%)	6.5%	57.9	67.9(16.5%)	6.2%	65.9
(D)Y8b_(A) δ 10a	3.5%	46.9(77%)	5.1%	47.2	43.6(15.6%)	6.3%	41	52(7.3%)	8.0%	48.5
(D) β 5c_(A) δ 23d	4.4%	44.4(67.4%)	5.0%	41.0	53.7(10.7%)	8.6%	58.8	62.4(21.9%)	5.8%	63
(D) β 8c_(A) δ 23d	4.3%	45.5(68.5%)	5.1%	46.9	53(21.9%)	7.2%	57.2	58.1(9.6%)	9.6%	60.1
(D) β 11c_(A) δ 23d	4.5%	43.9(66.6%)	5.1%	45.1	53(11.6%)	6.5%	55.5	62.6(21.8%)	5.8%	56.6
(D) β 14c_(A) δ 23d	4.5%	39.2(62.5%)	5.2%	35.5	46.8(18.7%)	11.3%	45	46.8(18.7%)	11.3%	44.9
(D) β 27b_(A) δ 23d	4.6%	53.8(82.1%)	5.1%	55.5	46.2(7.4%)	8.5%	44.9	53.8(10.5%)	11.7%	58.2
(D)Y8b_(A) δ 23d	3.8%	51(79.3%)	5.1%	53.4	43.6(9%)	7.0%	46.7	57.2(11.8%)	7.8%	55.2
(D)Y29a_(A) δ 23d	4.6%	65.6(84.9%)	6.7%	68.2	65.6(10.2%)	9.7%	67.6	53.1(4.9%)	6.1%	57.4
(D) β 5c_(A) δ 26d	5.6%	56.6(78.2%)	5.1%	52.2	69.5(10.3%)	7.2%	69.5	85.3(11.4%)	10.1%	76.7

DA-pair	σ_{app}	$\langle R_{\text{DA}} \rangle_{E, \text{Å}} (x_1)$	ΔR_{DA}	R_{model}	$\langle R_{\text{DA}} \rangle_{E, \text{Å}} (x_2)$	ΔR_{DA}	R_{model}	$\langle R_{\text{DA}} \rangle_{E, \text{Å}} (x_3)$	ΔR_{DA}	R_{model}
(D)β8c_(A)δ26d	5.5%	62.5(59.9%)	5.3%	59.9	71.2(20%)	10.3%	70.1	71.2(20%)	10.3%	71.6
(D)β11c_(A)δ26d	5.4%	58.8(79.4%)	5.1%	58.7	70.6(10.3%)	11.8%	66.9	70.6(10.3%)	11.8%	66.3
(D)β14c_(A)δ26d	5.5%	51.6(78%)	5.1%	47.0	62.7(11%)	13.4%	54.3	62.7(11%)	13.4%	55.3
(D)β27b_(A)δ26d	4.7%	63(74.4%)	5.4%	63.0	52.4(13.9%)	9.2%	48.1	63(11.7%)	13.1%	61.1
(D)Y8b_(A)δ26d	6.0%	56.8(82.4%)	5.5%	61.1	44.4(4.3%)	18.2%	50.8	56.8(13.3%)	15.9%	57.8
(D)Y29a_(A)δ26d	5.5%	78(69%)	5.9%	76.6	66(27.4%)	7.3%	74.3	52.6(3.6%)	10.8%	49.6
(D)β5c_(A)δ28d	5.6%	53.6(78.2%)	5.0%	58.2	67.6(9.9%)	8.1%	69.9	85(11.8%)	22.5%	80.6
(D)β8c_(A)δ28d	5.6%	58.1(76.8%)	5.1%	66.0	73.6(11.6%)	11.7%	72.2	73.55(11.6%)	11.7%	75.7
(D)β11c_(A)δ28d	5.8%	57.7(83.9%)	5.1%	62.8	73(8.1%)	20.6%	70.2	73(8.1%)	20.6%	68.7
(D)β14c_(A)δ28d	5.0%	52.1(69.4%)	5.2%	50.3	62.2(17.3%)	12.7%	57	71.2(13.4%)	13.0%	57.5
(D)β27b_(A)δ28d	5.6%	67.4(64.8%)	5.3%	62.3	56.8(23.1%)	5.4%	54.3	67.4(12%)	13.8%	60.2
(D)Y8b_(A)δ28d	6.0%	61.4(81.6%)	5.3%	60.3	47.2(6.3%)	7.7%	56.9	61.4(12.1%)	21.1%	56.6
(D)Y29a_(A)δ28d	6.0%	80.1(76.5%)	5.7%	79.0	64.9(11.7%)	15.3%	77.3	64.9(11.7%)	15.3%	51.1

4.4 Fit parameters for all PDA fits

Table 4-4 PDA fit parameters for the datasets measured at 20 mM MgCl₂

DA-pair	$\langle R_{\text{DA}} \rangle_{E,1, \text{Å}} (x_1)$	$\langle R_{\text{DA}} \rangle_{E,2, \text{Å}} (x_2)$	$\langle R_{\text{DA}} \rangle_{E,3, \text{Å}} (x_3)$	σ_{app}	D-only	impurities $R_{\text{DA}}, \text{Å} (x)$	$\langle B_G \rangle, \text{kHz}$	$\langle B_R \rangle, \text{kHz}$	α	g_G/g_R	χ_r^2
(D)β5c_(A)α12d	53.4(62.1%)	53.4(8%)	60.2(8.6%)	4.1%	20%	75.9(1.3%)	1.19	0.79	2.4%	0.4	1.48
(D)β8c_(A)α12d	58.7(59%)	48.3(11.6%)	58.7(9.4%)	3.9%	18.9%	74.6(1.1%)	1.13	0.75	2.5%	0.4	1.14
(D)β11c_(A)α12d	56.4(67.1%)	46.6(7.6%)	41(3.7%)	3.8%	20.8%	77.8(0.8%)	1.13	0.75	2.3%	0.4	0.97
(D)β14c_(A)α12d	44.2(62.1%)	44.2(8.6%)	44.2(8.6%)	5.9%	19%	80.5(1.8%)	1.13	0.75	2.9%	0.4	3.05
(D)Y8b_(A)α12d	46.8(56.2%)	36.9(6.1%)	36.9(6.1%)	4.9%	30.6%	57.9(1.1%)	0.99	0.67	2.8%	0.32	2.11
(D)Y29a_(A)α12d	55.1(66.5%)	55.1(8%)	47.7(3.1%)	3.6%	20%	74.8(2.5%)	1.44	0.99	2.3%	0.4	1.89
(D)δ7a_(A)α12d	39.8(51.1%)	49.1(10.7%)	49.1(10.7%)	4.6%	26.7%	69.9(8%)	1.12	0.73	3.2%	0.35	1.78
(D)Y8b_(A)β14c	44.3(60%)	50.5(5.6%)	44.3(8.7%)	4.3%	24.9%	67.3(0.7%)	1.13	0.75	2.7%	0.4	1.06
(D)Y29a_(A)β14c	52.2(59.5%)	42.2(2.2%)	56.1(9.3%)	4.5%	25.9%	72.3(3.1%)	1.27	0.53	1.5%	0.85	1.2
(D)δ7a_(A)β14c	43.9(57.9%)	38.9(4.1%)	51(3.8%)	4.3%	33.1%	62.4(1%)	1.42	0.61	1.6%	0.8	1.02
(D)Y29a_(A)β27b	42.1(44.6%)	50.8(4.9%)	60.5(14.7%)	3.7%	33.3%	68.9(2.5%)	1.43	0.62	2%	0.8	1.03
(D)δ7a_(A)β27b	39.3(38.4%)	46.3(7.8%)	51.8(11.6%)	3.7%	38.7%	71.1(3.5%)	1.43	0.62	1.9%	0.8	1.21
(D)Y29a_(A)β33b	49.5(55.5%)	59.8(5.6%)	75.1(7.7%)	4.3%	28.4%	90.7(2.8%)	1.43	0.62	1.9%	0.8	1.42
(D)δ7a_(A)β33b	52.7(57%)	59.7(10.6%)	73.4(16.3%)	4.4%	13.7%	95.1(2.3%)	1.50	0.56	1.5%	0.69	2.11
(D)β11c_(A)Y24a	48.7(51.4%)	62.4(8.5%)	62.4(8.5%)	5.5%	28.7%	78.9(2.9%)	1.54	0.59	1.5%	0.73	1.48
(D)β8c_(A)Y24a	48.3(42.5%)	54.6(7.6%)	63.2(14%)	4%	35.8%		1.81	0.75	1.4%	0.74	2.4
(D)β14c_(A)Y24a	51.7(59.4%)	44.9(4.9%)	51.7(7.9%)	3.5%	26.4%	65.6(1.3%)	1.74	0.64	1.4%	0.71	1.73
(D)β27b_(A)Y24a	41.7(32.5%)	48.7(12.6%)	53.9(11.9%)	3.5%	41.7%	64.8(1.3%)	1.73	0.72	1.4%	0.74	1.22
(D)β5c_(A)Y24a	45.3(38.6%)	41.6(17.1%)	50.4(16.7%)	3.5%	26.5%	64(1%)	1.54	0.59	1.3%	0.73	1.71
(D)β5c_(A)Y8b	64.1(39.6%)	64.1(6.6%)	54.3(7.7%)	4.8%	44%	77.4(2.1%)	1.13	0.75	2.2%	0.4	0.78
(D)β8c_(A)Y8b	59.3(47.1%)	59.3(8.2%)	48.9(9.3%)	4.8%	32.8%	75.9(2.6%)	1.29	0.55	1.3%	0.7	1.99

DA-pair	$\langle R_{DA} \rangle_{E,1}, \text{ \AA}$ (x_1)	$\langle R_{DA} \rangle_{E,2}, \text{ \AA}$ (x_2)	$\langle R_{DA} \rangle_{E,3}, \text{ \AA}$ (x_3)	σ_{app}	D-only	impurities $R_{DA}, \text{ \AA} (x)$	$\langle B_6 \rangle,$ kHz	$\langle B_R \rangle,$ kHz	α	g_G/g_R	χ_r^2
(D) β 11c_(A)Y8b	50.7(46.9%)	50.7(6.2%)	41.2(6.7%)	4.8%	38.6%	63.8(1.6%)	1.57	0.69	1.3%	0.8	1.38
(D) β 14c_(A)Y8b	46.1(47%)	53.1(3.8%)	46.1(6.2%)	4.7%	41.5%	66.5(1.5%)	1.57	0.69	1.3%	0.8	1.42
(D) δ 7a_(A)Y8b	41.2(41.4%)	48.9(10.2%)	48.9(10.2%)	4.3%	37.1%	61.9(1%)	1.29	0.67	1.3%	0.75	2.74
(D) β 11c_(A) δ 10a	49.8(43.3%)	56.4(14.5%)	67.9(11.4%)	4.5%	28%	83.4(2.8%)	1.95	0.64	1.2%	0.71	2.32
(D) β 8c_(A) δ 10a	51.6(25.2%)	61.6(9.8%)	51.6(5.8%)	5.7%	57.6%	69.1(1.7%)	1.81	0.75	1.3%	0.74	0.78
(D) β 27b_(A) δ 10a	52.8(64.8%)	46.5(6.7%)	52.8(7.9%)	3.5%	19.7%	66.3(0.9%)	1.74	0.64	1.5%	0.71	1.32
(D) β 14c_(A) δ 10a	41.9(35.5%)	47.7(12.7%)	53(13%)	3.5%	37.7%	61.1(1%)	1.73	0.72	1.9%	0.74	1.69
(D) β 5c_(A) δ 10a	45.8(31.2%)	42.1(8.6%)	50.5(13.9%)	3.5%	43.1%	66(3.1%)	1.73	0.72	1.5%	0.74	1.74
(D)Y8b_(A) δ 10a	46.9(56.6%)	52(5.4%)	43.6(11.5%)	3.5%	24%	70.9(2.5%)	1.73	0.72	1.5%	0.73	0.83
(D) β 5c_(A) δ 23d	44.4(32.7%)	53.7(5.2%)	62.4(10.6%)	4.4%	51.5%		1.19	0.79	2.4%	0.4	1.36
(D) β 8c_(A) δ 23d	45.5(38.5%)	53(12.3%)	58.1(5.4%)	4.3%	42%	73.2(1.8%)	1.30	0.54	1.4%	0.85	0.84
(D) β 11c_(A) δ 23d	44.2(45.1%)	51.2(4.9%)	58.8(11.9%)	3.8%	36.2%	71.5(1.9%)	1.66	0.65	1.3%	0.67	1.88
(D) β 14c_(A) δ 23d	39.2(31.7%)	46.8(9.5%)	46.8(9.5%)	4.5%	45.9%	72.6(3.5%)	1.29	0.54	1.5%	0.85	1.5
(D) β 27b_(A) δ 23d	53.8(50.9%)	46.2(4.6%)	53.8(6.5%)	4.6%	36.4%	66.5(1.5%)	1.27	0.53	1.2%	0.85	1.17
(D)Y8b_(A) δ 23d	51(31%)	43.6(3.5%)	57.2(4.6%)	3.8%	60.1%	72(0.9%)	0.99	0.67	2.6%	0.32	0.68
(D)Y29a_(A) δ 23d	65.6(65.7%)	65.6(7.9%)	53.1(3.8%)	4.6%	22.6%		1.57	0.69	1.3%	0.8	1.25
(D) β 5c_(A) δ 26d	56.6(56.7%)	69.5(7.5%)	85.3(8.3%)	5.6%	26.3%	90(1.3%)	0.99	0.67	3%	0.32	1.17
(D) β 8c_(A) δ 26d	62.5(46.1%)	71.2(15.4%)	71.2(15.4%)	5.5%	21%	101(2.2%)	0.99	0.67	2.6%	0.32	0.98
(D) β 11c_(A) δ 26d	58.8(64.7%)	70.6(8.4%)	70.6(8.4%)	5.4%	18.5%		0.99	0.67	2.6%	0.32	1.05
(D) β 14c_(A) δ 26d	51.6(39.1%)	62.7(5.5%)	62.7(5.5%)	5.5%	47.2%	73.5(2.8%)	1.19	0.79	2.7%	0.4	0.78
(D) β 27b_(A) δ 26d	63(53.6%)	52.4(10%)	63(8.4%)	4.7%	28%		2.19	1.15	0.7%	0.75	0.99
(D)Y8b_(A) δ 26d	56.8(51.9%)	44.4(2.7%)	56.8(8.4%)	6%	33.6%	72.9(3.4%)	1.44	0.99	2.2%	0.4	1.29
(D)Y29a_(A) δ 26d	78(51.4%)	66(20.4%)	52.6(2.7%)	5.5%	25.6%		1.44	0.99	2.3%	0.4	0.7
(D) β 5c_(A) δ 28d	53.6(58.2%)	67.6(7.4%)	85(8.8%)	5.6%	23.2%	103(2.3%)	1.44	0.91	2.4%	0.4	1.15
(D) β 8c_(A) δ 28d	58.1(58.2%)	73.6(8.8%)	73.6(8.8%)	5.6%	23%	82.3(1.2%)	1.30	0.54	1.4%	0.85	0.67
(D) β 11c_(A) δ 28d	57.7(62.5%)	73(6%)	73(6%)	5.8%	22.6%	84.2(3%)	1.29	0.55	1.2%	0.7	0.78
(D) β 14c_(A) δ 28d	52.1(49.8%)	62.2(12.4%)	71.2(9.6%)	5%	27%	86.3(1.3%)	1.30	0.54	1.3%	0.85	0.9
(D) β 27b_(A) δ 28d	67.4(46.3%)	56.8(16.5%)	67.4(8.6%)	5.6%	28.6%		1.29	0.67	1.2%	0.75	2.06
(D)Y8b_(A) δ 28d	61.4(54%)	47.2(4.2%)	61.4(8%)	6%	32.1%	72.4(1.6%)	1.83	0.79	1.4%	0.67	1.51
(D)Y29a_(A) δ 28d	80.1(58.1%)	64.9(8.9%)	64.9(8.9%)	5.5%	24.2%		1.57	0.69	1.3%	0.8	1.75

Table 4-5 PDA fit parameters for the datasets of the Mg^{2+} -titrations

DA-pair	$[Mg^{2+}]$, mM	$\langle R_{DA} \rangle_{E,1}$, Å (x_1)	$\langle R_{DA} \rangle_{E,2}$, Å (x_2)	$\langle R_{DA} \rangle_{E,3}$, Å (x_3)	σ_{app}	D-only	impurities R_{DA} , Å (x)	$\langle B_G \rangle$, kHz	$\langle B_R \rangle$, kHz	α	g_G/g_R	χ_r^2
(D)b5c_(A)a12d	20	52.7(49.6%)	52.7(7.5%)	58.8(9%)	4%	32.6%	71.8(1.3 %)	2.66	1.56	0.7%	0.75	1.04
	1	55.8(56.8%)	55.8(2.7%)	58.8(10.4%)	3.6%	26.5%	73.6(3.6 %)	2.66	1.56	0.5%	0.75	2.02
	0.1	57.5(61.3%)	51(2.5%)	57.5(7.4%)	3.3%	27.5%	69.5(1.2 %)	2.66	1.46	0.5%	0.75	0.9
(D)b8c_(A)a12d	20	58.7(60%)	58.7(8.5%)	48.3(11.6%)	3.9%	18.9%	74.6(1.1 %)	1.13	0.75	2.5%	0.4	1.16
	1	61.6(49.4%)	61.6(2.5%)	51(20.1%)	4%	25.3%	77.8(2.7 %)	1.66	0.70	1.5%	0.73	1.71
	0.1	60.3(49.9%)	60.3(1%)	52.2(22.2%)	4%	24.1%	71.5(2.8 %)	1.44	0.60	1.5%	0.71	1.26
(D)b11c_(A)a12d	20	56.6(33.6%)	48.2(3.1%)	43.2(3.8%)	2.5%	41.9%	81.7(17.7 %)	1.91	0.63	1.5%	0.68	1.42
	1	56.5(42.8%)	49.9(9.2%)	44(9.7%)	3.72%	36.8%	68.9(1.4 %)	1.91	0.63	1.3%	0.68	1.7
	0.1	54.8(26.1%)	48(9.4%)	43.8(2.8%)	3.7%	61.6%		1.91	0.63	1.3%	0.68	1.17
(D)g29a_(A)a12d	20	54.1(53.3%)	49.6(9.7%)	54.1(7.4%)	2.7%	26.6%	66.5(3 %)	2.79	1.31	0.7%	0.76	1.24
	1	56(54.5%)	49.6(5.1%)	56(6.7%)	2.8%	29.2%	67(4.5 %)	2.79	1.31	0.7%	0.76	0.85
	0.1	56.6(34.3%)	49.6(2%)	56.6(4.5%)	3.2%	52.8%	71.2(6.3 %)	2.79	1.41	0.8%	0.76	1.37
(D)d7a_(A)a12d	20	39.8(51.1%)	49.1(10.7%)	49.1(10.7%)	4.6%	26.7%	69.9(0.8 %)	1.12	0.73	3.2%	0.35	1.78
	1	41.1(41.6%)	49.7(7.7%)	49.7(7.7%)	4.1%	42.3%	65.4(0.7 %)	1.44	0.60	1.6%	0.71	1.59
	0.1	42.5(36.9%)	49.6(8.6%)	49.6(8.6%)	4.5%	44.3%	66.3(1.6 %)	1.44	0.60	1.6%	0.71	1.25
(D)g29a_(A)b33b	20	49.8(49.8%)	55.8(7.6%)	72.1(6.6%)	3.2%	32.7%	80.4(3.3 %)	1.79	0.74	1.7%	0.71	1
	1	59.3(50.7%)	53.1(3.3%)	67.4(13.4%)	3.2%	29.6%	76.5(3 %)	1.79	0.74	1.7%	0.71	1.51
	0.1	62.7(45%)	55.8(2.2%)	68.6(22.5%)	3.2%	27.5%	74.7(2.8 %)	1.79	0.74	1.7%	0.71	1.43
(D)d7a_(A)b33b	20	51.9(48.8%)	59.8(8.6%)	73.1(14%)	4.2%	25.8%	86.6(2.7 %)	1.43	0.62	2%	0.8	1.43
	1	53.4(42.1%)	61.5(7%)	72(12.5%)	4%	38.5%		1.43	0.62	1.4%	0.8	1.12
	0.2	53.8(40.9%)	61.4(20.1%)	71.1(13%)	4.1%	26%		1.66	0.70	1.4%	0.73	1.04
(D)b5c_(A)g24a	20	48.7(51.4%)	62.4(8.5%)	62.4(8.5%)	5.5%	28.7%	78.9(2.9 %)	1.54	0.59	1.5%	0.73	1.66
	1	45.8(24.3%)	56.2(7.7%)	68.8(18.9%)	4.6%	44.6%	87.1(4.5 %)	1.64	0.62	1.6%	0.7	1.27
	0.1	46.9(23.4%)	55.2(11%)	66.8(14.1%)	5.1%	46.8%	77.8(4.7 %)	1.64	0.62	1.6%	0.7	0.94
(D)b8c_(A)g24a	20	48.3(42.5%)	54.6(7.6%)	63.2(14%)	4%	35.8%		1.81	0.75	1.4%	0.74	2.4
	1	47.6(23.7%)	56.4(10.6%)	65.7(12.9%)	4%	52.7%		1.66	0.70	1.2%	0.73	1.09
	0.2	48.3(18.3%)	56(18%)	64.3(15.3%)	4%	48.4%		1.66	0.70	1.5%	0.73	1.09
(D)b11c_(A)g24a	20	51.7(59.4%)	44.9(4.9%)	51.7(7.9%)	3.5%	26.4%	65.6(1.3 %)	1.74	0.64	1.4%	0.71	1.73
	1	51.6(63.4%)	46.1(1.7%)	51.6(7.6%)	2.9%	25.3%	61.5(2 %)	1.66	0.70	1.3%	0.73	1.11
	0.2	52.4(63.7%)	45.8(1.1%)	52.4(7.8%)	2.6%	26.1%	62.6(1.3 %)	1.66	0.70	1.4%	0.73	1.41
(D)b14c_(A)g24a	20	41.7(32.5%)	48.7(12.6%)	53.9(11.9%)	3.5%	41.7%	64.8(1.3 %)	1.73	0.72	1.4%	0.74	2.28
	1	42.4(11.4%)	50.3(10.8%)	57.2(6.7%)	3.5%	60.1%	74.7(11 %)	1.64	0.62	1.6%	0.7	1.49
	0.1	41.5(5.3%)	48.6(8.7%)	56.1(9.8%)	3.5%	63%	69.5(13.2 %)	1.64	0.62	1.8%	0.7	2.07
(D)b27b_(A)g24a	20	45.1(22.8%)	41.4(7.2%)	50.3(10.3%)	3.5%	58.8%	64.8(0.9 %)	2.79	1.41	0.6%	0.76	2.02
	1	45.1(27.9%)	41.4(6.6%)	50.3(9.5%)	3.5%	54.3%	66.7(1.7 %)	2.79	1.21	0.6%	0.76	1.47
	0.1	45.1(20.2%)	41.4(1%)	50.3(7.1%)	3.5%	68.7%	69.2(3 %)	2.79	1.51	0.8%	0.76	2.26
(D)b5c_(A)g8b	20	63.2(41.6%)	63.2(4.9%)	52.7(4.7%)	4.7%	48.9%		1.82	0.77	1%	0.73	0.81

	1	65.5(37.7%)	65.5(1.4%)	54.3(5%)	4.7%	55.9%		1.82	0.77	1%	0.73	1.05
	0.1	67.2(45.9%)	67.2(0.4%)	57.7(10.6%)	4.7%	43.1%		1.82	0.77	1%	0.73	1.7
DA-pair	[Mg ²⁺], mM	$\langle R_{DA} \rangle_{E,1}, \text{\AA}$ (x_1)	$\langle R_{DA} \rangle_{E,2}, \text{\AA}$ (x_2)	$\langle R_{DA} \rangle_{E,3}, \text{\AA}$ (x_3)	σ_{app}	D-only	impurities $R_{DA}, \text{\AA} (x)$	$\langle B_G \rangle$, kHz	$\langle B_R \rangle$, kHz	α	g_G/g_R	χ_r^2
(D)b8c_(A)g8b	20	59.2(48.9%)	59.2(7.7%)	49(9.7%)	4.8%	30.8%	76.3(2.9 %)	1.29	0.55	1.3%	0.7	1.76
	0.1	59(42.2%)	59(0.7%)	51.2(19.3%)	4.8%	35%	72.2(2.7 %)	1.35	0.63	1.7%	0.75	2.06
(D)b11c_(A)g8b	20	52.2(49%)	52.2(6%)	42.8(5.5%)	4.4%	37.3%	68.5(2.3 %)	1.49	0.58	1.5%	0.69	1.3
	3	53.2(49.1%)	53.2(3.8%)	40.6(4.9%)	4.4%	37.7%	71(4.5 %)	1.49	0.58	1.5%	0.69	1.82
	1	53(49.1%)	53(2.7%)	41.2(4.8%)	4.4%	38.1%	73.4(5.3 %)	1.49	0.58	1.5%	0.69	1.37
	0.3	53.4(46.2%)	53.4(1.7%)	45(6.9%)	4.4%	42.2%	69(3 %)	1.69	0.62	1.6%	0.66	2.14
	0.1	52.6(40.8%)	52.6(0.5%)	44.7(7.3%)	5%	47.8%	71.7(3.6 %)	1.62	0.54	1.5%	0.71	2.01
	0.03	52.4(41.5%)	52.4(0.2%)	44.7(6.8%)	5%	46%	71(5.6 %)	1.62	0.54	1.5%	0.71	2.12
(D)b5c_(A)d10a	20	49.8(43.3%)	56.4(14.5%)	67.9(11.4%)	4.5%	28%	83.4(2.8 %)	1.95	0.64	1.2%	0.71	2.32
	1	46.6(8.9%)	55.7(10.6%)	68.9(3.7%)	4.5%	69.8%	74.8(7 %)	1.64	0.62	1.5%	0.7	1.26
	0.1	47.1(6.6%)	55.5(13.7%)	69.2(6.6%)	4.5%	69.6%	79.5(3.5 %)	1.64	0.62	1.7%	0.7	1.17
(D)b8c_(A)d10a	20	51.3(20.7%)	61.6(9.4%)	51.3(4.6%)	4.5%	65.3%		1.82	0.81	0.9%	0.75	1.19
	1	58(15.5%)	58(1.4%)	50(9.9%)	3.9%	69.7%	66.7(3.6 %)	1.82	0.81	1%	0.75	0.9
	0.1	58.7(19.5%)	58.7(0.6%)	51.1(5.7%)	4.7%	71.2%	66.8(3 %)	1.82	0.81	1.1%	0.75	1.51
(D)b11c_(A)d10a	20	52.8(64.8%)	46.5(6.7%)	52.8(7.9%)	3.5%	19.7%	66.3(0.9 %)	1.74	0.64	1.5%	0.71	1.32
	1	52.6(56.6%)	46.5(2%)	52.6(7%)	3.5%	30.3%	64.3(4.2 %)	1.44	0.60	1.6%	0.71	1.46
	0.1	53.7(43.7%)	46.5(1.5%)	53.7(5.9%)	3.5%	45.4%	65.3(3.6 %)	1.44	0.60	1.6%	0.71	0.92
(D)b14c_(A)d10a	20	41.9(35.5%)	47.7(12.7%)	53(13%)	3.5%	37.7%	61.1(1 %)	1.73	0.72	1.9%	0.74	1.69
	1	42(11.6%)	49.9(9%)	56.2(8.8%)	3.5%	56.4%	73.1(14.2 %)	1.64	0.62	2.1%	0.7	1.77
	0.1	43.2(6.5%)	50.9(10.9%)	59.6(7%)	3.5%	59.7%	74.5(15.8 %)	1.64	0.62	2.1%	0.7	3.57
(D)b27b_(A)d10a	20	46.7(29.8%)	42.8(14.4%)	51.4(9.3%)	3.5%	43.6%	64.3(2.9 %)	1.73	0.72	1.6%	0.74	1.51
	1	47.8(41.3%)	42.8(4.8%)	51.8(11%)	3.5%	41.2%	65.4(1.8 %)	1.73	0.72	1.4%	0.74	1.29
	0.1	47.9(34.4%)	42.8(2%)	52.4(7.5%)	3.5%	54.5%	65.7(1.6 %)	1.73	0.72	1.3%	0.74	1.99
(D)g8b_(A)d10a	20	46.4(59.3%)	50.8(9.1%)	42.4(4.9%)	3.5%	24.1%	69.7(2.6 %)	1.73	0.72	1.4%	0.74	0.86
	1	51.4(46.2%)	42.4(1.2%)	51.4(5.5%)	3.5%	46.2%	62.7(0.9 %)	1.73	0.72	1.6%	0.74	2.17
	0.1	52.8(65.4%)	42.4(0.9%)	52.8(7.5%)	7.5%	25%	64.8(1.1 %)	1.73	0.72	1.6%	0.74	1.15
(D)b5c_(A)d23d	20	44.1(28.3%)	51.1(5%)	62.5(15.7%)	4.4%	51%		1.56	0.65	1.6%	0.71	0.78
	1	50.5(36.1%)	50.5(1.5%)	62.6(13.6%)	4.4%	48.8%		1.56	0.65	1.6%	0.73	0.93
	0.1	53.5(30.5%)	53.5(0.6%)	61(13.5%)	4.4%	55.4%		1.56	0.65	1.5%	0.71	0.83
(D)b11c_(A)d23d	20	44.2(45.1%)	51.2(4.9%)	58.8(11.9%)	3.8%	36.2%	71.5(1.9 %)	1.66	0.65	1.3%	0.67	1.88
	1	43.1(40.5%)	50.9(9%)	58.7(12.1%)	3.7%	38.4%		1.82	0.81	0.9%	0.75	1.16
	0.1	44.3(25.4%)	51.2(15.7%)	58.5(12.4%)	4.1%	46.4%		1.82	0.81	1.1%	0.75	1.16
(D)b27b_(A)d23d	20	53.7(52.4%)	46.3(5.3%)	53.7(6.9%)	4.5%	34.3%	68.2(1.1 %)	1.27	0.53	1.4%	0.85	1.02
	0.1	53.4(60.2%)	46.3(1.8%)	53.4(7.1%)	4.5%	28.3%	75(2.7 %)	1.27	0.53	1.8%	0.85	0.83
(D)g8b_(A)d23d	20	49.9(34.5%)	42.9(3.8%)	57.5(2.8%)	3.1%	55.5%	69.9(3.4 %)	1.82	0.77	1.1%	0.73	1.19
	1	46.9(20.7%)	42.1(1.4%)	51.5(9.9%)	3%	65.6%	62(2.4 %)	1.82	0.69	1.1%	0.73	1.54

	0.1	47.5(20.1%)	43(1.8%)	51.4(14.5%)	3%	62.6%	62.1(1.1 %)	1.82	0.67	1.1%	0.73	0.88
DA-pair	[Mg ²⁺], mM	$\langle R_{DA} \rangle_{E,1}, \text{Å}$ (x_1)	$\langle R_{DA} \rangle_{E,2}, \text{Å}$ (x_2)	$\langle R_{DA} \rangle_{E,3}, \text{Å}$ (x_3)	σ_{app}	D-only	impurities $R_{DA}, \text{Å} (x)$	$\langle B_G \rangle$, kHz	$\langle B_R \rangle$, kHz	α	g_G/g_R	χ_r^2
(D)b27b_(A)d26d	20	63.1(55.4%)	52.6(10%)	63.1(8%)	4.9%	26.7%		2.19	1.15	0.7%	0.75	1.03
	1	62.6(61.6%)	52.6(3%)	62.6(7.1%)	4%	28.3%		2.19	1.11	0.5%	0.75	0.69
	0.1	62.8(64.2%)	52.6(0.8%)	62.8(8.1%)	4.6%	26.9%		2.19	1.11	0.5%	0.75	1.8
(D)b14c_(A)d28d	20	52.2(52.4%)	63.1(12.8%)	71.1(8.2%)	5.1%	25.2%	89.6(1.4 %)	1.29	0.54	1.3%	0.85	1.26
	0.1	54.8(23.6%)	62(43.1%)	71.4(12.9%)	4.7%	20.4%		1.35	0.63	1.7%	0.75	0.73
(D)b27b_(A)d28d	20	68.4(57.7%)	56.8(14.1%)	68.4(8.1%)	5.5%	17.5%	90.7(2.7 %)	1.49	0.55	1.5%	0.68	1.27
	3	66.8(68.9%)	55.7(4.1%)	66.8(8.1%)	5.33%	16.2%	102(2.7 %)	1.69	0.62	1.6%	0.66	0.96
	1	67.2(62.4%)	57(2.8%)	67.2(7.4%)	5.4%	21%	86.5(6.3 %)	1.49	0.55	1.5%	0.68	1.23
	0.3	67.7(54.8%)	56.7(2.7%)	67.6(6.2%)	5.3%	15.5%	96.9(20.9 %)	1.49	0.55	1.5%	0.68	1.09
	0.1	66.9(68%)	56.9(3.8%)	66.9(7.9%)	4.4%	16.9%	96.6(3.4 %)	1.49	0.55	1.5%	0.68	0.93
	0.03	67.3(67.4%)	57.4(3.7%)	67.3(8%)	4.7%	19.1%	92.5(1.8 %)	1.49	0.55	1.5%	0.68	1.38
(D)g8b_(A)d28d	20	61.4(54.1%)	47.2(4.2%)	61.4(8%)	6%	32.1%	72.4(1.6 %)	1.83	0.79	1.4%	0.67	1.5
	1	59.5(54.7%)	51.5(1.8%)	59.5(8.7%)	5.4%	34.8%		2.00	0.90	0.9%	0.72	1.85
	0.1	59.7(52.7%)	48.9(0.7%)	59.7(7.8%)	4.5%	38.8%		2.00	0.90	0.9%	0.72	1.3
(D)g29a_(A)d28d	20	77.5(62.3%)	65.1(13.1%)	65.1(13.1%)	5.5%	11.4%		1.54	0.59	1.1%	0.73	1.15
	1	75.9(64.9%)	66.2(12.4%)	66.2(12.4%)	5.5%	10.3%		1.54	0.59	1.4%	0.73	1.49
	0.1	79.4(48.2%)	70(18.6%)	70(18.6%)	5.5%	14.7%		1.54	0.59	1.5%	0.73	0.9

4.5 Conformer (*ad*)_a at 0.1 mM MgCl₂

Docking the RNA4WJ with the 30 distances for conformer (*ad*)_a measured at 0.1 mM MgCl₂ (see Table 4-6) yields two solutions (see Figure 4.2 and Figure 4.3) with very similar values for χ_r^2 ($\chi_r^2 = 1.9$ and 2.0, respectively, $\chi_{r,max}^2 = 2.2$). For the docking an error of 7 % was assumed for each distance. The solution with the second lowest χ_r^2 is very similar to the one found for 20 mM MgCl₂ (RMSD over all P atoms is 7.8 Å, see Figure 4.2) with the same helix stacking but with angles between the stacked helix pairs closer to 90° compared to the structure at 20 mM MgCl₂. The solution with the lowest χ_r^2 (see Figure 4.3) resembles a “rough” quasi-mirrored image of the second best one with clearly distorted helix stacking. In our opinion it results from an instability in the rigid body docking due to an insufficient amount of distance restraints.

Table 4-6. $\langle R_{DA} \rangle_E$ measured at 0.1 mM $MgCl_2$ and corresponding ones at 20 mM.

DA pair	0.1 mM $MgCl_2$	20 mM $MgCl_2$	DA pair	0.1 mM $MgCl_2$	20 mM $MgCl_2$
	$\langle R_{DA} \rangle_{E_f}$ Å	$\langle R_{DA} \rangle_{E_f}$ Å		$\langle R_{DA} \rangle_{E_f}$ Å	$\langle R_{DA} \rangle_{E_f}$ Å
(D) $\beta 5c_{(A)}\alpha 12d$	57.5	53.4	(D) $\beta 5c_{(A)}\delta 10a$	47.1	49.8
(D) $\beta 8c_{(A)}\alpha 12d$	60.3	58.7	(D) $\beta 8c_{(A)}\delta 10a$	58.7	51.6
(D) $\beta 11c_{(A)}\alpha 12d$	54.8	56.4	(D) $\beta 11c_{(A)}\delta 10a$	53.7	52.8
(D) $\gamma 29a_{(A)}\alpha 12d$	56.6	55.1	(D) $\beta 14c_{(A)}\delta 10a$	43.2	41.9
(D) $\delta 7a_{(A)}\alpha 12d$	42.5	39.8	(D) $\beta 27b_{(A)}\delta 10a$	47.9	45.8
(D) $\gamma 29a_{(A)}\beta 33b$	62.7	49.5	(D) $\gamma 8b_{(A)}\delta 10a$	52.8	46.9
(D) $\delta 7a_{(A)}\beta 33b$	53.8	52.7	(D) $\beta 5c_{(A)}\delta 23d$	53.5	44.4
(D) $\beta 5c_{(A)}\gamma 24a$	46.9	48.7	(D) $\beta 11c_{(A)}\delta 23d$	44.3	44.2
(D) $\beta 8c_{(A)}\gamma 24a$	48.3	48.3	(D) $\beta 27b_{(A)}\delta 23d$	53.4	53.8
(D) $\beta 11c_{(A)}\gamma 24a$	52.4	51.7	(D) $\gamma 8b_{(A)}\delta 23d$	47.5	51
(D) $\beta 14c_{(A)}\gamma 24a$	41.5	41.7	(D) $\beta 27b_{(A)}\delta 26d$	62.8	63
(D) $\beta 27b_{(A)}\gamma 24a$	45.1	45.1	(D) $\beta 14c_{(A)}\delta 28d$	54.8	52.1
(D) $\beta 5c_{(A)}\gamma 8b$	67.2	64.1	(D) $\beta 27b_{(A)}\delta 28d$	66.9	67.4
(D) $\beta 8c_{(A)}\gamma 8b$	59	59.3	(D) $\gamma 8b_{(A)}\delta 28d$	59.7	61.4
(D) $\beta 11c_{(A)}\gamma 8b$	52.6	50.7	(D) $\gamma 29a_{(A)}\delta 28d$	79.4	80.1

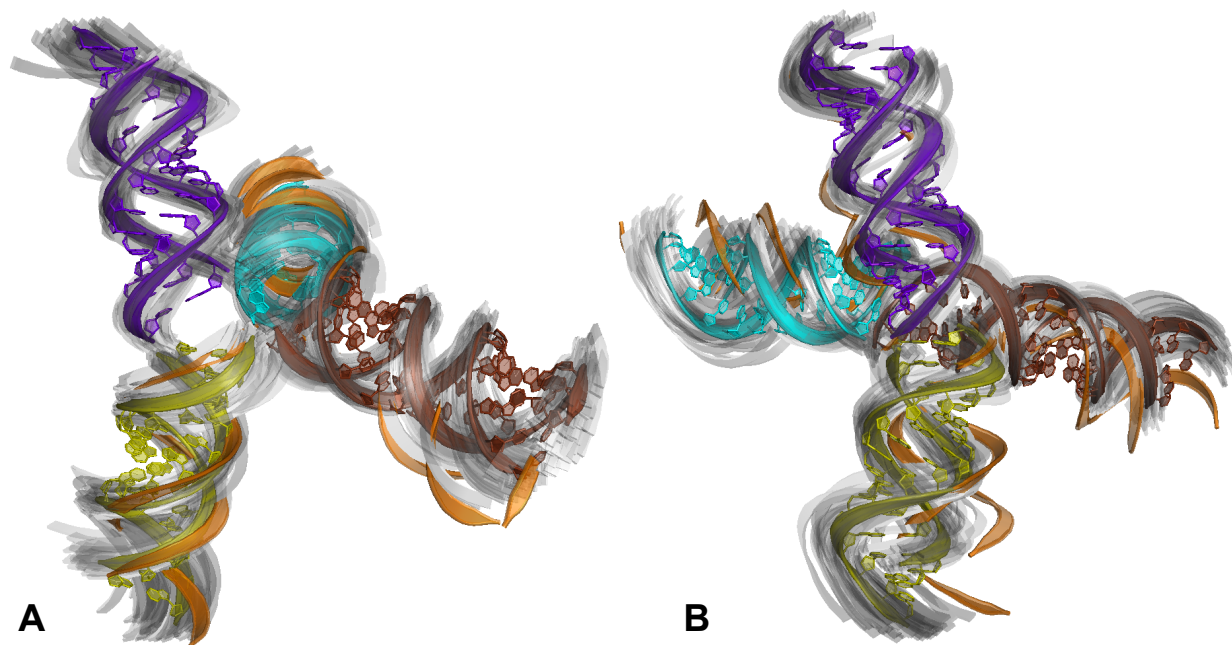


Figure 4.2. (A and B) Docking result with second lowest χ_r^2 of conformer $(ad)_a$ with distances measured at 0.1 mM $MgCl_2$ (colored) overlaid with 100 structures (grey transparent) indicating the uncertainties of the helix positions and orientations resulting from bootstrapping and with best solution for 20 mM $MgCl_2$ (orange).



27

4.6 Rigid body models

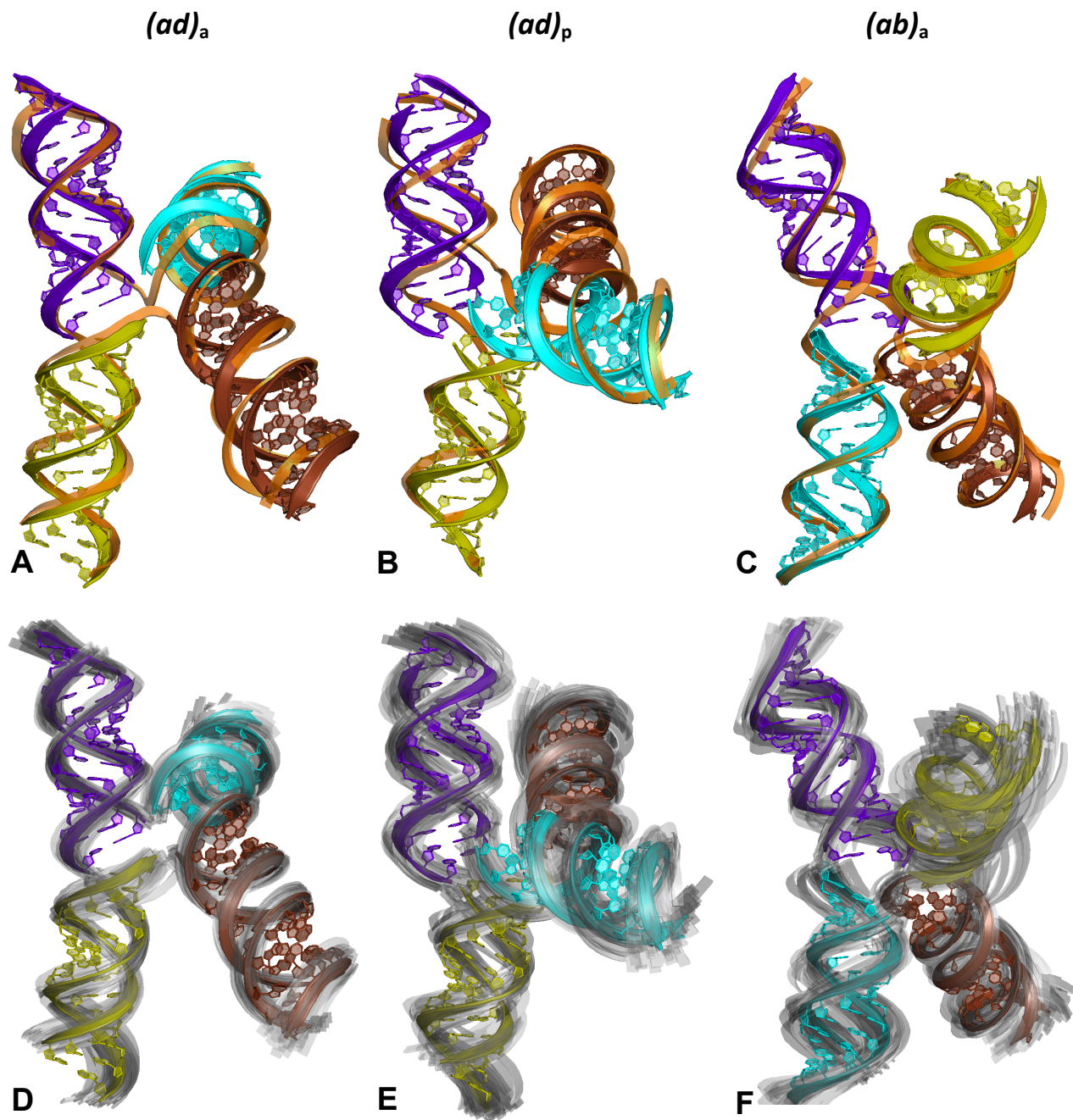


Figure 4.4. Structures (cartoon representation) of the rigid body models with the lowest χ_r^2 for conformer $(ad)_a$ (A and D), $(ad)_p$ (B and E) and $(ab)_a$ (C and F) overlaid with the respective MD models with lowest χ_r^2 (orange, A - C) and with 100 structures (D - F, grey transparent) indicating the uncertainties of the helix positions and orientations resulting from bootstrapping.

4.7 “2nd best” docking results

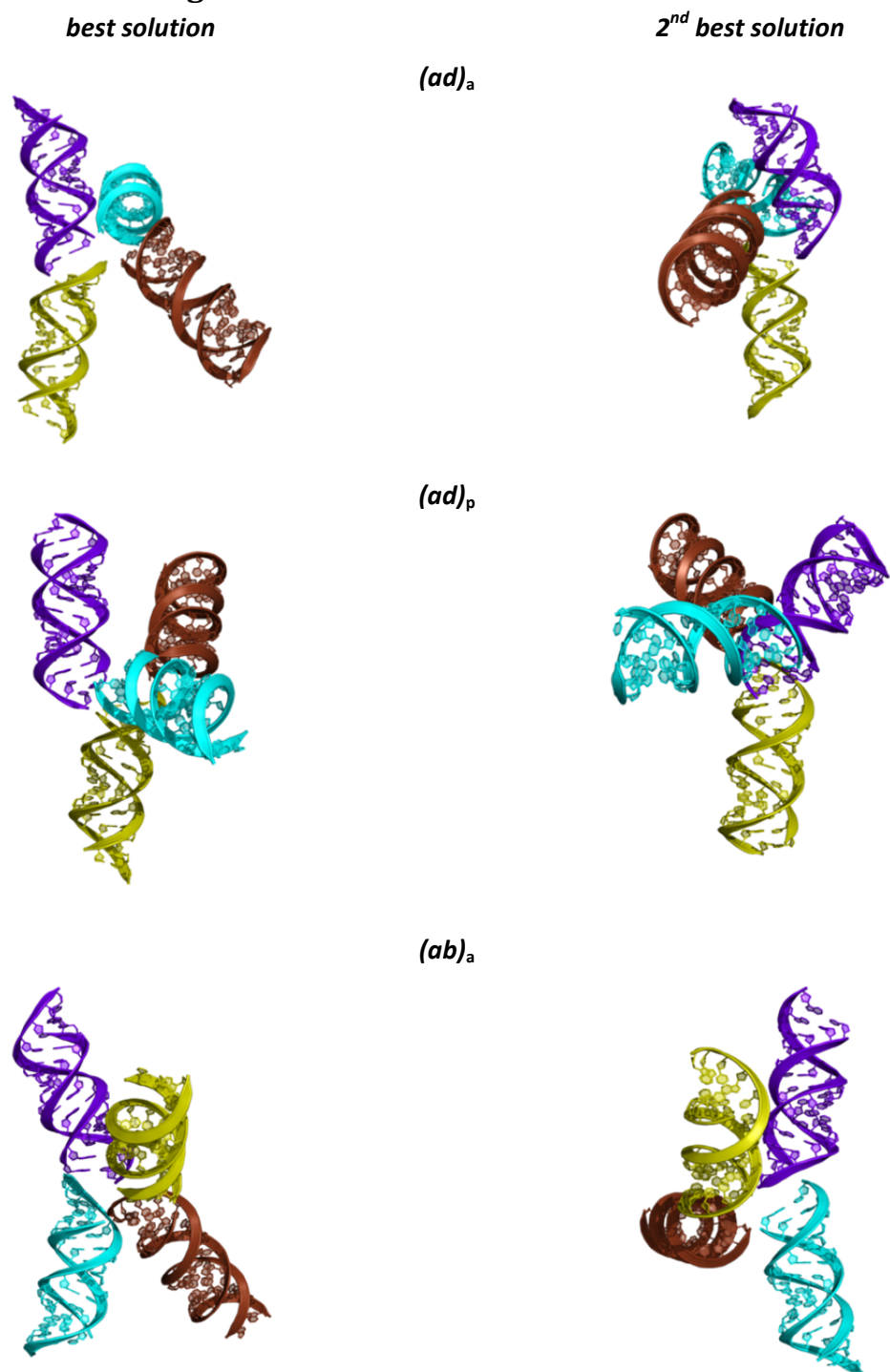


Figure 4.5. Comparison of structures with lowest (left) and 2nd lowest (right) χ_r^2 resulting from rigid body docking for conformers $(ad)_a$ (top), $(ad)_p$ (middle) and $(ab)_a$ (bottom).

4.8 SimRNA models

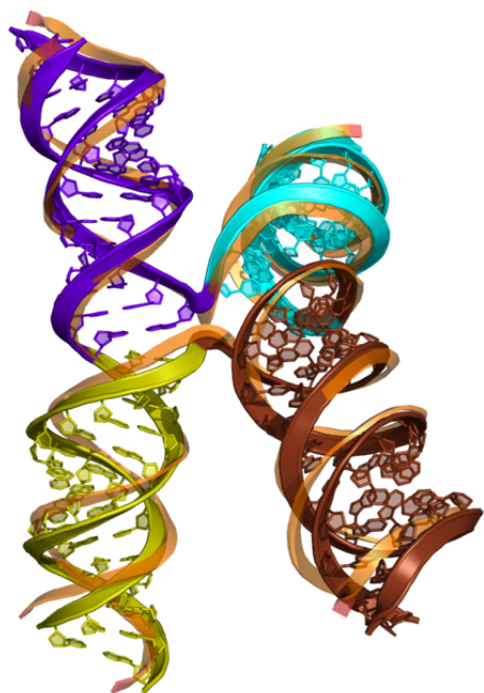


Figure 4.6. Structures (cartoon representation) determined by coarse-grained modeling with SimRNA for $(ad)_8$ overlaid with MD model with best agreement to FRET restraints (orange, transparent). RMSD over all P atoms 3.52 Å.

5 References

- Antonik, M., S. Felekyan, et al. (2006). "Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis." Journal of Physical Chemistry B **110**(13): 6970-6978.
- Banas, P., D. Hollas, et al. (2010). "Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins." Journal of Chemical Theory and Computation **6**(12): 3836-3849.
- Cai, Q., A. K. Kusnetzow, et al. (2007). "Nanometer distance measurements in RNA using site-directed spin Labeling." Biophysical Journal **93**(6): 2110-2117.
- Case, D. A., T. E. Cheatham, et al. (2005). "The Amber biomolecular simulation programs." Journal of Computational Chemistry **26**(16): 1668-1688.
- Darden, T., D. York, et al. (1993). "Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems." Journal of Chemical Physics **98**(12): 10089-10092.
- Efron, B. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion." Annals of Statistics **14**(4): 1301-1304.
- Eggeling, C., J. Widengren, et al. (2006). "Analysis of Photobleaching in Single-Molecule Multicolor Excitation and Förster Resonance Energy Transfer Measurements." Journal of Physical Chemistry A **110**(9): 2979-2995.
- Hornak, V., R. Abel, et al. (2006). "Comparison of multiple amber force fields and development of improved protein backbone parameters." Proteins **65**(3): 712-725.
- Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparison of Simple Potential Functions for Simulating Liquid Water." Journal of Chemical Physics **79**(2): 926-935.
- Joung, I. S. and T. E. Cheatham (2008). "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations." Journal of Physical Chemistry B **112**(30): 9020-9041.
- Kalinin, S., S. Felekyan, et al. (2007). "Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer." Journal of Physical Chemistry B **111**(34): 10253-10262.
- Kalinin, S., S. Felekyan, et al. (2008). "Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis." Journal of Physical Chemistry B **112**(28): 8361-8374.
- Kalinin, S., E. Sisamak, et al. (2010). "On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits." Journal of Physical Chemistry B **114**: 6197-6206.
- Kalinin, S., A. Valeri, et al. (2010). "Detection of structural dynamics by FRET: A photon distribution and fluorescence lifetime analysis of systems with multiple states." J. Phys. Chem. B **114**: 7983-7995.
- Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. (2012). "FRET restrained high-precision structural modeling resolves the configuration of primer/template DNA in complex with HIV-1 reverse transcriptase including the 5'-overhang."
- Muschielok, A., J. Andrecka, et al. (2008). "A nano-positioning system for macromolecular structural analysis." Nature Methods **5**(11): 965-971.
- O'Connor, D. V. and D. Phillips (1984). Time-correlated Single Photon Counting. New York, Academic Press.
- Perez, A., I. Marchan, et al. (2007). "Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers." Biophysical Journal **92**(11): 3817-3829.

- Rasnik, I., S. A. McKinney, et al. (2006). "Nonblinking and longlasting single-molecule fluorescence imaging." Nature Methods **3**(11): 891-893.
- Rothwell, J. P. (2002). Structural Investigation of the HIV-1 RT using single pair Fluorescence Energy Transfer. Max-Planck-Institut für Molekulare Physiologie. Dortmund, University of Dortmund.
- Rupert, P. B., A. P. Massey, et al. (2002). "Transition state stabilization by a catalytic RNA." Science **298**(5597): 1421-1424.
- Ryckaert, J. P., G. Ciccotti, et al. (1977). "Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes." Journal of Computational Physics **23**(3): 327-341.
- Sindbert, S., S. Kalinin, et al. (2011). "Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity." Journal of the American Chemical Society **133**(8): 2463-2480.
- Sisamak, E., A. Valeri, et al. (2010). "Accurate single-molecule FRET studies using multiparameter fluorescence detection." Methods in Enzymology **475**: 456-514.
- Soong, T. T. (2004). Fundamentals of probability and statistics for engineers. West Sussex, England, John Wiley & Sons.
- Woźniak, A. K., G. Schröder, et al. (2008). "Single molecule FRET measures bends and kinks in DNA." Proc.Natl.Acad.Sci.USA. **105**: 18337-18342.

IV

Complementary measurement data and PDA fits for “RNA4WJ manuscript”

1 2D-Histograms

1.1 Measurements at 20 mM MgCl₂

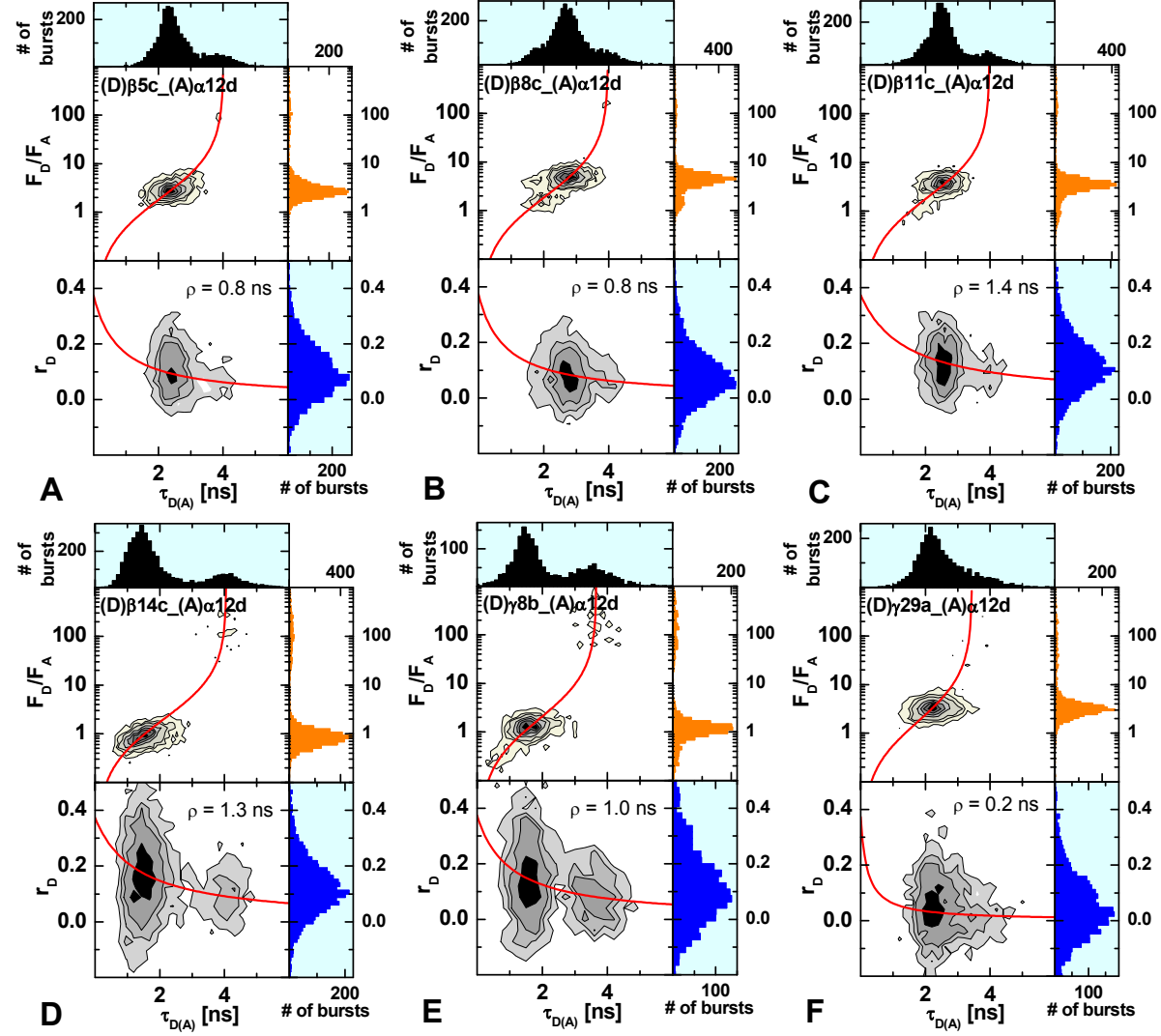


Figure 1.1. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\beta 5c_{-}(A)\alpha 12d$ (A), (D) $\beta 8c_{-}(A)\alpha 12d$ (B), (D) $\beta 11c_{-}(A)\alpha 12d$ (C), (D) $\beta 14c_{-}(A)\alpha 12d$ (D), (D) $\gamma 8b_{-}(A)\alpha 12d$ (E), and (D) $\gamma 29a_{-}(A)\alpha 12d$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), 0.8 ns (B), 1.4 ns (C), 1.3 ns (D), 1.0 ns (E) and 0.2 ns (F).

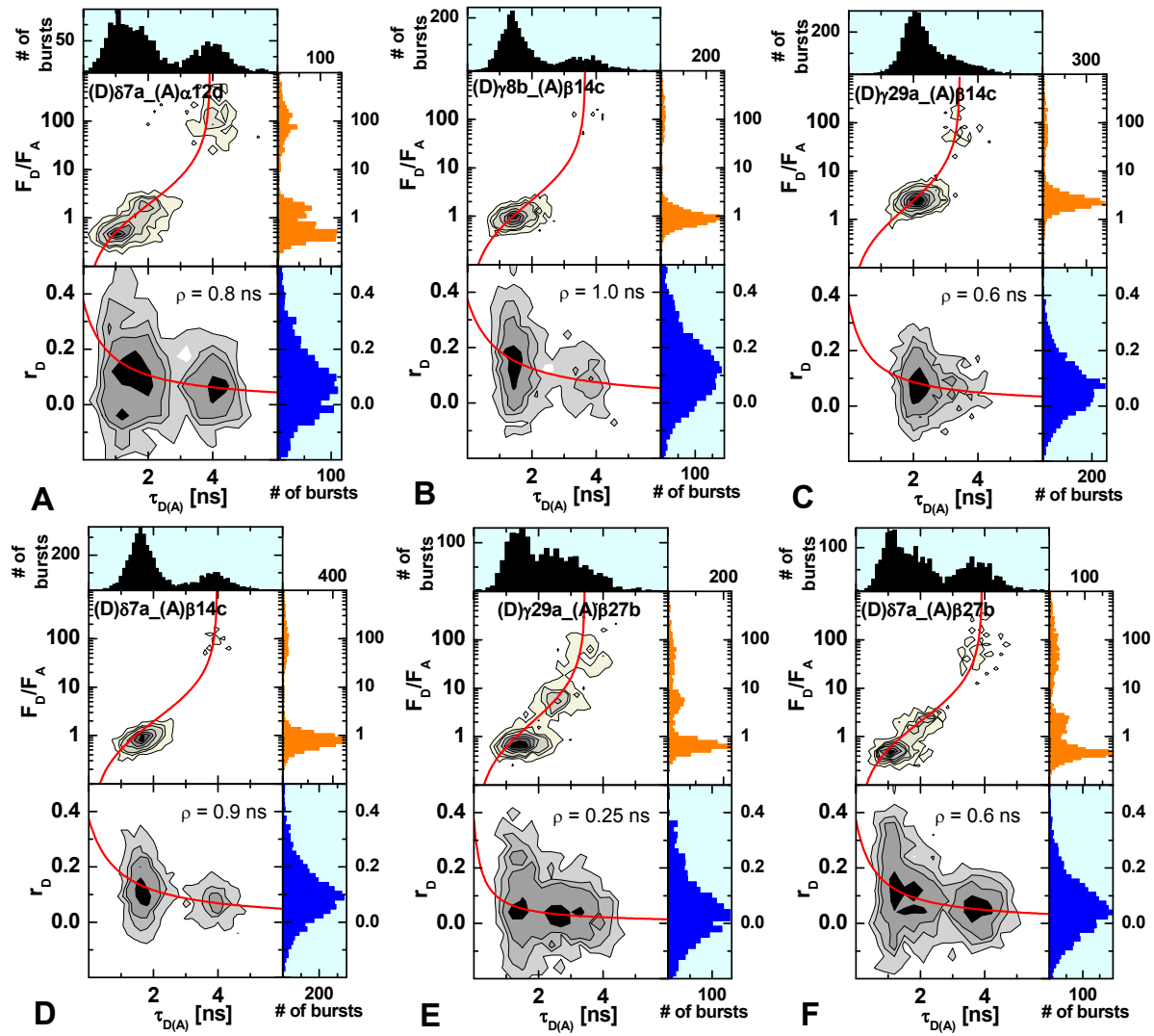


Figure 1.2. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\delta 7a$ _(A) $\alpha 12d$ (A), (D) $\gamma 8b$ _(A) $\beta 14c$ (B), (D) $\gamma 29a$ _(A) $\beta 14c$ (C), (D) $\delta 7a$ _(A) $\beta 14c$ (D), (D) $\gamma 29a$ _(A) $\beta 27b$ (E), and (D) $\delta 7a$ _(A) $\beta 27b$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), 1.0 ns (B), 0.6 ns (C), 0.9 ns (D), 0.25 ns (E) and 0.6 ns (F).

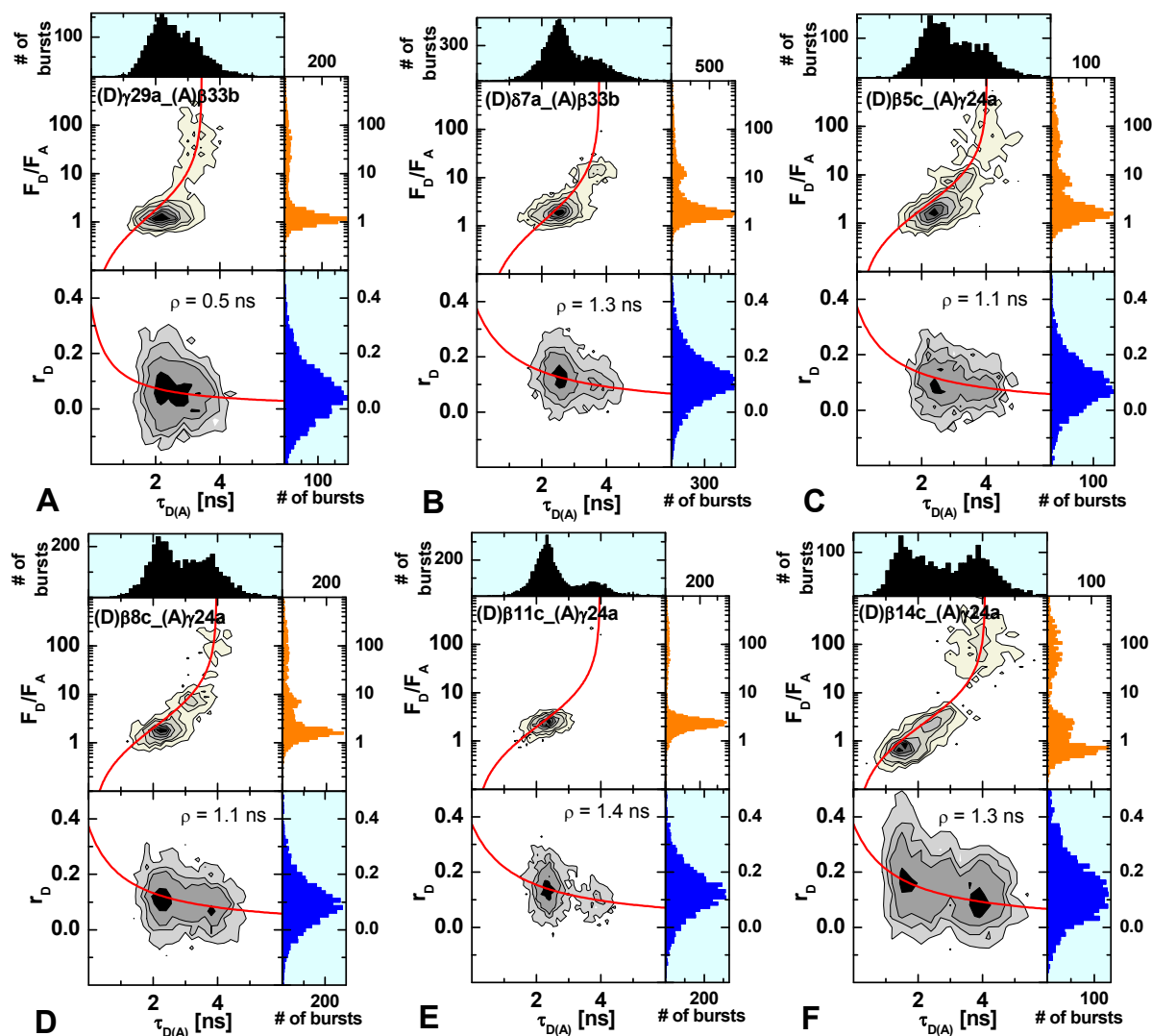


Figure 1.3. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) γ 29a_(A) β 33b (A), (D) δ 7a_(A) β 33b (B), (D) β 5c_(A) γ 24a (C), (D) β 8c_(A) γ 24a (D), (D) β 11c_(A) γ 24a (E), and (D) β 14c_(A) γ 24a (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.5$ ns (A), 1.3 ns (B), 1.1 ns (C), 1.1 ns (D), 1.4 ns (E) and 1.3 ns (F).

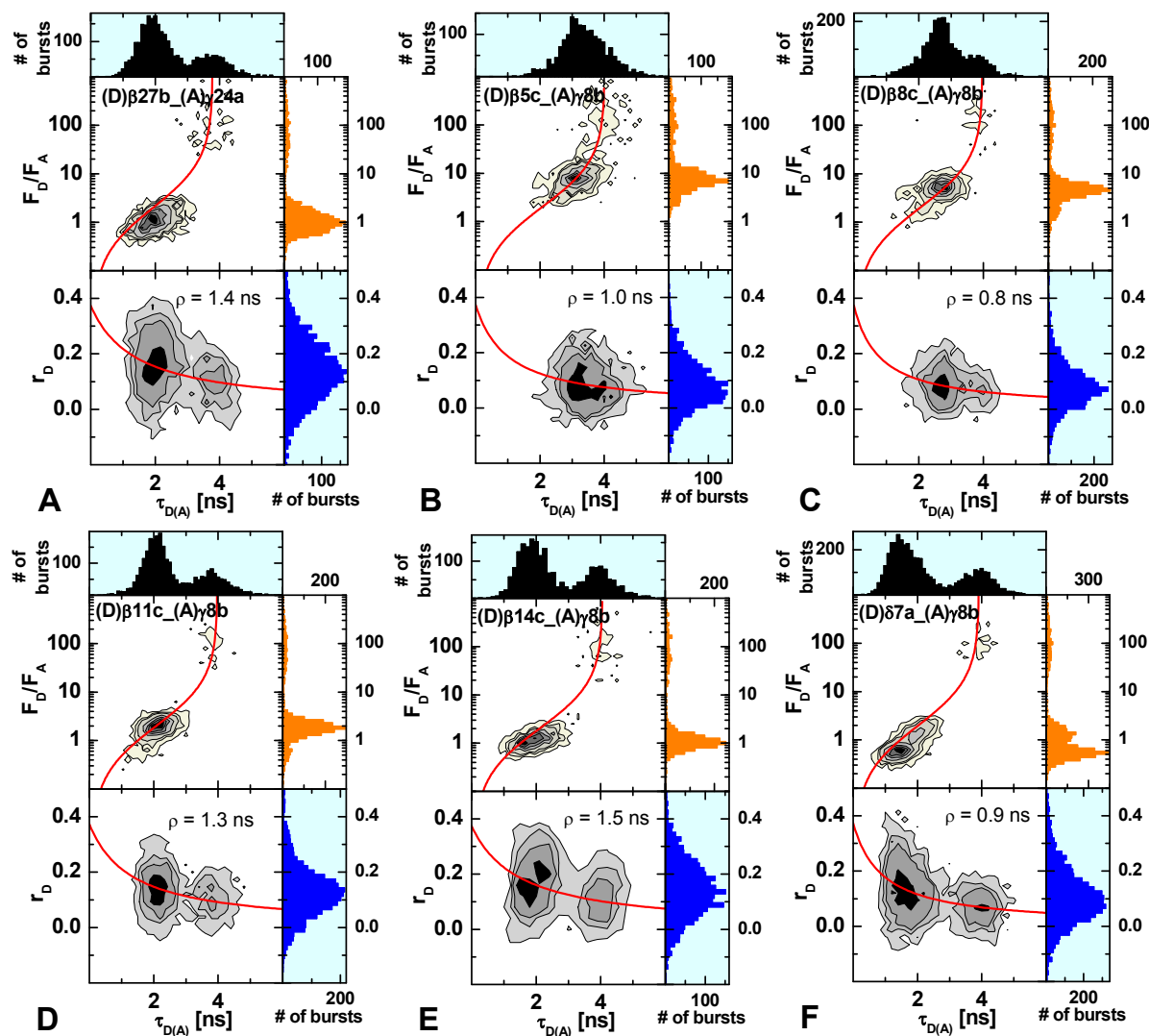


Figure 1.4. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\beta 27b_A$) $\gamma 24a$ (A), (D) $\beta 5c_A$) $\gamma 8b$ (B), (D) $\beta 8c_A$) $\gamma 8b$ (C), (D) $\beta 11c_A$) $\gamma 8b$ (D), (D) $\beta 14c_A$) $\gamma 8b$ (E), and (D) $\delta 7a_A$) $\gamma 8b$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.4$ ns (A), 1.0 ns (B), 0.8 ns (C), 1.3 ns (D), 1.5 ns (E) and 0.9 ns (F).

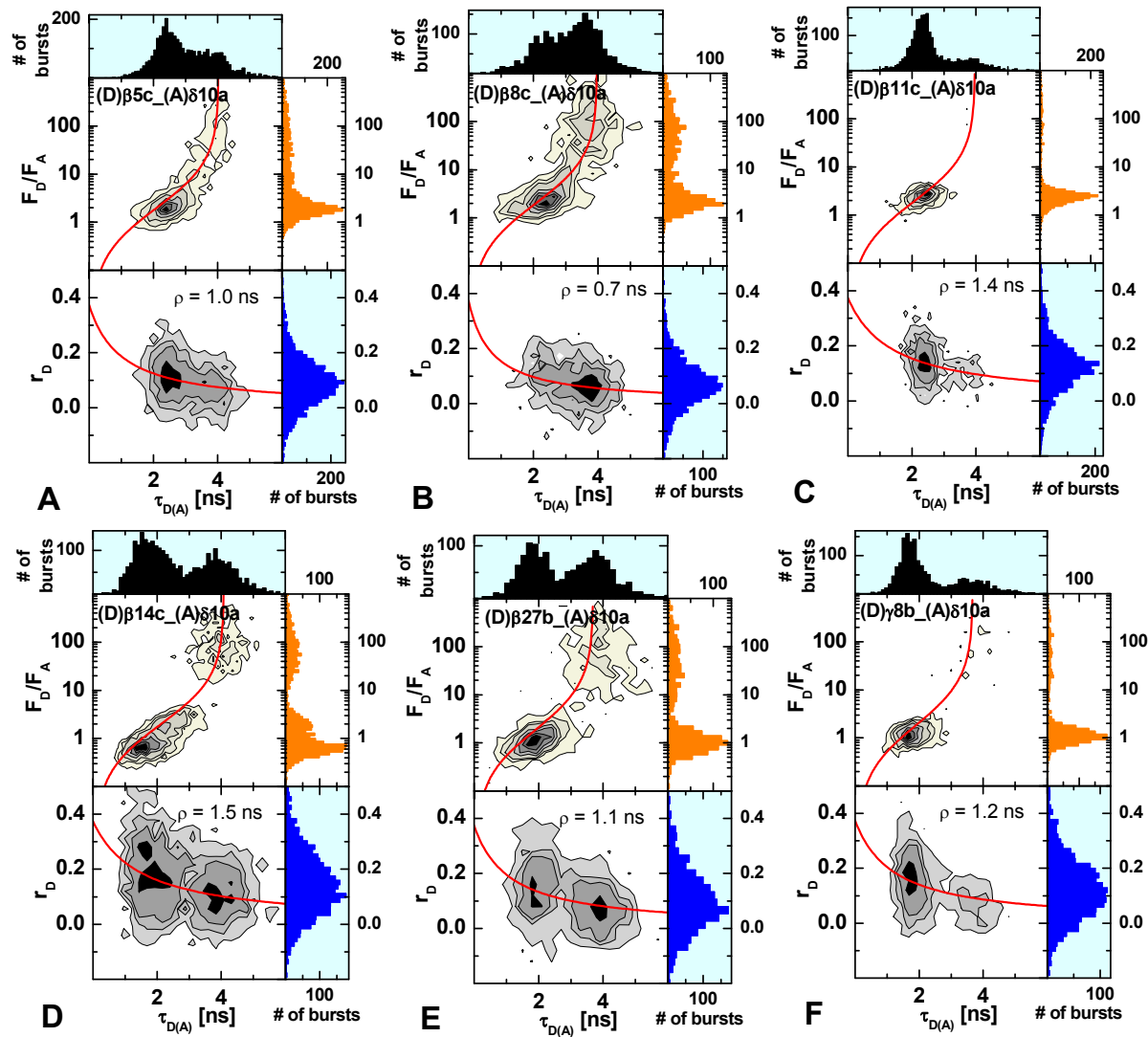


Figure 1.5. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples $(D)\beta 5c_{-}(A)\delta 10a$ (A), $(D)\beta 8c_{-}(A)\delta 10a$ (B), $(D)\beta 11c_{-}(A)\delta 10a$ (C), $(D)\beta 14c_{-}(A)\delta 10a$ (D), $(D)\beta 27b_{-}(A)\delta 10a$ (E), and $(D)\gamma 8b_{-}(A)\delta 10a$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red, see SI for details) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.0$ ns (A), 0.7 ns (B), 1.4 ns (C), 1.5 ns (D), 1.1 ns (E) and 1.2 ns (F).

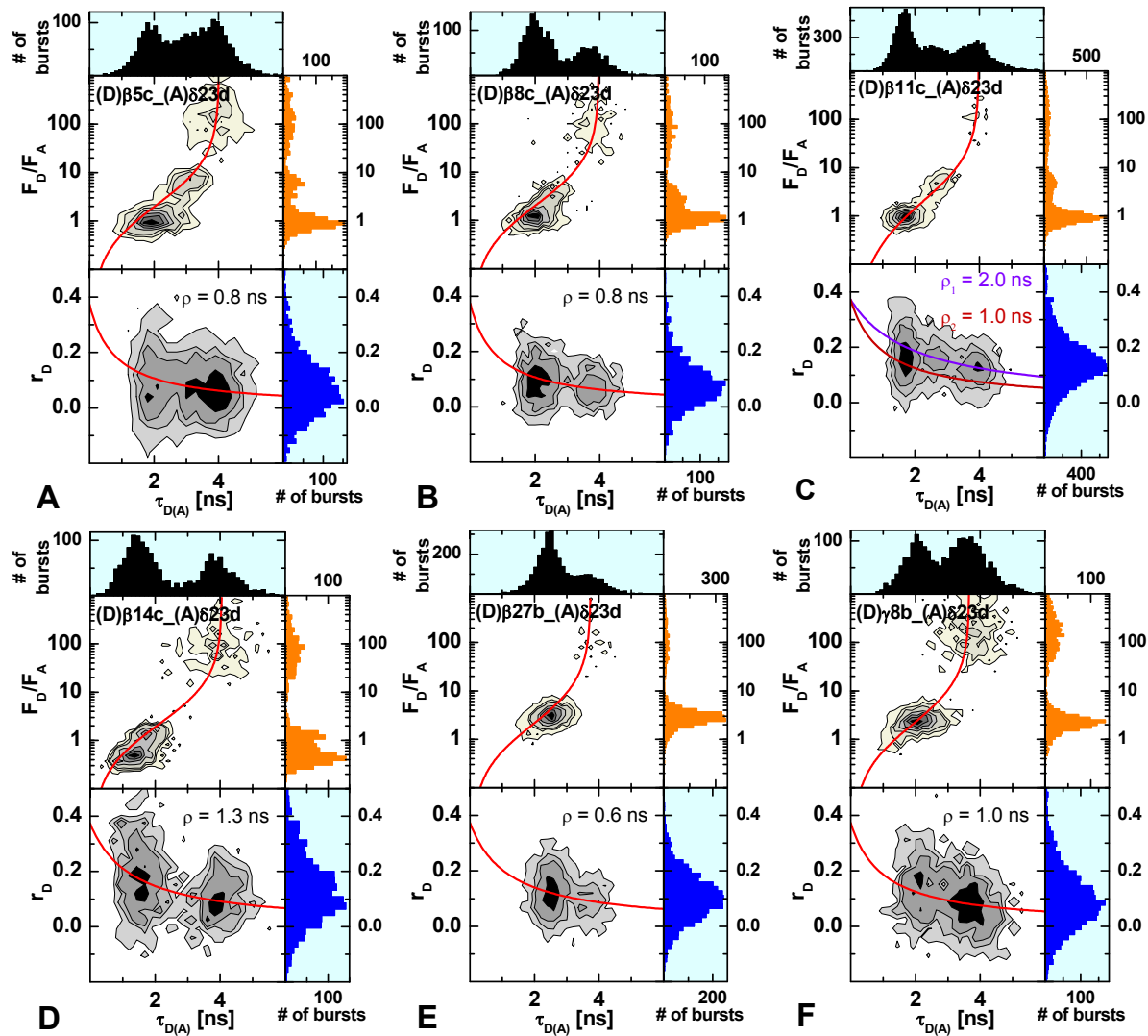


Figure 1.6. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\beta 5c_{-}(A)\delta 23d$ (A), (D) $\beta 8c_{-}(A)\delta 23d$ (B), (D) $\beta 11c_{-}(A)\delta 23d$ (C), (D) $\beta 14c_{-}(A)\delta 23d$ (D), (D) $\beta 27b_{-}(A)\delta 23d$ (E), and (D) $\gamma 8b_{-}(A)\delta 23d$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red, see SI for details) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), 0.8 ns (B), $\rho_1 = 2.0$ ns and $\rho_2 = 1.0$ ns (C), $\rho = 1.3$ ns (D), 0.6 ns (E) and 1.0 ns (F).

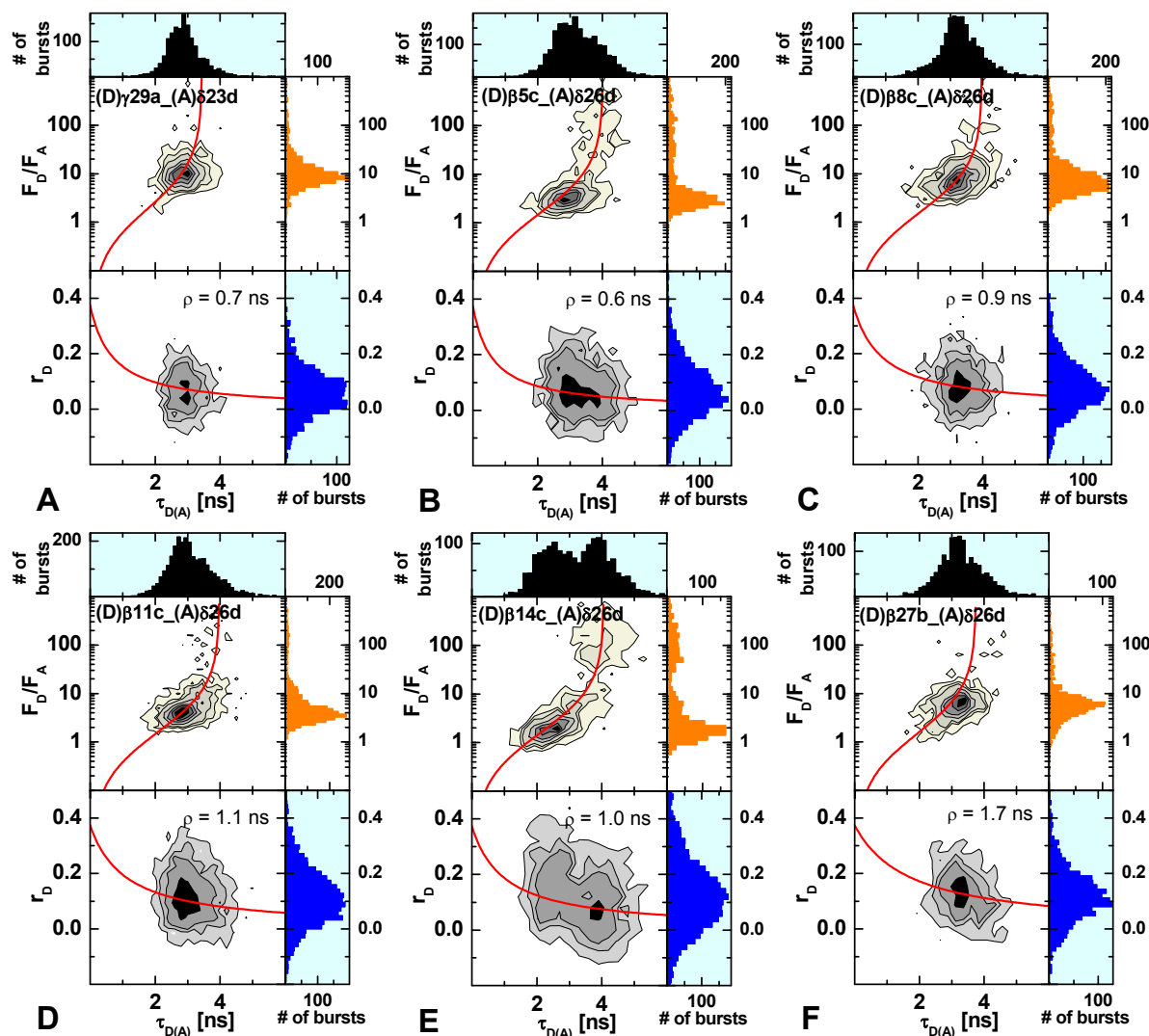


Figure 1.7. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) γ 29a_(A) δ 23d (A), (D) β 5c_(A) δ 26d (B), (D) β 8c_(A) δ 26d (C), (D) β 11c_(A) δ 26d (D), (D) β 14c_(A) δ 26d (E), and (D) β 27b_(A) δ 26d (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.7$ ns (A), 0.6 ns (B), 0.9 ns (C), 1.1 ns (D), 1.0 ns (E) and 1.7 ns (F).

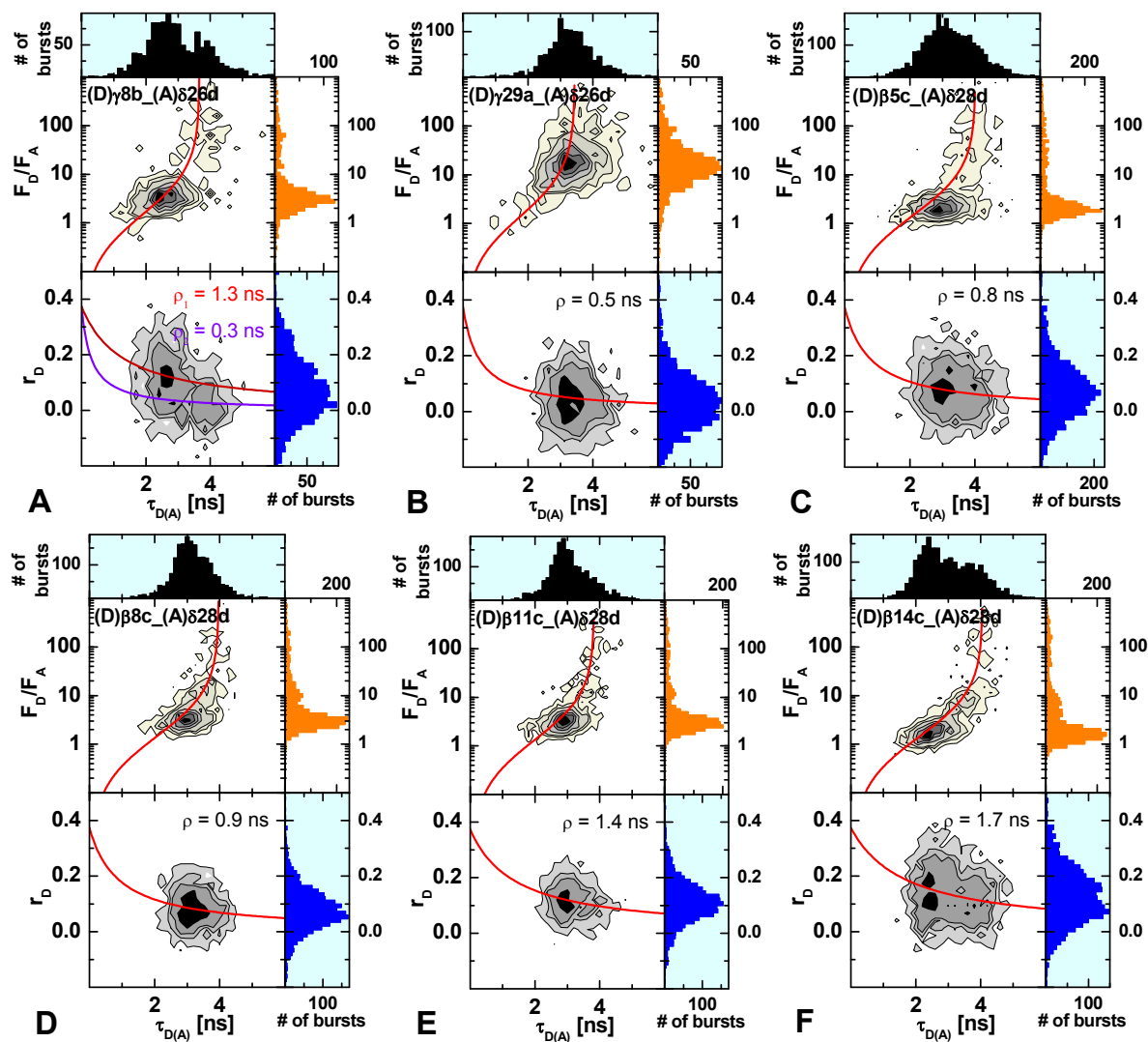


Figure 1.8. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) γ 8b_(A) δ 26d (A), (D) γ 29a_(A) δ 26d (B), (D) β 5c_(A) δ 28d (C), (D) β 8c_(A) δ 28d (D), (D) β 11c_(A) δ 28d (E), and (D) β 14c_(A) δ 28d (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho_1 = 1.3$ ns and $\rho_2 = 0.3$ ns (A), $\rho = 0.5$ ns (B), 0.8 ns (C), 0.9 ns (D), 1.4 ns (E) and 1.7 ns (F).

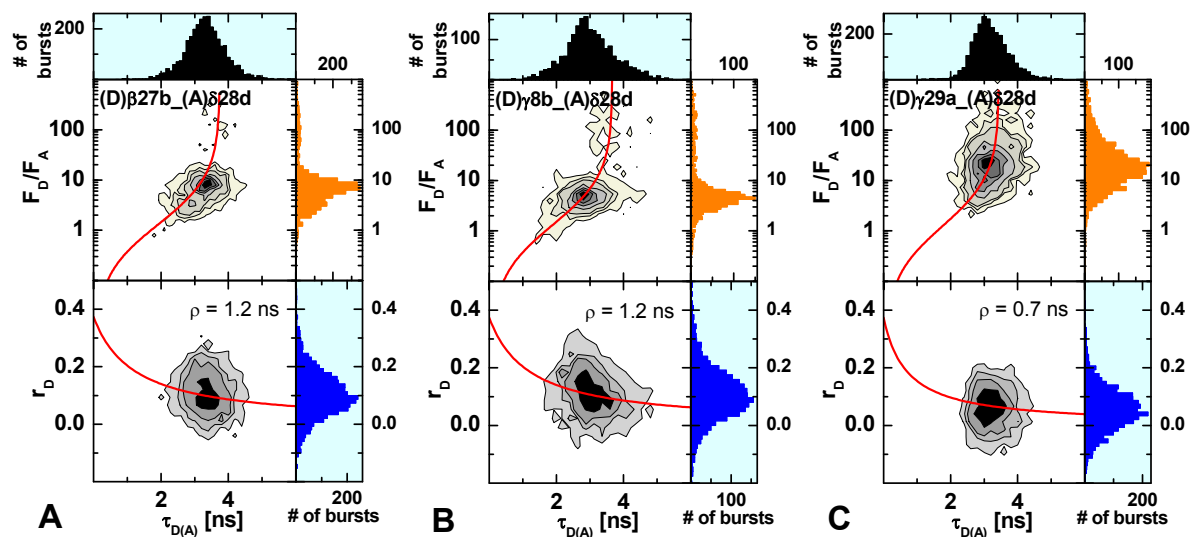


Figure 1.9. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples $(D)\beta 27b_ (A)\delta 28d$ (A), $(D)\gamma 8b_ (A)\delta 28d$ (B) and $(D)\gamma 29a_ (A)\delta 28d$ (C). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.2$ ns (A), $\rho = 1.2$ ns (B) and $\rho = 0.7$ ns (C).

1.2 Mg²⁺-titrations

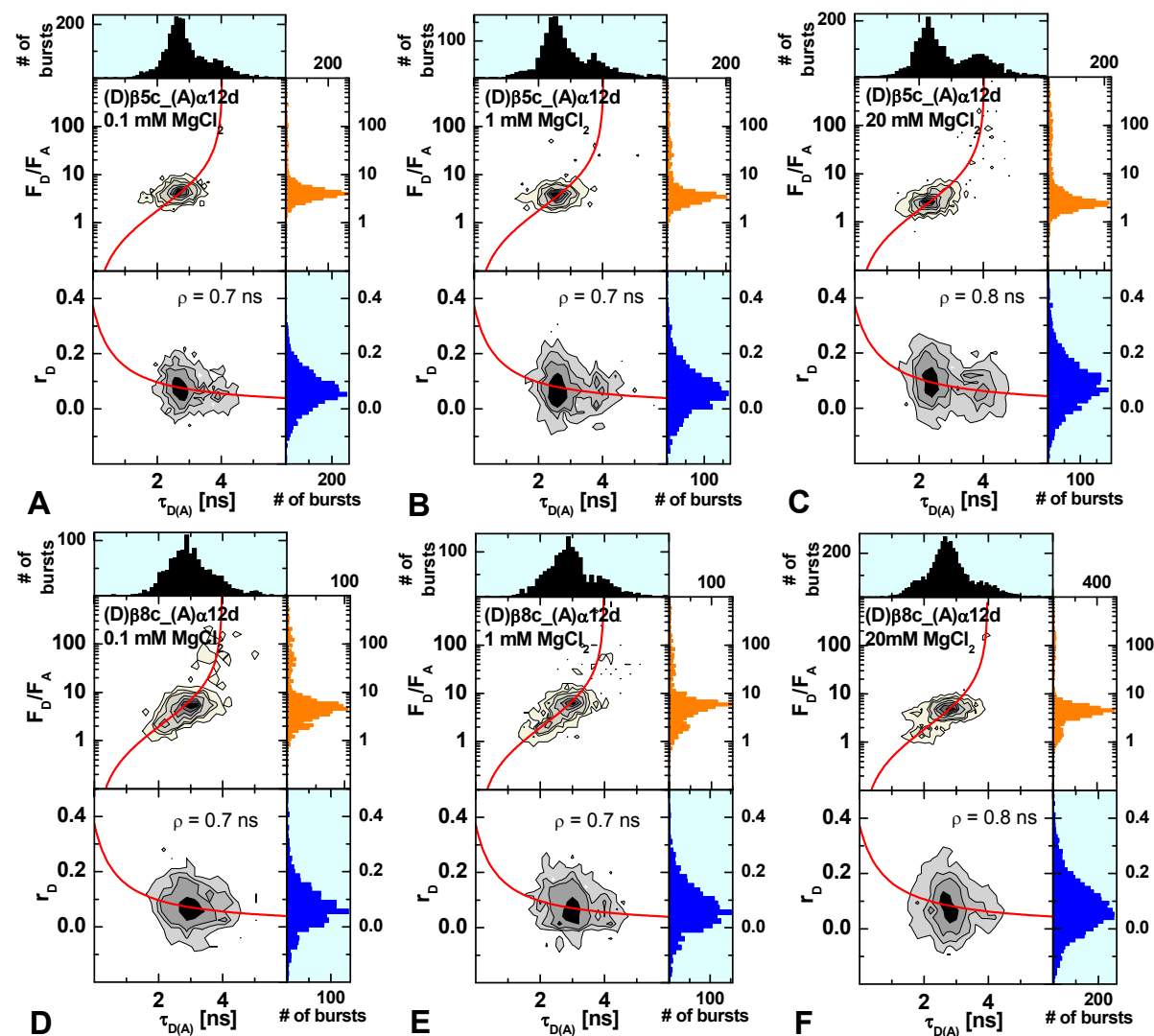


Figure 1.10. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β5c_(A)α12d (A - C) and (D)β8c_(A)α12d (D - F) measured at 0.1 (left), 1 (middle) and 20 mM MgCl₂ (right). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.7$ ns (A), $\rho = 0.7$ ns (B), $\rho = 0.8$ ns (C), $\rho = 0.7$ ns (D), $\rho = 0.7$ ns (E) and $\rho = 0.8$ ns (F).

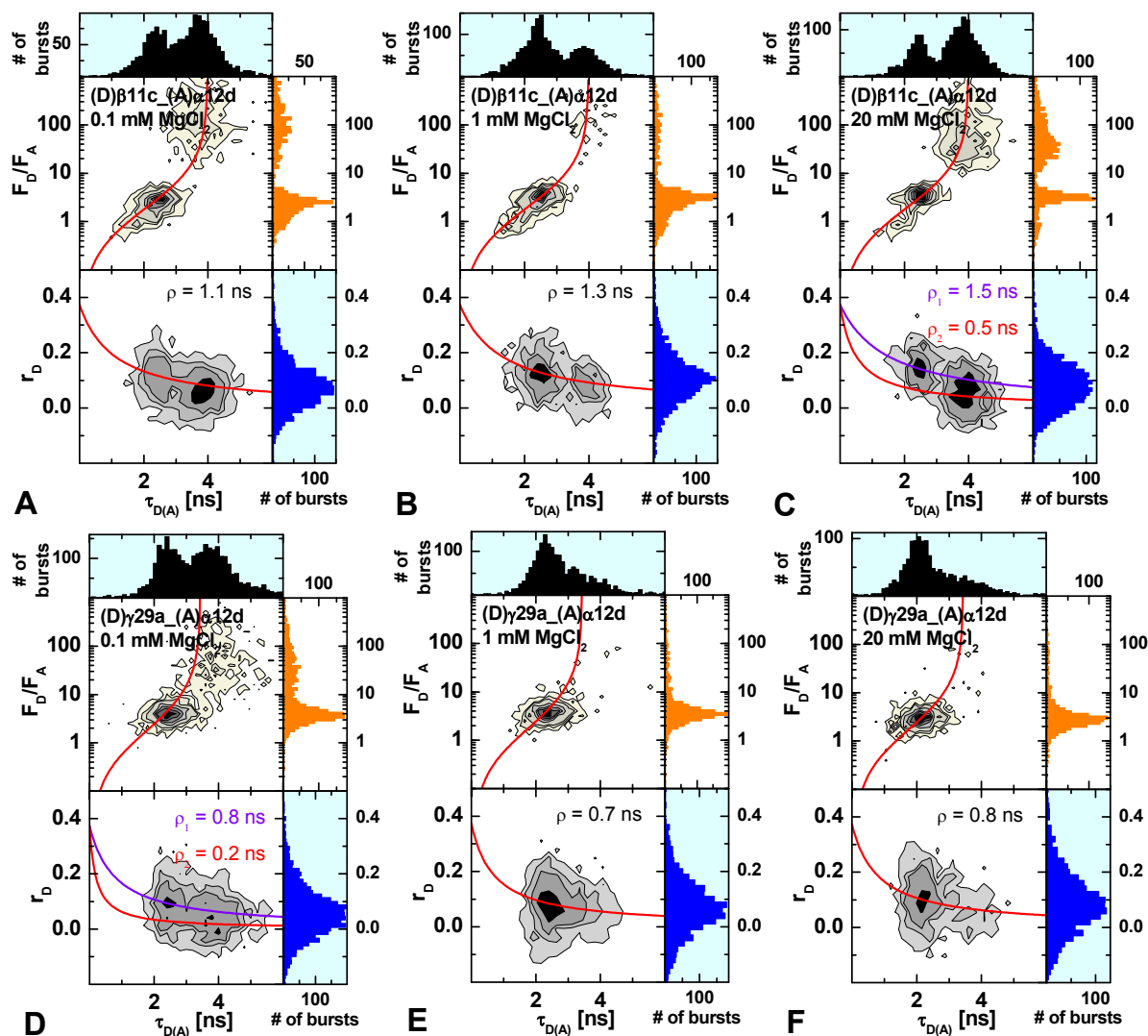


Figure 1.11. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples $(D)\beta11c_ (A)\alpha12d$ (A - C) and $(D)\gamma29a_ (A)\alpha12d$ (D - F) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red and purple lines in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.1$ ns (A), $\rho = 1.3$ ns (B), $\rho_1 = 0.5$ ns and $\rho_2 = 1.5$ ns (C), $\rho_1 = 0.2$ ns and $\rho_2 = 0.8$ ns (D), $\rho = 0.7$ ns (E) and $\rho = 0.8$ ns (F).

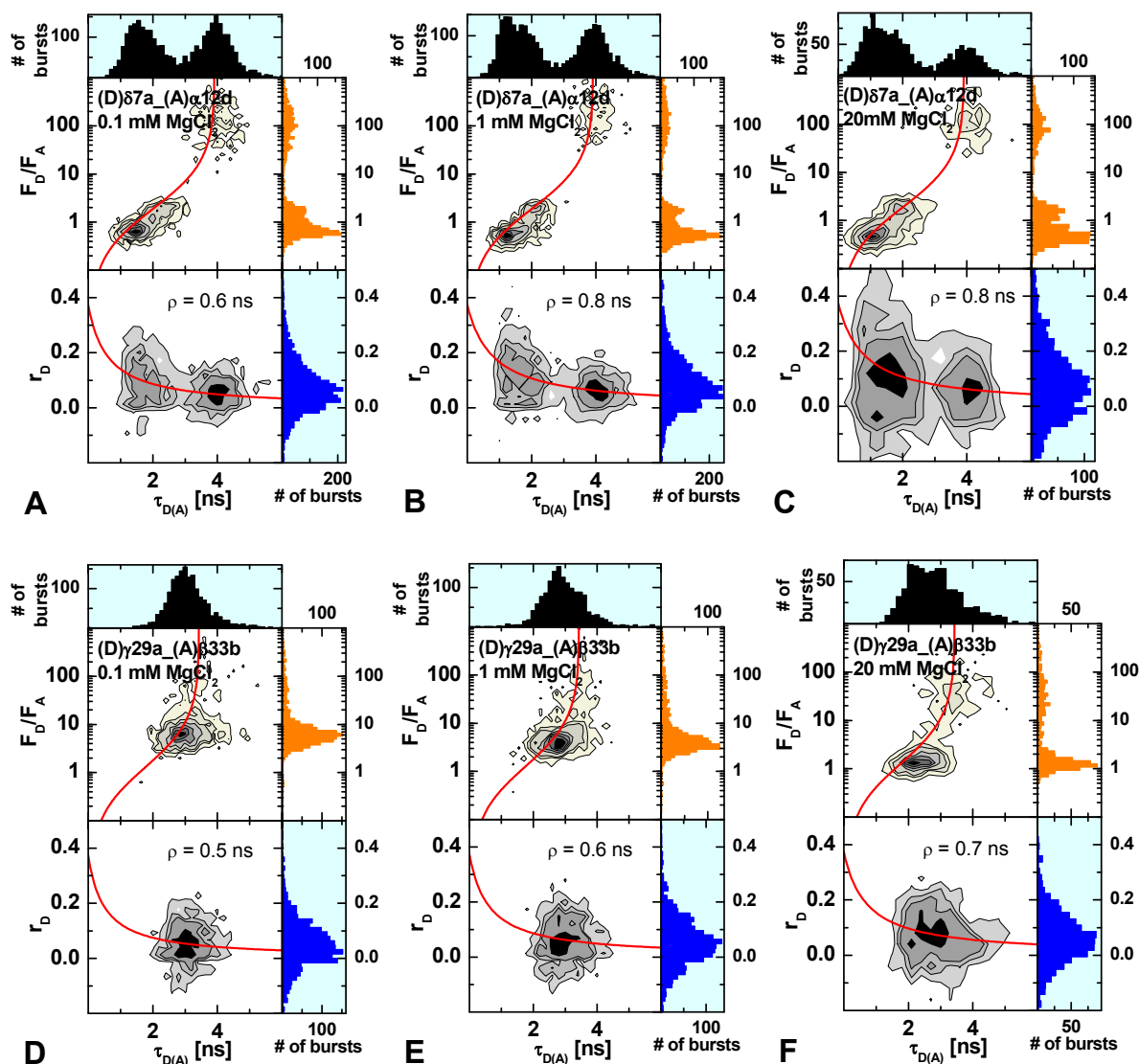


Figure 1.12. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples $(D)\delta7a_ (A)\alpha12d$ (A - C) and $(D)\gamma29a_ (A)\beta33b$ (D - F) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (right). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.6$ ns (A), $\rho = 0.8$ ns (B), $\rho = 0.8$ ns (C), $\rho = 0.5$ ns (D), $\rho = 0.6$ ns (E) and $\rho = 0.7$ ns (F).

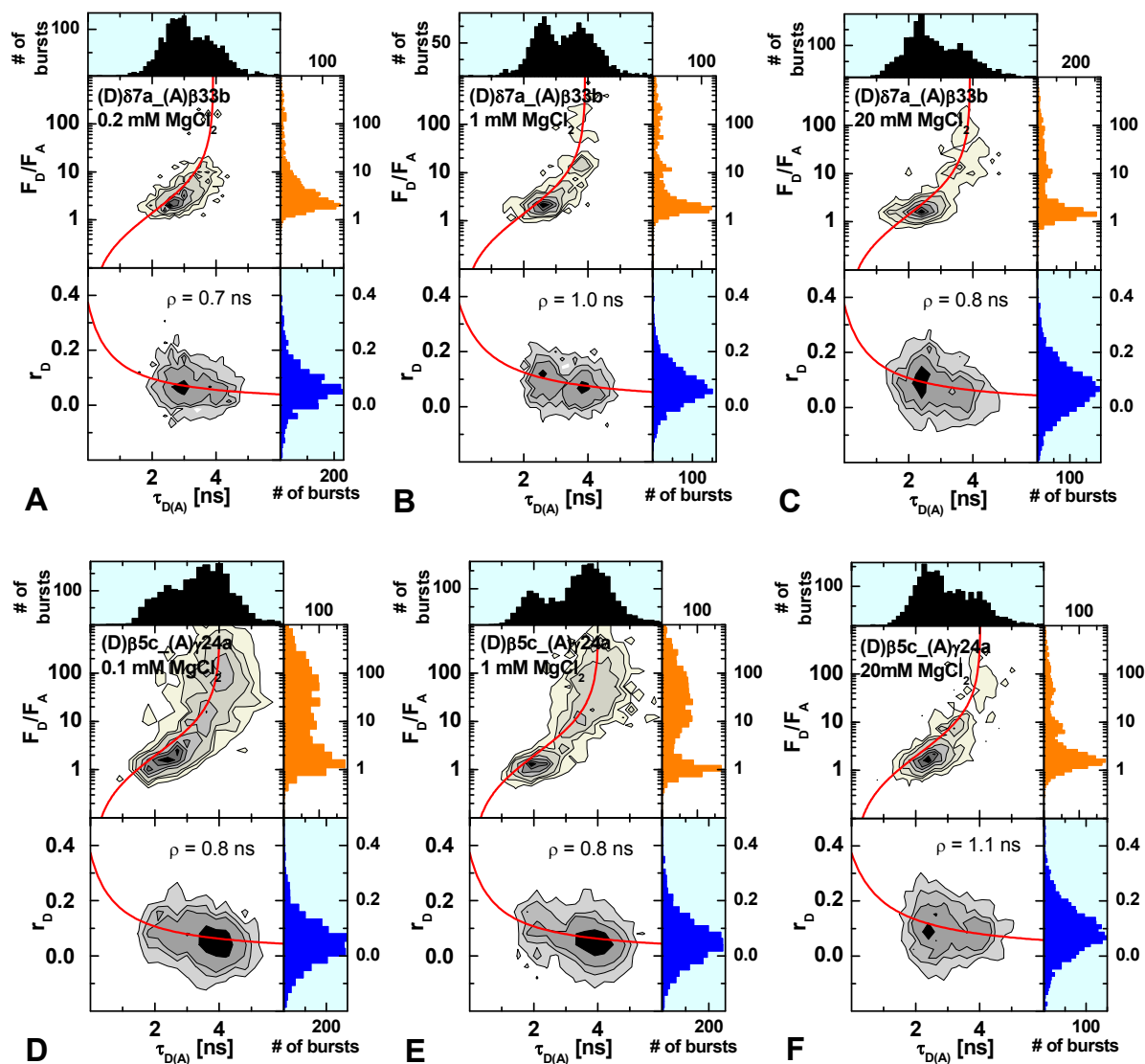


Figure 1.13. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)δ7a_(A)β33b (A - C) and (D)β5c_(A)γ24a (D - F) measured at 0.1/0.2 (left), 1 (middle) and 20 mM MgCl₂ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.7$ ns (A), $\rho = 1.0$ ns (B), $\rho = 0.8$ ns (C), $\rho = 0.8$ ns (D), $\rho = 0.8$ ns (E) and $\rho = 1.1$ ns (F).

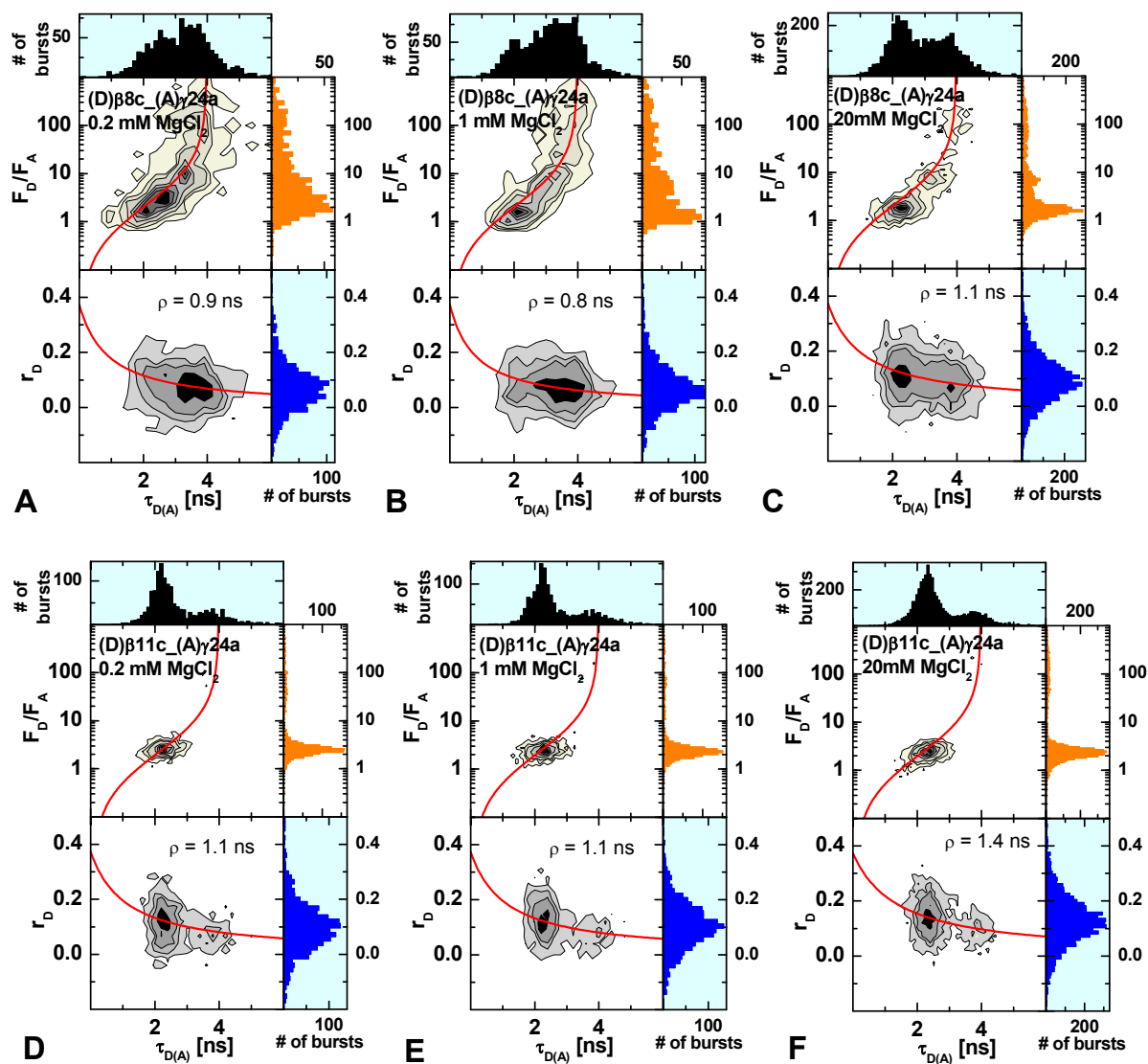


Figure 1.14. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β8c_(A)γ24a (A - C) and (D)β11c_(A)γ24a (D - F) measured at 0.2 (left), 1 (middle) and 20 mM MgCl₂ (right). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.9$ ns (A), $\rho = 0.8$ ns (B), $\rho = 1.1$ ns (C), $\rho = 1.1$ ns (D), $\rho = 1.1$ ns (E) and $\rho = 1.4$ ns (F).

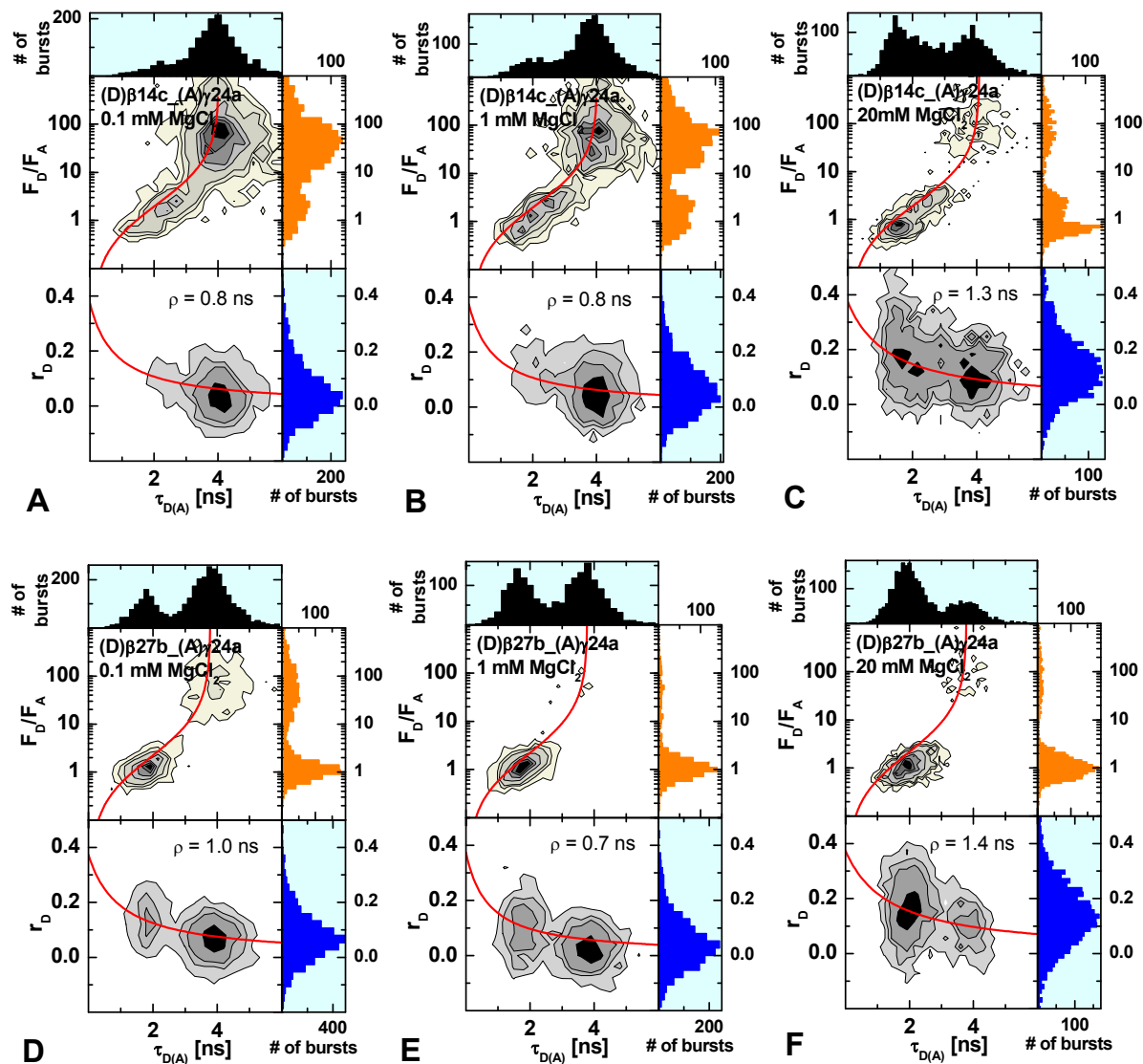


Figure 1.15. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) β 14c_(A) γ 24a (A - C) and (D) β 27b_(A) γ 24a (D - F) measured at 0.1 (left), 1 (middle) and 20 mM MgCl_2 (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), $\rho = 0.8$ ns (B), $\rho = 1.3$ ns (C), $\rho = 1.0$ ns (D), $\rho = 0.7$ ns (E) and $\rho = 1.3$ ns (F).

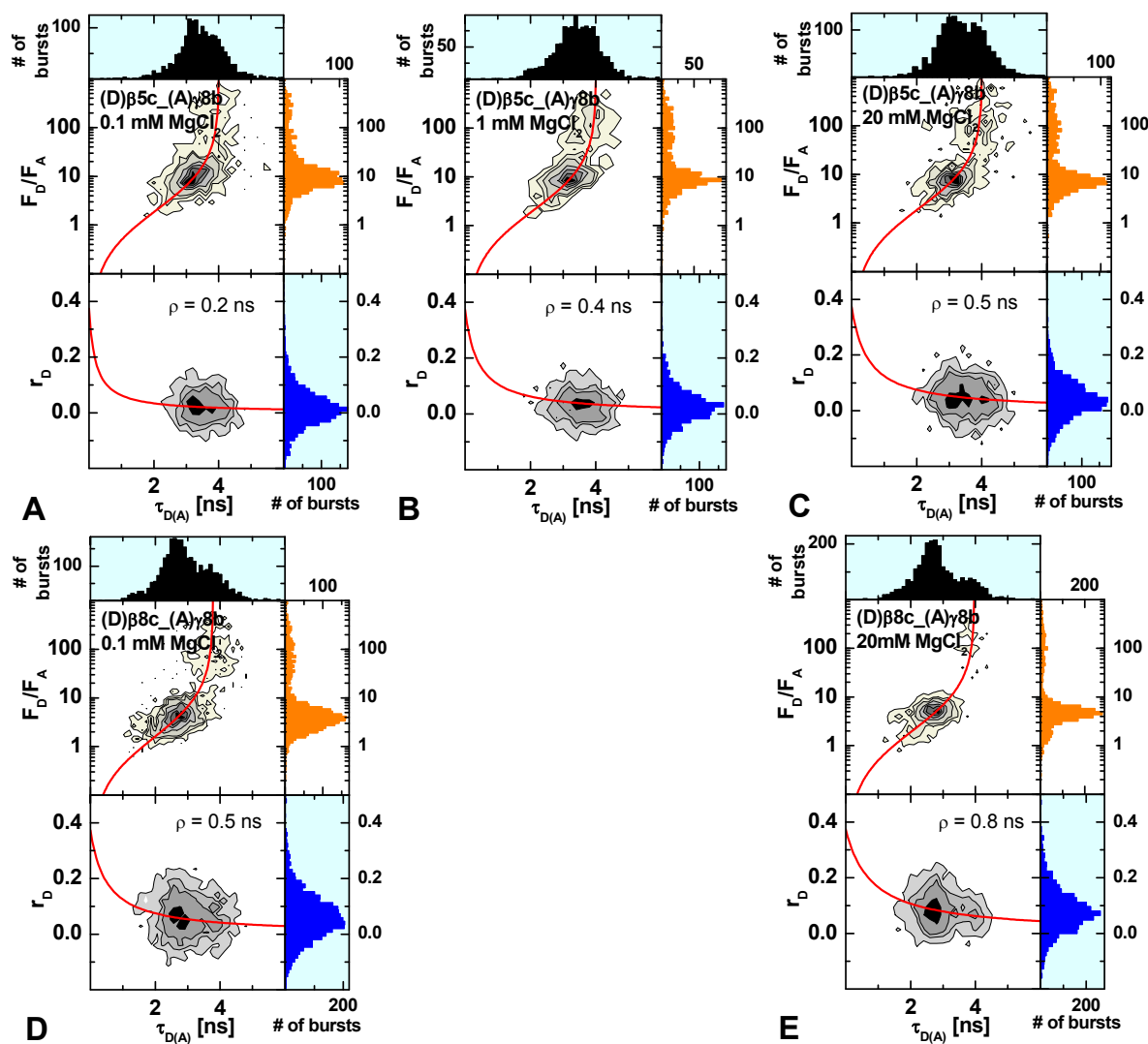


Figure 1.16. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\beta 5c_{-}(A)\gamma 8b$ (A - C) and (D) $\beta 8c_{-}(A)\gamma 8b$ (D and E) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.2$ ns (A), $\rho = 0.4$ ns (B), $\rho = 0.5$ ns (C), $\rho = 0.5$ ns (D) and $\rho = 0.8$ ns (E).

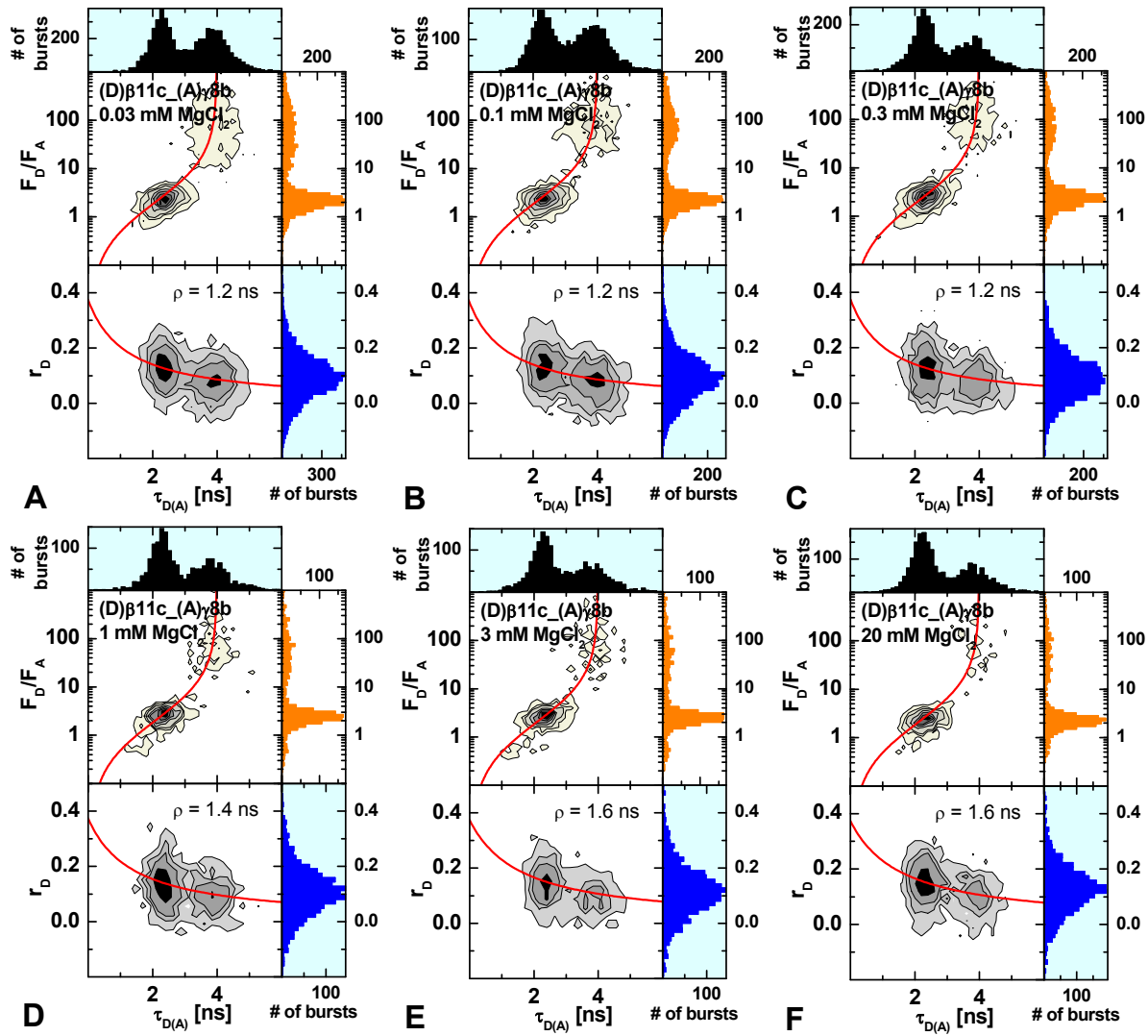


Figure 1.17. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for sample (D)β11c_(A)γ8b measured at 0.03 (A), 0.1 (B), 0.3 (C), 1 (D), 3 (E) and 20 mM MgCl₂ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.2$ ns (A), $\rho = 1.2$ ns (B), $\rho = 1.2$ ns (C), $\rho = 1.4$ ns (D), $\rho = 1.6$ ns (E) and $\rho = 1.6$ ns (F).

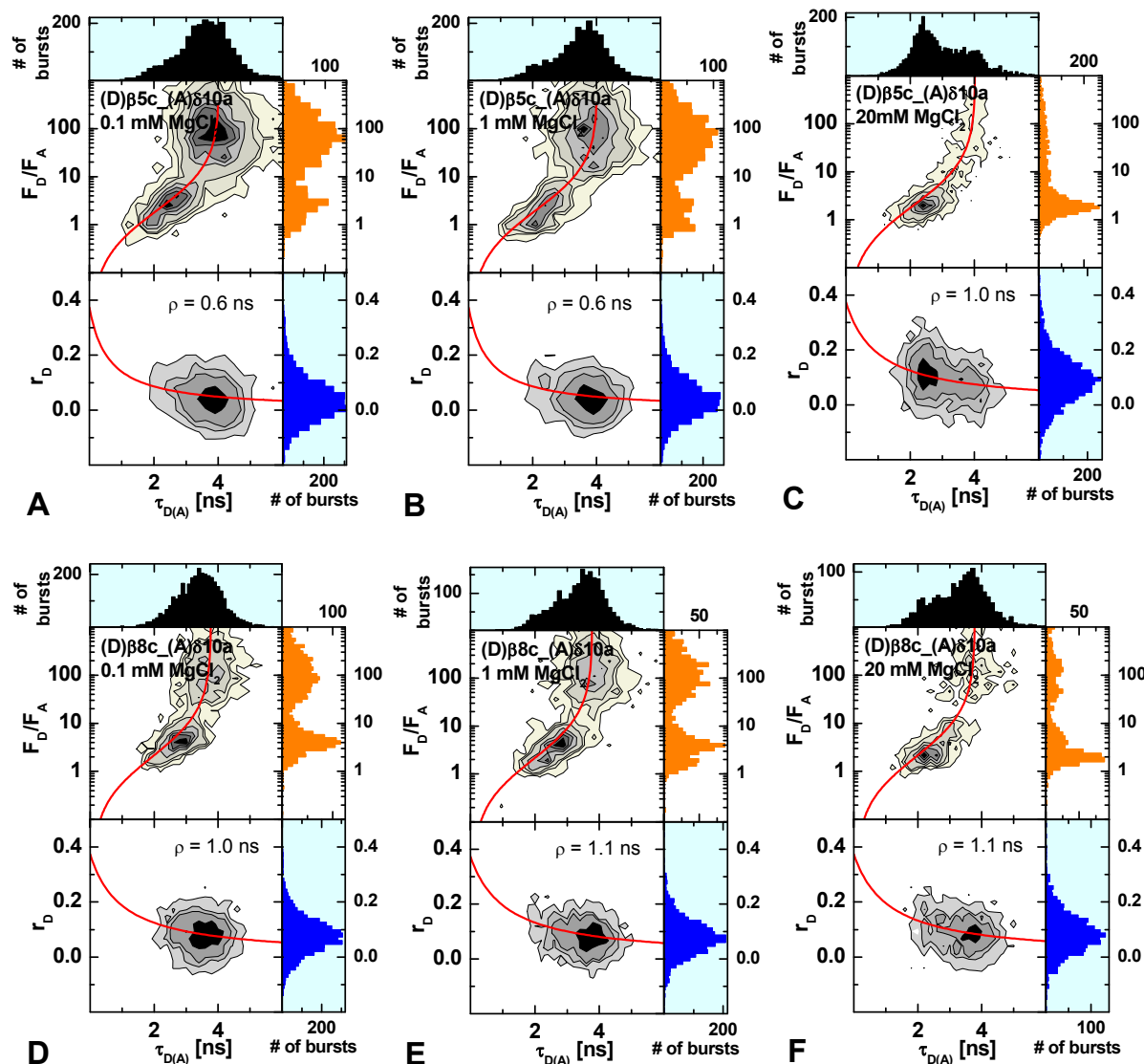


Figure 1.18. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D) $\beta 5c$ _(A) $\delta 10a$ (A - C) and (D) $\beta 8c$ _(A) $\delta 10a$ (D - F) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.6$ ns (A), $\rho = 0.6$ ns (B), $\rho = 1.0$ ns (C), $\rho = 1.0$ ns (D), $\rho = 1.1$ ns (E) and $\rho = 1.1$ ns (F).

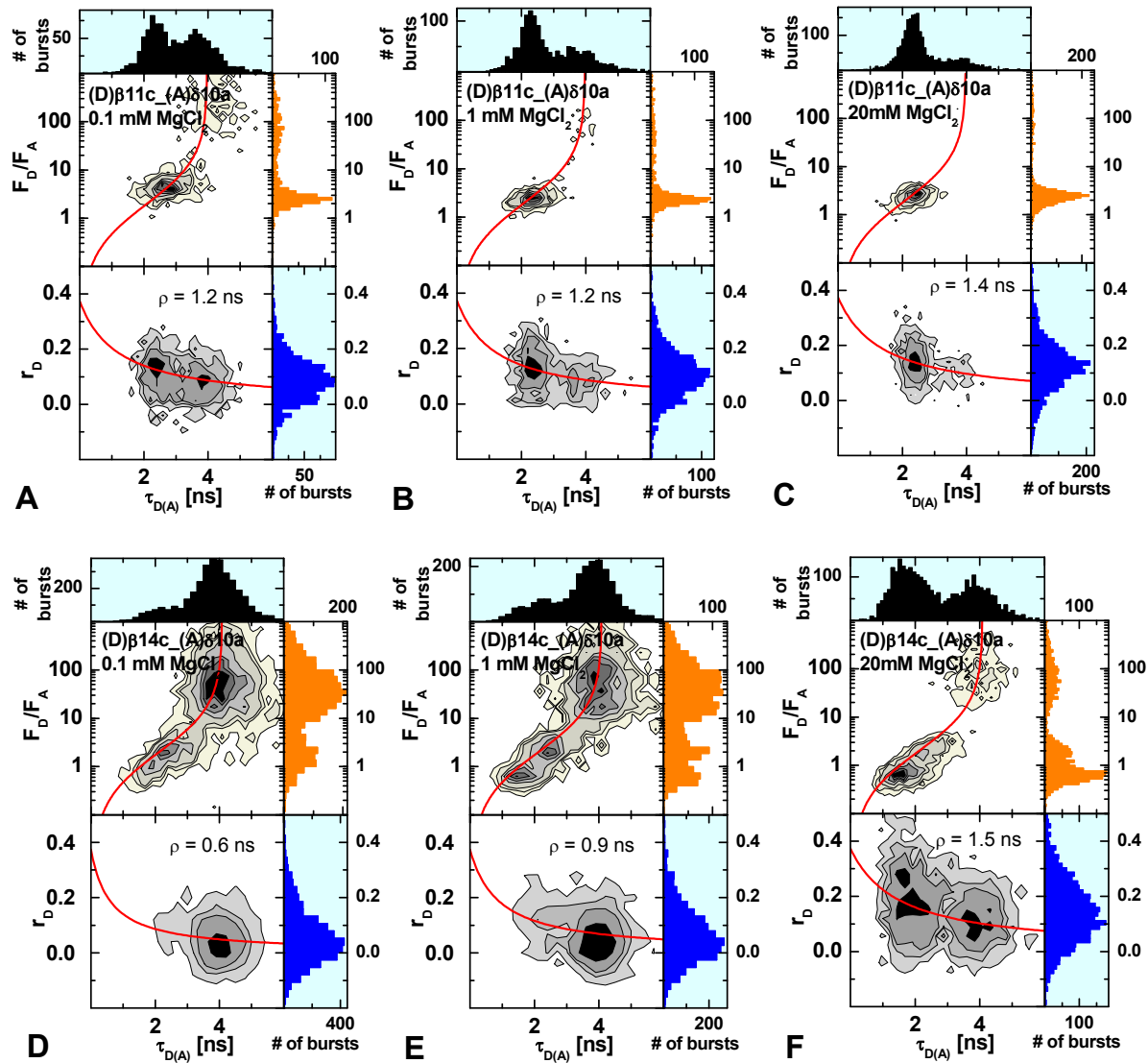


Figure 1.19. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β11c_(A)δ10a (A - C) and (D)β14c_(A)δ10a (D - F) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 1.2$ ns (A), $\rho = 1.2$ ns (B), $\rho = 1.4$ ns (C), $\rho = 0.6$ ns (D), $\rho = 0.9$ ns (E) and $\rho = 1.5$ ns (F).

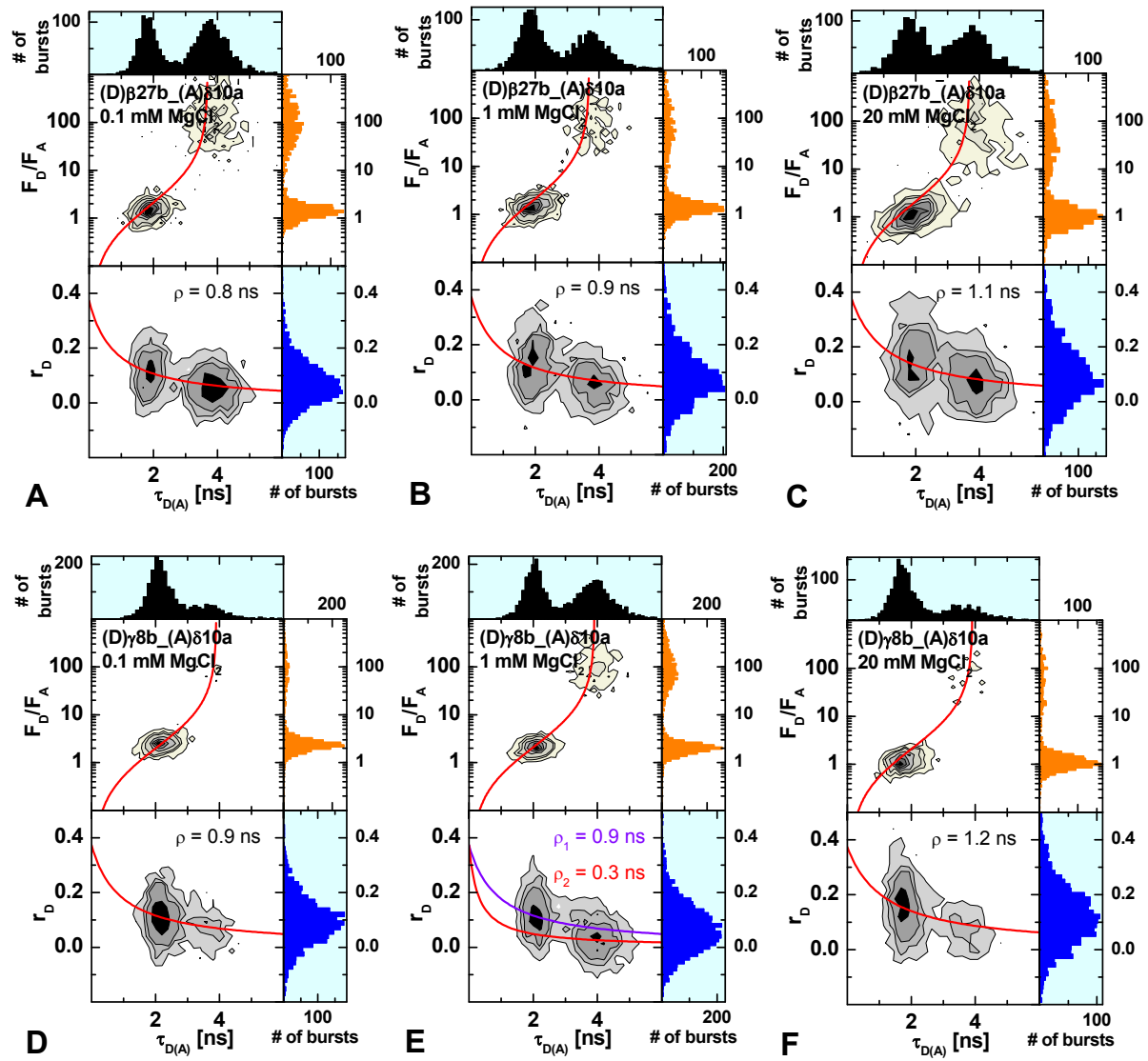


Figure 1.20. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β27b_(A)δ10a (A - C) and (D)γ8b_(A)δ10a (D - F) measured at 0.1 (left), 1 (middle) and 20 mM MgCl₂ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red and purple lines in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), $\rho = 0.9$ ns (B), $\rho = 1.1$ ns (C), $\rho = 0.9$ ns (D), $\rho_1 = 0.9$ ns and $\rho_2 = 0.3$ ns (E) and $\rho = 1.2$ ns (F).

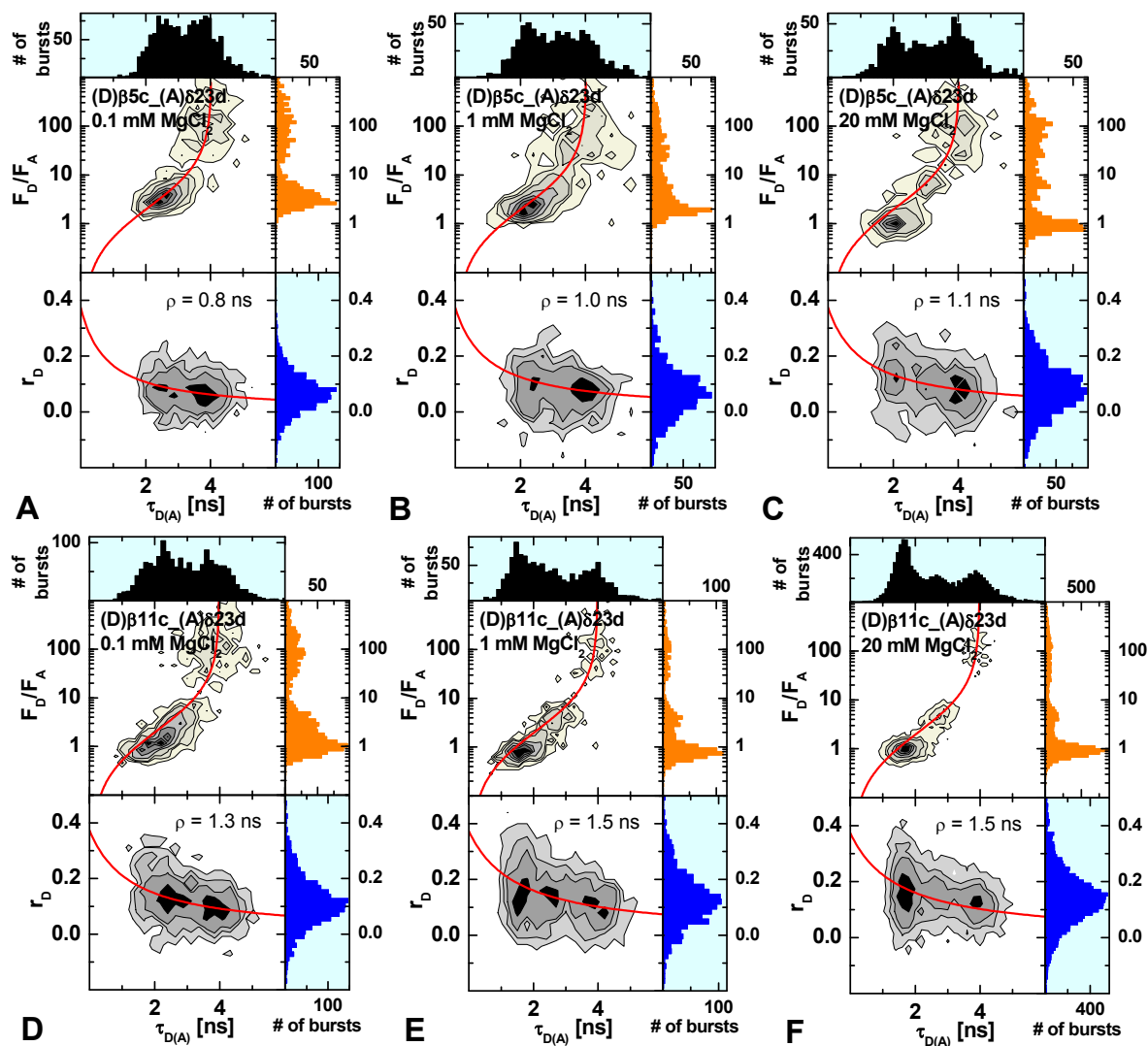


Figure 1.21. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples $(D)\beta5c_{(A)}\delta23d$ (A - C) and $(D)\beta11c_{(A)}\delta23d$ (D - F) measured at 0.1 (left), 1 (middle) and 20 mM $MgCl_2$ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 0.8$ ns (A), $\rho = 1.0$ ns (B), $\rho = 1.1$ ns (C), $\rho = 1.3$ ns (D), $\rho = 1.5$ ns (E) and $\rho = 1.5$ ns (F).

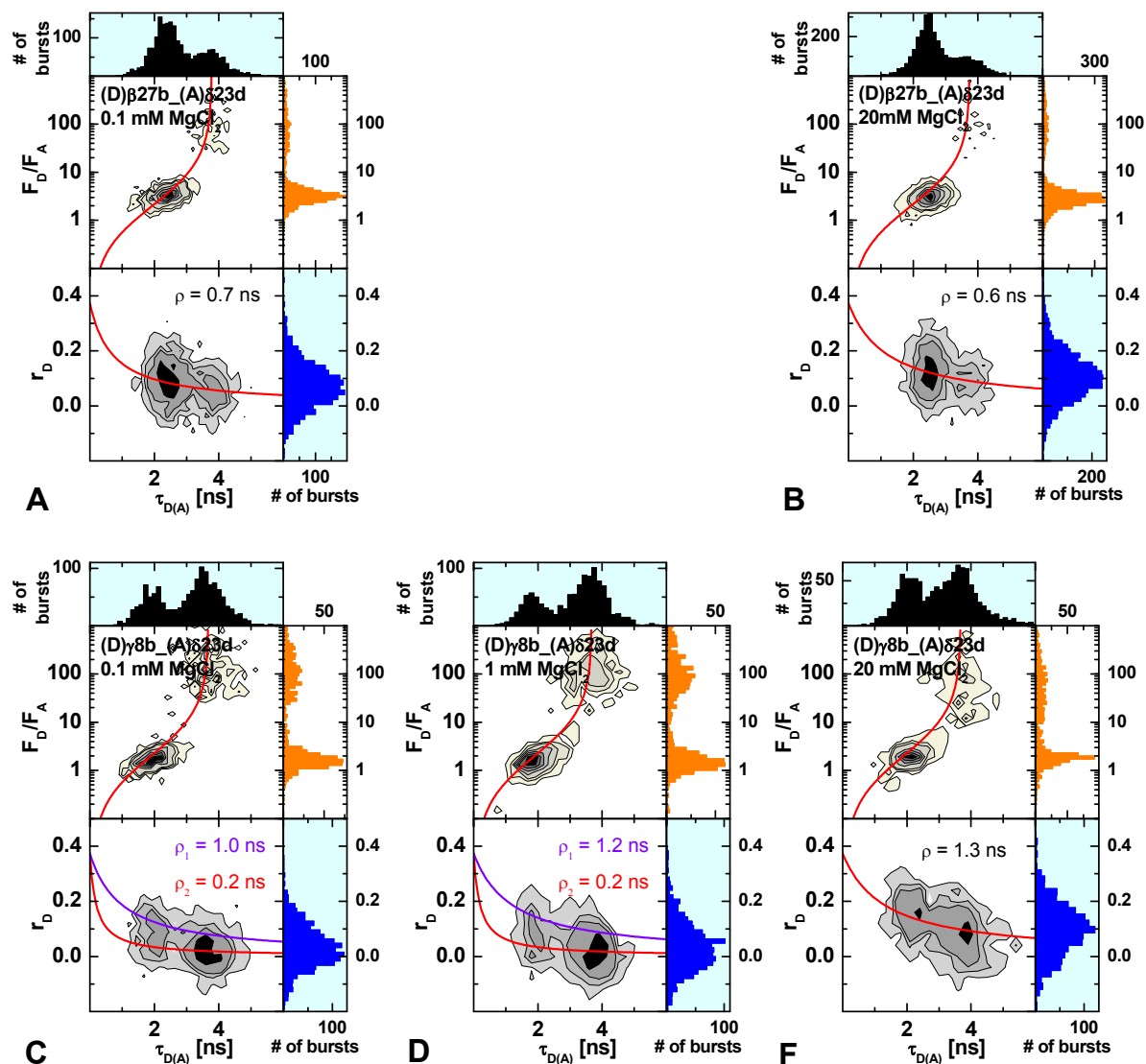


Figure 1.22. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β27b_(A)δ23d (A and B) and (D)γ8b_(A)δ23d (C - E) measured at 0.1 (left), 1 (middle) and 20 mM MgCl₂ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red and purple lines in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 0.7$ ns (A), $\rho = 0.6$ ns (B), $\rho_1 = 1.0$ ns and $\rho_2 = 0.2$ ns (C), $\rho_1 = 1.2$ ns and $\rho_2 = 0.2$ ns (D) and $\rho = 1.3$ ns (E).

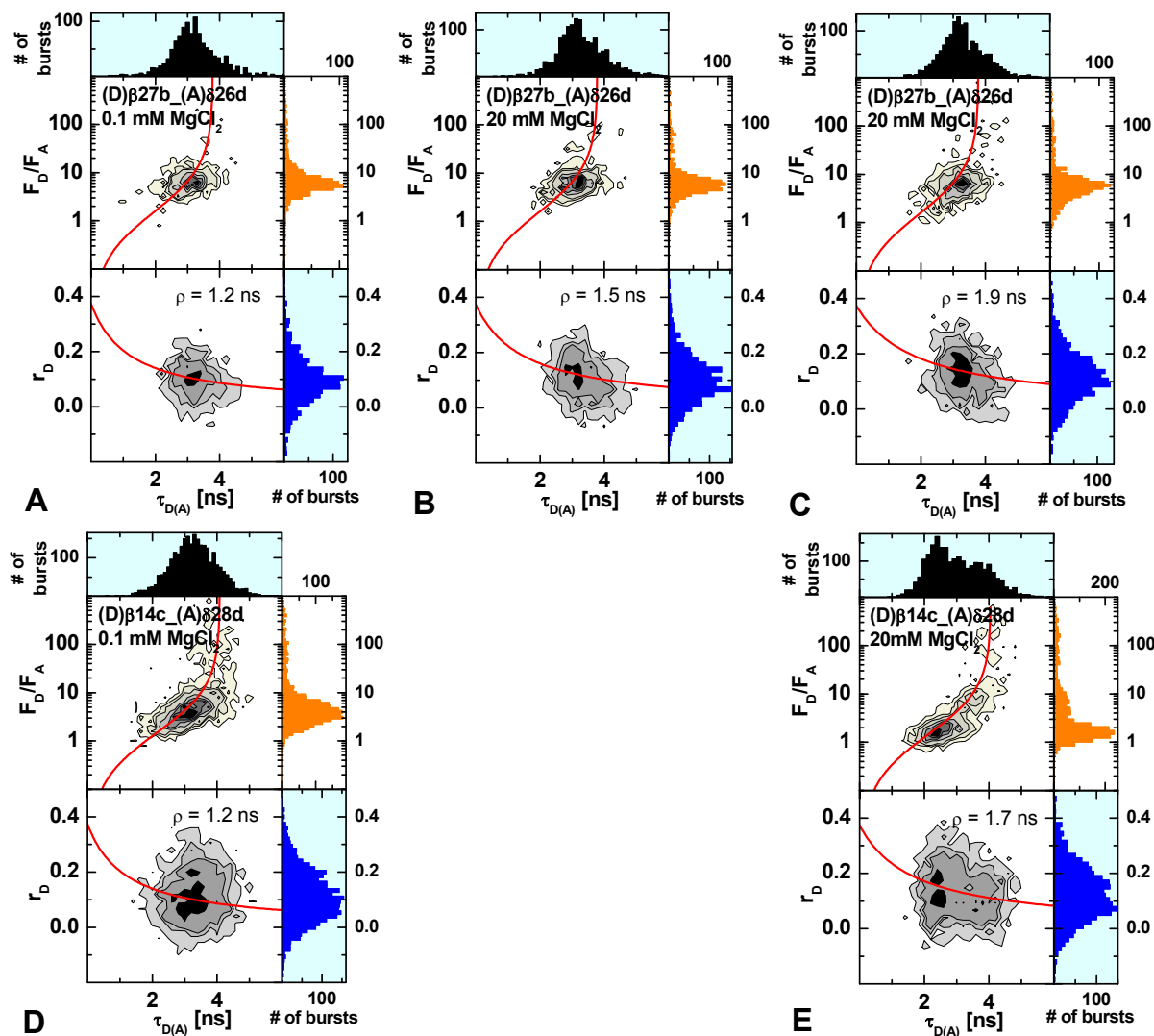


Figure 1.23. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)β27b_(A)δ26d (A - C) and (D)β14c_(A)δ28d (D and E) measured at 0.1 (left), 1 (middle) and 20 mM MgCl₂ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs. $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1 + \tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 1.2$ ns (A), $\rho = 1.5$ ns (B), $\rho = 1.9$ ns (C), $\rho = 1.2$ ns (D) and $\rho = 1.7$ ns (E).

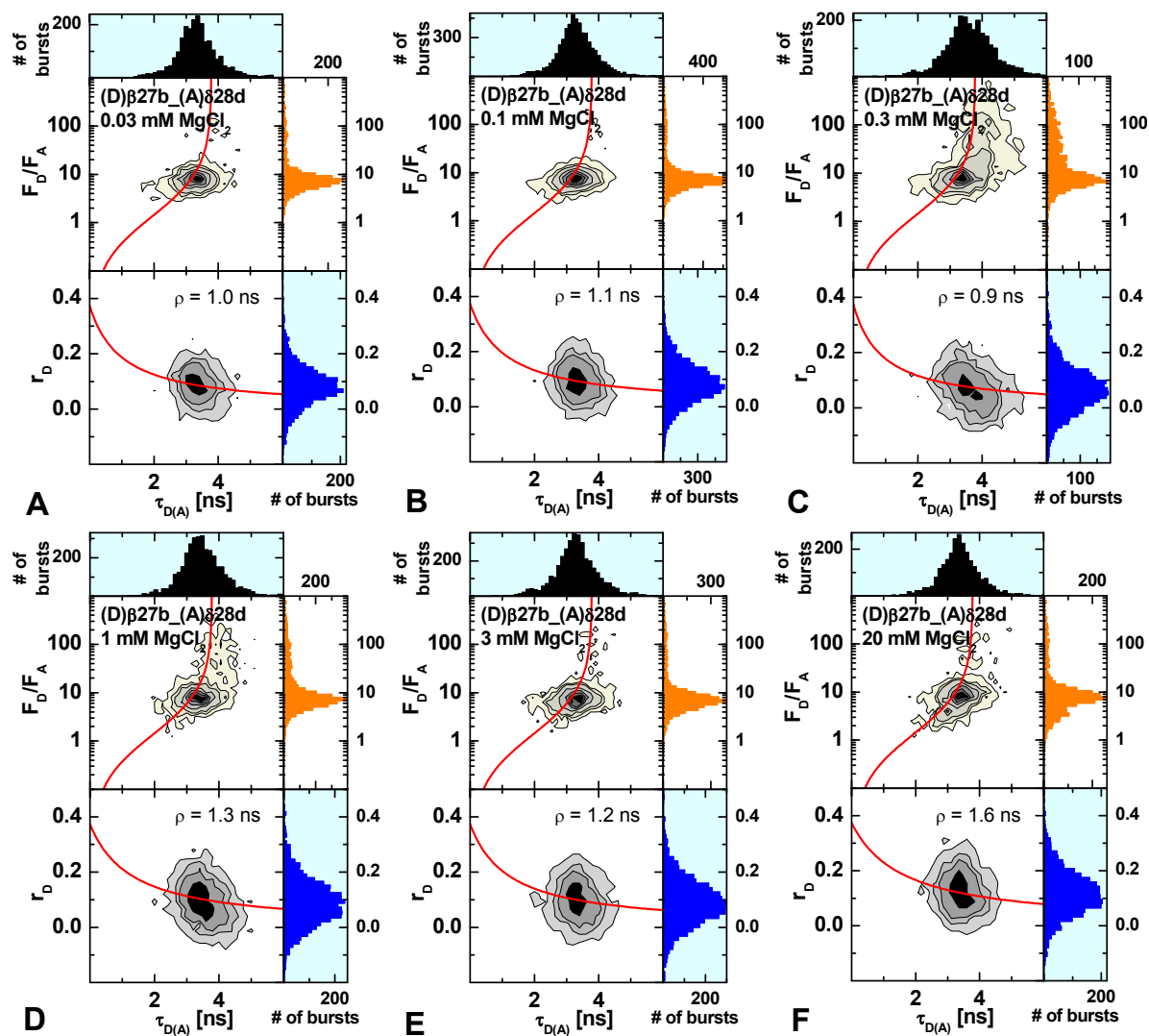


Figure 1.24. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for sample (D) β 27b_(A) δ 28d measured at 0.03 (A), 0.1 (B), 0.3 (C), 1 (D), 3 (E) and 20 mM $MgCl_2$ (F). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$, with $r_0 = 0.374$ and $\rho = 1.0$ ns (A), $\rho = 1.1$ ns (B), $\rho = 0.9$ ns (C), $\rho = 1.3$ ns (D), $\rho = 1.2$ ns (E) and $\rho = 1.6$ ns (F).

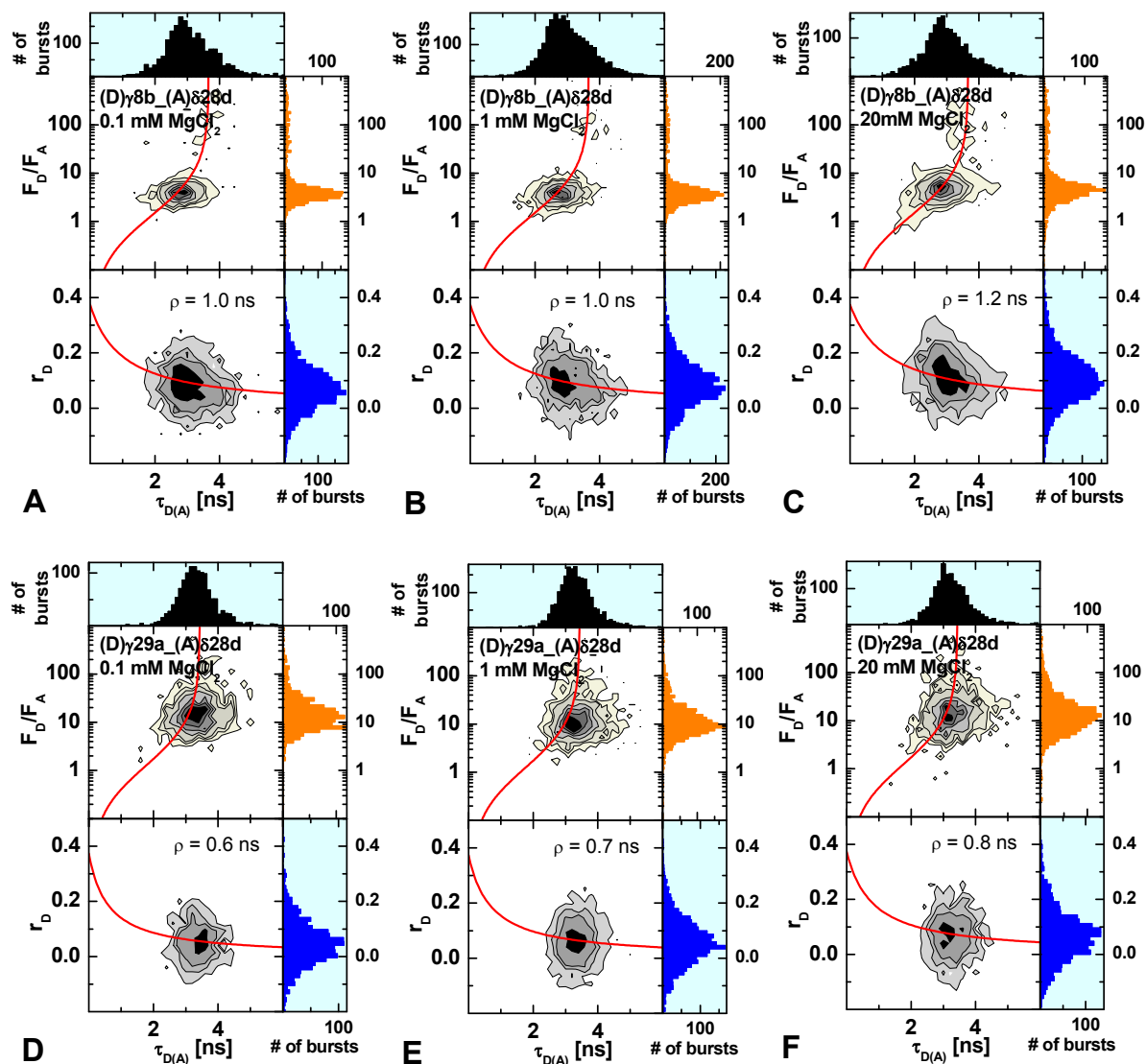


Figure 1.25. 2D burst frequency histograms of F_D/F_A versus the donor fluorescence lifetime $\tau_{D(A)}$ (upper panel) and the donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (lower panel) for samples (D)γ8b_(A)δ28d (A - C) and (D)γ29a_(A)δ28d (D and E) measured at 0.1 (left), 1 (middle) and 20 mM MgCl₂ (left). The number of molecules (fluorescence bursts) in each bin is gray scale, shaded from white (lowest) to black (highest). 1D histograms are shown as projections. In the F_D/F_A vs $\tau_{D(A)}$ plot, the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; red) is overlaid. The solid red line in the r_D - $\tau_{D(A)}$ diagram is given by the Perrin equation $r_D = r_0 / (1 + \tau_{D(A)} / \rho)$, with $r_0 = 0.374$ and $\rho = 1.0$ ns (A), $\rho = 1.0$ ns (B), $\rho = 1.2$ ns (C), $\rho = 0.6$ ns (D), $\rho = 0.7$ ns (E) and $\rho = 0.8$ ns (F).

2 PDA plots

2.1 Measurements at 20 mM MgCl₂

2.1.1 Labeling on helix pairs ab and cd

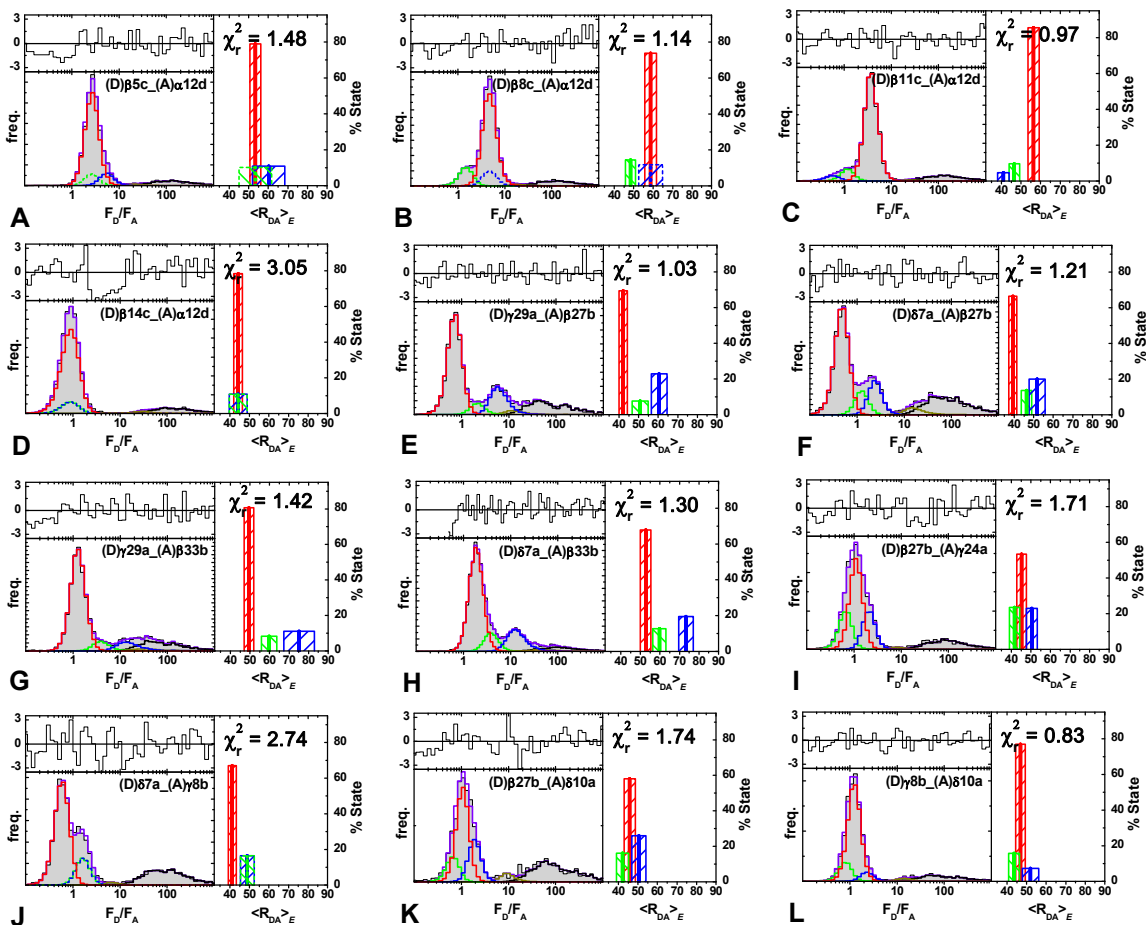


Figure 2.1. PDAs for samples (D) β 5c_(A) α 12d (A), (D) β 8c_(A) α 12d (B), (D) β 11c_(A) α 12d (C), (D) β 14c_(A) α 12d (D), (D) γ 29a_(A) β 27b (E), (D) δ 7a_(A) β 27b (F), (D) γ 29a_(A) β 33b (G), (D) δ 7a_(A) β 33b (H), (D) β 27b_(A) γ 24a (I), (D) δ 7a_(A) γ 8b (J), (D) β 27b_(A) δ 10a (K), (D) γ 8b_(A) δ 10a (L) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances, relative amplitudes (solid lines, dashed if overlapped) and the confidence intervals (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-4 in the SI of “RNA4WJ manuscript” for all parameters used.

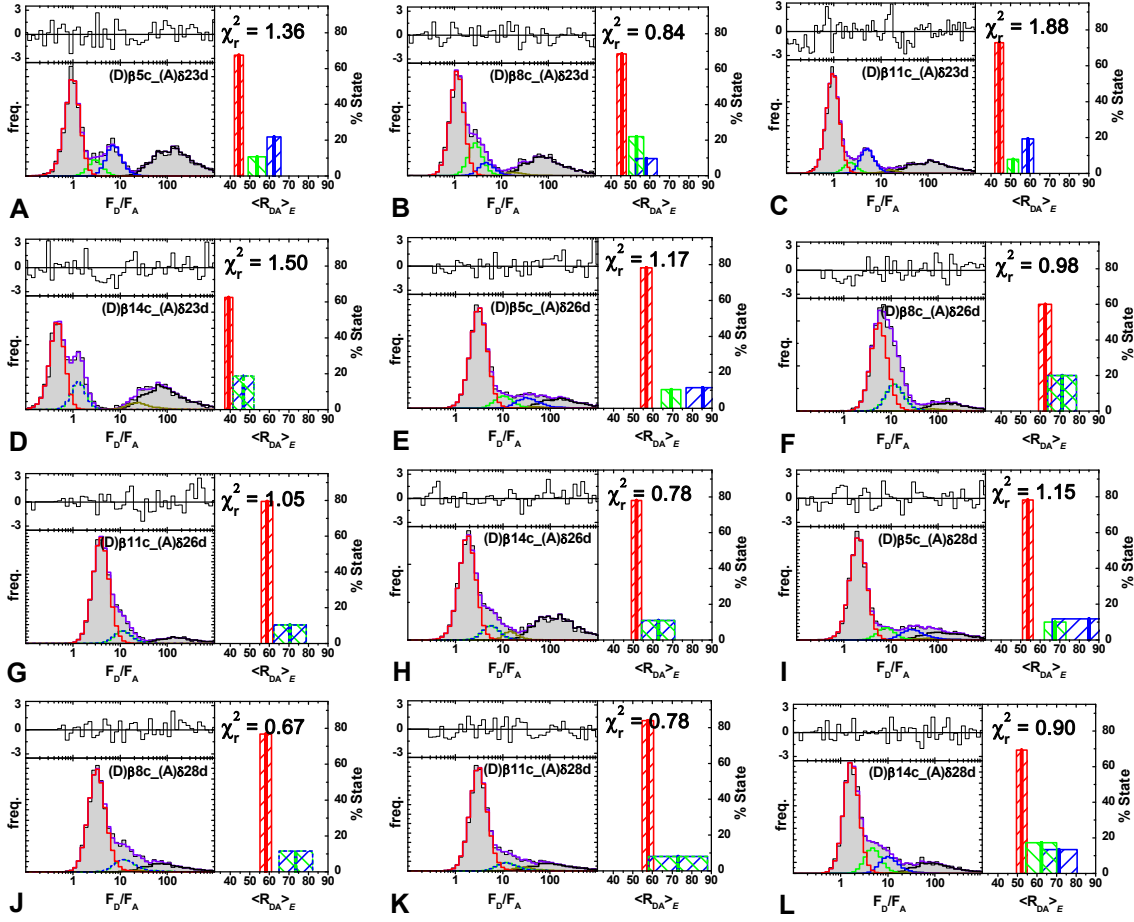


Figure 2.2. PDAs for samples (D) $\beta 5c$ _(A) $\delta 23d$ (A), (D) $\beta 8c$ _(A) $\delta 23d$ (B), (D) $\beta 11c$ _(A) $\delta 23d$ (C), (D) $\beta 14c$ _(A) $\delta 23d$ (D), (D) $\beta 5c$ _(A) $\delta 26d$ (E), (D) $\beta 8c$ _(A) $\delta 26d$ (F), (D) $\beta 11c$ _(A) $\delta 26d$ (G), (D) $\beta 14c$ _(A) $\delta 26d$ (H), (D) $\beta 5c$ _(A) $\delta 28d$ (I), (D) $\beta 8c$ _(A) $\delta 28d$ (J), (D) $\beta 11c$ _(A) $\delta 28d$ (K), (D) $\beta 14c$ _(A) $\delta 28d$ (L) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$, a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances, relative amplitudes (solid lines, dashed if overlapped) and the confidence intervals (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-4 in the SI of “RNA4WJ manuscript” for all parameters used.

2.1.2 Labeling on helix pairs ac and bd

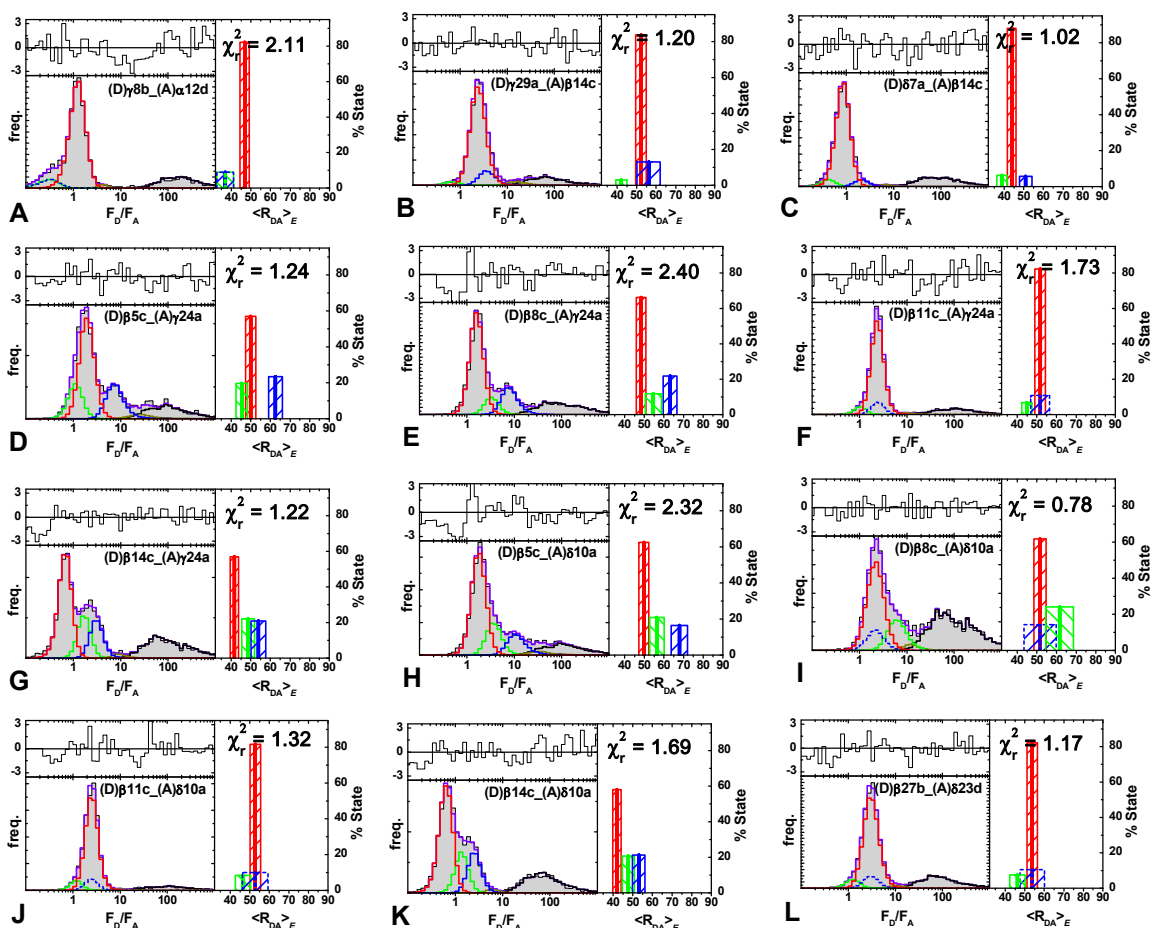


Figure 2.3. PDAs for samples (D) γ 8b_(A) α 12d (A), (D) γ 29a_(A) β 14c (B), (D) δ 7a_(A) β 14c (C), (D) β 5c_(A) γ 24a (D), (D) β 8c_(A) γ 24a (E), (D) β 11c_(A) γ 24a (F), (D) β 14c_(A) γ 24a (G), (D) β 5c_(A) δ 10a (H), (D) β 8c_(A) δ 10a (I), (D) β 11c_(A) δ 10a (J), (D) β 14c_(A) δ 10a (K), (D) β 27b_(A) δ 23d (L) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances, relative amplitudes (solid lines, dashed if overlapped) and the confidence intervals (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-4 in the SI of “RNA4WJ manuscript” for all parameters used.

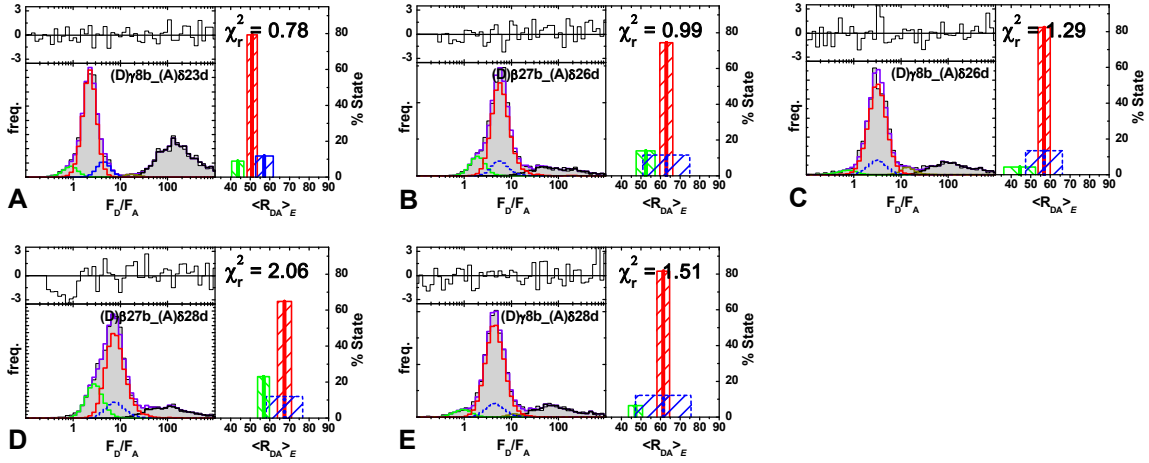


Figure 2.4. PDAs for samples (D) γ 8b_(A) δ 23d (A), (D) β 27b_(A) δ 26d (B), (D) γ 8b_(A) δ 26d (C), (D) β 27b_(A) δ 28d (D) and (D) γ 8b_(A) δ 28d (E) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$, a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances, relative amplitudes (solid lines, dashed if overlapped) and the confidence intervals (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-4 in the SI of “RNA4WJ manuscript” for all parameters used.

2.1.3 Labeling on helix pairs ad and bc

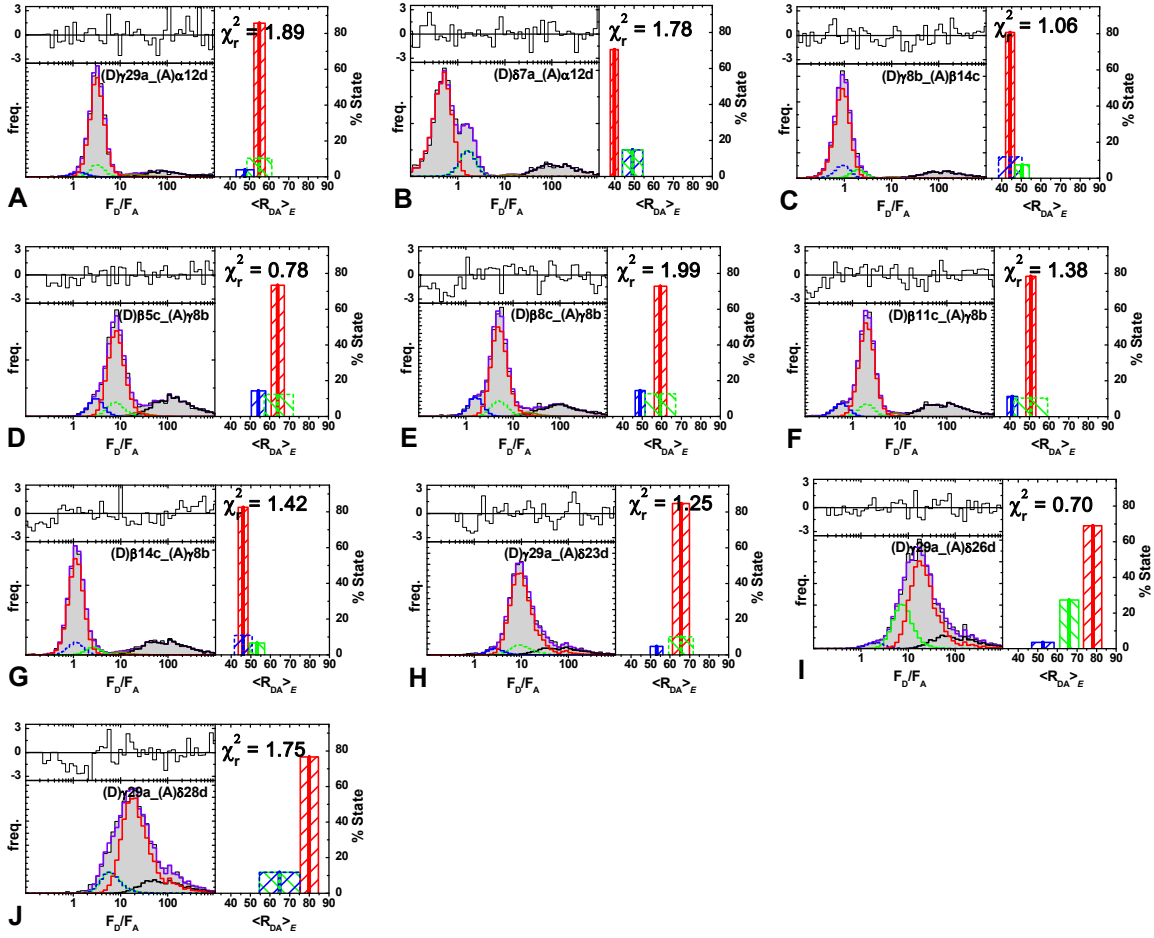


Figure 2.5. PDAs for samples (D) γ 29a_(A) α 12d (A), (D) δ 7a_(A) α 12d (B), (D) γ 8b_(A) β 14c (C), (D) β 5c_(A) γ 8d (D), (D) β 8c_(A) γ 8b (E), (D) β 11c_(A) γ 8b (F), (D) β 14c_(A) γ 8b (G), (D) γ 29a_(A) δ 23d (H), (D) γ 29a_(A) δ 26d (I) and (D) γ 29a_(A) δ 28d (J) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances, relative amplitudes (solid lines, dashed if overlapped) and the confidence intervals (striped boxes) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-4 in the SI of “RNA4WJ manuscript” for all parameters used.

2.2 Mg²⁺-titrations

2.2.1 Labeling on helix pairs ab or cd

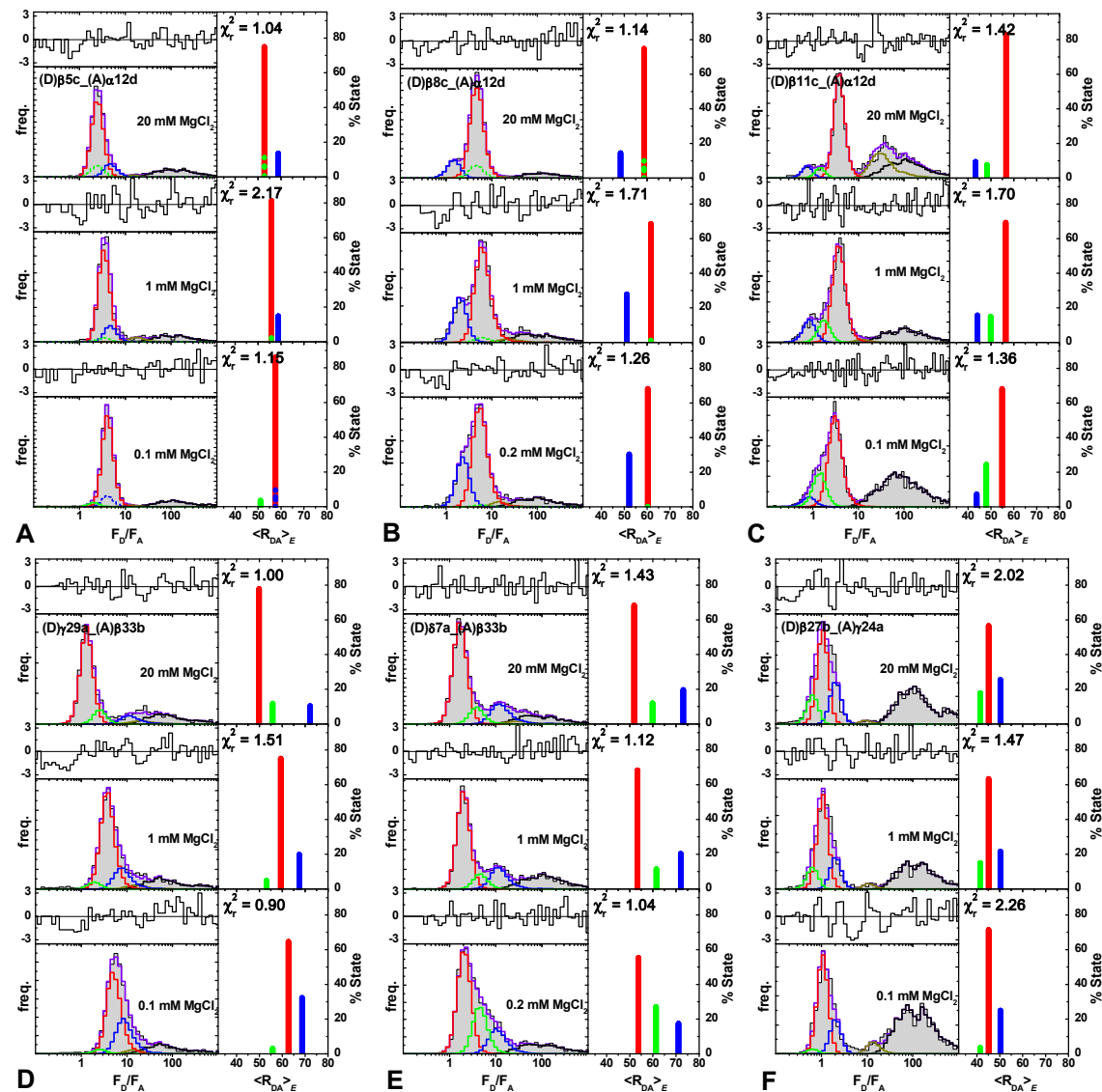


Figure 2.6. PDAs of Mg-titrations for samples (D) β 5c_(A) α 12d (A), (D) β 8c_(A) α 12d (B), (D) β 11c_(A) α 12d (C), (D) γ 29a_(A) β 33b (D), (D) δ 7a_(A) β 33b (E) and (D) β 27b_(A) γ 24a (F) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

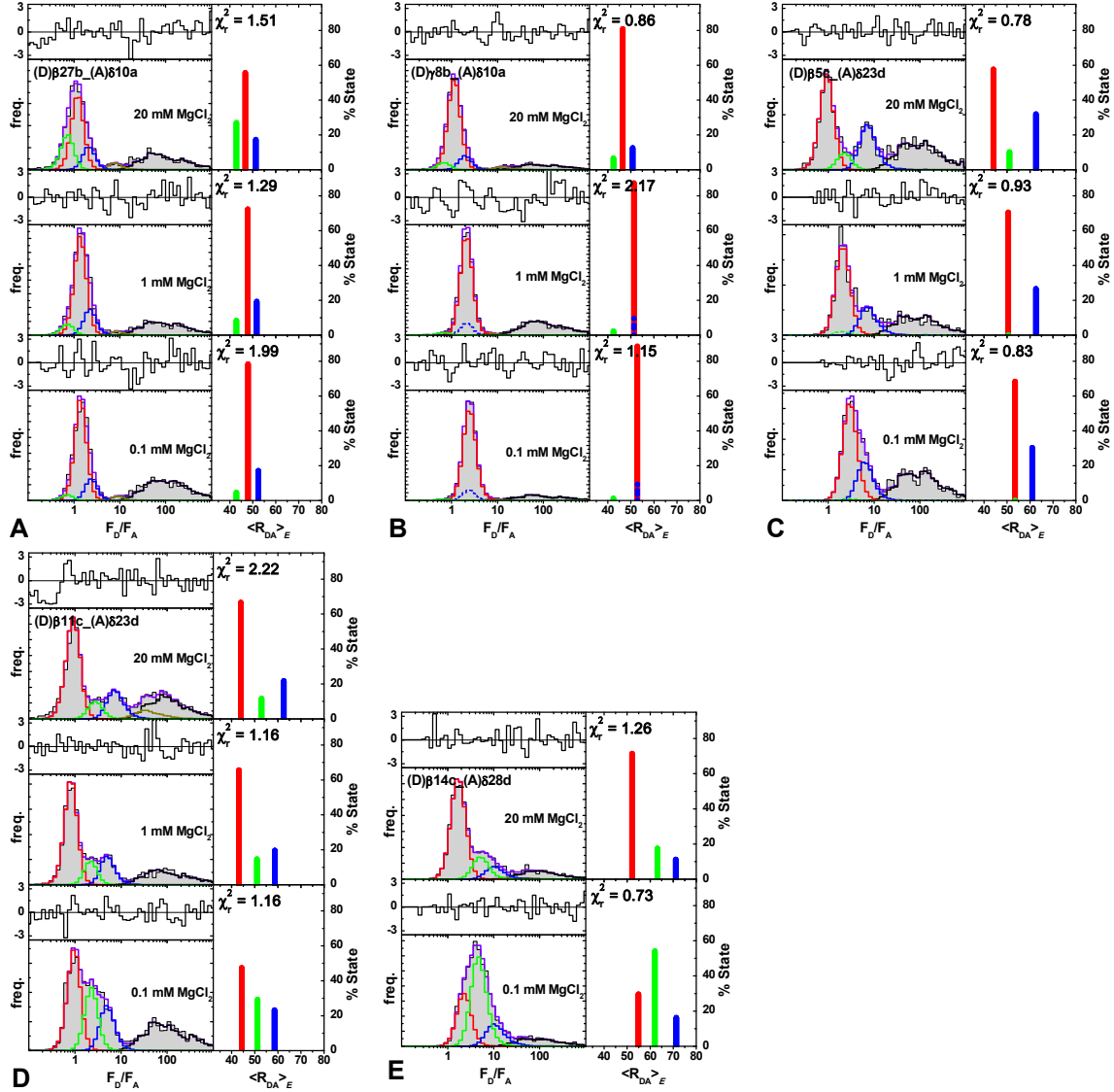


Figure 2.7. PDAs of Mg-titrations for samples (D) β 27b_(A) δ 10a (A), (D) γ 8b_(A) δ 10a (B), (D) β 5c_(A) δ 23d (C), (D) β 11c_(A) δ 23d (D) and (D) β 14c_(A) δ 28d (E) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_{av}$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$, a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_{av}$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

2.2.2 Labeling on helix pairs ac and bd

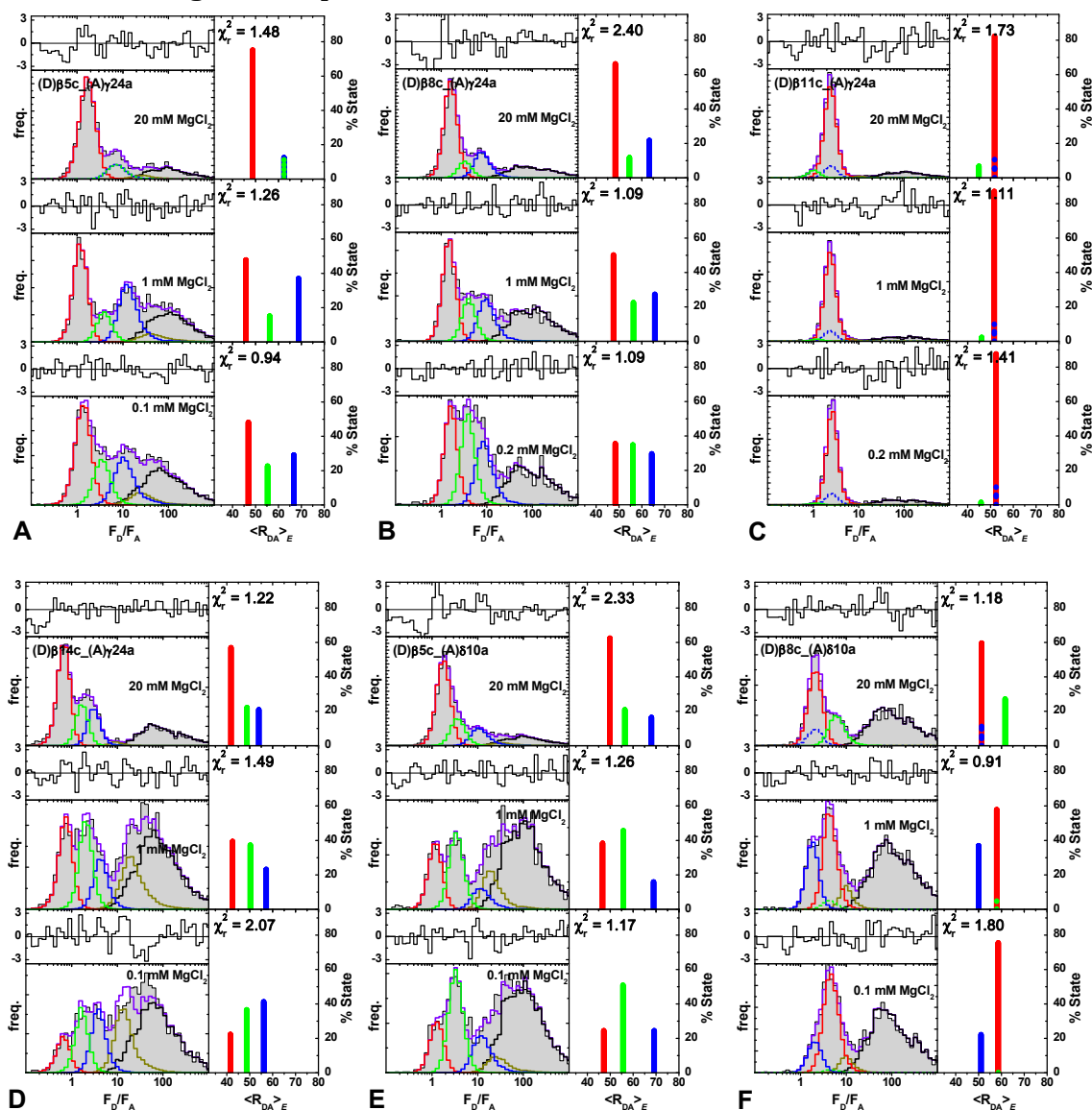


Figure 2.8. PDAs of Mg-titrations for samples $(D)\beta 5c_{-}(A)\gamma 24a$ (A), $(D)\beta 8c_{-}(A)\gamma 24a$ (B), $(D)\beta 11c_{-}(A)\gamma 24a$ (C), $(D)\beta 14c_{-}(A)\gamma 24a$ (D), $(D)\beta 5c_{-}(A)\delta 10a$ (E) and $(D)\beta 8c_{-}(A)\delta 10a$ (F) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

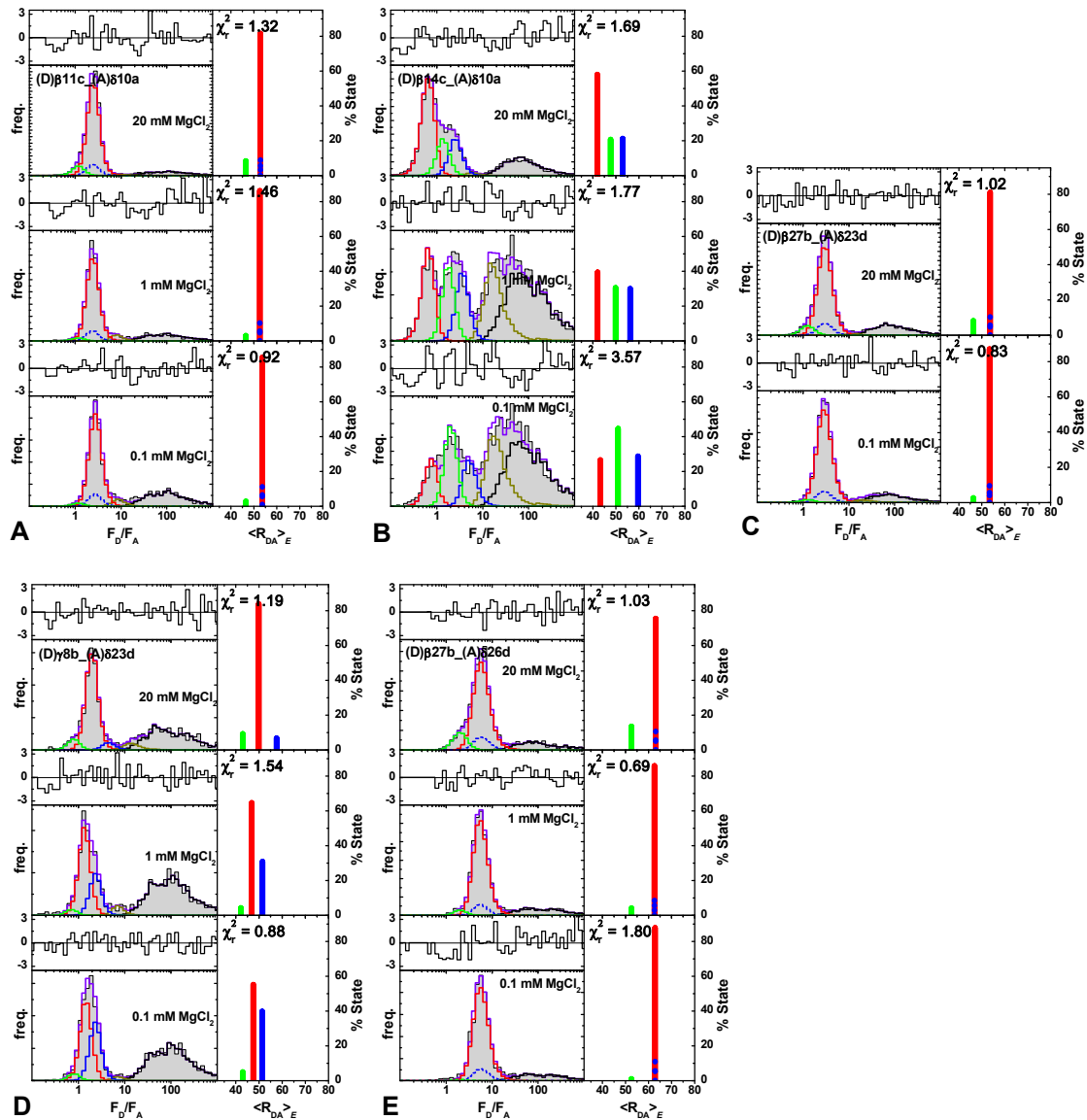


Figure 2.9. PDAs of Mg-titrations for samples (D) β 11c_(A) δ 10a (A), (D) β 14c_(A) δ 10a (B), (D) β 27b_(A) δ 23d (C), (D) γ 8b_(A) δ 23d (D) and (D) β 27b_(A) δ 26d (E) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

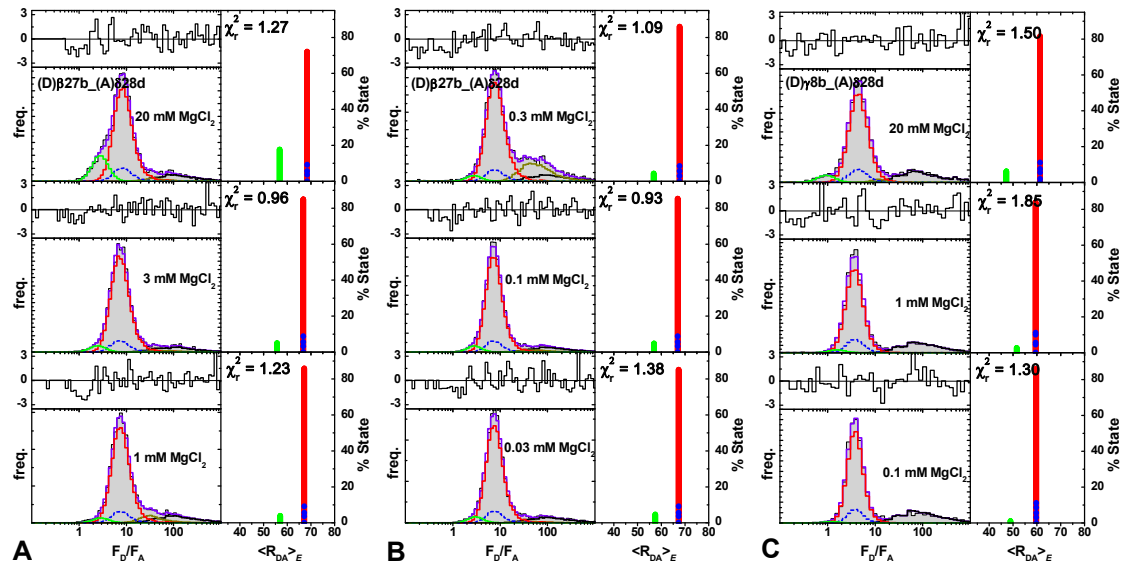


Figure 2.10. PDA of Mg-titrations for sample (D) β 27b_(A) δ 28d (A and B) and (D) γ 8b_(A) δ 28d (C) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue dashed line for $(ab)_a$), a D-only state (black solid line) and a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed for overlapped minor states) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

2.2.2.1 Labeling on helix pairs *ad* and *bc*

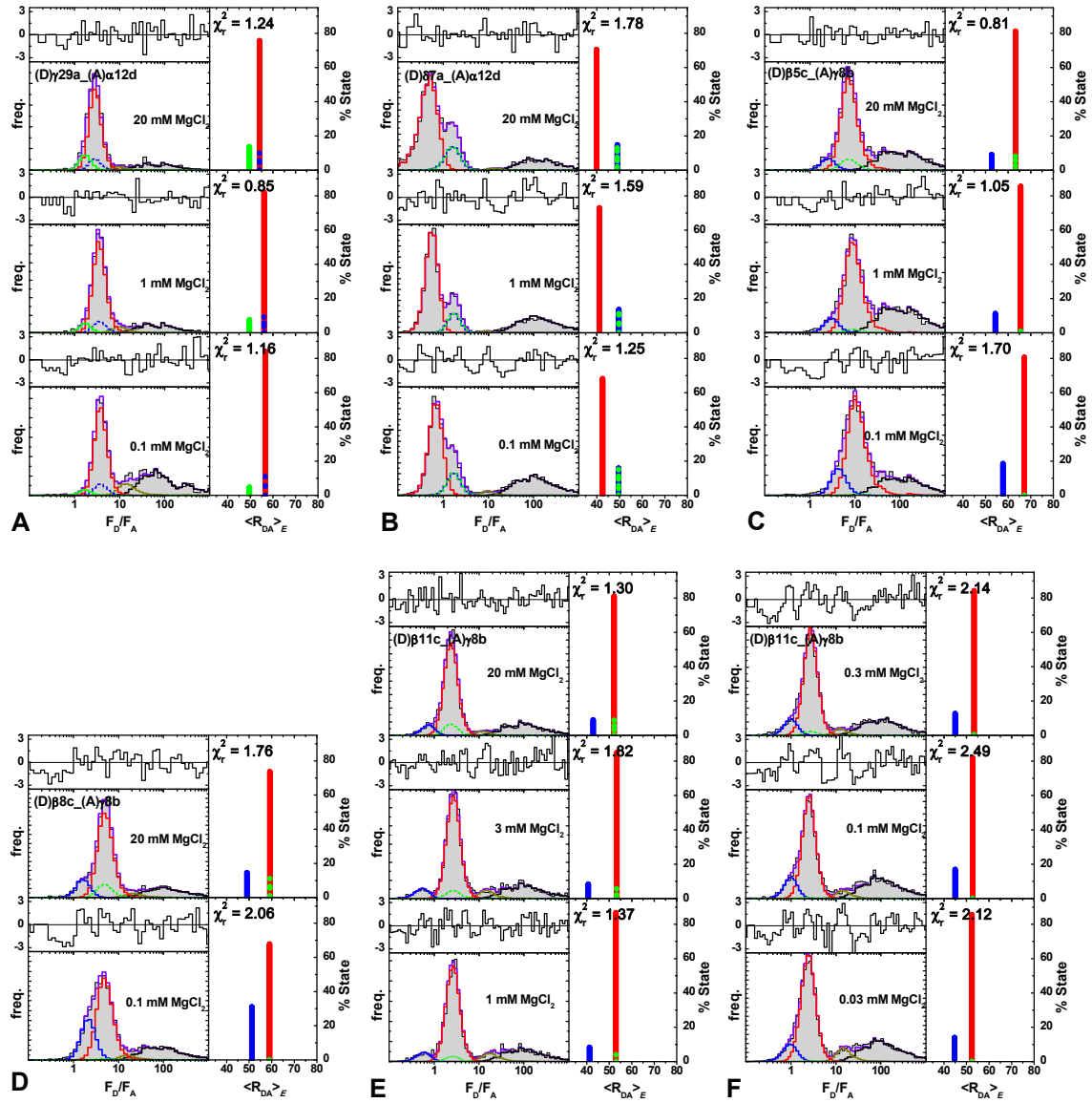
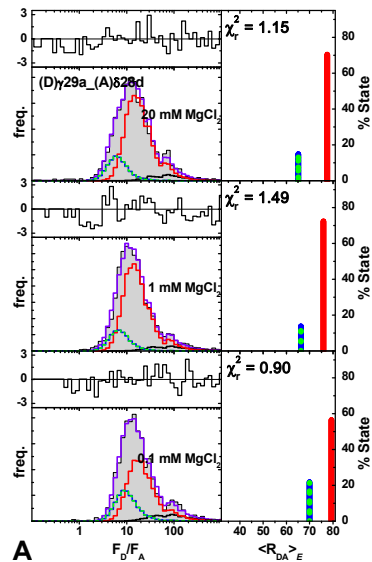


Figure 2.11. PDAs of Mg-titrations for samples (D) γ 29a_(A) α 12d (A), (D) δ 7a_(A) α 12d (B), (D) β 5c_(A) γ 8b (C), (D) β 8c_(A) γ 8b (D) and (D) β 11c_(A) γ 8b (E and F) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_{ar}$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_{ar}$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.



A Figure 2.12. PDAs of Mg-titrations for sample *(D)γ29a_(A)δ28d* (A) (selected bursts). F_D/F_A histograms of experimental data (gray area) are fitted (violet solid line) with three FRET states ($\langle R_{DA} \rangle_{E(1)}$, red solid line for $(ad)_a$, $\langle R_{DA} \rangle_{E(2)}$, green solid line (dashed if overlapped) for $(ad)_p$ and $\langle R_{DA} \rangle_{E(3)}$, blue solid line (dashed if overlapped) for $(ab)_a$), a D-only state (black solid line) and, if necessary, a state accounting for impurities (dark yellow solid line). Weighted residuals are shown in the upper plot. The right panel shows the distances and relative amplitudes (solid lines, dashed if overlapped) of the three FRET states (green for $(ad)_a$, red for $(ad)_p$ and blue for $(ab)_a$). See Table 4-5 in the SI of “RNA4WJ manuscript” for all parameters used.

3 eTCSPC fluorescence decay measurements of single labeled RNA4WJs

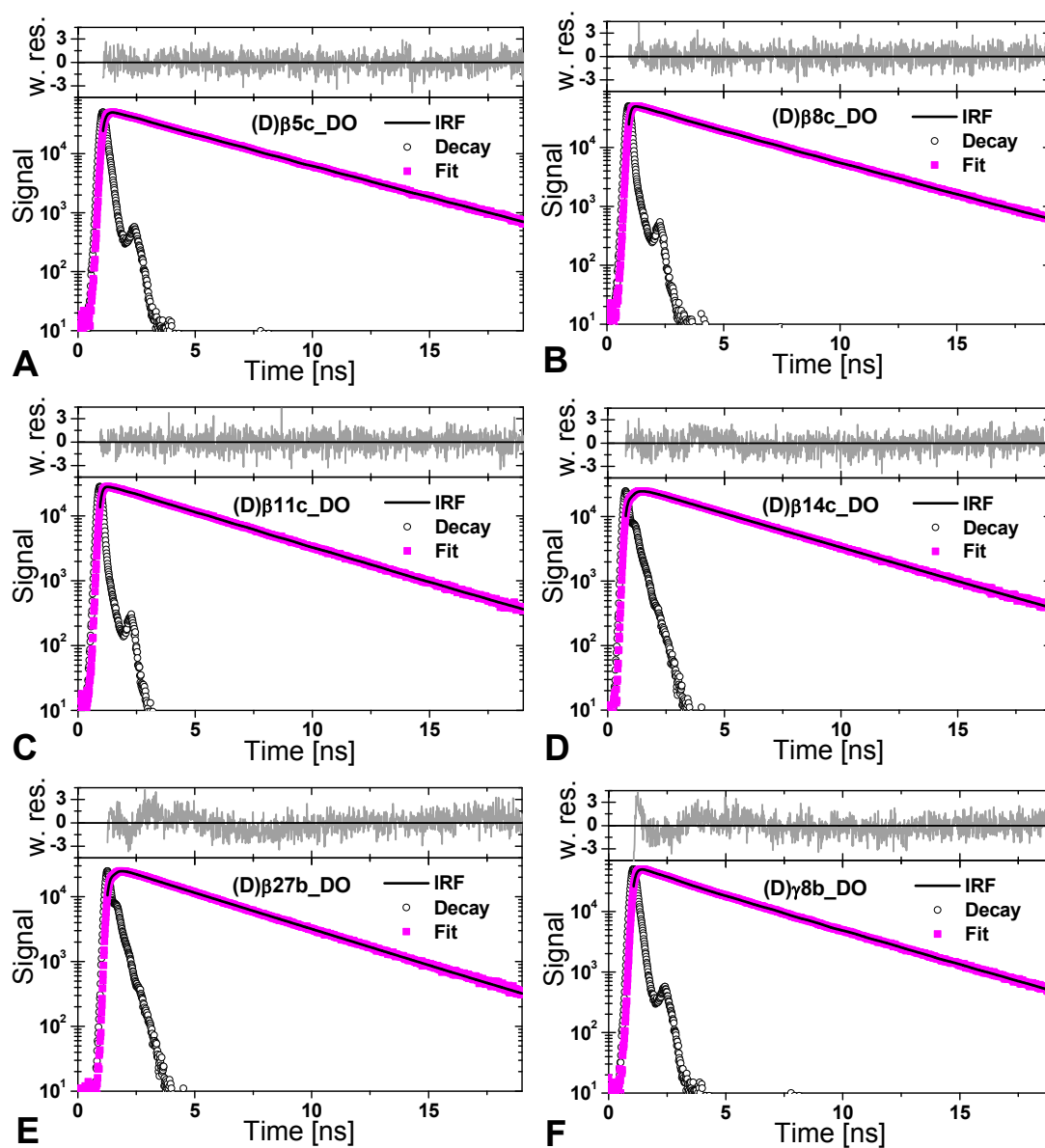


Figure 3.1. eTCSPC measurements of fluorescence decay for RNA4WJs single-labeled with Alexa488. Labeling at positions (D) β 5c (A), (D) β 8c (B), (D) β 11c (C), (D) β 14c (D), (D) β 27b (E) and (D) γ 8b (F). Experimental data (purple filled squares), instrument response function (IRF, black open circles) and fits to the data (black solid lines) are shown. Weighted residuals are presented above each plot (gray solid lines). See Table 4-1 in the SI of “RNA4WJ manuscript” for all fit parameters.

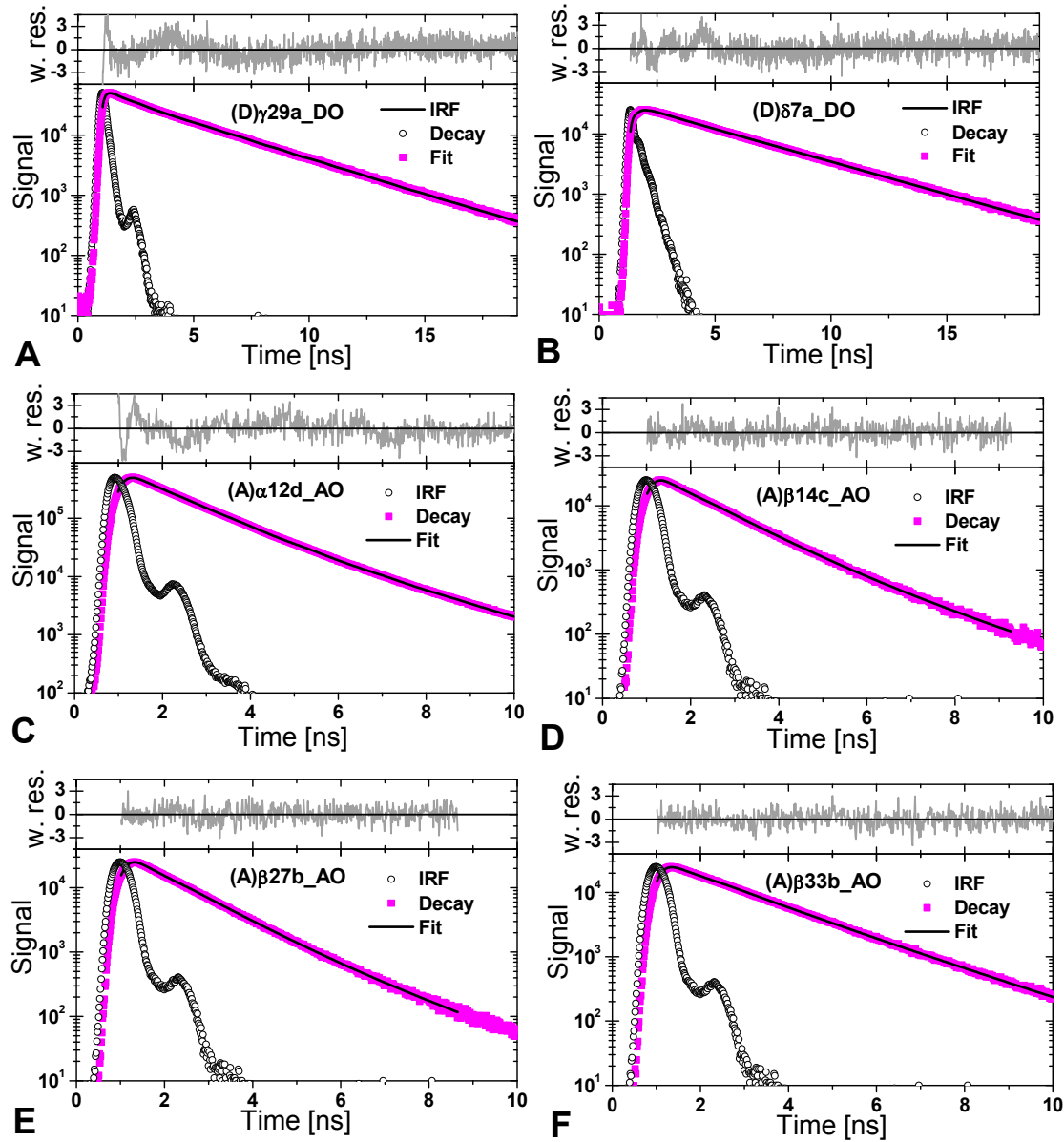


Figure 3.2. eTCSPC measurements of fluorescence decay for RNA4WJs single-labeled with Alexa488 (A and B) and Cy5 (C - F). Labeling at positions (D) γ 29a (A), (D) δ 7a (B), (A) α 12d (C), (A) β 14c (D), (A) β 27b (E) and (A) β 33b (F). Experimental data (purple filled squares), instrument response function (IRF, black open circles) and fits to the data (black solid lines) are shown. Weighted residuals are presented above each plot (gray solid lines). See Table 4-1 in the SI of “RNA4WJ manuscript” for all fit parameters.

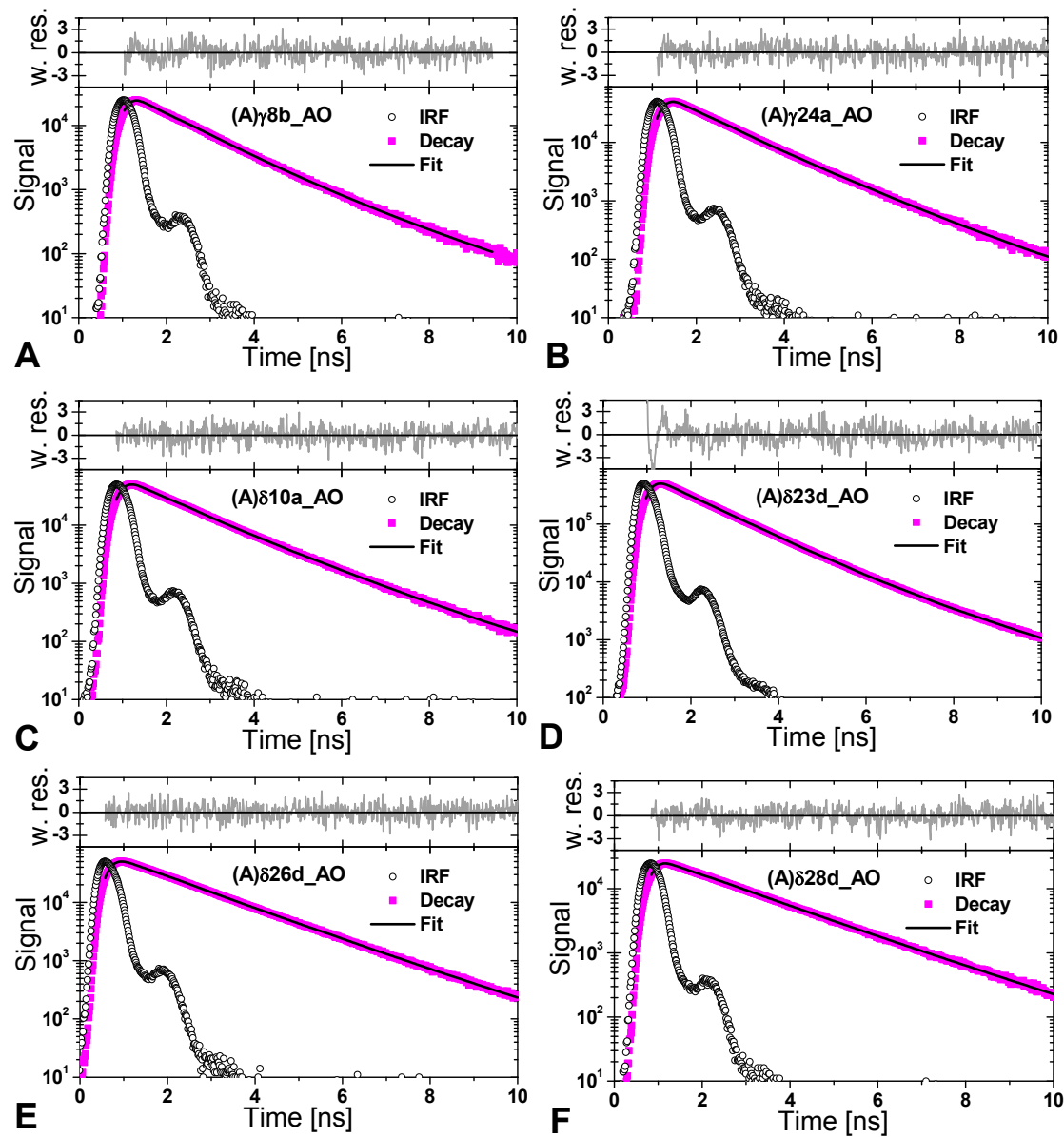


Figure 3.3. eTCSPC measurements of fluorescence decay for RNA4WJs single-labeled with Cy5. Labeling at positions (A) $\gamma 8b$ (A), (A) $\gamma 24a$ (B), (A) $\delta 10a$ (C), (A) $\delta 23d$ (D), (A) $\delta 26d$ (E) and (A) $\delta 28d$ (F). Experimental data (purple filled squares), instrument response function (IRF, black open circles) and fits to the data (black solid lines) are shown. Weighted residuals are presented above each plot (gray solid lines). See Table 4-1 in the SI of “RNA4WJ manuscript” for all fit parameters.

Erklärung

Hiermit erkläre ich, dass ich die Promotion mit dem Thema

“FRET restrained high-precision structural modeling of biomolecules”

am Institut für Physikalische Chemie II der Heinrich-Heine- Universität Düsseldorf unter der Leitung von Prof. Dr. Claus A. M. Seidel eigenständig und ohne unerlaubte Hilfe angefertigt und in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht habe.

Es existieren keine vorherigen Promotionsversuche.

Düsseldorf, den 02.05.2012

