

---

# **Spectroscopic Tools for Quantitative Studies of DNA Structure and Dynamics**

---

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Søren Preus

The PhD School of Science

Academic advisors:

Kristine Kilså<sup>1</sup>, L. Marcus Wilhelmsson<sup>2</sup> & Bo Albinsson<sup>2</sup>

<sup>1</sup>Department of Chemistry  
Faculty of Science  
University of Copenhagen  
Denmark

<sup>2</sup>Department of Chemical  
and Biological Engineering  
Chalmers University of Technology  
Sweden

Submitted: 7/12/2012

## Assessment committee:

**Assoc. Prof. Tom Vosch**

Department of Chemistry, University of Copenhagen, Denmark

**Prof. Lennart B.-Å. Johansson**

Chemistry Department, Umeå University, Sweden

**Researcher Reiner Lomoth**

Department of Chemistry, Ångström Laboratory, Uppsala University,  
Sweden

Copyright © 2013 by Søren Preus (spreus@gmail.com)

ISBN 978-87-91963-23-0

---

## **Spectroscopic Tools for Quantitative Studies of DNA Structure and Dynamics**

### **Short abstract:**

The main objective of this thesis is to develop quantitative fluorescence-based, spectroscopic tools for probing the 3D structure and dynamics of DNA and RNA. The thesis is founded on six peer-reviewed papers covering mainly the development, characterization and use of fluorescent nucleobase analogues. In addition, four software packages is presented for the simulation and quantitative analysis of time-resolved and steady-state UV-Vis absorption and fluorescence experiments.

**Keywords:** DNA, FRET, Fluorescence, DNA base analogues

---



# LIST OF PAPERS

---

- I Preus, S. and Wilhelmsson, L. M.  
*Advances in Quantitative FRET-Based Methods for Studying Nucleic Acids*  
ChemBioChem (**2012**), 13, 1990-2001
- II Preus, S., Börjesson, K., Kilså, K., Albinsson, B., and Wilhelmsson, L. M.  
*Characterization of Nucleobase Analogue FRET Acceptor tC<sub>nitro</sub>*  
J. Phys. Chem. B (**2010**), 114, 1050-1056
- III Preus, S., Kilså, K., Miannay, F. A., Albinsson, B., and Wilhelmsson, L. M.  
*FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids*  
Nucleic Acids Res. (**2012**), doi: 10.1093/nar/gks856
- IV Burns, J. R., Preus, S., Singleton, D. G., and Stulz, E.  
*A DNA based five-state switch with programmed reversibility*  
Chem. Commun. (**2012**), 48, 11088-11090
- V Preus, S., Kilså, K., Wilhelmsson, L. M., and Albinsson, B.  
*Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family* Phys. Chem. Chem. Phys. (**2010**), 12, 8881-8892
- VI Dierckx, A., Miannay, F. A., Gaied, N. B., Preus, S., Björck, M., Brown, T., and Wilhelmsson, L. M.  
*Quadracyclic Adenine: A Non-Perturbing Fluorescent Adenine Analogue*  
Chem. Eur. J. (**2012**), 18, 5987-5997

## CONTRIBUTION REPORT

Papers I, II, III and V are mainly the work of the author. Paper IV is the main work of J. Burns and E. Stulz where the author contributed with simulations, data analysis and discussion of results. Paper VI is the main work of A. Dierckx and F. A. Miannay where the author contributed with design, calculations and discussion of results.

Papers not included in the thesis:

- Preus, S., Jönck, S., Pittelkow, M., Monhaphol, T., Albinsson, B., and Wilhelmsson, L. M.  
*The photoproduct of fluorescent nucleobase analogue tC in the presence of O<sub>2</sub>*  
In preparation. (2012)
- Preus, S., Börjesson, K., McPhee, S.A., Lilley, D.M.J., and Wilhelmsson, L.M.  
*tC<sup>O</sup>: a highly fluorescent RNA base analogue*  
In preparation. (2012)

Papers published prior PhD:

- Börjesson, K., Preus, S., El-Sagheer, A. H., Brown, T., Albinsson, B., and Wilhelmsson, L. M.  
*Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems*  
J. Am. Chem. Soc. (2009), 131, 4288-4293
- Shanks, D., Preus, S., Qvortrup, K., Hassenkam, T., Nielsen, M. B., and Kilså, K.  
*Excitation energy transfer in novel acetylenic perylene diimide scaffolds*  
New J. Chem. (2009), 33, 507-516

Software published during PhD ([www.fluortools.com](http://www.fluortools.com)):

- *FRETmatrix*: Simulation and analysis of FRET (published with Paper III)
- *a|e*: UV-Vis absorption and emission spectra analysis software
- *FluorFit*: Analysis of time-resolved fluorescence decays
- *AniFit*: Analysis of time-resolved fluorescence anisotropy

# CONTENTS

---

<b>I Description of Thesis</b>	<b>1</b>
<b>1 Background and Motivation</b>	<b>3</b>
1.1 Nucleic Acids . . . . .	4
1.2 Fluorescence and Nucleic Acids . . . . .	5
1.2.1 Biocompatible Fluorescent Dyes . . . . .	5
1.2.2 Fluorescent Nucleobase Analogues . . . . .	6
1.2.3 FRET and Nucleic Acids . . . . .	8
<b>2 Concepts and Methods</b>	<b>13</b>
2.1 Photophysics of Organic Dyes . . . . .	14
2.2 FRET - Förster Resonance Energy Transfer . . . . .	16
2.3 Computational Chemistry . . . . .	18
2.3.1 Density Functional Theory. . . . .	19
2.4 Optical Spectroscopy . . . . .	21
2.4.1 Steady-State Spectroscopy . . . . .	21
2.4.2 Polarized Spectroscopy . . . . .	22
2.4.3 Time-Resolved Emission Spectroscopy . . . . .	23
2.5 Quantitative Data Analysis . . . . .	24
2.6 Making User-Interfaced Software. . . . .	25
2.7 Experimental Details. . . . .	26
<b>3 Overview of Published Work</b>	<b>27</b>
3.1 Paper I . . . . .	28
3.2 Paper II . . . . .	30
3.3 Paper III . . . . .	32
3.4 Paper IV . . . . .	34
3.5 Paper V . . . . .	36
3.6 Paper VI . . . . .	38
3.7 Published Software . . . . .	41
3.7.1 FRETmatrix . . . . .	41
3.7.2 a e . . . . .	43
3.7.3 FluorFit . . . . .	43
3.7.4 AniFit . . . . .	46
<b>4 Conclusions and Outlook</b>	<b>49</b>
4.1 Summary . . . . .	50

4.2 Outlook . . . . .	51
4.2.1 Novel Fluorescent Base Analogues . . . . .	51
4.2.2 Quantitative FRET . . . . .	52
<b>5 Acknowledgements</b>	<b>55</b>
<b>Bibliography</b>	<b>57</b>
<b>A Appendix</b>	<b>67</b>
<b>II Appended Papers</b>	<b>89</b>

---

## **Resumé:**

The goal of this thesis is to develop new quantitative fluorescence-based tools for studying the three-dimensional structure and dynamics of DNA and RNA. The work may be divided into three interconnected sub-topics: 1) Development, 2) characterization, and 3) use of synthetic fluorescent DNA modifications. In addition, a continuous goal was to develop general methodologies for the quantitative evaluation of fluorescence-based experiments, including the development of user-interfaced software.

A central theme throughout the thesis is Förster resonance energy transfer (FRET), an energy transfer phenomenon which is widely used as a "molecular ruler" for monitoring distances and interactions at the nanoscale level. However, measuring quantitative nanoscale distances using FRET is highly challenging. This research field is reviewed in Paper I. In the pursuit for an improved quantitative FRET toolbox we developed "base-base FRET": a FRET pair system consisting of two DNA base analogues. This technique facilitates a very high control of both the position and orientation of the FRET probes relative to the nucleic acid which allows more information to be obtained from the data. Paper II reports the characterization of base analogue FRET acceptor, tC<sub>nitro</sub>. The information gained from this study, such as the direction of the lowest energy electronic transition dipole moment, was vital in order to use base-base FRET quantitatively.

Paper III concerns the development of a new generic method called FRETmatrix for analysing FRET experiments in nucleic acids quantitatively. Paper III demonstrates how base-base FRET in combination with FRETmatrix can provide quantitative information about the three-dimensional structure and dynamics of nucleic acids. In relation to Paper III, Paper IV reports a reversible five-state DNA switch with readable fluorescence output. This paper additionally demonstrates how FRETmatrix can be used to model any type of FRET system in nucleic acids.

The development of new fluorescent DNA base analogues with improved brightness and photostability is a challenging field. Paper V reports new insight into the quenching processes of the tC base analogues as well as into their potential energy surfaces important for their adaptability into various constrained (biological) environments. Paper VI reports a new fluorescent adenine analogue, qA, and describes its photophysical and nucleobase mimicking properties.

---

---

## Resumé:

Formålet med denne afhandling er at udvikle nye kvantitative fluorescens-baserede værktøjer til at studere den tre-dimensionelle struktur og dynamik af DNA og RNA. Afhandlingen kan groft inddeltes i tre forbundne temaer: 1) Udvikling, 2) karakterisering, og 3) anvendelse af syntetiske fluorescerende DNA modifikationer. Derudover er det et løbende formål at udvikle mere generelle metoder til at evaluere fluorescens-baserede eksperimenter kvantitativt, hvilket inkluderer udvikling af brugerflade-styret software.

Et centralt tema gennem afhandlingen er Förster's resonans-energiorførsel (FRET), et energiorførselsfænomen der kan anvendes som en "molekylær lineal" til at overvåge afstande og interaktioner i nanoskala-størrelsesordenen. Det er imidlertid ret udfordrende at anvende FRET til at måle kvantitative afstande. Artikel I giver en gennemgang af det nyeste inden for dette forskningsfelt. I jagten efter en forbedret kvantitativ FRET-værktøjsboks udviklede vi "base-base FRET": et FRET-par system bestående af to DNA base analoger. Denne teknik faciliterer en øget kontrol over både positionen og orienteringen af FRET-proberne relativt til DNA molekylet, hvilket betyder at der kan opnås mere information fra eksperimenterne. Artikel II rapporterer karakteriseringen af base-analogen, tC<sub>nitro</sub>, med henblik på dens anvendelse som FRET probe. Informationen opnået under dette studie, såsom retningen af overgangsmomentet, var vital for at anvende base-base FRET kvantitativt.

Artikel III omhandler udviklingen af en ny generisk metode kaldet FRETmatrix til at analysere FRET eksperimenter i DNA kvantitativt. Artikel III demonstrerer hvordan base-base FRET i kombination med FRETmatrix kan give kvantitativ information omkring den tre-dimensionelle struktur og dynamik af DNA. I relation til artikel III rapporterer artikel IV en reversibel kontakt med fem stationære tilstande og et aflæseligt fluorescens-output. Denne artikel demonstrerer yderligere hvordan FRETmatrix kan anvendes til at modellere et hvilken som helst FRET system i DNA.

Udviklingen af nye fluorescerende DNA base analoger med forbedret lysstyrke og stabilitet er et udfordrende forskningsfelt. Artikel V rapporterer ny indsigt i quenching-processerne af tC base analogerne samt deres potential energi-overflader. Sidstnævnte spiller en rolle for hvordan disse prober tilpasser sig forskellige biologiske miljøer. Artikel VI rapporterer en ny fluorescerende adenin analog, qA, og beskriver dennes fotofysiske egenskaber samt dens evne til at efterligne adenin i DNA.

---

# PREFACE

---

This thesis was submitted to the Faculty of Science, University of Copenhagen, as a partial fulfillment of the requirements to obtain the PhD degree. The work presented was carried out in the years 2009-2012 partially in the laboratory of Prof. Kristine Kilså at the Department of Chemistry University of Copenhagen and partially at the Department of Chemical and Biological Engineering at Chalmers in the laboratories of Assoc. Prof. Marcus Wilhelmsson and Prof. Bo Albinsson. I've always enjoyed my time at Chalmers where I spent about six months combined during the PhD. I additionally spent three extremely experiencing months in the biochemical laboratory of Prof. Daniel Herschlag at Stanford University, CA, during the spring of 2010. Since no papers resulted from the collaboration with the Herschlag lab the work carried out while I was at Stanford and during the following months is not included in the thesis.

The thesis is a direct continuation of my Master's thesis titled "*Characterization and Use of Fluorescent Nucleobase Analogues*" which was carried out in part at Chalmers and defended in 2009. The paper resulting from my Master's work, published in JACS in 2009, is highly relevant and frequently referred to in this thesis and it has therefore been included as Appendix 1.

## THESIS OBJECTIVES

The objectives of the work reported in this thesis are primarily to 1) design new fluorescent probes based on synthetic DNA base analogues, 2) characterize their photophysical properties in DNA and 3) demonstrate how such probes can be exploited for studying the structure and dynamics of nucleic acids. In addition, a continuing goal is to improve and expand the quantitative Förster resonance energy transfer (FRET) toolbox in a more general sense, which includes the development of generic methodologies and software packages for the simulation and analysis of quantitative time-resolved fluorescence and FRET experiments.

## THESIS OUTLINE

This thesis is in the form of a synopsis with attached published papers. While Part I of the thesis provides background information and an overview of the attached papers Part II constitutes the papers.

In Part I, the first chapter introduces the rapidly progressing research field of fluorescence-based biophysical technologies and explains why this field is important particularly in the life sciences. The next chapter provides an introductory level foundation to the theoretical, experimental and computational concepts used throughout the thesis. Finally, a short explanatory summary of each published paper and software is provided in chapter 3. Part II presents the papers in the order of theme: First papers regarding FRET are presented (Paper I, II, III, and IV), followed by the characterization of existing DNA base analogues (Paper V, as well as paper II and VI), finalizing with the development of new fluorescent base analogues (Paper VI).

# Part I

DESCRIPTION OF THESIS



# 1

## BACKGROUND AND MOTIVATION

---

*The first chapter introduces fluorescence-based DNA technology and highlights the motivation of the research conducted in the thesis*

### Contents

---

1.1	Nucleic Acids . . . . .	4
1.2	Fluorescence and Nucleic Acids . . . . .	5
1.2.1	Biocompatible Fluorescent Dyes . . . . .	5
1.2.2	Fluorescent Nucleobase Analogues . . . . .	6
1.2.3	FRET and Nucleic Acids . . . . .	8

---

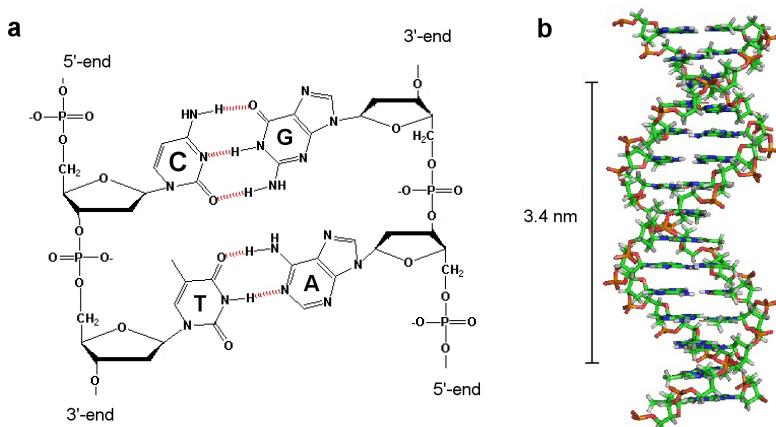
## 1.1 Nucleic Acids

---

The nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are central biological molecules in all living organisms.<sup>1</sup> Since the function of DNA and RNA in various biological environments are directly related to their three-dimensional structure and time-dependent structural changes (dynamics), insight into these properties is a vital requirement in order to understand genetic diseases, develop cures, exploit biological processes in various technologies, and simply to understand life itself.

The DNA molecule is a linear polymer built up from four different subunits called nucleotides (Figure 1.1a). Each nucleotide consists of a five-carbon, ring-structured sugar, a negatively charged phosphate group and one of the four natural nucleobases; adenine (A), thymine (T), cytosine (C) and guanine (G). The nucleobases are planar, aromatic, heterocyclic compounds capable of forming highly specific H-bonds to their complementary counterpart (A to T and C to G). Through base-pairing, two complementary strands of DNA may be hybridized resulting in a DNA duplex of two antiparallel strands and a net negative charge. DNA duplexes may adopt a number of different conformations, of which the most stable form at biological conditions is the famous B-DNA double helix (Figure 1.1b).<sup>2</sup> In B-DNA the base pairs are located in a hydrophobic core while the backbone is exposed on the outside of the helix where the negative charges are neutralized by counter ions in the solution. The B-DNA helix is right-handed with a helical twist of  $\sim 36^\circ$  per base pair, and an average of  $\sim 10.5$  base-pairs per turn corresponding to a distance of 3.4 nm. Alternative DNA conformations such as triplexes<sup>3,4</sup> and G-rich quadruplex structures<sup>5</sup> also exist. The role of these more complex structures in chromosomal DNA have recently received attention since they may serve as targets for therapeutic drugs.<sup>6-11</sup> The role of RNA enzymes (ribozymes) and switches (riboswitches) has also received attention due their role in *e.g.* gene control.<sup>12</sup>

The high specificity of Watson-Crick hydrogen bonding allows DNA molecules to be taken out of their biological context and exploited as building blocks in the construction of self-assembled nanoscale structures and functional devices (for recent reviews see *e.g.* <sup>13-19</sup>). The rigidity and chemical stability of the B-DNA helix makes this self-assembly approach highly attractive, in particular because it is relatively easy to design intermolecular nanostructures with predefined 2D- and 3D-shapes<sup>20-24</sup> and even a functionality such as mechanical movement,<sup>25-31</sup> catalytic activity<sup>32-34</sup> or drug transport.<sup>35,36</sup>



**Figure 1.1:** The DNA molecule. a) The chemical structure of DNA in which all four canonical nucleotides are shown in their base-pairing environment. b) The double-stranded B-DNA helix consisting of two antiparallel DNA strands hybridized as a result of Watson-Crick H-bonding between complementary bases in the two strands.

## 1.2 Fluorescence and Nucleic Acids

---

Over the years, fluorescence-based techniques have become invaluable tools in the molecular biosciences, and several new classes of fluorescent probes have recently been created to study biomolecules. The main advantage of fluorescence-based methods is that the natural nucleobases are virtually non-fluorescent<sup>37</sup> enabling an outstanding signal-to-noise ratio when introducing fluorescence into DNA.

### 1.2.1 Biocompatible Fluorescent Dyes

---

Fluorescent dyes may be introduced into DNA either non-covalently or covalently. Non-covalent DNA-binding dyes such as ethidium bromide bind to DNA by unspecific electrostatic interactions with the negatively charged backbone and/or by intercalation in between the hydrophobic bases.<sup>38,39</sup> In contrast, covalently attached fluorophores are usually tethered via a flexible linker to the DNA. This approach allows practically any fluorophore to be attached to the DNA at several different sites. The most popular probes are commercially

available dyes such as the Cy-dyes and the Alexa-dyes which offer an unsurpassable overall brightness. These dyes are available in a variety spanning the entire visible spectrum.<sup>40</sup> Alternative luminescent probes are e.g. lanthanides characterized by long excited state lifetimes<sup>41</sup> and inorganic semi-conductor quantum dots<sup>42,43</sup> characterized by their extremely high brightness.

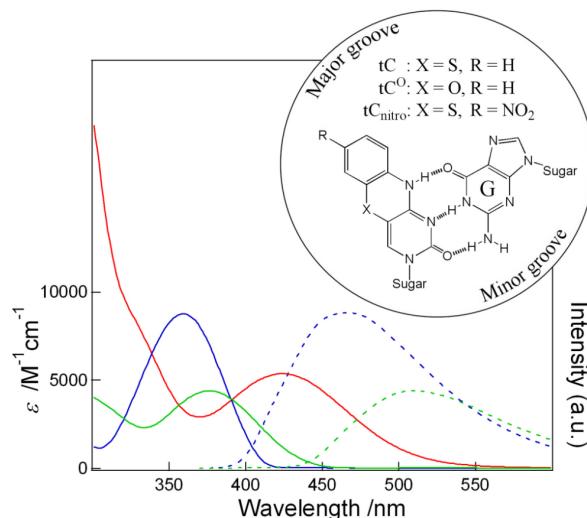
Using dyes attached externally to biomolecules has a number of disadvantages. Firstly, dyes often interact with their surrounding environment which may hamper the interpretation of experiments or change the fluorescence properties of the dye.<sup>44–48</sup> Secondly, since external dyes reside on the outside of the biomolecular surface there is an inherent limitation in the kind of information obtainable from quantitative fluorescence experiments since the probes respond and report indirectly on the structure and dynamics of the biomolecule.

### 1.2.2 Fluorescent Nucleobase Analogues

---

For studies involving nucleic acids, fluorescent nucleobase analogues constitute a different class of fluorophores that overcomes these problems (for reviews see<sup>49–56</sup>). These probes can be inserted into DNA as a replacement for one of the natural bases mimicking the properties of the substituted base, in most cases without significantly affecting the DNA structure and function. In addition, when using nucleobase analogues the reporter can be positioned close to or even in the very site of interest when investigating DNA. Traditionally, the most utilized fluorescent nucleobase analogue has been 2-aminopurine (2-AP), an isomer of adenine capable of forming a base-pair with thymine and a less stable base-pair with cytosine.<sup>57</sup> 2-AP is highly fluorescent in its free, monomeric form but is almost completely quenched in the base-stacking environment provided by dsDNA.<sup>58</sup> This property has been exploited in probing the local structure and dynamics of DNA<sup>59–61</sup> and in DNA-protein interactions that result in locally unwound strands.<sup>61–63</sup> Other reported fluorescent base analogues include, but is not limited to, the pteridines,<sup>64</sup> the expanded DNA bases,<sup>65</sup> the wide DNA bases,<sup>66</sup> the base-discriminating fluorescent base analogues<sup>53,67–69</sup> and the emissive RNA alphabet by Tor and co-workers.<sup>70</sup>

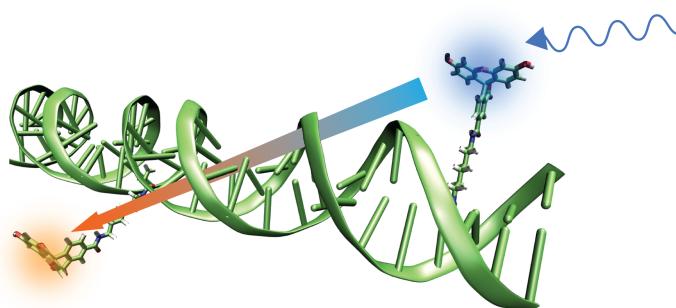
The strict environment set by the DNA double helix in terms of H-bonding, base-stacking and steric hindrances, however, often affects the fluorescence properties of the fluorophore. As a result, fluorescent nucleobase analogues lag behind commercially available dye molecules in terms of overall brightness and spectral properties. Developing new fluorescent base analogues thus remains somewhat challenging. Common for all the fluorescent nucleobase



**Figure 1.2:** UV-Vis absorption (full-drawn) and emission spectra (dashed) of the tricyclic cytosine analogues, tC (green), tC<sup>O</sup> (blue) and tC<sub>nitro</sub> (red). Insert: Chemical structures of the tC probes in the base-pairing environment with guanine.

analogues mentioned above are their sensitivity to the surrounding micro-environment, most often a significant quenching of fluorescence in dsDNA and a strong destabilizing effect of the DNA double helix. The highly relevant exception to these properties is the tricyclic cytosine (tC) family, 1,3-diaza-2-oxophenothiazine (tC), 1,3-diaza-2-oxophenoxazine (tC<sup>O</sup>) and 7-nitro-1,3-diaza-2-oxophenothiazine (tC<sub>nitro</sub>).<sup>49,71-78</sup> These base analogues are directly built upon the molecular framework of cytosine and extended by two extra rings (Figure 1.2). tC and tC<sup>O</sup> both possess high fluorescence quantum yields ( $\sim 0.2$ ) and single-exponential lifetimes ( $\sim 4\text{-}5$  ns) in double-stranded DNA, relatively insensitive of neighbouring bases.<sup>73,75</sup> In contrast, tC<sub>nitro</sub> is virtually non-fluorescent at room-temperature but possess a red-shifted lowest energy absorption band compared to tC and tC<sup>O</sup> which makes it perfectly suited as an energy transfer acceptor with tC or tC<sup>O</sup> serving as the donor (Figure 1.2).<sup>77,78</sup> All of the tC probes adapt the position and orientation of the substituted cytosine base and stabilize the B-DNA double helix compared to regular cytosine.<sup>75-77</sup>

The tC family has a central role in this thesis. In Paper II and V they are subjected to a thorough photophysical characterization, while Paper III demonstrates their use in probing DNA structure and dynamics.



**Figure 1.3:** Illustration of FRET between two dyes tethered to DNA. After excitation of the donor this dye transfers its excitation energy to the acceptor located in close proximity. The FRET efficiency is highly distance dependent which makes the technique suited as a "molecular ruler" for probing nanoscale distances and molecular interactions.

### 1.2.3 FRET and Nucleic Acids

---

FRET is an energy transfer phenomenon that is widely used as a molecular ruler in the biosciences for probing nanoscale distances and molecular interactions (Figure 1.3).<sup>79</sup> The technique is usually performed by covalently labelling one or more biomolecules with two dyes: an energy donor and an acceptor. If the two probes are located within a distance related to the critical Förster distance of the pair ( $<10$  nm) energy transfer can be induced between the two dyes which is observed as a decreased donor fluorescence yield, decreased donor excited state lifetime and, if the acceptor is emissive, an increased acceptor fluorescence yield. FRET may thus either serve as an "on/off" sensor of the relative location of the two dyes or as a quantitative ruler for determining the distance in between the FRET-pair. One of the aims of this thesis is to demonstrate that much more quantitative information, than a mere relative distance, can be gained from FRET experiments.

#### 1.2.3.1 Historical View

---

FRET was first described by Förster in the 1940's as an interaction between two dipoles separated by a fixed distance,  $r$ , and oscillating in resonance.<sup>80</sup> By multiplying the probability that the two dipoles have the same resonance frequency with the rate of energy transfer when they do have the same resonance frequency, Förster derived the characteristic  $r^{-6}$  dependency of the energy

transfer rate constant (section 2.2).<sup>81</sup> The FRET theory was continued to be discussed up to and during the 1970's,<sup>82–84</sup> however, it was not until late 1980's and early 1990's that FRET was exploited as a quantitative ruler to probe distances within nucleic acid structures.<sup>85,86</sup> In these studies, Lilley and co-workers determined the stereochemical arrangement of the four helices making up the Holliday junction, a four-stranded, four-way DNA junction which is an intermediate in genetic recombination. This late entrance of quantitative FRET in DNA was mainly due to difficulties in labeling nucleic acids site-specifically at that time. However, the DNA fluorescence toolbox has advanced greatly since then and custom dye-labelled oligonucleotides are now ordered and delivered on a day-to-day basis.<sup>40,87–90</sup>

---

### 1.2.3.2 FRET in Modern Applications

---

Today, FRET is used routinely as a technique to probe the structural states of nucleic acids. Particular advancement has occurred in recent years in the specific case where quantitative information about FRET-pair distances and nucleic acid three-dimensional structures is wanted from FRET experiments. The combination of multiple FRET-pair positions has provided quantitative insight into the three-dimensional structure of nucleic acids.<sup>91–98</sup> Increasingly sophisticated techniques have been developed to model the uncertainty in the position and orientation of the dyes relative to the DNA.<sup>91,99,100</sup> In addition to these improvements, the properties of many popular commercial FRET dyes attached to DNA are now known to some extent, such as the Cy-dyes,<sup>44,46,47,101–103</sup> FAM<sup>104–107</sup> and TAMRA.<sup>105–109</sup> This knowledge is not only insightful but in fact vital in order to use FRET quantitatively on nucleic acid containing systems.

An exciting development within FRET is the ability to monitor one molecule at a time, called single-molecule FRET (smFRET).<sup>110–114</sup> Single-molecule FRET techniques provide real-time insight into biomolecular structures and dynamics without hiding molecular heterogeneity behind an ensemble average. However, interpreting smFRET experiments into quantitative distances and dynamics is extremely difficult due to the low signal-to-noise ratio in such experiments, and methods are still being developed for analysing data from smFRET experiments quantitatively.<sup>115–122</sup> Multiparameter single-molecule fluorescence techniques, in which all possible ascertainable fluorescence parameters are monitored at the single-molecule level (intensity, anisotropy and lifetime), are extremely interesting in the context of quantitative smFRET.<sup>123–129</sup> Because all information required to interpret the measured signal into quantitative distances are acquired, such state-of-the art single-molecule techniques offer

vast potential for providing detailed insight into the structure and dynamics of large biomolecules. Recently, single-molecule FRET has been used to probe the three-dimensional geometry of three-way DNA junctions,<sup>93,94</sup> RNA polymerase II complexes<sup>96,97,130</sup> and a helicase-DNA complex<sup>98</sup> to name but a few.

More references and information on recent advances within quantitative FRET is found in Paper I of this thesis.

**Base-base FRET.** In a previous study of ours (Appendix 1), a fluorescent nucleobase analogue FRET pair system was developed as an alternative to traditional, external FRET probes (Figure 1.4).<sup>77</sup> Since the base probes adapt the position and orientation of the canonical bases within double-stranded DNA this technique provides a number of advantages compared to traditional probes: First of all, the base probes can be positioned inside the very site of interest mimicking the behaviour of the substituted base and, thus, provides a means to probe the local base orientation and position at specific sites. Secondly, because both the orientation and position of the probes are highly constrained at the timescale of energy transfer this allows a higher degree of control of the orientation factor in the energy transfer process, and thus more detailed studies to be performed of three-dimensional nucleic acid structures without complications associated with linker flexibility.

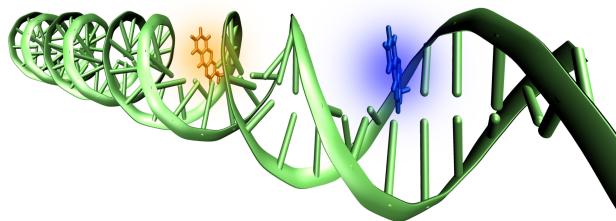
This latter point is demonstrated by the FRET efficiencies measured between the base analogue FRET-pairs positioned in B-DNA with distances varying from 2-13 base-pairs in between the FRET-pair (Figure 1.5).<sup>77</sup> As the number of bases in between the pair increases both the distance as well as the relative orientation between the donor and acceptor gradually changes in a predictable manner. The result is a FRET curve decreasing with distance and, additionally, oscillating with a periodicity in phase with the helical periodicity of the B-DNA helix due to the change in orientation between the two transition dipole moments of the probes. The oscillation of the measured energy transfer efficiency in Figure 1.5 demonstrates the well-defined orientation of the base probes inside the DNA helix. The main disadvantage of base-base FRET is the complexity involved in the quantitative analysis of data from such experiments. A solution to this limitation is provided in this thesis.

### 1.2.3.3

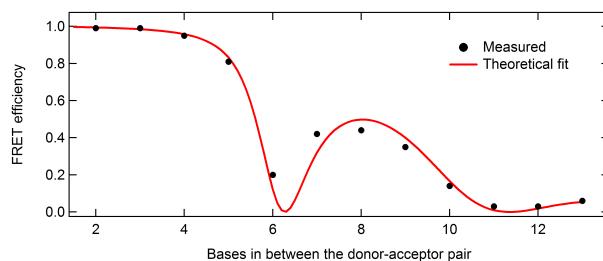
#### Advantages and Drawbacks

---

Compared to higher resolution structural techniques like NMR spectroscopy<sup>131</sup> and X-ray crystallography,<sup>132</sup> FRET has the advantage of not being limited by the size of the system under investigation (unlike NMR), is solution-based



**Figure 1.4:** Illustration of two base analogues positioned in B-DNA providing the foundation for base-base FRET.



**Figure 1.5:** The FRET efficiencies measured between two base analogues positioned inside double-stranded B-DNA with varying inter-pair distance.<sup>77</sup> The red line is a theoretical fit based on the assumption of static transition dipoles positioned inside a perfect B-form helix with a twist of  $36^\circ$  and rise of  $3.4 \text{ \AA}$  per dinucleotide step.

(unlike crystallography) and, in addition, FRET is generally cheaper, faster, simpler and more sensitive. The drawbacks of FRET, however, are the dependency of the energy transfer on 1) probe orientation, which is very hard to control, and 2) the fluorescence quantum yield of the donor, which strongly depends on the interaction between the probe and its micro-environment and thus vary from sample to sample. While the second drawback can be accounted for using proper reference measurements (i.e. by using reference samples labelled with the donor only) the first drawback is more difficult to account for. Lower and upper boundaries of  $\kappa^2$  can be estimated based on fluorescence anisotropy measurements,<sup>83,84</sup> however, the exact  $\kappa^2$ -value or  $\kappa^2$ -distribution is practically impossible to obtain. This problem leads to an uncertainty in the distance deduced from FRET measurements.

In this thesis, the novel FRET technique base-base FRET is developed and refined. It is demonstrated how the method facilitates an unprecedentedly high control of the position and orientation of the FRET probes and thus leads to more accurate and versatile quantitative FRET-based experiments.



# 2

## CONCEPTS AND METHODS

---

*This chapter introduces the theoretical, experimental and computational concepts used throughout the thesis*

### Contents

---

<b>2.1</b>	<b>Photophysics of Organic Dyes . . . . .</b>	<b>14</b>
<b>2.2</b>	<b>FRET - Förster Resonance Energy Transfer . . . . .</b>	<b>16</b>
<b>2.3</b>	<b>Computational Chemistry . . . . .</b>	<b>18</b>
2.3.1	Density Functional Theory. . . . .	19
<b>2.4</b>	<b>Optical Spectroscopy . . . . .</b>	<b>21</b>
2.4.1	Steady-State Spectroscopy . . . . .	21
2.4.2	Polarized Spectroscopy . . . . .	22
2.4.3	Time-Resolved Emission Spectroscopy . . . . .	23
<b>2.5</b>	<b>Quantitative Data Analysis . . . . .</b>	<b>24</b>
<b>2.6</b>	<b>Making User-Interfaced Software. . . . .</b>	<b>25</b>
<b>2.7</b>	<b>Experimental Details. . . . .</b>	<b>26</b>

---

## 2.1 Photophysics of Organic Dyes

---

The theory of photophysics describes why dyes are fluorescent and what determines their color and their brightness. This section provides a short explanation of the most relevant terms.

**Electronically excited states.** For the sake of simplicity we shall not go into details about the description of electronically excited states. For now it is sufficient to describe an electronic state as the distribution of electrons within the molecular orbitals of the molecule. Another important approximation (which is made in all modern molecular electronic structure calculations) is the Born-Oppenheimer (BO) approximation in which the nuclear and electronic part of the wavefunction is separated.<sup>133</sup> In the BO approximation the electrons are pictured as moving in a constant potential field from all the nuclei and the positions of the electrons are immediately adjusted when a small perturbation is applied to the nuclear coordinates. The BO approximation can successfully be applied to molecules because the electron movement is much faster compared to a nucleus carrying the same momentum due to the high ratio between the masses of the nuclei and electrons.

**Electronic transitions** Since light, or electromagnetic radiation in general, may be characterized as an electromagnetic wave of oscillating magnetic and electric dipoles, it is very intuitive that light interacts with the negatively charged electrons of a molecule. The perturbation caused by the electromagnetic wave may lead to a coupling between the initial ground state and electronic states of higher energies. In order for a transition to occur from the ground state,  $S_0$ , to an excited state,  $S_n$ , the energy difference between the two states must be equal to the energy of the incoming photon:

$$\Delta E = E_{S_n} - E_{S_0} = h\nu = h\frac{c_0}{\lambda} \quad (2.1)$$

where  $h$  is Planck's constant,  $c_0$  is the speed of light and  $\lambda$  is the wavelength of the electromagnetic wave. The energy,  $\Delta E$ , of a given transition is theoretically a  $\delta$ -function at 0 K for molecules in the gas phase, however, spectral broadening is always observed for molecules in solution due to solvent-chromophore interactions, often combined with excitations into more than one vibrational state.

**The transition dipole moment.** The transition dipole moment,  $\vec{\mu}_{n0}$ , is a vector depicting the electric dipole moment associated with the transition between two states. The direction of  $\vec{\mu}_{n0}$  gives the polarization of the transition, which determines how the system will interact with a polarized electromagnetic wave, while the square of the magnitude gives the strength of the interaction.

**The oscillator strength.** Another measure of the probability of a transition is the oscillator strength,  $f_{n0}$ , which can be obtained experimentally directly from the integral of the absorption band:

$$f_{n0} = \frac{4\pi m_e}{3\hbar e^2} \tilde{\nu}_{n0} |\mu_{n0}|^2 = \frac{4.3 \cdot 10^{-9}}{n} \int \varepsilon(\tilde{\nu}) d\tilde{\nu} \quad (2.2)$$

where  $m_e$  is the electron mass,  $\hbar$  is the reduced Planck constant,  $\tilde{\nu}$  is the frequency in  $\text{cm}^{-1}$  (wavenumbers),  $n$  is the refractive index and  $\varepsilon(\tilde{\nu})$  is the molar absorptivity describing the frequency dependence of the absorption spectrum. The magnitude of the oscillator strength is roughly speaking  $f_{n0} < 0.05$  for weakly allowed transitions, such as  $(n, \pi^*)$ , and  $f_{n0} > 0.05$  for strongly allowed transitions, such as  $(\pi, \pi^*)$ . The rate of fluorescence and the magnitude of the oscillator strength of the lowest energy electronic transition are directly proportional.

**Radiative vs. non-radiative decay.** After excitation the molecule normally decays to the lowest vibrational level of  $S_1$  where a number of decay processes can occur. If  $f_{n0}$  is high, the molecule may decay to  $S_0$  via fluorescence. The fluorescence process, however, competes with non-radiative decay processes from  $S_1$  such as internal conversion (IC), intersystem crossing (ISC), or a bimolecular quenching process such as FRET. Each process is associated with a rate constant,  $k_i$ . The process with the largest value of  $k_i$  dominates the decay. The competition between the individual decay processes from  $S_1$  is described quantitatively by the quantum yield,  $\Phi_r$ , of process  $r$ . The quantum yield is defined as

$$\Phi_r = \frac{k_r}{\sum k_i} = k_r \cdot \tau, \quad (2.3)$$

where  $\tau$  is the excited state lifetime and the sum is over all decay processes from  $S_1$ .

**Excited state lifetime.** The lifetime of a dye is the average time the molecule spends in the excited state before emitting a photon and is given by:

$$\tau = \frac{1}{\sum k_i} \quad (2.4)$$

Excited molecules typically decay by first order kinetics resulting in an exponential intensity decay of the fluorescence signal following excitation:

$$I(t) = \sum_j \alpha_j \cdot e^{-\frac{t}{\tau_j}} \quad (2.5)$$

where  $\alpha_j$  is the pre-exponential factor and represents the fraction of fluorophores with lifetime  $\tau_j$  (*i.e.*  $\sum \alpha_j = 1$ ).

**Temperature-dependent excited state decay.** The non-radiative rate constants may be separated into a temperature dependent and a temperature independent term. If the temperature dependent rate constant is approximated as an Arrhenius expression, then  $\Phi_f$  becomes

$$\Phi_f = \frac{k_f}{k_f + k_{\text{ISC}} + k_{\text{IC}}} = \frac{k_f}{k_f + k_0 + B \cdot \exp(-\frac{E_a}{RT})} \quad (2.6)$$

where  $k_0$  is the temperature independent non-radiative decay rate constant,  $B$  is the frequency factor,  $E_a$  is the activation energy of the temperature dependent decay process,  $R$  is the gas constant and  $T$  is the temperature.

**Potential Energy Surfaces.** The total energy of a molecule is a function of its nuclear and electronic coordinates. This dependency can be represented by a multidimensional surface with the atomic coordinates as variables, which is called the potential energy surface. The individual electronic states of a molecule are each represented by a potential energy surface defined within the BO approximation described above.

## 2.2

## FRET - Förster Resonance Energy Transfer

---

**FRET theory.** The rate of dipole-dipole resonance energy transfer between a donor fluorophore and an acceptor chromophore separated by a fixed distance,  $r$ , was shown by Förster to be<sup>81</sup>

$$k_T = \frac{1}{\tau_D} \cdot \left( \frac{R_0}{r} \right)^6 \quad (2.7)$$

where  $\tau_D$  is the lifetime of D in absence of A,  $r$  is the distance separating D and A, and  $R_0$  is the critical distance – the distance at which the efficiency of

energy transfer is 50%. If the wavelength is in nm then the critical distance is in Ångström given by

$$R_0 = 0.211 \left[ \frac{\kappa^2 \Phi_D J(\lambda)}{n^4} \right]^{\frac{1}{6}} \quad (2.8)$$

where  $\kappa$  is an orientation factor between the donor and acceptor,  $n$  is the refractive index,  $\Phi_D$  is the fluorescence quantum yield of the donor in absence of the acceptor and  $J(\lambda)$  is the overlap integral between the normalized emission spectrum of D and the absorption spectrum of A in units of  $M^{-1} cm^{-1} nm^4$ .

The orientation factor is given by

$$\kappa = \vec{e}_1 \cdot \vec{e}_2 - 3(\vec{e}_1 \cdot \vec{e}_{12})(\vec{e}_{12} \cdot \vec{e}_2) \quad (2.9)$$

Here,  $\vec{e}_1$  and  $\vec{e}_2$  are the unit vectors of the donor and acceptor transition dipoles and  $\vec{e}_{12}$  the unit vector between their centres. The value of  $\kappa^2$  can range from 0 to 4 and it is thus highly important to have an accurate estimate of  $\kappa^2$  when calculating a distance based on a FRET experiment. The most used  $\kappa^2$  is 2/3 corresponding to freely rotating transition moments, however, this value is often misused due to lack of knowledge of the orientation of donor and acceptor chromophores. Detailed discussions of the serious problems associated with the value of  $\kappa^2$  have been provided by Dale and Eisinger.<sup>83,84</sup>

**Measuring FRET.** The degree of FRET is typically expressed as the quantum yield of the process, called the FRET efficiency:

$$E \equiv \frac{k_T}{k_T + \sum k_i} = \frac{R_0^6}{R_0^6 + r^6} \quad (2.10)$$

where the sum is over all *intra*-molecular decay processes from  $S_1$ . The second step comes from inserting  $\frac{1}{\tau_D}$  from equation (2.7) into the quantum yield equation. In ensemble measurements the FRET efficiency can be determined in a number of ways.<sup>134</sup> The most common are: 1) Using a reference sample of donor in absence of acceptor. Here the FRET efficiency is determined either from the ratio of the donor lifetime or fluorescence quantum yield in the presence (DA) and absence (D) of acceptor:

$$E = 1 - \frac{\tau_{DA}}{\tau_D} \quad (2.11)$$

or

$$E = 1 - \frac{\Phi_{DA}}{\Phi_D} \quad (2.12)$$

In the simple setup where the concentration of D is the same in the DA and D samples  $\Phi$  may be replaced by the (integrated) donor intensity. 2) If the acceptor is emissive, an internal reference is obtained by exciting at a wavelength where only the acceptor absorbs and then determine  $E$  from the increase in acceptor emission in the presence of FRET. In this case, one emission spectrum is measured at a donor excitation wavelength (1) and a second, reference-spectrum, at an acceptor-only excitation wavelength (2).  $E$  may then be determined using the following expression for  $I_1$ :

$$\begin{aligned} I_1(\lambda) &= I_D(\lambda) + I_{A,\text{direct}}(\lambda) + I_{A,\text{FRET}}(\lambda) \\ &= C \times I_{D,\text{ref}}(\lambda) + A_{A,1} \frac{I_2(\lambda)}{A_{A,2}} + E \times A_{D,1} \frac{I_2(\lambda)}{A_{A,2}} \\ &= C \times I_{D,\text{ref}}(\lambda) + \left( \frac{\varepsilon_{A,1}}{\varepsilon_{A,2}} + \frac{\varepsilon_{D,1}}{\varepsilon_{A,2}} E \right) I_2(\lambda) \end{aligned} \quad (2.13)$$

where  $I_1$  is the emission spectrum measured at excitation wavelength 1,  $I_D$  is the D intensity spectrum from excitation at wavelength 1,  $I_{A,\text{direct}}$  is the A emission spectrum resulting from direct excitation at wavelength 1,  $I_{A,\text{FRET}}$  is the A emission spectrum resulting from FRET following excitation of the donor at wavelength 1,  $C$  is a constant (fitting parameter),  $I_{D,\text{ref}}$  is the D monomeric spectrum,  $A_{A,1}$  is the A absorbance at excitation wavelength 1,  $I_2$  is the emission spectrum measured at excitation wavelength 2,  $A_{A,2}$  is the A absorbance at excitation wavelength 2,  $E$  is the FRET efficiency,  $A_{D,1}$  is the absorbance of D at excitation wavelength 1, and  $\varepsilon_x$  are the molar absorptivities of D and A at the respective wavelengths. The second step assumes that D and A are present in a 1:1 ratio.

## 2.3 Computational Chemistry

---

**Geometry Optimization.** In a geometry optimization the atomic coordinates,  $R$ , of a molecule are optimized to a minimum on the potential energy surface (PES),<sup>135</sup> *i.e.*:

$$\frac{dE}{dR} = 0 \quad (2.14)$$

The calculation of the energy as a function of atomic coordinates is what distinguishes the different methods from each other. Since the vibrational frequencies of the molecule are proportional to the second derivative of the potential energy with respect to coordinate, calculating the IR spectrum of an optimized molecular geometry provides a means to determine if the geometry

is in fact a minimum on the PES. For geometries located away from minimum,  $\frac{\partial^2 E}{\partial R^2}$  is negative and at least one of the calculated vibrational frequency is an imaginary number. In a transition state (TS) optimization the geometry is optimized to the nearest saddle point on the PES. A TS calculation thus results in a single imaginary vibrational frequency, corresponding to the maximum on the intrinsic reaction coordinate (IRC) that connects two local minima on the PES.<sup>135</sup>

**Solvation shell.** Solvation effects may be modelled using either explicit solvent molecules or an implicit solvent model such as the CPCM method used in this thesis.<sup>135–137</sup> The CPCM method treats the interaction between solute and solvent by a polarizable dielectric continuum surrounding the cavity shell formed by the solute. The solvation effect is modelled by means of apparent polarization charges distributed on the cavity surface, which are determined by imposing that the total electrostatic potential cancels out on the surface. The induced charges on the cavity surface, in turn, affect the electric charges of the solute. The CPCM model is especially successful in representing polar liquids compared to gas phase conditions.

### 2.3.1 Density Functional Theory.

---

Density functional theory (DFT) is a quantum mechanical method that includes electron correlation (electron-electron interactions) in the calculation of the ground state electronic structure of molecules and solids [135, chapter 6] (for a review aimed at researchers from other fields see<sup>138</sup>). DFT is conceptually different compared to wave function based methods as it replaces the electronic wave function with the total electron density,  $\rho(\vec{r})$ , as the fundamental quantity from which all other observables can be extracted. Since  $\rho(\vec{r})$  itself is a function of only 3 variables (the three spatial coordinates) DFT calculations are in principle considerably faster compared to the wave function based electron correlation methods that depend on the coordinates of all  $N$  electrons of the system. In addition, DFT calculations often provide very accurate results and it is currently the most reliable, and hence the most preferred, method to calculate the ground state electronic structure of large many-body systems ( $>100$  electrons).

**The Kohn-Sham scheme.** DFT is usually, and in this thesis, performed using the Kohn-Sham (KS) approach. In the KS scheme the electron density is calculated using single-particle wave functions called KS orbitals, being directly

analogues to molecular orbitals, and each electron is considered moving in an average field of all the other electrons. The energy of the system is split into four contributions:

$$E[\rho] = E_{\text{Ne}}[\rho] + T_s + J[\rho] + E_{\text{xc}}[\rho] \quad (2.15)$$

where  $E_{\text{Ne}}$  is the electron-nuclear interaction energy,  $T_s$  is the kinetic energy of an independent-particle reference system,  $J$  is the Coulombic electron-electron repulsion and  $E_{\text{xc}}[\rho]$  is called the exchange-correlation energy which contains the correlation contribution to the total energy.<sup>139</sup>

**Exchange-correlation functionals.** The  $E_{\text{xc}}[\rho]$  is the only part of the energy (equation 2.15) that is still unknown in DFT theory. Approximate  $E_{\text{xc}}[\rho]$  functionals (a functional is a function of a function) are even today frequently being developed, however, the most popular and reliable models are still the hybrid functionals such as B3LYP developed in the early 1990's.<sup>140–142</sup> Hybrid functionals are developed by fitting simulations to experimental data or accurate wave function based methods, which is why DFT to some extend may be classified as a semi-empirical method.

**Basis set.** A basis set is a collection of functions for various atoms that is used to create the molecular orbitals.<sup>135</sup> In order to calculate the energy of a molecule each molecular orbital, or Kohn-Sham orbital, is written as a linear combination of basis functions:

$$\chi = \sum_m^M c_m \varphi_m \quad (2.16)$$

Each basis function, in turn, is usually written as a linear combination of Gaussian-type functions (primitive Gaussian-type orbitals). Basis set classifications, such as 6-31G(d,p), are used to denote the number of basis functions in a given basis set and the number of primitive Gaussian-type orbitals representing core and valence electrons. In general, the larger the basis set becomes the more exact the calculated orbitals (and thus the wavefunction) becomes.

**Prediction of Electronic Spectra.** In this work, all calculations of electronic excitations were performed using time-dependent density functional theory (TDDFT) on the DFT optimized geometries and assuming the Franck-Condon principle (vertical excitations). TDDFT is the extension of DFT to time-dependent phenomena.<sup>143,144</sup> In the specific case of calculating the electronic absorption spectrum, the system (molecule + light) can be approximated within

linear response theory due to the weak intensity of the external electromagnetic field,<sup>143</sup> which basically describes how an equilibrium system changes in response to an applied potential. Similarly to ground state DFT calculations, TDDFT is in practice implemented using a time-dependent KS scheme. In general, the popular hybrid functional B3LYP has been shown to be fairly reliable in predicting excitation energies of local valence-excited states.<sup>145, 146</sup>

## 2.4 Optical Spectroscopy

### 2.4.1 Steady-State Spectroscopy

**UV-Vis absorption.** An absorption spectrum is a plot of the UV-Vis wavelength-dependent absorbance of a sample and is recorded using a UV-Vis absorption spectrophotometer, typically in the wavelength range 200-800 nm. In this wavelength range, it is the relative radiative transition probability between different electronic states that is being monitored.

**Emission spectroscopy.** An emission spectrum is a plot of the relative wavelength-dependent intensity emitted from an excited sample. Emission spectra are recorded using a fluorescence spectrophotometer, usually in the 300-800 nm range (samples emitting in the visible range are highly beautiful, so go for that!). The intensity of the excitation beam and the measured emission signal is controlled by passing the light through slits with adjustable bandwidths. The intensity of the light source, as well as the sensitivity of the photodetector, is highly wavelength-dependent and the lamp profile additionally changes with time. The final spectrum is therefore calculated by correcting the measured intensity spectrum with a wavelength-dependent sensitivity factor.<sup>79</sup>

**Quantum yield determination.** The fluorescence quantum yield,  $\Phi_f$ , is the fraction of photons absorbed that results in emission of fluorescence.  $\Phi_f$  is determined relative to a reference compound of known quantum yield. If the same excitation wavelength and slit bandwidths are applied for the two samples, then  $\Phi_f$  is calculated as:

$$\Phi_f = \Phi_{\text{ref}} \cdot \frac{n_{\text{solvent}}^2}{n_{\text{ref.solv.}}^2} \frac{I}{A} \frac{A_{\text{ref}}}{I_{\text{ref}}} \quad (2.17)$$

where  $I$  is the integrated fluorescence. The absorbance is kept below 0.05 in order to ensure linear response and avoid inner filter effects.<sup>79</sup>

**Solvatochromism.** Solvatochromism describes the solvent-dependency of a chromophore's absorption spectrum and a fluorophore's emission spectrum. The solvatochromic effect is the result of an electrostatic interaction between the ground and excited state dipole moments of the dye and the permanent and temporary dipole moments of the surrounding solvent molecules.

### 2.4.2 Polarized Spectroscopy

**Circular dichroism.** In CD the difference in absorption of left- (L) and right-handed (R) circular polarized light is measured using a CD spectrometer.<sup>147</sup> The CD is then defined as

$$\text{CD}(\lambda) = A_{\text{L}}(\lambda) - A_{\text{R}}(\lambda) = (\varepsilon_{\text{L}}(\lambda) - \varepsilon_{\text{R}}(\lambda))cl \quad (2.18)$$

where  $\Delta\varepsilon = \varepsilon_{\text{L}} - \varepsilon_{\text{R}}$  is the molar circular dichroism. A CD signal is only observed in the absorption band of chiral chromophores or chromophores perturbed by a chiral environment.

**Fluorescence anisotropy.** Fluorescence anisotropy is a measure of the degree of polarized emission emanating from a sample excited using linearly polarized light. In this setup, fluorophores having their absorption transition dipole moment aligned with incident electromagnetic wave will be preferentially excited. If the fluorophore rotates in between time of absorption and emission of a photon, the emission from a population of fluorophores becomes depolarized to an extent depending on the ratio between the rotational speed and the excited state lifetime of the fluorophore. Dyes in slowly rotating environments, such as a large biomolecule, will possess higher fluorescence anisotropy compared to fluorophores tumbling free in solution.

If the molecules are completely immobilized, *e.g.* in a hydrocarbon glass at low temperatures, the anisotropy only depends on the relative angle,  $\beta$ , between the absorbing and emitting transition moments according to

$$r_{\text{A},0} = \frac{1}{5} (3 \cos^2 \beta - 1) \quad (2.19)$$

which is called the fundamental anisotropy.

### 2.4.3 Time-Resolved Emission Spectroscopy

**Time-correlated single photon counting.** In this work time-resolved intensity decays were recorded using time-correlated single photon counting (TCSPC). In TCSPC the sample is repeatedly excited using short light pulses and the subsequent emission of photons is detected at a photodetector positioned perpendicularly to the excitation beam. The time between each excitation pulse and the first detected photon is measured and stored in a statistical histogram corresponding to the measured fluorescence decay profile of the sample.<sup>79</sup>

**Time-resolved fluorescence anisotropy.** Time-resolved fluorescence anisotropy is the fluorescence anisotropy measured as a function of time following pulsed excitation.<sup>79</sup> The time-resolved fluorescence anisotropy is thus a direct measure of the rotational decay of the fluorophore. In order to acquire the time-resolved fluorescence anisotropy, the sample is excited using vertically polarized light pulses and the intensity decay of the sample is measured through a polarizer oriented vertically,  $I_{VV}(t)$ , and horizontally,  $I_{VH}(t)$ , to the sample. The anisotropy decay,  $r(t)$ , is then calculated as

$$r(t) = \frac{I_{\text{polarized}}(t)}{I_{\text{total}}(t)} = \frac{I_{VV}(t) - GI_{VH}(t)}{I_{VV}(t) + 2GI_{VH}(t)} \quad (2.20)$$

were  $G$  is the instrument sensitivity ratio towards vertically and horizontally polarized light. The  $G$  factor is not related to the properties of the sample, but is purely an experimental correction for the polarization bias of the detection system.  $G$  is measured by exciting the sample using horizontally polarized light and subsequently measuring the horizontally and vertically polarized components of the emission intensity ( $I_{HH}$  and  $I_{HV}$ ), each for the same period of time. Since there is no difference between the number of photons coming towards the HH and HV channels from the sample,  $G$  is calculated as the ratio between the measured total intensities (counts) in each channel:

$$G = \frac{\int I_{HV}(t)dt}{\int I_{HH}(t)dt} \quad (2.21)$$

**Analysis of time-resolved decays.** In practice, the excitation pulse is not a  $\delta$ -function and the instrumentation additionally has a certain electronic response time. This is quantified by the instrument response function,  $\text{IRF}(t)$ , which is the response profile of the instrument to a purely scattering solution. If  $\text{IRF}(t)$

is considered to be a series of  $\delta$ -excitation pulses with varying amplitude, the measured intensity at time  $t$ ,  $N(t)$ , is the sum of responses to each  $\delta$ -excitation pulse up until  $t$ . Thus

$$N(t) = \int_0^t \text{IRF}(t'_\delta) I(t - t'_\delta) dt' \quad (2.22)$$

where  $I(t - t'_\delta)$  denotes the fluorescence intensity from the sample at time  $t$ , originating as a response to a  $\delta$ -excitation pulse at time  $t_\delta$  and with amplitude  $L(t_\delta)$ . Equation (2.22) is called the convolution integral and the task is to determine the function,  $I(t)$ , which yields the best overall fit between  $N(t)$  and  $L(t)$ . The fitting procedure is very often performed using least squares analysis in which the "goodness-of-fit" parameter  $\chi^2$  is minimized by iteratively optimizing the model parameters (see section 2.5).<sup>79</sup>

## 2.5 Quantitative Data Analysis

---

This short section is limited to the analysis of data from fluorescence experiments.

**$\chi^2$  analysis and  $\chi^2$  surfaces.** In quantitative data analysis, a theoretical model is fitted to a set of experimental data.  $\chi^2$  denotes the sum of squared differences between the modelled and measured data, weighted according to the standard deviation of each data point and (for the reduced chi-square,  $\chi^2_R$ ) the total number of data points:

$$\chi^2_R = \sum_{k=1}^n \left[ \frac{\text{actual deviation}}{\text{expected deviation}} \right]^2 \times \frac{1}{n} = \left[ \sum_{k=1}^n \frac{(N(t_k) - N_c(t_k))^2}{N(t_k)} \right] \times \frac{1}{n}$$

Here  $N(t_k)$  is the number of measured counts at time  $t_k$ ,  $N_c(t_k)$  is the calculated number of counts at time  $t_k$ , and  $n$  is the number of datapoints.<sup>79</sup> A good fit is characterized by a value of  $\chi^2$  close to 1 and the residual between the measured and calculated decay curve should be randomly distributed around 0. This procedure is called a least-squares analysis.

**Global decay analysis.** In a global decay analysis multiple decays are fitted simultaneously with one or more fitting parameters being constrained globally (i.e. the parameter is set to be the same for all decays). In this case the global

reduced chi-square is the sum of the individual chi-squares weighted according to the total number of data points of all decays.

**Optimization algorithms.** For large datasets and complex models, optimizing a set of parameters to experimental data can easily take hours and days and often results in the algorithm being stuck in a local  $\chi^2$  minimum far from the correct, global, minimum. It may also be the case that there are multiple desirable minima on the (multidimensional)  $\chi^2$ -surface. Optimization algorithms is thus a research field in continuous development and there is no such thing as a single, universally applicable optimization algorithm.

If it is possible to evaluate the  $\chi^2$ -gradient the fastest optimization algorithm is usually steepest-descent. For many purposes in this work, the simplex search method of Lagarias *et al.* was used.<sup>148</sup> This algorithm is a direct search method that does not use gradients but searches around the current point looking for a new point where  $\chi^2$  is lower than the value at the current point.

## 2.6 Making User-Interfaced Software.

---

Software operated through a graphical user-interface (GUI) is an extremely convenient way to share applications and knowledge among researchers as it greatly reliefs the burden of reading and learning all the code behind the program. For the simulation and analysis of large amounts of data MATLAB is a particularly powerful programming environment as it combines the advantage of storing data in multidimensional matrices with the ease of making user-interfaces using the Java programming language.<sup>149</sup>

**GUIDE.** Making a user-interface in MATLAB is accomplished either by 1) hand (code only) or 2) using a MATLAB figure-file created by the GUIDE (Graphical User Interface Development Environment). While the first option is more flexible the second option is by far easier. GUIDE is run by typing `>>guide` in the MATLAB command window. The GUIDE is used to set the appearance of the user-interface such as inserting and arranging objects (buttons, text-boxes, etc.), setting the GUI size, adding menu icons, etc.. The properties of the user-interface, such as the action caused by pressing a button, is programmed in the .m-file associated with the constructed FIG-file.

**Creating stand-alone software.** A software written in MATLAB can be deployed into a stand-alone application using a MATLAB compiler tool. The deployment tool is run by typing `>>deploytool` in the MATLAB command window. The successful deployment process produces an executable program (.exe) that can be run outside of MATLAB provided that a MATLAB Compiler Runtime (MCR) environment has been installed on the target computer (which is free from MathWorks's website).

## **2.7** Experimental Details.

---

The experimental details of the work carried out in this thesis is provided in the appended papers.

# 3

## OVERVIEW OF PUBLISHED WORK

---

*This chapter briefly describes the appended papers and developed softwares*

### Contents

---

3.1	Paper I . . . . .	28
3.2	Paper II . . . . .	30
3.3	Paper III . . . . .	32
3.4	Paper IV . . . . .	34
3.5	Paper V . . . . .	36
3.6	Paper VI . . . . .	38
3.7	Published Software . . . . .	41
3.7.1	FRETmatrix . . . . .	41
3.7.2	a e . . . . .	43
3.7.3	FluorFit . . . . .	43
3.7.4	AniFit . . . . .	46

---

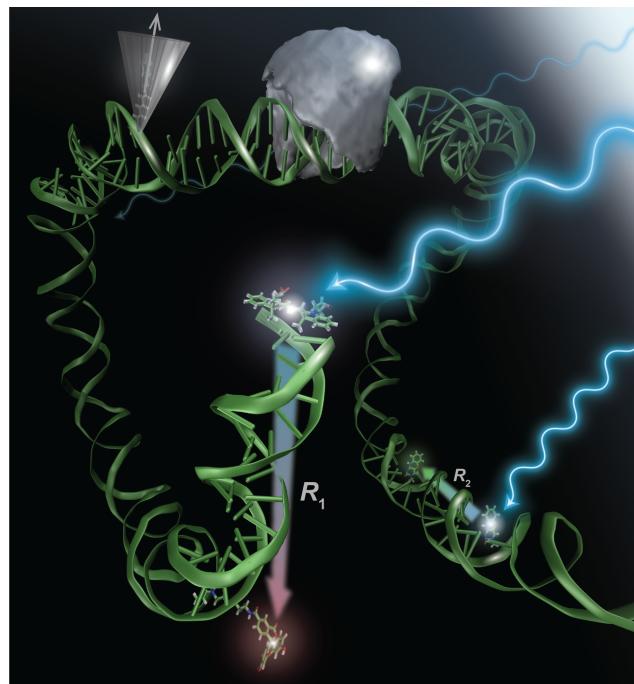
### **3.1 Paper I**

This review paper provides a comprehensive overview of studies aiming to advance FRET as a quantitative molecular ruler in studies of nucleic acid structure and dynamics. Although originally intended to be a MiniReview focusing on ensemble techniques we quickly realised during the initial literature search that this topic could not be covered satisfactorily without including single-molecule techniques. The paper includes all aspects of performing quantitative FRET from initial experiment design to data acquisition and analysis and was one of the five most accessed papers from the whole archive of ChemBioChem during September and October 2012.

**Summary of most interesting topics.** The review starts by comparing the advantages and limitations of FRET to those of other structural characterization techniques. Section 2 then provides a short overview of the properties of most popular commercial FRET dyes characterized in nucleic acid environments. Here, Table 1 provides a comprehensive source of references of dye lifetimes, rotational dynamics and interactions with DNA which are all relevant for the design and interpretation of FRET experiments. For many researchers, it may come as a surprise that practically all commercial dyes are somehow influenced by their local environment when positioned in DNA. Section 3 describes the advantages and limitations of alternative FRET probes, including the nucleobase analogues relevant for this thesis.

Focusing mostly on single-molecule methods, Section 4 describes experimental techniques developed to measure FRET quantitatively. FRET can be measured quantitatively at the single-molecule level both in experiments where the molecules are surface-immobilized and when they are diffusing freely in the solution. However, the quantitative analysis of single-molecule FRET is complex as a result of the low signal-to-noise ratio in such experiments. References to most important papers in this research field is provided in Section 5.

In order to interpret measured FRET signals into quantitative structural information, the local dynamics of the dyes must be taken into account. Most, if not all, techniques developed to model local probe rotational and diffusional dynamics is described in Section 6 of the paper. These methods include various theoretical models of fluorophore orientational distributions and algorithms optimized for the prediction of dye position and diffusion. In relation to this section, the final section highlights advanced software developed to analyse



**Figure 3.1:** Frontispiece for Paper I.

FRET quantitatively. Most of this software is freely available from the home pages of various group leaders active in this field.

Section 7 describes recent advanced techniques which exploit FRET to probe three-dimensional nucleic acid structures quantitatively. Most noteworthy is the triangulation-based techniques which exploit the combination of multiple FRET-pair positions in a DNA molecule in order to reconstruct part of the three-dimensional structure.<sup>91–98</sup>

**Preparation of figures.** Information or courses on how scientific illustrations are prepared are rarely available. Students are thus often oblivious when preparing such visuals (including myself), which is unfortunate since illustrative figures greatly reduces the time required to understand the paper. This short section addresses the issue in the context of the frontispiece of Paper I (Figure 3.1). A detailed step-by-step tutorial including all the raw data files and sessions is found at

[www.fluortools.com/misc/scientific-illustration](http://www.fluortools.com/misc/scientific-illustration)

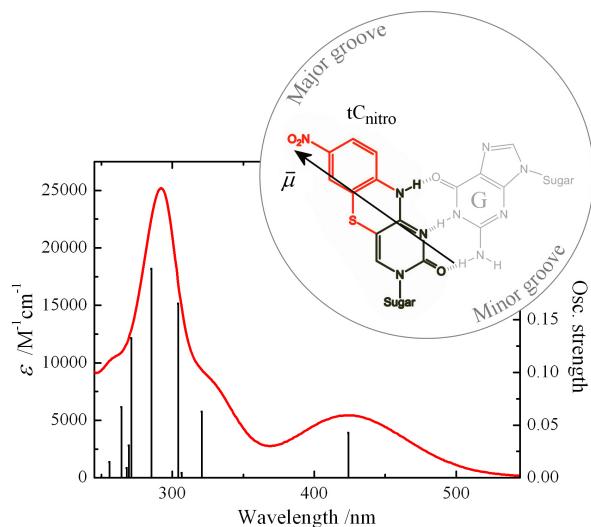
In summary, a DNA coordinate file was first constructed by w3DNA<sup>150</sup> using a manually edited input file describing all base-pair step parameters in the

structure. Using Avogadro,<sup>151</sup> the DNA molecule file was then combined with an accessible volume (AV) coordinate file of Alexa488 attached to DNA. Individual dye coordinates were constructed using ChemBio3D<sup>152</sup> and positioned at the desired sites in the molecule file using Avogadro. The DNA molecule file, including the AV and dyes, was then visualized in VMD<sup>153</sup> producing a POV-Ray file for high resolution rendering in POV-Ray.<sup>154</sup> The rendered png file was then edited using Adobe Illustrator in order to add light waves, dye-glow, the transition moment-cone and FRET arrows. To get the feeling that the arrows lie within the molecule, several figure frames with and without arrows were combined using Paint Shop Pro.

## **3.2 Paper II**

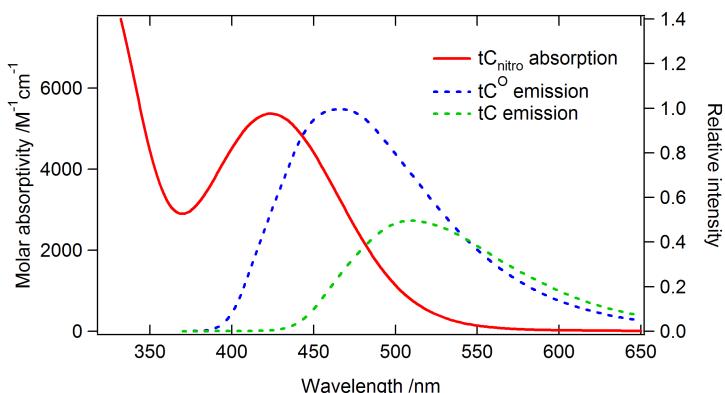
Paper II describes the characterization of nucleobase analogue tC<sub>nitro</sub> with particular focus put on its use as a FRET acceptor with tC or tC<sup>O</sup> serving as the donor. The paper builds upon the findings of the 2009 JACS paper provided in Appendix 1 by further elucidating the properties of tC<sub>nitro</sub> under biological conditions.

**Motivation and most important findings.** In order to simulate and interpret base-base FRET experiments quantitatively, the number of electronic transitions within the lowest energy absorption band of tC<sub>nitro</sub> had to be determined as well as the orientation of the transition dipole moment vectors within the molecular framework. In other words, the number of antennas involved in FRET and their directions had to be known. Paper II reports in particular these two findings (shown in Figure 3.2). Using a combination of TDDFT calculations, MCD and the fundamental fluorescence anisotropy recorded in low temperature glass, the lowest energy absorption band of tC<sub>nitro</sub> was shown to be the result of a single electronic transition. From the LD of tC<sub>nitro</sub> aligned in stretched PVA film combined with TDDFT calculations, the direction of the lowest energy transition dipole moment of tC<sub>nitro</sub> was found to be oriented 27° towards the NO<sub>2</sub>-group from the molecular long axis. The paper additionally shows that the pK<sub>a</sub> of tC<sub>nitro</sub> is 11.1 (deprotonation at the central enamine involved in Watson-Crick H-bonding) which led us to the conclusion that the neutral, Watson-Crick active form is totally predominant at biological conditions. While tC<sub>nitro</sub> is virtually non-fluorescent in aqueous solution at room temperature, the absorption spectrum of tC<sub>nitro</sub> neatly overlaps the emission spectra of both tC and tC<sup>O</sup> (Figure 3.3). The spectral overlap integrals in DNA were found to be  $J = 5.4 \times 10^{13}$  and  $J = 1.2 \times 10^{14}$  resulting in critical Förster



**Figure 3.2:** The absorption spectrum of tC<sub>nitro</sub> in H<sub>2</sub>O along with the TDDFT predicted electronic transitions. The most important finding of Paper II is the direction of the lowest energy transition dipole moment of tC<sub>nitro</sub> (insert)

distances of  $R_0 = 23.4 \text{ \AA}$  and  $R_0 = 27.2 \text{ \AA}$ , respectively, with the assumption of  $\kappa^2 = \frac{2}{3}$ .



**Figure 3.3:** Spectral overlap between the absorption spectrum of tC<sub>nitro</sub> and the emission spectra of tC and tC<sup>O</sup>.

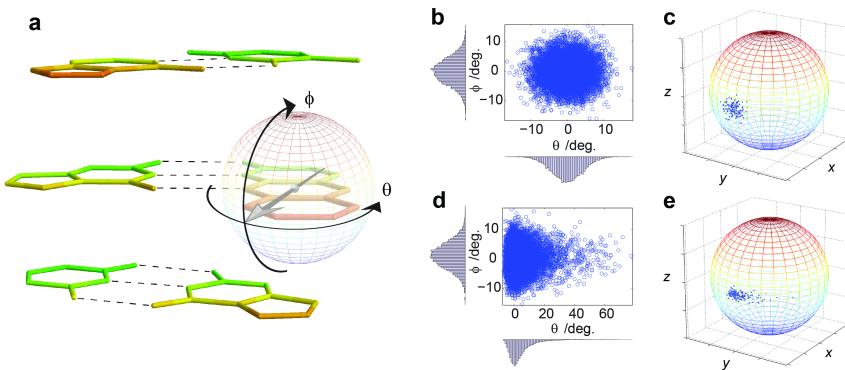
### **3.3 Paper III**

---

Paper III describes a new computational method for the simulation and analysis of FRET in nucleic acids. The method is generic in the sense that it can be used to simulate and analyse FRET quantitatively between any two probes positioned in any nucleic acid geometry. However, the method is particularly powerful for modelling constrained probes such as base-base FRET.

**Motivation.** In the first paper reporting base-base FRET (Appendix 1), the data was analysed by expressing the FRET efficiencies between multiple donor-acceptor pairs in dsDNA as a function of donor-acceptor separation.<sup>77</sup> This modelling was possible due to the simple sequential 5'-3' increase in donor-acceptor separation combined with the well-known helical periodicity of regular B-DNA. This model system results in a sinusoidal oscillation of the FRET efficiency as the number of bases in between the FRET-pair increases and, indeed, the experimental results showed that the structure in which the probes were positioned corresponded to a regular B-DNA helix (Figure 1.5). However, in an actual FRET experiment the structure is rarely a simple periodic helix but rather completely or partly unknown. This makes it inherently complicated to model FRET between probes in constrained environments where the assumption of freely rotating fluorophores (i.e.  $\kappa^2 = \frac{2}{3}$ ) is no longer valid. In other words, there are always two unknown parameters for each measured FRET efficiency:  $R$  and  $\kappa^2$  (in the method used in Appendix 1 there were essentially only one parameter since  $R$  was known in advance). Without a fundamentally new methodology capable of simulating both probe orientation and position in any nucleic acid structure, it would be impossible to interpret actual base-base FRET experiments quantitatively.

**The principle of the method.** FRETmatrix is build on a rigorous, matrix-based nucleic acid structure scheme called the Cambridge University Engineering Department Helix Computation Scheme (CEHS). In short, a 3D nucleic acid geometrical model is build by defining six rigid-body parameters describing each base-pair in the structure: three rotations (buckle, propeller, opening) and three translations (shear, stretch, stagger), as well as six rigid-body parameters describing each dinucleotide step in the structure: three rotations (tilt, roll, twist) and three translations (shift, slide, rise). By replacing one of the canonical bases in the modelled structure with a base probe, the position and orientation of the transition dipole moment of the dye can be accurately described at any site within the nucleic acid. Then, knowing the 3D coordinates



**Figure 3.4:** Representing local dye rotational dynamics by directional vector distributions. a) Definition of in-plane and out-of-plane movement of the base. b-c) Representing  $\theta$  and  $\phi$  by Gaussian distributions. d-e) Representing the out-of-plane coordinate by a Gaussian distribution and in-plane coordinate by the Boltzmann distribution for a harmonic potential.

describing the position and orientation of all transition dipole moments within the modelled nucleic acid structure, the energy transfer rate constants are simulated using the well-known Förster equations.

In order to interpret base-base FRET experiments quantitatively the rotational dynamics of the base probes must be known or included in the analysis. Such fast dynamics is an inherent property of the nucleobases and not a unique property of the dyes. In FRETmatrix, rotational dynamics is modelled by assigning two different energy potentials to the in-plane and out-of-plane movement of the bases in their base-pairing environment within dsDNA (Figure 3.4). The distribution of transition dipole moments in a given ensemble experiment is then described by the Boltzmann distribution along each of the two coordinates. The result is a directional vector distribution representing the orientation of a probe during an energy transfer event (Figure 3.4).

**Experimental demonstration studies.** Since any set of structural or dynamical parameters can be defined as the unknowns in a given experiment, the method is highly flexible. In demonstration study 1 of Paper III donor-acceptor pairs were positioned in regular dsDNA in which case the overall nucleic acid geometry was known in advance. This allowed us to analyze the experimental data, being all the donor intensity decays, in terms of the local probe dynamics. The results revealed directional vector distributions in excellent agreement with those expected for bases positioned within dsDNA. In addition, the results showed that the probes were bent towards the 3' end and into the major groove as predicted by TDDFT calculations in Paper V.

In demonstration study 2 of Paper III we used the information gained in demonstration study 1 and probed the 3D geometry of two local sites in DNA. Here, the model structures were deliberately chosen to be as simple as possible representing more complex geometries such as a protein binding site, a DNA lesion, a mismatch site, an RNA junction or some other local structural perturbation.

**The software.** To many people seeking to use base-base FRET for their own studies, the FRETmatrix methodology was in large parts too laborious to implement. Large amounts of effort was therefore put into making the method available to other researchers but ourselves and the result was a MATLAB-based software equipped with a user-interface and a user-guide (Supplementary Note 3 in Paper III). The software is described further in section 3.7. A one page brief summary of the software was made for the 2012 Autumn newsletter of Glen Research (Appendix 2).

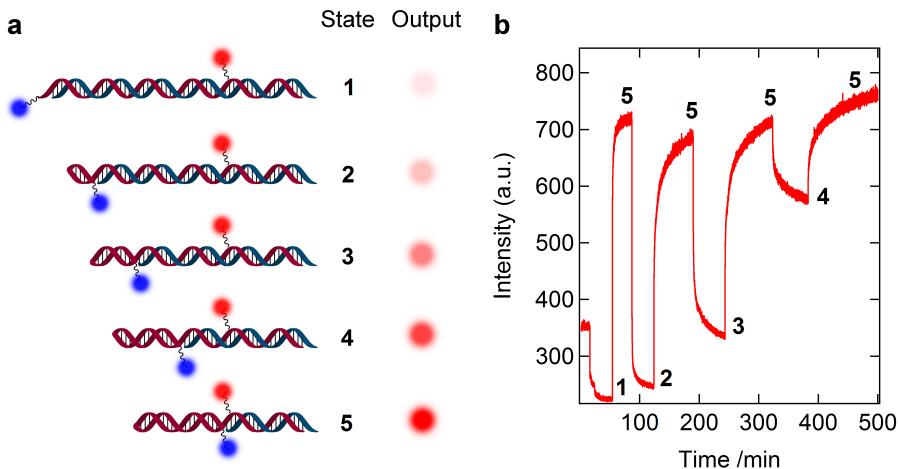
## 3.4 Paper IV

---

In Paper IV we report a simple DNA based adjustable strap capable of existing in five different states, each state accessed reversibly and with distinct fluorescence readouts.

**Motivation.** The ability to program readable information into nanoscale devices is an attractive technological advance in several aspects. Firstly, as a long-term information storage medium, DNA is stable and extremely dense (think of the amount of information coded into the human genome!). Methods capable of rapidly programming and reading information from such media are thus highly attractive.<sup>155–157</sup> Secondly, functional molecular machines capable of responding to an external stimuli are attractive as sensors,<sup>158</sup> long-range transport<sup>29–31, 35, 36</sup> and even computing.<sup>159–161</sup>

**Principles of the switch.** The switch consists of a 45 bp long single DNA strand divided into four repeating TAAT-ATTA box sequences plus an additional 13 base long tail. A FRET donor and acceptor is attached at either end of the TAAT-ATTA region which serves to signal the state of the switch. The repeating TAAT-ATTA region is self-complementary and thus has four different local equilibrium hairpin (stem-loop) structures characterized by different stem lengths (Figure 3.5a). The four different hairpin structures constitute four



**Figure 3.5:** The DNA switch of Paper IV. a. The five different states giving rise to a change in the fluorescence output. The red strand is the main strand functionalized with a FRET donor (blue) and a FRET acceptor (red) at either end of a TAAT-ATTA repeat region. The blue strands are complementary strands of different lengths. The output signal is the acceptor intensity upon donor excitation. b. Demonstration of accessing the five different states of switch (from Figure 3 in Paper IV). The graph shows the intensity from the acceptor upon donor excitation. The reversibility is obtained by elongating the blue complementary strands with an overhang which is the target of a third strand (release-strand) added at each of the lower plateaus.

of the five states. The fifth state is the fully opened hairpin obtained by the addition of the 45 bp long fully complementary strand. The state of the switch is the state with the lowest overall free energy of the system, which is controlled by adding the appropriate complementary strands of different lengths to the switch (blue strands in Figure 3.5a). If no complementary strands are added the main strand (red strand in Figure 3.5a) is folded into the longest possible hairpin structure, i.e. state 5. This DNA-based strategy for creating a dynamic switch is called strand-displacement and has been exploited in numerous previous reconfigurable nanostructures.<sup>162</sup>

**Demonstration studies.** All experiments were performed by J. Burns and E. Stulz. The demonstration studies show that the state of the switch can be set and altered at will by the addition of the proper navigation strands to the solution containing the switch. The combination of theoretical FRET calculations performed using FRETmatrix combined with FRET efficiencies measured for the different states of the switch proves that the switch exists in the expected pre-defined states. The reversibility of the switch is demonstrated by monitoring the relative fluorescence intensity from the acceptor upon donor

excitation as the complementary strands are added sequentially to the solution (Figure 3.5b). When the signal is high the donor and acceptor is brought close together, which corresponds to a long stem hairpin, while the signal is low when the donor and acceptor is located far from one another, which corresponds to a short stem hairpin.

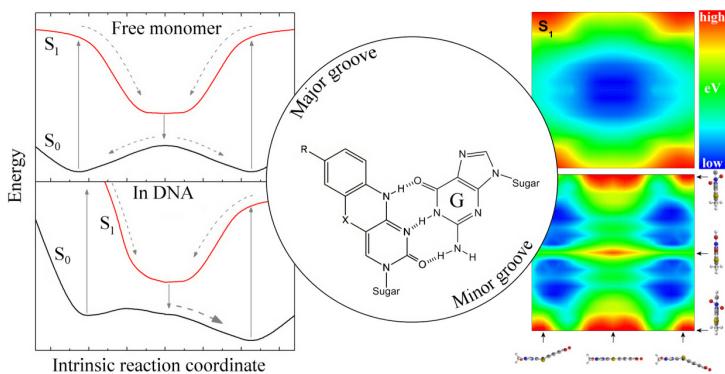
### **3.5 Paper V**

---

Paper V reports new insight into the electronic and structural properties of the tC bases ( $tC$ ,  $tC^O$  and  $tC_{nitro}$ ).

**Motivation.** This work was motivated by two driving forces. Firstly, the high, relatively stable fluorescence properties and single-exponential fluorescence decays of the  $tC$  and  $tC^O$  bases in DNA are unique features of these dyes compared to almost all other dyes positioned in DNA. It is thus not just of fundamental interest but also highly useful when designing new and improved fluorescent DNA modification to understand why  $tC$  and  $tC^O$  exhibit such desirable properties while  $tC_{nitro}$  is virtually non-fluorescent at room temperature in aqueous solution. Secondly, in order to understand the behaviour of the  $tC$  probes in various confined biological environments, as well as to interpret FRET and fluorescence anisotropy experiments involving these probes, the molecular geometries and structural dynamics of the probes had to be known.

**Insight into the excited state decay pathways.** The excited state relaxation processes of the  $tC$  probes were investigated by a combination of temperature-dependent fluorescence quantum yield measurements, TDDFT calculations and solvatochromic experiments. Using an Arrhenius expression to describe the temperature-dependent non-radiative decay process of  $tC_{nitro}$ , we found that the reason  $tC_{nitro}$  is non-fluorescent at room temperature is due to an efficient IC from the first excited state, very likely being directly associated with twisting of the  $NO_2$  group. This decay pathway was characterized by means of TDDFT calculations which provided insight into the electronic structure and potential energy surfaces along the  $NO_2$  rotational coordinate of  $tC_{nitro}$  (Figure 8 in Paper V). The results showed that the excited state dipole moment of  $tC_{nitro}$  increases as the  $NO_2$  group rotates leading to a higher activation energy for this process in non-dipolar solvents. In other words, as the dipolarity of the solvent increases the activation barrier for IC decreases leading to a lower fluorescence quantum yield of  $tC_{nitro}$ . The result is that  $tC_{nitro}$  is non-fluorescent in dipolar



**Figure 3.6:** The most important finding of Paper V is the potential energy surfaces of tC, tC<sup>O</sup> and tC<sub>nitro</sub>. Left: Proposed potential energy surfaces of tC<sub>nitro</sub> in its free monomeric form (predicted by TDDFT) and after insertion into double-stranded, B-form DNA (proposed model). Right: Potential energy surfaces of tC<sub>nitro</sub> following the coordinate corresponding to folding along the middle S-N axis and the coordinate corresponding to twisting of the NO<sub>2</sub>-group.

solvents such as H<sub>2</sub>O and fluorescent in less dipolar solvents such as THF and dioxane (Figure 9 in Paper V).

The temperature-dependent experiments were performed using a temperature-controlled Oxford Cryostat mounted in a fluorescence spectrophotometer. At this point the reader is warned: Doing quantitative spectroscopy at low temperatures can be time consuming and non-trivial. At low temperatures the sample often precipitates and the cuvette may break as a result of tensions in the formed glass. In such cases the cryostat can get severely contaminated if not properly evacuated. In addition, it is a tedious task to obtain perfect baselines of the cryostat at all temperatures.

**Insight into the molecular structures.** Insight into the potential energy surfaces of all three probes were obtained by means of TDDFT calculations (Figure 3.6). It was found that all three bases possess a low energy intrinsic reaction coordinate corresponding to folding along the central S-N axis of tC and tC<sub>nitro</sub> and O-N axis of tC<sup>O</sup> (Figures 5-7 in Paper V). In short this means that the tC bases are highly flexible in terms of bending along this central axis. While this result at first glance may seem uninteresting, this low energy IRC of the tC bases very likely dictates the geometrical appearance of the bases in constrained biological environments and thus their ability to adapt sterically into various biological systems such as DNA-protein complexes (Figure 10 in Paper V). In addition the bending property of the tC probes has a profound effect in the interpretation of FRET experiments, and this model was confirmed experimentally in Paper III.

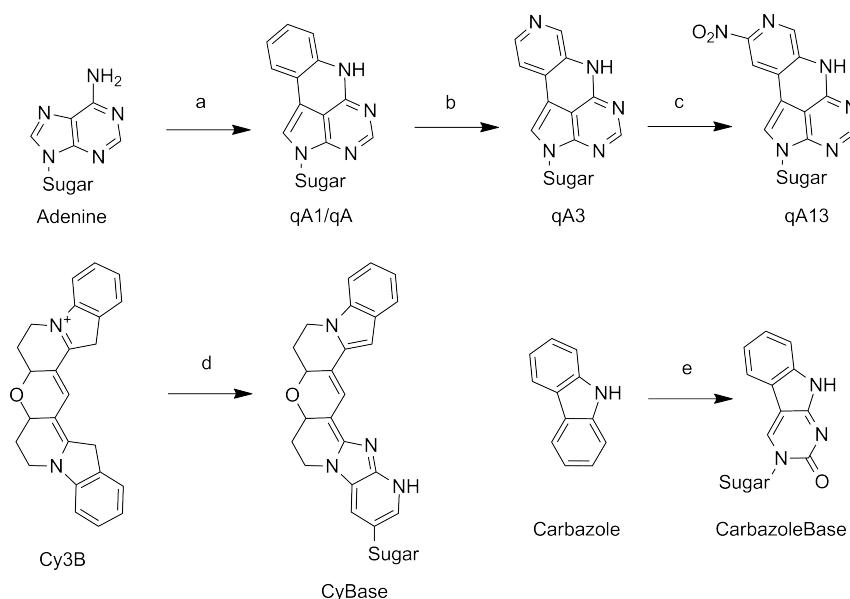
**3.6 Paper VI**

Paper VI reports the development of a new fluorescent adenine analogue, qA, and describes its photophysical and base-mimicking properties in DNA.

**Motivation.** Since we had fluorescent cytosine base analogues only, namely tC and tC<sup>O</sup>, we were highly interested in expanding the fluorescent nucleobase arsenal with dyes mimicking the base-pairing of adenine, guanine, or thymine. Such additions would facilitate a more versatile fluorescence toolbox and simplify systems design in e.g. FRET experiments.

**Design of fluorescent DNA base analogues.** Prior synthesizing qA, a theoretical nucleobase analogue library was constructed which served as a source of inspiration in the development of new, potentially fluorescent, DNA base analogues (Appendix 3). Due to synthetic difficulties, qA was the only structure from the library which resulted in a phosphoramidite for oligonucleotide synthesis. Several other molecular structures from the library, however, are equally or even more interesting as seen from a fluorescence application point of view. The qA structure was also derived by Marcus independently of the base library at the same time qA was added to the library.

The construction of the base library was achieved from a number of rational considerations and design steps. First, core structures were designed using the canonical bases, adenine, thymine, cytosine and guanine, as starting points. This step is exemplified by qA in Figure 3.7a. In the design of the core structures, the goal was to increase the  $\pi$ -system of the molecule but without causing sterical interactions between the base and its micro environment when positioned in double-stranded DNA, such as interactions with the backbone and neighbouring bases. In addition, the Watson-Crick hydrogen bonding pattern of the bases was maintained. Then, most of the C atoms in the aromatic frameworks of these core structures were systematically substituted by N without affecting the  $\pi$ -conjugation throughout the molecule (Figure 3.7b). Finally, NO<sub>2</sub> side-groups were systematically attached to the core structures (Figure 3.7c). Aromatic molecules with NO<sub>2</sub> side-groups usually cause a red-shifted absorption, as in the case of tC<sub>nitro</sub>, which is interesting in the development of novel FRET acceptors. This design strategy resulted in about 90% of the structures in the library. The other 10% were designed using known dye structures as starting points functionalized with Watson-Crick hydrogen



**Figure 3.7:** Design of fluorescent nucleobase analogues.

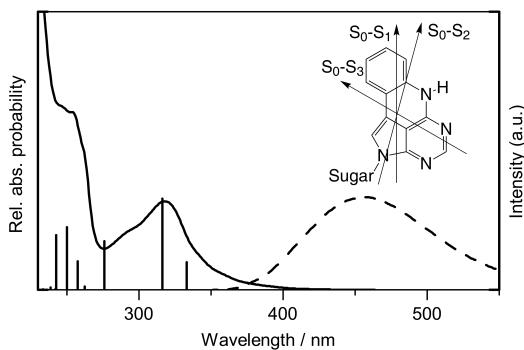
bonds which includes CyBase based on the Cy3B framework and CarbazoleBase based on the carbazole framework (Figure 3.7d and e, respectively).

It is inherently difficult, practically impossible, to predict all possible excited state decay rate constants of a molecule theoretically, especially in the case where the dye is positioned in demanding environments such as inside the base-stack of double-stranded DNA. This limitation makes it impossible to accurately predict the fluorescence quantum yield or lifetime of a dye that has not yet been synthesized. However, it is possible to increase the probability of a molecular structure to be fluorescent by the combination of different design considerations and predictions. First of all, since fluorescence occurs from the lowest energy electronic transition,  $S_1$ , the fluorescence rate constant is directly related to the probability of the  $S_0-S_1$  transition which is reflected in the magnitude of the oscillator strength of this transition.<sup>163</sup> Thus, estimating the magnitude of the lowest energy electronic transition theoretically provides a means to estimate the magnitude of the fluorescence rate constant thus increasing the probability that the molecule is fluorescent. However, in order to increase the fluorescence quantum yield the rate of the non-radiative decay pathways must also be minimized. This can be achieved by rigidifying the molecular framework and limiting the number of "floppy" side-groups contributing to the electronic ground and first excited states.<sup>164</sup> In addition, the presence of heavy atoms should be minimized in order to limit intersystem crossing.

Overall, the most interesting base analogue structures from the library, as seen from a fluorescence application point of view, are identified as those with a lowest energy electronic transition well separated from higher energy transitions, a high oscillator strength of this lowest energy transitions and as red-shifted an excitation energy as possible. The first of these criteria is important in order to use the probe in FRET and fluorescence anisotropy applications since multiple transition dipole moments in the lowest energy absorption band will complicate quantitative data analysis in such experiments. The second criteria is in order to increase the fluorescence rate constant thus increasing the probability of a high fluorescence quantum yield as discussed above. The third criteria is a convenience property since many microscopy setups and laser sources are cheaper and easier to operate for excitation in the visible region of the spectrum and this is thus the standard setup in many laboratories. Furthermore, red-shifted transitions are interesting in the development of FRET acceptors and limits the presence of auto-fluorescence when working with biological samples.

**Most important findings.** Interestingly, the predicted energies and oscillator strengths of the lowest energy electronic transitions of qA were in good agreement with the measured UV-Vis absorption bands of the synthesized probe (Figure 3.8). The results showed that the lowest energy absorption band of qA contains at least two electronic transitions with the lowest in energy characterized by a relatively small oscillator strength compared to the second lowest in energy. This is an interesting observation because it may explain why qA has a fairly modest fluorescence quantum yield ( $\Phi_f = 6.8\%$ ) compared to e.g. tC and tC<sup>O</sup>: The low fluorescence quantum yield is the result of a low-probability S<sub>0</sub>-S<sub>1</sub> electronic transition "hidden" in the red part of the measured lowest energy absorption band of qA. The predicted directions of the transition dipole moments are also in fairly good agreement with the measured fundamental fluorescence anisotropy excitation spectrum which showed more than one transition hidden in the lowest energy absorption band of qA (Figure 3 in Paper VI).

Paper VI is an excellent example of how the combination of quantum chemical predictions and rational structure design is extremely powerful in the development of novel fluorophores with predefined properties.



**Figure 3.8:** Predicted lowest energy electronic transitions of qA with measured absorption and emission spectra in  $\text{H}_2\text{O}$ .

## 3.7 Published Software

---

It is a goal to make the small and larger programs and scripts I develop for myself available to others. All of the programs are written in MATLAB but they are operated through user-interfaces and can be run as stand-alone executables (exe) only requiring a free MATLAB Compiler Runtime (MCR) installed. All the software can be downloaded from:

[www.fluortools.com](http://www.fluortools.com)

A comprehensive source of documentation is provided on the home page.

### 3.7.1 FRETmatrix

---

FRETmatrix is used for the simulation and analysis of base-base FRET (Figure 3.9). The software has undergone many iterations and a second version that did not make it into the thesis is underway. The theory behind the program is described in Paper III and a User Guide is provided in the Appendix of Paper III. FRETmatrix has since the end of September 2012 been downloaded >20 times from the Chalmers website.

### 3. Overview of Published Work

42

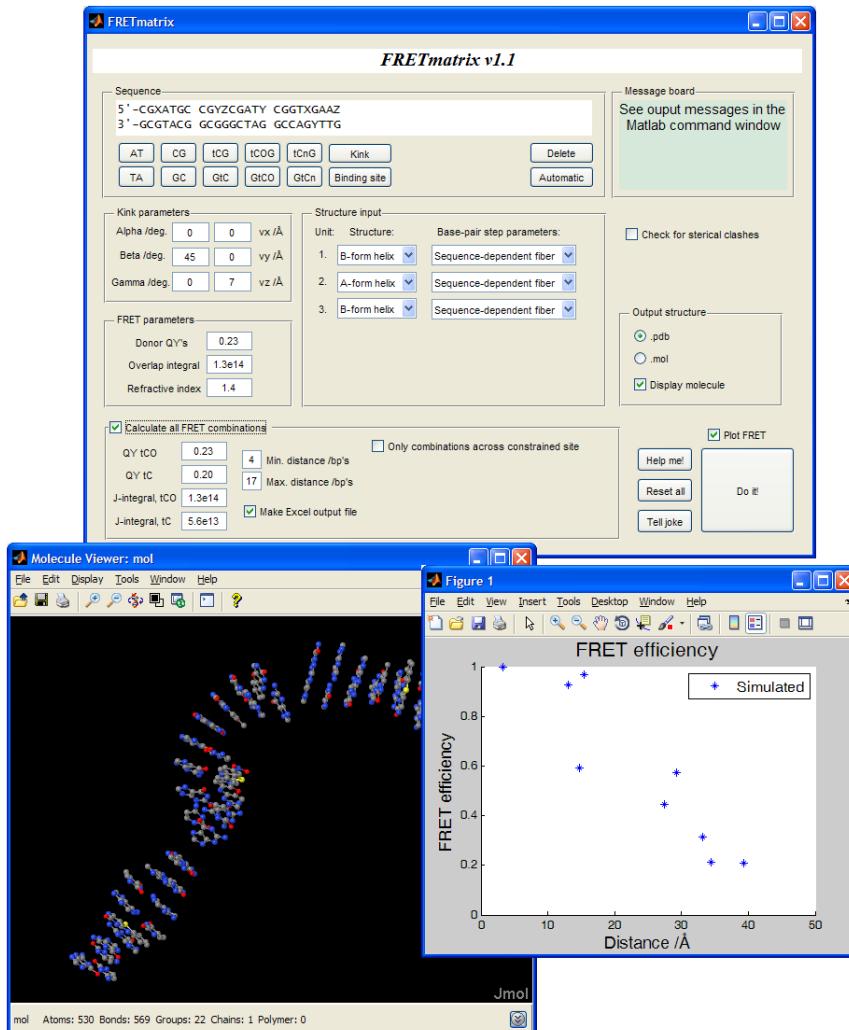


Figure 3.9: Screenshot of FRETmatrix v1.1.

**3.7.2** a|e

---

a|e (pronounced ae) is a software for analyzing, editing and plotting UV-Vis absorption and emission spectra. a|e was made as a free MATLAB-based contrast to commercial data plotting and analysis softwares such as Igor and Origin Pro and is designed particularly for handling and processing multiple UV-Vis spectral datasets simultaneously.

- As a spectral analysis software, a|e can be used to calculate fluorescence quantum yields and spectral overlap integrals (for FRET) independent of input data ranges and sizes.
- Reference UV-Vis absorption and fluorescence spectra of most commercial dyes can be downloaded directly from Invitrogen's homepage and into a|e for further analysis.
- Binding constants or  $pK_a$  values can be obtained from UV-Vis titration measurements by means of singular value decomposition (SVD).
- As a spectral editing software a|e can be used to perform mathematical operations to multiple spectra independent of the individual wavelength ranges and stepsizes of the loaded data.
- Baselines and light scattering can be subtracted from UV-Vis absorption spectra, the latter *via* a fit to the far-red end of the spectrum (see online documentation page at [www.fluortools.com/software/ae/documentation/scatter](http://www.fluortools.com/software/ae/documentation/scatter) for more information).
- As with FluorFit and AniFit, the message board can be used as a MATLAB command window for more specialized actions.
- For general plotting purposes a|e is meant to handle multiple spectral datasets in a convenient and fast way.
- The abscissa is easily alternated in between units of wavelengths (nm) or energy ( $\text{cm}^{-1}$  or eV).

**3.7.3** FluorFit

---

FluorFit is a general fluorescence decay fitting software for both standard and more specialized intensity decay analysis (Figure 3.11). FluorFit was made because in my experience popular intensity decay deconvolution programs either

### 3. Overview of Published Work

44

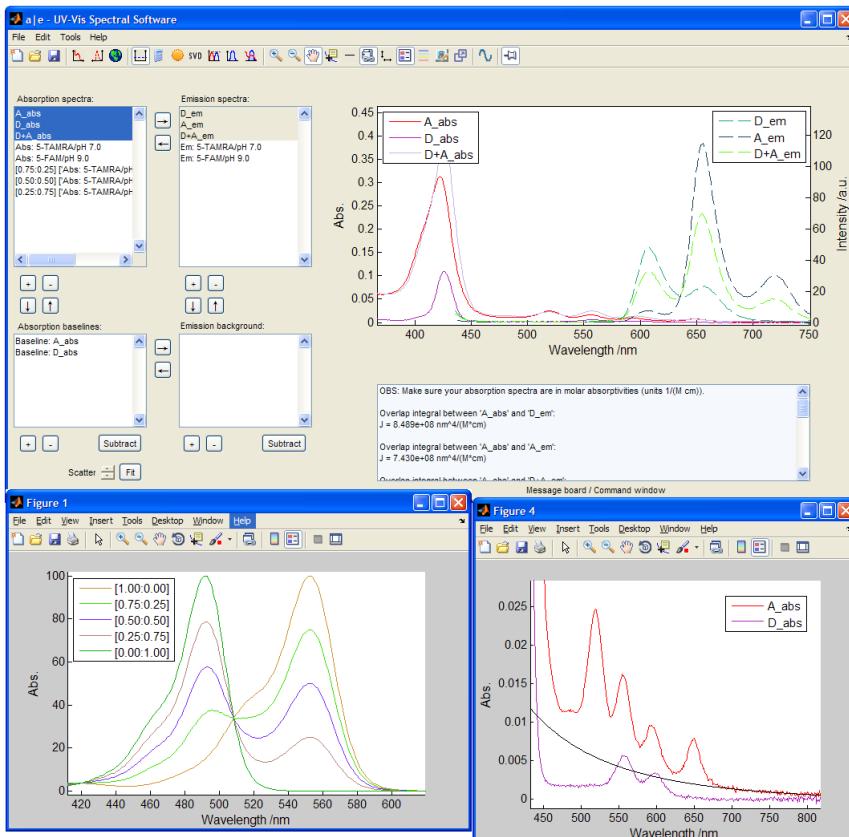
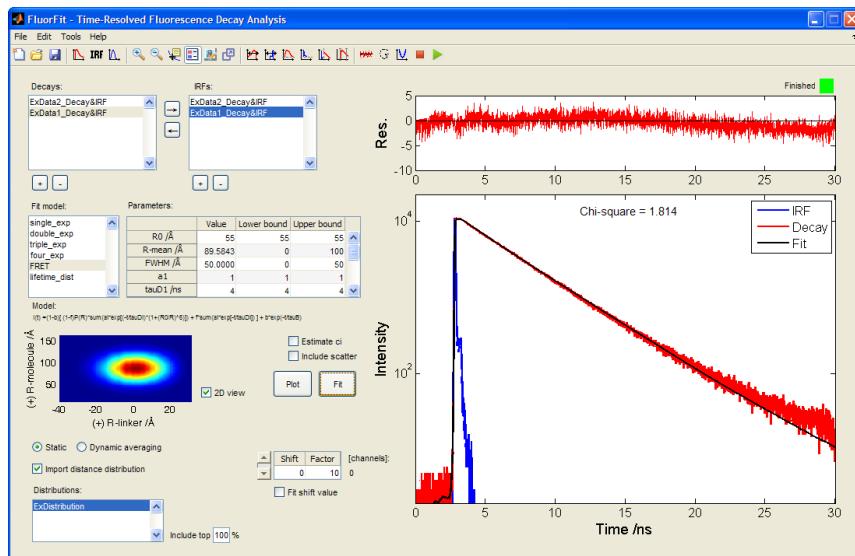


Figure 3.10: Screenshot of a|e v1.0.

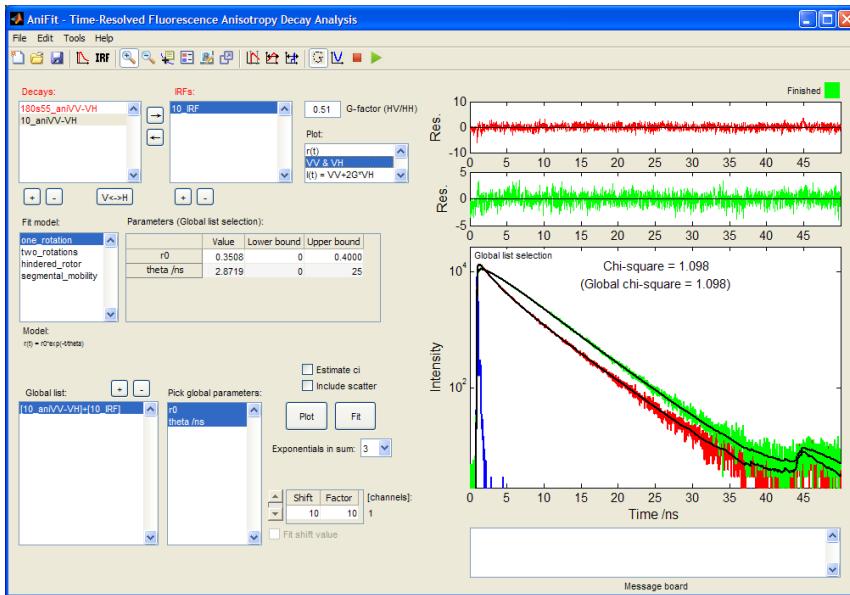


**Figure 3.11:** Screenshot of FluorFit v1.0.

come with limited features or their license require that they are installed on lab-based computers only. This was inconvenient to me since the lab computer was located in Gothenburg while my office was located in Copenhagen. In addition, using an in-house build software allows for greater flexibility in terms of more specialized and custom data analysis.

FluorFit provides most if not all of the fitting features found in commercial programs like Fluofit from PicoQuant and DAS6 from Horiba. FluorFit accepts ASCII-based input files and can handle multiple decays, IRFs and fitting models simultaneously. A FluorFit session can handle multiple graphics window simultaneously and sessions can be saved and reopened. The message board displaying output messages can additionally be used as a MATLAB command window for more specialized actions. Some of the analysis features of FluorFit:

- Reconvolution or tailfitting analysis
- Imported or user-defined, Gaussian-shaped Instrument Response Functions
- Global analysis of multiple decays
- Exponential decays or lifetime distributions
- FRET analysis using a distribution of donor-acceptor distance: Gaussian or imported distributions
- User-defined fitting models



**Figure 3.12:** Screenshot of AniFit v1.0.

- Parameter confidence-interval estimation from the Jacobian matrix
- Parameter confidence-interval estimation by calculation of  $\chi^2$ -surfaces
- Autocorrelation plot analysis
- Scattering, background and IRF shifting can be taken into account

#### 3.7.4 AniFit

AniFit is a time-resolved fluorescence anisotropy decay analysis software (Figure 3.12). AniFit implements a form of global direct vector deconvolution of the vertical and horizontal intensity decay components recorded in an anisotropy experiment in order to optimize the parameters of the anisotropy decay model. First the isotropic fluorophore intensity decay

$$I(t) = I_{VV} + 2G \times I_{VH}$$

is optimized using iterative deconvolution with a multiexponential decay. The fitted sum,  $I(t)$ , is then used as a constraint to simulate each of the  $I_{VV}$  and  $I_{VH}$  components:

$$I_{VV,\text{sim}} = \frac{1}{3} I(t) [1 + 2r_{\text{model}}(t)]$$

$$I_{\text{VH,sim}} = \frac{1}{3} I(t) [1 - r_{\text{model}}(t)]$$

The simulated decays are convolved with the instrument response function:

$$I_{\text{VV,conv}} = \int_{-\infty}^t \text{IRF}(t') \times I_{\text{VV,sim}}(t - t') dt'$$

$$I_{\text{VH,conv}} = G \int_{-\infty}^t \text{IRF}(t') \times I_{\text{VH,sim}}(t - t') dt'$$

and iteratively optimized to the measured decays by optimizing the parameter values in the theoretical expression for  $r(t)$ . This means that AniFit requires a single IRF only (usually VV and VH IRFs are identical).

AniFit is built on the framework of FluorFit and provides many of the same features listed above but with focus on the analysis of time-resolved fluorescence anisotropy parameters rather than lifetimes.



# 4

## CONCLUSIONS AND OUTLOOK

---

*This chapter concludes and discuss future development within quantitative fluorescence probing of biomolecular structures*

### Contents

---

4.1	Summary	50
4.2	Outlook	51
4.2.1	Novel Fluorescent Base Analogues	51
4.2.2	Quantitative FRET	52

---

## 4.1 Summary

---

The work presented in this thesis aimed at developing new tools based on fluorescence for studying the structure and dynamics of nucleic acids. In particular, quantitative FRET-based methods relying on orientationally constrained nucleobase analogues were developed. In addition, a library of fluorescent nucleobase analogue candidates was constructed which served as a source of inspiration for developing new and improved fluorescent DNA-mimicking probes.

**Appended papers.** The appended papers I-IV concern the development of quantitative FRET-based techniques. Paper I reviews the field of quantitative FRET with particular focus on recent advancements. Paper II reports the characterization of nucleobase analogue FRET acceptor  $tC_{nitro}$  and it was found that the lowest energy absorption band of  $tC_{nitro}$  is the result of a single electronic transition with an in-plane transition dipole moment oriented towards the nitro-group. Paper III describes a new generic methodology for the simulation and analysis of FRET in nucleic acids and demonstrates the power of the method to model base-base FRET. In Paper IV a DNA-based switch with five discrete states and a fluorescence readout is presented, where FRETmatrix was used to model the theoretical FRET efficiencies in the different states.

Paper V and VI are focused on the characterization and development of fluorescent nucleobase analogues. In Paper V we obtain insight into the excited state decay processes of  $tC$ ,  $tC^O$  and  $tC_{nitro}$  and explain why  $tC_{nitro}$  is non-fluorescent while  $tC$  and  $tC^O$  are highly fluorescent. In Paper VI we report a new fluorescent adenine analogue and describe its photophysical and base-mimicking properties in DNA.

**Published software.** Four general software packages for the simulation and analysis of data from spectroscopic measurements are presented. All of the programs are compiled into stand-alone executables, however, an important consideration is that they are all MATLAB-based which makes them extendable by end-users:

FRETmatrix is a software for the simulation and analysis of FRET in highly constrained nucleic acid geometries and was published with Paper III. The software implements a new methodology based on the ability to construct any three-dimensional nucleic acid model followed by a FRET simulation between two dyes positioned in the modelled structure. FluorFit is a general time-resolved intensity decay fitting software which contains both everyday

reconvolution features and more specialized capabilities, such as time-resolved FRET and global fitting. AniFit is the time-resolved fluorescence anisotropy counterpart to FluorFit. AniFit can be used to analyse multiple fluorescence anisotropy decays, both locally and globally, using a modified version of global iterative vector deconvolution. a|e is a software for the plotting, editing and analysis of UV-Vis absorption and emission spectra. Since a|e is designed specifically for spectral analysis it has a number of convenient shortcuts compared to more general data analysis and graphing software like IGOR Pro and Origin. Such features include subtraction of background scatter in absorption measurements, calculation of spectral overlap integrals in FRET, calculation of fluorescence quantum yields, singular value decomposition analysis, decomposition of spectra resulting from multiple absorbing or emitting species, and mathematical operations independent of the wavelength grids of the individual spectra.

## 4.2 Outlook

---

Based on the findings of this thesis and my experience working in the covered areas these are my thoughts on the future development within fluorescent nucleobases and quantitative FRET.

### 4.2.1 Novel Fluorescent Base Analogues

---

Fluorescent probes are traditionally developed based on a trial-and-error strategy, often inspired by known dye structures. However, the work reported in this thesis has shown that it is possible to rationally design molecules with DNA base-mimicking properties and at the same time to provide good predictions of their optical properties. Since high level quantum chemical methods like DFT and TDDFT are now being performed on standard laptop computers on an everyday basis, in principle without any requirements of programming skills or deep theoretical insight, this paves the way for a fundamentally new strategic approach for developing state-of-the-art fluorescent probes. The nucleobase analogue library initiated in Appendix 3 is meant to provide a source of inspiration in the development of new and improved probes, such as red-shifted excitation energies, higher brightness and higher stability. The omnipresent limitation of this design-based strategy, however, is the synthetic difficulties involved in making the most promising candidates.

In my eyes, the holy grail in the field of fluorescent nucleobase analogues is the development of probes sufficiently bright and photostable for single-molecule detection. The combination of single-molecule FRET and fluorescent nucleobase analogues could facilitate completely new possibilities to investigate the three-dimensional structure, dynamics and thermodynamics of nucleic acids with unprecedented temporal and spatial resolution. This, however, is a more complicated task than originally imagined: Single molecule setups usually require excitation sources located well into the visible region, however, such excitation energies ideally require large  $\pi$ -conjugated systems thus compromising the nucleobase-mimicking properties of the probe, in particular if the probe is to be positioned in sterically demanding biological environments. It is likely that fluorescent DNA-modifications for single-molecule applications will become a new technology in the near future, however, there will probably always be a compromise between the bio-mimicking properties of the probes and their fluorescent properties.

### 4.2.2 Quantitative FRET

---

While quantitative FRET has served as a molecular ruler for decades, the recent progress in this technique has shown that this field is only at its beginning. The ability to design experiments with multiple well-defined donor-acceptor pairs positioned at various sites within a biomolecule opens up new possibilities to probe the detailed three-dimensional position and geometry of unknown structural elements. This technique will be further aided by the well-defined positions and orientations of the base probes utilized in this thesis. Quantitative FRET-based methods are and continue to be important alternatives to more complex, time-consuming, expensive and less sensitive higher resolution techniques like biomolecular NMR spectroscopy and X-ray crystallography.

#### 4.2.2.1 The Outlook of FRETmatrix

---

Since FRETmatrix constitutes a general framework for simulating FRET in nucleic acids it is anticipated that the methodology will serve as a scaffold for even more sophisticated studies of nucleic acids structure and dynamics.

An interesting potential extension to FRETmatrix is the probing of nucleic acid dynamics occurring on timescales longer than the lifetimes of the probes themselves, i.e. in the  $\mu$ s-ms timescales. Dynamics occurring in these timescales are, for example, the conformational dynamics across helix-junction-helix motifs

in large functional RNAs, which are highly important in the folding of RNAs as well as for the transition between different functional states.<sup>165</sup> Insight into this type of dynamics is a rapidly progressing field and currently relies on state-of-the-art NMR with residual dipolar couplings.<sup>166</sup> Using FRETmatrix in combination with base-base FRET local base-dynamics occurring on pico- and subnanosecond timescales may be separated from longer time-scale dynamics by proper fluorescence decay analysis. This is theoretically achievable because fast probe dynamics, such as those probed in demonstration study 1 of Paper III, leads to dynamic averaging while slow dynamics leads to static averaging of  $\kappa^2$  and  $R$ . Dynamic averaging, in turn, leads to a single-exponential decay while static averaging leads to a multi-exponential decay, and these two types of dynamics can therefore be distinguished in the analysis of fluorescence decays. In other words, if the donor decay in a FRET experiment is single-exponential then the averaging regime is dynamic while if the decay is multi-exponential the averaging regime is static. This is of course true only when the donor decay is single-exponential in absence of acceptor, which is the case for tC<sup>O</sup> and tC in double-stranded DNA. In demonstration study 1 of Paper III, the measured donor decays of tC<sup>O</sup> in the presence of tC<sub>nitro</sub> were well fitted using a single lifetime as a result of the rigidity of B-DNA combined with the fast timescales of local rotational dynamics of the probes which leads to dynamic averaging. This experimental observation of single-exponential donor decays in the presence of FRET supports the suggested proposal.

A final note: FRETmatrix simulates signals, i.e. fluorescence decays in the case of Paper III, from three-dimensional nucleic acid geometrical models. It is thus equally imaginable to combine FRETmatrix with the simulation and analysis of alternative experimental techniques such as single-molecule FRET, FRET microscopy or Fluorescence Lifetime Imaging (FLIM).



# 5

## ACKNOWLEDGEMENTS

---

It is a pleasure for me to express my gratitude to the many people who have helped me or have been an important part of my life the last three and a half years.

First of all, I am grateful to my supervisor in Copenhagen, Kristine Kilså, for supporting me all the way from my Bachelor work to pursuing the PhD degree. Although circumstances hindered an explicit supervision during most of the PhD I am incredibly thankful for the time and responsibility you trusted me with all this time.

I am equally thankful to my supervisor at Chalmers, Bo Albinsson, for all of the supervision and for the encouragement to follow my own ideas. I have always been free to pursue the projects I had in mind yet still been guided in the right direction, and I am grateful for that.

Of all the important people involved in my thesis, I am particularly grateful to Marcus Wilhelmsson at Chalmers for the very close collaboration we have had. You have been the weekly contact and supervisor in almost all of my projects and I will always be thankful for that. Most of these projects could not have happened without you. I admire your skills and talents and I am very happy for the friendship we have build up during the years.

I am thankful to the previous group members in Copenhagen, Eduardo Della Pia, Jeppe Fock and Martin Rosenberg, for the good coffee breaks and discussions we have had. Thank you Søren, for working on my projects and for testing the software. Thank you Erling, for nice discussions and software testing.

A warm thanks goes to all of the nice people I have met at Chalmers, both previous and current members. I am grateful to you Anke Dierckx for being a friend and close collaborator. Jakob Woller for the friendship and the many table tennis matches we fought at Chalmers. Also, thank you Jonas Hannestad for the time spent helping with time-resolved measurements and data conversion.

I am thankful to all of my co-authors, in particular Karl B, Francois and Anke for the work you have put into these projects. Thank you Jon Burns and Eugen Stulz at Southampton University for our fruitful collaboration.

I am thankful to the Herschlag lab at Stanford for inviting me in as a visiting researcher, in particular to Dan and Xuesong for our collaboration and to Rishi for the many great discussions. I had a truly rewarding stay that I would not have been without.

Finally, I am forever thankful to my wife Susan for all of your love and support, even when I have brought home work and I am mentally distracted or my working hours have exceeded the maximum healthy dose. Thanks to my wonderful son Sigurd for putting a huge smile on my face every single day. Jeg kunne ikke være mere lykkelig, end når jeg er sammen med jer.

*Søren Preus*

*Copenhagen, Denmark, December 2012*

# BIBLIOGRAPHY

---

- [1] V. A. Bloomfield, D. M. Crothers, and I. Tinoco Jr. *Nucleic Acids. Structures, Properties, and Functions*. University Science Books, 2000.
- [2] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [3] G. Felsenfeld, D. R. Davies, and A. Rich. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.*, 79:2023–2024, 1957.
- [4] M. Duca, P. Velkhoff, K. Oussedik, L. Halby, and P. B. Arimondo. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.*, 36:5123–5138, 2008.
- [5] J. T. Davis and G. P. Spada. Supramolecular architectures generated by self-assembly of guanosine derivatives. *Chem. Soc. Rev.*, 36:296–313, 2007.
- [6] A. Jain, G. Wang, and K. M. Vasquez. Dna triple helices: Biological consequences and therapeutic potential. *Biochimie*, 90:1117–1130, 2008.
- [7] J. Y. Chin, E. B. Schleifman, and P. M. Glazer. Repair and recombination induced by triple helix dna. *Front. Biosci.*, 12:4288–4297, 2007.
- [8] R. Besch, C. Giovannangeli, and K. Degitz. Triplex-forming oligonucleotides - sequence-specific dna ligands as tools for gene inhibition and for modulation of dna associated functions. *Current Drug Targets*, 5:691–703, 2004.
- [9] J. L. Huppert. Hunting g-quadruplexes. *Biochimie*, 90:1140–1148, 2008.
- [10] Y. Qin and L. H. Hurley. Structures, folding patterns, and functions of intramolecular dna g-quadruplexes found in eukaryotic promoter regions. *Biochimie*, 90:1149–1171, 2008.
- [11] B. Chen, J. Liang, X. Tian, and X. Liu. G-quadruplex structure: a target for anticancer therapy and a probe for detection of potassium. *Biochemistry*, 73:853–861, 2008.
- [12] A. Serganov and D. J. Patel. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.*, 8(10):776–90, 2007.
- [13] N. C. Seeman. An overview of structural dna nanotechnology. *Mol. Biotechnol.*, 37:246–257, 2007.
- [14] N. Seeman. Nanomaterials based on dna. *Annu. Rev. Biochem.*, 79:65–87, 2010.
- [15] A. Kuzuya and M. Komiyama. Dna origami: fold, stick and beyond. *Nanoscale*, 2:310–322, 2010.
- [16] J. Nangreave, D. Han, Y. Liu, and H. Yan. Dna origami: a history and current perspective. *Curr. Opin. Chem. Biol.*, 14:608–615, 2010.
- [17] T. Torring, N. V. Voigt, J. Nangreave, H. Yan, and K. V. Gothelf. Dna origami: a quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.*, 40:5636–5646, 2011.
- [18] W. M. Shih. Knitting complex weaves with dna origami. *Curr. Opin. Struct. Biol.*, 20:276–282, 2010.
- [19] F. A. Aldaye, A. L. Palmer, and H. F. Sleiman. Assembling materials with dna as the guide. *Science*, 321:1795–1799, 2008.

- [20] J. Tumpane, P. Sandin, R. Kumar, V. E. C. Powers, E. P. Lundberg, N. Gale, P. Baglioni, J.-M. Lehn, B. Albinsson, P. Lincoln, L. M. Wilhelmsson, T. Brown, and B. Norden. Addressable high-information-density dna nanostructures. *Chem. Phys. Lett.*, 440:125–129, 2007.
- [21] P. W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.
- [22] F. C. Simmel. Three-dimensional nanoconstruction with dna. *Angew. Chem. Int. Ed.*, 47:5884–5887, 2008.
- [23] H. Dietz, S. M. Douglas, and W. Shih. Folding dna into twisted and curved nanoscale shapes. *Science*, 325:725–730, 2009.
- [24] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih. Self-assembly of dna into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.
- [25] J. S. Shin and N. A. Pierce. A synthetic dna walker for molecular transport. *J. Am. Chem. Soc.*, 126:10834–10835, 2004.
- [26] Y. Chen and C. Mao. Putting a brake on an autonomous dna nanomotor. *J. Am. Chem. Soc.*, 126:8626–8627, 2004.
- [27] W. B. Sherman and N. C. Seeman. A precisely controlled dna biped walking device. *Nano Lett.*, 4:1203–1207, 2004.
- [28] J. J. Li and W. Tan. A single dna molecule nanomotor. *Nano Lett.*, 2:315–318, 2002.
- [29] S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama, and A. J. Turberfield. A dna-based molecular motor that can navigate a network of tracks. *Nat. Nanotechnology*, 7:169–173, 2012.
- [30] S.F.J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, and A. J. Turberfield. Direct observation of stepwise movement of a synthetic molecular transporter. *Nat. Nanotechnology*, 66:166–169, 2011.
- [31] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan. Molecular robots guided by prescriptive landscapes. *Nature*, 465:206–210, 2010.
- [32] D. A. Baum and S. K. Silverman. Deoxyribozymes: useful dna catalysts in vitro and in vivo. *Cell. Mol. Life Sci.*, 65:2156–2174, 2008.
- [33] F. Chen, R. Wang, Z. Li, B. Liu, X. Wang, Y. Sun, D. Hao, and J. Zhang. A novel replicating circular dnazyme. *Nucleic Acids Res.*, 32:2336–2341, 2004.
- [34] Y. Xiao, V. Pavlov, T. Niazov, A. Dishon, M. Kotler, and I. Willner. Catalytic beacons for the detection of dna and telomerase activity. *J. Am. Chem. Soc.*, 126:7430–7431, 2004.
- [35] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems. Self-assembly of a nanoscale dna box with a controllable lid. *Nature*, 459:73–76, 2009.
- [36] S. M. Douglas, I. Bachelet, and G. M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335:831–834, 2012.
- [37] P. R. Callis. Electronic states and luminescence of nucleic acid systems. *Ann. Rev. Phys. Chem.*, 34:329–357, 1983.

- [38] H. Ihmels and D. Otto. Intercalation of organic dye molecules into double-stranded dna - general principles and recent developments. *Top. Curr. Chem.*, 258:161–204, 2005.
- [39] B. A. Armitage. Cyanine dye - dna interactions : Intercalation , groove binding , and aggregation. *Met. Enzym.*, 253:55–76, 2005.
- [40] Life Technologies Corporation:. <http://www.lifetechnologies.com>.
- [41] P. R. Selvin. Principles and biophysical applications of lanthanide-based probes. *Annu. Rev. Biophys. Biomol. Struct.*, 31:275–302, 2002.
- [42] B. N. Giepmans, S. R. Adams, M. H. Ellisman, and R. Y. Tsien. The fluorescent toolbox for assessing protein location and function. *Science*, 312(5771):217–24, 2006.
- [43] I. L. Medintz, H. T. Uyeda, E. R. Goldman, and H. Mattoussi. Quantum dot bioconjugates for imaging, labelling and sensing. *Nature Mater.*, 4:435–446, 2005.
- [44] A. Iqbal, L. Wang, K. C. Thompson, D. M. Lilley, and D. G. Norman. The structure of cyanine 5 terminally attached to double-stranded dna: implications for fret studies. *Biochemistry*, 47(30):7857–62, 2008.
- [45] H. Neubauer, N. Gaiko, S. Berger, J. Schaffer, C. Eggeling, J. Tuma, L. Verdier, C. A. Seidel, C. Griesinger, and A. Volkmer. Orientational and dynamical heterogeneity of rhodamine 6g terminally attached to a dna helix revealed by nmr and single-molecule fluorescence spectroscopy. *J. Am. Chem. Soc.*, 129(42):12746–55, 2007.
- [46] D. G. Norman, R. J. Grainger, D. Uhrin, and D. M. Lilley. Location of cyanine-3 on double-stranded dna: importance for fluorescence resonance energy transfer studies. *Biochemistry*, 39(21):6317–24, 2000.
- [47] M. E. Sanborn, B. K. Connolly, K. Gurunathan, and M. Levitus. Fluorescence properties and photophysics of the sulfoindocyanine cy3 linked covalently to dna. *J. Phys. Chem. B*, 111(37):11064–74, 2007.
- [48] C. R. Sabanayagam, J. S. Eid, and A. Meller. Using fluorescence resonance energy transfer to measure distances along individual dna molecules: corrections due to nonideal transfer. *J. Chem. Phys.*, 122(6):061103, 2005.
- [49] L. M. Wilhelmsson. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.*, 43(2):159–83, 2010.
- [50] R. W. Sinkeldam, N. J. Greco, and Y. Tor. Fluorescent analogs of biomolecular building blocks: design, properties, and applications. *Chem. Rev.*, 110(5):2579–619, 2010.
- [51] M. J. Rist and J. P. Marino. Fluorescent nucleotide base analogs as probes of nucleic acid structure, dynamics and interactions. *Curr. Org. Chem.*, 6(9):775–793, 2002.
- [52] M. E. Hawkins. *Fluorescent nucleoside analogues as DNA probes*. Kluwer Academic/Plenum Publishers, 2003.
- [53] A. Okamoto, Y. Saito, and I. Saito. Design of base-discriminating fluorescent nucleosides. *Photochem. Photobiol. C: Photochem. Rev.*, 6:108–122, 2005.
- [54] J. N. Wilson and E. T. Kool. Fluorescent dna base replacements: Reporters and sensors for biological systems. *Org. Biomol. Chem.*, 4(23):4265–74, 2006.
- [55] U. Asseline. Development and applications of fluorescent oligonucleotides. *Curr. Org. Chem.*, 10(4):491–518, 2006.

- [56] D. Loakes. Nucleotides and nucleic acids; oligo- and polynucleotides. *Organophosphorus Chem.*, 36:232–312, 2007.
- [57] E. Freese. Specific mutagenic effect of base analogues on phage t4. *J. Mol. Biol.*, 1:87–105, 1959.
- [58] D. C. Ward, E. Reich, and L. Stryer. Fluorescence studies of nucleotides and polynucleotides. *J. Biol. Chem.*, 244:1228–1237, 1969.
- [59] C. R. Guest, R. A. Hochstrasser, L. C. Sowers, and D. P. Millar. Dynamics of mismatched base pairs in dna. *Biochemistry*, 30:3271–3279, 1991.
- [60] D. Xu, K. O. Evans, and T. M. Nordlund. Melting and premelting transitions of an oligomer measured by dna base fluorescence and absorption. *Biochemistry*, 33:9592–9599, 1994.
- [61] D. P. Millar. Fluorescence studies of dna and rna structure and dynamics. *Curr. Opin. in Struct. Biol.*, 6:322–326, 1996.
- [62] R. A. Hochstrasser, T. E. Carver, L. C. Sowers, and D. P. Millar. Melting of a dna helix terminus within the active site of a polymerase. *Biochemistry*, 33:11971–11979, 1994.
- [63] K. D. Raney, L. C. Sowers, D. P. Millar, and S. J. Benkovic. A fluorescence-based assay for monitoring helicase activity. *Proc. Nat. Acad. Sci. USA*, 91:6644–6648, 1994.
- [64] M. E. Hawkins. Fluorescent pteridine nucleoside analogs - a window on dna interactions. *Cell Biochem. Biophys.*, 34:257–281, 2001.
- [65] B. Liu and G. C. Bazan. Homogeneous fluorescence-based dna detection with water-soluble conjugated polymers. *Chem Mater.*, 16(23):4467–4476, 2004.
- [66] H. Lu, K. He, and E. T. Kool. ydna: A new geometry for size-expanded base pairs. *Angew. Chem. Int. Ed.*, 43:5834–5836, 2004.
- [67] A. Okamoto, K. Tainaka, and I. Saito. Synthesis and properties of a novel fluorescent nucleobas, naphthopyridopyrimidine. *Tetrahedron Lett.*, 44:6871–6874, 2003.
- [68] A. Okamoto, K. Tainaka, T. Fukuta, and I. Saito. Design of base-discriminating fluorescent nucleoside and its application to t/c snp typing. *J. Am. Chem. Soc.*, 125:9296–9297, 2003.
- [69] A. Okamoto, K. Tainaka, and I. Saito. Clear distinction of purine bases on the complementary strand by a fluorescence change of a novel fluorescent nucleoside. *J. Am. Chem. Soc.*, 125:4972–4973, 2003.
- [70] D. Shin, R. W. Sinkeldam, and Y. Tor. Emissive rna alphabet. *J. Am. Chem. Soc.*, 133:14912–14915, 2011.
- [71] L. M. Wilhelmsson, A. Holmen, P. Lincoln, P. E. Nielson, and B. Norden. A highly fluorescent dna base analogue that forms watson-crick base pairs with guanine. *J. Am. Chem. Soc.*, 123(10):2434–2435, 2001.
- [72] L. M. Wilhelmsson, P. Sandin, A. Holmen, B. Albinsson, P. Lincoln, and B. Norden. Photophysical characterization of fluorescent dna base analogue, tc. *J. Phys. Chem. B*, 107(34):9094–9101, 2003.

- [73] P. Sandin, L. M. Wilhelmsson, P. Lincoln, V. E. Powers, T. Brown, and B. Albinsson. Fluorescent properties of dna base analogue tc upon incorporation into dna–negligible influence of neighbouring bases on fluorescence quantum yield. *Nucleic Acids Res.*, 33(16):5019–25, 2005.
- [74] P. Sandin, P. Lincoln, T. Brown, and L. M. Wilhelmsson. Synthesis and oligonucleotide incorporation of fluorescent cytosine analogue tc: a promising nucleic acid probe. *Nat. Protocols*, 2(3):615–23, 2007.
- [75] P. Sandin, K. Borjesson, H. Li, J. Martensson, T. Brown, L. M. Wilhelmsson, and B. Albinsson. Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent dna base analogue. *Nucleic Acids Res.*, 36(1):157–67, 2008.
- [76] K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Norden, B. Albinsson, and L. M. Wilhelmsson. Dna adopts normal b-form upon incorporation of highly fluorescent dna base analogue tc: Nmr structure and uv-vis spectroscopy characterization. *Nucleic Acids Res.*, 32(17):5087–95, 2004.
- [77] K. Borjesson, J. Tumpane, T. Ljungdahl, L. M. Wilhelmsson, B. Norden, T. Brown, J. Martensson, and B. Albinsson. Membrane-anchored dna assembly for energy and electron transfer. *J. Am. Chem. Soc.*, 131(8):2831–9, 2009.
- [78] S. Preus, K. Borjesson, K. Kilsa, B. Albinsson, and L. M. Wilhelmsson. Characterization of nucleobase analogue fret acceptor tcnitro. *J. Phys. Chem. B*, 114(2):1050–6, 2010.
- [79] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*, volume 3rd. Springer, New York, 3rd edition, 2006.
- [80] Th. Foerster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437:55–75, 1948.
- [81] R. M. Clegg. "Fluorescence resonance energy transfer" in "Fluorescence Imaging Spectroscopy and Microscopy". John Wiley and Sons, New York, 1996.
- [82] L. Stryer and R. P. Haugland. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A.*, 58:719–726, 1967.
- [83] R. E. Dale and J. Eisinger. Intramolecular distances determined by energy transfer. dependence on orientational freedom of donor and acceptor. *Biopolymers*, 13(8):1573–1605, 1974.
- [84] R. E. Dale, J. Eisinger, and W. E. Blumberg. Orientational freedom of molecular probes - orientation factor in intra-molecular energy-transfer. *Biophys. J.*, 26(2):161–193, 1979.
- [85] A. I. Murchie, R. M. Clegg, E. von Kitzing, D. R. Duckett, S. Diekmann, and D. M. Lilley. Fluorescence energy transfer shows that the four-way dna junction is a right-handed cross of antiparallel molecules. *Nature*, 341(6244):763–6, 1989.
- [86] R. M. Clegg, A. I. Murchie, A. Zechel, C. Carlberg, S. Diekmann, and D. M. Lilley. Fluorescence resonance energy transfer analysis of the structure of the four-way dna junction. *Biochemistry*, 31(20):4846–56, 1992.
- [87] GE Healthcare. <http://www.gelifesciences.com>.
- [88] Glen Research. <http://www.glenresearch.com>.
- [89] Integrated DNA Technologies (IDT):. [www.idtdna.com](http://www.idtdna.com).
- [90] ATD Bio:. [www.atdbio.com](http://www.atdbio.com).

- [91] A. Muschielok, J. Andrecka, A. Jawhari, F. Bruckner, P. Cramer, and J. Michaelis. A nano-positioning system for macromolecular structural analysis. *Nat. Methods*, 5(11):965–71, 2008.
- [92] A. Muschielok and J. Michaelis. Application of the nano-positioning system to the analysis of fluorescence resonance energy transfer networks. *J. Phys. Chem. B*, 115(41):11927–37, 2011.
- [93] T. Sabir, G. F. Schroder, A. Toulmin, P. McGlynn, and S. W. Magennis. Global structure of forked dna in solution revealed by high-resolution single-molecule fret. *J. Am. Chem. Soc.*, 133(5):1188–91, 2011.
- [94] T. Sabir, A. Toulmin, L. Ma, A. C. Jones, P. McGlynn, G. F. Schroder, and S. W. Magennis. Branchpoint expansion in a fully complementary three-way dna junction. *J. Am. Chem. Soc.*, 134(14):6280–5, 2012.
- [95] A. K. Wozniak, G. F. Schroder, H. Grubmuller, C. A. Seidel, and F. Oesterhelt. Single-molecule fret measures bends and kinks in dna. *Proc Natl Acad Sci U S A*, 105(47):18337–42, 2008.
- [96] J. Andrecka, R. Lewis, F. Bruckner, E. Lehmann, P. Cramer, and J. Michaelis. Single-molecule tracking of mrna exiting from rna polymerase ii. *Proc. Natl. Acad. Sci. U.S.A.*, 105(1):135–40, 2008.
- [97] J. Andrecka, B. Treutlein, M. A. Arcusa, A. Muschielok, R. Lewis, A. C. Cheung, P. Cramer, and J. Michaelis. Nano positioning system reveals the course of upstream and nontemplate dna within the rna polymerase ii elongation complex. *Nucleic Acids Res.*, 37(17):5803–9, 2009.
- [98] H. Balci, S. Arslan, S. Myong, T. M. Lohman, and T. Ha. Single-molecule nanopositioning: structural transitions of a helicase-dna complex during atp hydrolysis. *Biophys. J.*, 101(4):976–84, 2011.
- [99] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, and C. A. Seidel. Accurate distance determination of nucleic acids via forster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.*, 133(8):2463–80, 2011.
- [100] J. J. Rindermann, Y. Akhtman, J. Richardson, T. Brown, and P. G. Lagoudakis. Gauging the flexibility of fluorescent markers for the interpretation of fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, 133(2):279–285, 2011.
- [101] M. Levitus and S. Ranjit. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q Rev Biophys*, 44(1):123–51, 2011.
- [102] B. J. Harvey and M. Levitus. Nucleobase-specific enhancement of cy3 fluorescence. *J. Fluoresc.*, 19(3):443–8, 2009.
- [103] B. J. Harvey, C. Perez, and M. Levitus. Dna sequence-dependent enhancement of cy3 fluorescence. *Photochem. Photobiol. Sci.*, 8(8):1105–10, 2009.
- [104] J. E. Noble, L. Wang, K. D. Cole, and A. K. Gaigalas. The effect of overhanging nucleotides on fluorescence properties of hybridising oligonucleotides labelled with alexa-488 and fam fluorophores. *Biophys. Chem.*, 113(3):255–63, 2005.
- [105] J. R. Unruh, G. Gokulrangan, G. H. Lushington, C. K. Johnson, and G. S. Wilson. Orientational dynamics and dye-dna interactions in a dye-labeled dna aptamer. *Biophys. J.*, 88(5):3455–65, 2005.

- [106] J. R. Unruh, G. Gokulrangan, G. S. Wilson, and C. K. Johnson. Fluorescence properties of fluorescein, tetramethylrhodamine and texas red linked to a dna aptamer. *Photochem. Photobiol.*, 81(3):682–90, 2005.
- [107] L. Wang, A. K. Gaigalas, J. Blasic, and M. J. Holden. Spectroscopic characterization of fluorescein- and tetramethylrhodamine-labeled oligonucleotides and their complexes with a dna template. *Spectrochim Acta A Mol Biomol Spectrosc*, 60(12):2741–50, 2004.
- [108] G. Vamosi, C. Gohlke, and R. M. Clegg. Fluorescence characteristics of 5-carboxytetramethylrhodamine linked covalently to the 5' end of oligonucleotides: multiple conformers of single-stranded and double-stranded dye-dna complexes. *Bioophys. J.*, 71(2):972–94, 1996.
- [109] E. M. Stennett, G. Kodis, and M. Levitus. Photobleaching and blinking of tamra induced by mn(2+). *Chemphyschem*, 13(4):909–13, 2012.
- [110] E. T. Mollova. Single-molecule fluorescence of nucleic acids. *Curr. Opin. Chem. Biol.*, 6(6):823–8, 2002.
- [111] S. A. McKinney, E. Tan, T. J. Wilson, M. K. Nahas, A. C. Declais, R. M. Clegg, D. M. Lilley, and T. Ha. Single-molecule studies of dna and rna four-way junctions. *Biochem. Soc. Trans.*, 32(Pt 1):41–5, 2004.
- [112] P R Selvin and T Ha. *Single-Molecule Techniques: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, 2008.
- [113] R. Roy, S. Hohng, and T. Ha. A practical guide to single-molecule fret. *Nat. Methods*, 5(6):507–16, 2008.
- [114] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.*, 77:51–76, 2008.
- [115] L. P. Watkins, H. Chang, and H. Yang. Quantitative single-molecule conformational distributions: a case study with poly-(l-proline). *J. Phys. Chem. A*, 110(15):5191–203, 2006.
- [116] I. Gopich and A. Szabo. Theory of photon statistics in single-molecule forster resonance energy transfer. *J. Chem. Phys.*, 122(1):14707, 2005.
- [117] I. V. Gopich and A. Szabo. Single-molecule fret with diffusion and conformational dynamics. *J. Phys. Chem. B*, 111(44):12925–32, 2007.
- [118] I. V. Gopich and A. Szabo. Decoding the pattern of photon colors in single-molecule fret. *J. Phys. Chem. B*, 113(31):10965–73, 2009.
- [119] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, and S. Weiss. Shot-noise limited single-molecule fret histograms: comparison between theory and experiments. *J. Phys. Chem. B*, 110(44):22103–24, 2006.
- [120] M. Antonik, S. Felekyan, A. Gaiduk, and C. A. Seidel. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B*, 110(13):6970–8, 2006.
- [121] S. Kalinin, S. Felekyan, A. Valeri, and C. A. Seidel. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B*, 112(28):8361–74, 2008.
- [122] S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, and C. A. Seidel. Detection of structural dynamics by fret: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J. Phys. Chem. B*, 114(23):7983–95, 2010.

- [123] E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. Seidel. Accurate single-molecule fret studies using multiparameter fluorescence detection. *Methods Enzymol.*, 475:455–514, 2010.
- [124] R. Kuhnemuth and C. A. M. Seidel. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Mol.*, 2(4):251–254, 2001.
- [125] V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. Seidel, and D. C. Lamb. Combining mfd and pie for accurate single-pair forster resonance energy transfer measurements. *Chemphyschem*, 13(4):1060–78, 2012.
- [126] C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, and C. A. Seidel. Analysis of photobleaching in single-molecule multicolor excitation and forster resonance energy transfer measurements. *J. Phys. Chem. A*, 110(9):2979–95, 2006.
- [127] J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken, and C. A. Seidel. Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal. Chem.*, 78(6):2039–50, 2006.
- [128] T. A. Laurence, X. Kong, M. Jager, and S. Weiss. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 102(48):17348–53, 2005.
- [129] B. K. Muller, E. Zaychikov, C. Brauchle, and D. C. Lamb. Pulsed interleaved excitation. *Biophys. J.*, 89(5):3508–22, 2005.
- [130] B. Treutlein, A. Muschielok, J. Andrecka, A. Jawhari, C. Buchen, D. Kostrewa, F. Hog, P. Cramer, and J. Michaelis. Dynamic architecture of a minimal rna polymerase ii open promoter complex. *Mol Cell*, 46(2):136–46, 2012.
- [131] M. P. Foster, C. A. McElroy, and C. D. Amero. Solution nmr of large molecules and assemblies. *Biochemistry*, 46(2):331–40, 2007.
- [132] S. R. Holbrook. Structural principles from large rnas. *Annu. Rev. Biophys.*, 37:445–64, 2008.
- [133] M. Born and J. R. Oppenheimer. Zur quantentheorie der molekeln. *Ann. Phys. (Leipzig)*, 84:457–484, 1927.
- [134] R. M. Clegg. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.*, 211:353–88, 1992.
- [135] F. Jensen. *Introduction to Computational Chemistry 2<sup>nd</sup> Ed.* John Wiley and Sons Ltd, 2007.
- [136] V. Barone and M. Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A*, 102:1995–2001, 1998.
- [137] A. Klamt and G. Schuurman. Cosmo - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.*, 2:799–805, 1993.
- [138] K. Capelle. A bird’s-eye view on density-functional theory. *Braz. J. Phys.*, 36:1318–1343, 2006.
- [139] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

- [140] A. D. Becke. Density-functional thermochemistry .3. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.
- [141] C. T. Lee, W. T. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B*, 37:785–789, 1988.
- [142] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. b-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *J. Phys. Chem.*, 98:11623–11627, 1994.
- [143] M. A. L. Marques and E. K. U. Gross. Time-dependent density functional theory. *Annu. Rev. Phys. Chem.*, 55:427–455, 2004.
- [144] K. Burke, J. Werschnik, and E. K. U. Gross. Time-dependent density functional theory: Past, present and future. *J. Chem. Phys.*, 123:062206, 2005.
- [145] S. Grimme. Calculation of the electronic spectra of large molecules. *Reviews in Computational Chemistry*, 20:153–218, 2004.
- [146] M. R. Silva-Junior, M. Schreiber, S. P. A. Sauer, and W. Thiel. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *J. Chem. Phys.*, 129:104103, 2008.
- [147] A. Rodger and B. Norden. *Circular Dichroism and Linear Dichroism*. Oxford University Press, 1997.
- [148] JC; Lagarias, JA; Reeds, MH; Wright, and PE Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *Siam. J. Optim.*, 9:112–147, 1998.
- [149] MATLAB. *version 8 (R2012b)*. The MathWorks Inc., Natick, Massachusetts, 2012.
- [150] G. Zheng, X. J. Lu, and W. K. Olson. Web 3dna—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, 37(Web Server issue):W240–6, 2009.
- [151] Marcus D Hanwell, Donald E Curtiss, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4:17, 2012.
- [152] ChemBio3D:. [www.cambridgesoft.com/Ensemble\\_for\\_Biology/ChemBio3D/Default.aspx](http://www.cambridgesoft.com/Ensemble_for_Biology/ChemBio3D/Default.aspx).
- [153] VMD Visual Molecular Dynamics:. <http://www.ks.uiuc.edu/Research/vmd/>.
- [154] POV-ray:. <http://www.povray.org/>. Persistence of Vision Pty. Ltd. (2004). Persistence of Vision (TM) Raytracer. Persistence of Vision Pty. Ltd., Williamstown, Victoria, Australia.
- [155] G. M. Church, Y Gao, and S. Kosuri. Next-generation digital information storage in dna. *Science*, 337:1628, 2012.
- [156] P. Wong, K. Wong, and H. Foote. Organic data memory using the dna approach. *Commun. ACM*, 46:95–98, 2003.
- [157] N. G. Portney, Y. Wu, L. K. Quezada, S. Lonardi, and M. Ozkan. Length-based encoding of binary data in dna. *Langmuir*, 24:1613, 2008.
- [158] D. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree. Engineering entropy-driven reactions and networks catalyzed by dna. *Science*, 318:1121–1125, 2007.

- [159] D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree. Engineering entropy-driven reactions and networks catalyzed by dna. *Science*, 318:1121–1125, 2007.
- [160] M. N. Stojanovic, T. E. Mitchell, and D. Stefanovic. Deoxyribozyme-based logic gates. *J. Am. Chem. Soc.*, 124:3555–3561, 2002.
- [161] L. Qian and E. Winfree. Scaling up digital circuit computation with dna strand displacement cascades. *Science*, 332:1196–1201, 2011.
- [162] D. Y. Zhang and J. Seelig. Dynamic dna nanotechnology using strand-displacement reactions. *Nat. Chemistry*, 3:103–113, 2011.
- [163] S. J. Strickler and R. A. Berg. Relationship between absorption intensity and fluorescence lifetime of molecules. *J. Chem. Phys.*, 37:814–822, 1962.
- [164] B. Valeur. *Molecular Fluorescence. Principles and Applications*. Wiley-VCH, Weinheim, 2002.
- [165] E. A. Dethoff, J. Chugh, A. M. Mustoe, and H. M. Al-Hashimi. Functional complexity and regulation through rna dynamics. *Nature*, 482(7385):322–30, 2012.
- [166] J. R. Bothe, E. N. Nikolova, C. D. Eichhorn, J. Chugh, A. L. Hansen, and H. M. Al-Hashimi. Characterizing rna dynamics at atomic resolution using solution-state nmr spectroscopy. *Nat. Methods*, 8(11):919–31, 2011.

# A

## APPENDIX

---

### Contents:

---

**Appendix 1** Paper on base-base FRET published in JACS 2009.

**Appendix 2** Technical note in Glen Research, November 2012.

**Appendix 3** Library of potential fluorescent nucleobase analogues.

---

## Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems

Karl Börjesson,<sup>†</sup> Søren Preus,<sup>‡</sup> Afaf H. El-Sagheer,<sup>§,||</sup> Tom Brown,<sup>§</sup> Bo Albinsson,<sup>†</sup> and L. Marcus Wilhelmsson<sup>\*,†</sup>

*Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom, and Chemistry Branch, Department of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez Canal University, Suez, Egypt*

Received September 2, 2008; E-mail: marcus.wilhelmsson@chalmers.se

**Abstract:** We present the first nucleobase analog fluorescence resonance energy transfer (FRET)-pair. The pair consists of tC<sup>O</sup>, 1,3-diaza-2-oxophenoxazine, as an energy donor and the newly developed tC<sub>nitro</sub>, 7-nitro-1,3-diaza-2-oxophenothiazine, as an energy acceptor. The FRET-pair successfully monitors distances covering up to more than one turn of the DNA duplex. Importantly, we show that the rigid stacking of the two base analogs, and consequently excellent control of their exact positions and orientations, results in a high control of the orientation factor and hence very distinct FRET changes as the number of bases separating tC<sup>O</sup> and tC<sub>nitro</sub> is varied. A set of DNA strands containing the FRET-pair at wisely chosen locations will, thus, make it possible to accurately distinguish distance- from orientation-changes using FRET. In combination with the good nucleobase analog properties, this points toward detailed studies of the inherent dynamics of nucleic acid structures. Moreover, the placement of FRET-pair chromophores inside the base stack will be a great advantage in studies where other (biomacro)molecules interact with the nucleic acid. Lastly, our study gives possibly the first truly solid experimental support to the dependence of energy transfer efficiency on orientation of involved transition dipoles as predicted by the Förster theory.

### Introduction

In the search for improved methods for more accurate and detailed investigations on the structure and dynamics of nucleic acids as well as their interactions with other (biomacro)molecules we present the first base analog fluorescence resonance energy transfer (FRET)-pair. FRET is a technique frequently utilized to detect structural changes in biomacromolecular systems.<sup>1–5</sup> The strong dependence of the transfer efficiency (*E*; see eq 1) between an excited donor (D) and a ground-state acceptor (A) on distance ( $R_{DA}$ )<sup>–6</sup>, makes FRET the obvious choice for monitoring conformational changes and interactions between molecules. The efficiency of energy transfer is also governed by the Förster critical distance,  $R_0$ , (eq 2; at which the *E* is 0.5) which in turn depends on the quantum yield of the donor ( $\phi_D$ ), the donor/acceptor spectral overlap integral ( $J_{DA}$ ), the refractive index of the medium (*n*), and importantly the geometric factor ( $\kappa$ ).<sup>1,2,6</sup>

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

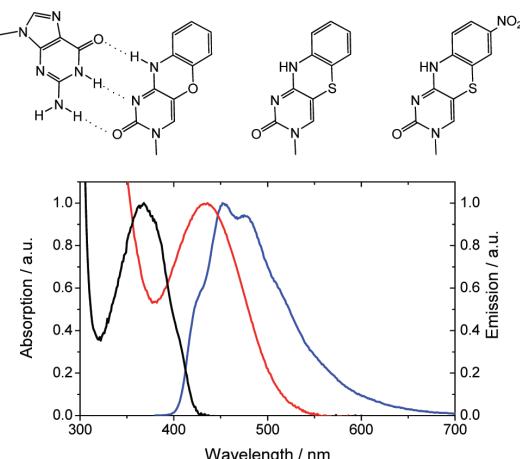
$$R_0 = 0.211 (J_{DA} \kappa^2 n^{-4} \phi_D)^{1/6} \text{Å} \quad (2)$$

The geometric factor takes the direction of the donor and acceptor transition dipoles into consideration and is described by eq 3:

$$\kappa = \mathbf{e}_1 \cdot \mathbf{e}_2 - 3(\mathbf{e}_1 \cdot \mathbf{e}_{12})(\mathbf{e}_{12} \cdot \mathbf{e}_2) \quad (3)$$

where  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are the unit vectors of the donor and acceptor transition dipoles and  $\mathbf{e}_{12}$  the unit vector between their centers. The value of  $\kappa^2$  can range from 0 to 4. Thus, to be able to extract detailed structural information from the measured FRET efficiency, an accurate estimate of  $\kappa^2$  is required. Such estimates of  $\kappa^2$  are rarely available due to the lack of knowledge of orientation of the donor or acceptor molecules themselves and/or their interacting transition dipole moments.<sup>7</sup> The most frequently used (both correctly and incorrectly)  $\kappa^2$  is 2/3, which corresponds to freely rotating donor and acceptor transition dipoles.

- (3) Clegg, R. M.; Murchie, A. I. H.; Zechel, A.; Lilley, D. M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2994–2998.
- (4) Murchie, A. I. H.; Clegg, R. M.; von Kitzing, E.; Duckett, D. R.; Diekmann, S.; Lilley, D. M. J. *Nature* **1989**, *341*, 763–766.
- (5) Stengel, G.; Gill, J. P.; Sandin, P.; Wilhelmsson, L. M.; Albinsson, B.; Nordén, B.; Millar, D. P. *Biochemistry* **2007**, *46*, 12289–12297.
- (6) Dale, R. E.; Eisinger, J. *Biopolymers* **1974**, *13*, 1573–1605.
- (7) Dolghih, E.; Roitberg, A. E.; Krause, J. L. *J. Photochem. Photobiol. A* **2007**, *190*, 321–327.



**Figure 1.** Top: Structure of G-tC° base pair (left), fluorescent cytosine analog tC (middle), and newly developed cytosine analog tC<sub>nitro</sub> (right). Bottom: Representative normalized absorption (black) and emission (blue) spectra of FRET donor tC° and absorption spectrum (red) of virtually nonfluorescent acceptor tC<sub>nitro</sub> within dsDNA showing the donor/acceptor spectral overlap. Measurements performed at 22 °C in 25 mM phosphate buffer (pH 7.5) and [Na<sup>+</sup>] = 100 mM.

When monitoring conformational changes or interaction processes in nucleic acid containing systems using FRET, the most common method is to covalently attach donor and acceptor molecules *via* flexible linkers to two different positions and to assume that  $\kappa^2$  is 2/3. However, many donor/acceptor chromophores interact with the nucleic acid structure<sup>7–9</sup> and, thus, the use of a  $\kappa^2$  of 2/3 is an inaccurate assumption that may result in considerable errors in structural interpretations. Better control of the  $\kappa^2$  in nucleic acid systems was presented in the excellent study by Lewis et al. in which a donor was rigidly attached to one end of the DNA helix resulting in a very good orientation control.<sup>10</sup> However, the acceptor on the opposite end was attached only to one of the strands and had considerable motional freedom. Other recent excellent studies trying to achieve better control of the FRET orientation factor or investigating in detail the nature of energy transfer in nucleic acid systems are those by Iqbal et al.<sup>11</sup> and Hurley et al.<sup>12</sup> In an attempt to achieve the highest possible control of donor/acceptor orientation we here present a novel FRET-pair composed of two cytosine analogs, tC° (1,3-diaza-2-oxophenoxyxazine) and the newly synthesized tC<sub>nitro</sub> (7-nitro-1,3-diaza-2-oxophenothiazine) (Figure 1).

## Results and Discussion

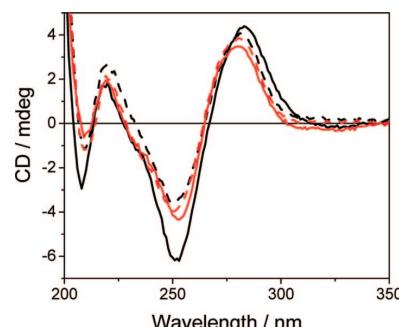
In the design of a nucleic acid base analog FRET-pair, our goal was to utilize tC or tC° (Figure 1). We have previously established that both analogs have a high and stable quantum yield in dsDNA as well as being rigidly stacked within the

- (8) Norman, D. G.; Grainger, R. J.; Uhrin, D.; Lilley, D. M. J. *Biochemistry* **2000**, *39*, 6317–6324.
- (9) Iqbal, A.; Wang, L.; Thompson, K. C.; Lilley, D. M. J.; Norman, D. G. *Biochemistry* **2008**, *47*, 7857–7862.
- (10) Lewis, F. D.; Zhang, L. G.; Zuo, X. B. *J. Am. Chem. Soc.* **2005**, *127*, 10002–10003.
- (11) 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.
- (12) Hurley, D. J.; Tor, Y. *J. Am. Chem. Soc.* **2002**, *124*, 13231–13241.

**Table 1.** DNA Melting Temperatures of tC<sub>nitro</sub>-Containing Duplexes

DNA sequence <sup>a</sup>	base opposite tC <sub>nitro</sub> /C			
	G/C <sup>b</sup>	A/G <sup>b</sup>	C/G <sup>b</sup>	T/A <sup>b</sup>
5'-CGTCYTTGC-3'	47 (45)	32 (21)	28 (20)	32 (27)
5'-CGTTYCTTGC-3'	43 (41)	29 (23)	29 (17)	30 (23)

<sup>a</sup> Y denotes tC<sub>nitro</sub> or a normal C. <sup>b</sup> Temperatures in parenthesis are for the unmodified duplexes.



**Figure 2.** Circular dichroism (CD) spectra of tC<sub>nitro</sub>-modified DNA duplexes and the corresponding unmodified duplexes. Modified duplex with CT neighboring tC<sub>nitro</sub> (black solid line), modified duplex with TC neighboring tC<sub>nitro</sub> (red solid line) and corresponding unmodified duplexes (dashed lines). Measurements performed in 25 mM phosphate buffer (pH 7.5) and [Na<sup>+</sup>] = 100 mM at a duplex concentration of approximately 3.5  $\mu$ M.

duplex and, thus, are excellent donor candidates.<sup>13–16</sup> With the objective of red-shifting the absorption of tC/tC° and maintaining their nucleobase properties we synthesized tC<sub>nitro</sub> (Figure 1) as a FRET acceptor.

Before investigating tC<sub>nitro</sub> as a spectroscopic tool its properties as a cytosine analogue needed to be established. In Table 1 a DNA melting temperature study for duplexes composed of tC<sub>nitro</sub>-containing strands and the corresponding unmodified complements with G, A, C, and T, respectively, opposite tC<sub>nitro</sub>/C is presented.

The duplexes where tC<sub>nitro</sub> is paired with G have melting temperatures that are 13–19 °C higher than when it pairs with A, C, or T on the opposite strand. The corresponding differences for unmodified cytosine are 18–25 °C. Although the differences are slightly lower for the tC<sub>nitro</sub>/G base-pair than for the normal C/G base-pair, this result shows that tC<sub>nitro</sub> is highly selective for base-pairing with guanine. In addition, tC<sub>nitro</sub> increases the melting temperature compared to C by 2 °C in the fully complementary “GC” case. This slight increase in duplex stability is in good agreement with our previous studies on tC and tC°.<sup>13,15</sup>

To further establish the suitability of tC<sub>nitro</sub> as a cytosine analogue when positioned inside the DNA double helix we performed circular dichroism (CD) measurements on the duplexes in Table 1 (see Figure 2). Comparing the modified duplexes (solid lines) to the corresponding unmodified ones

- (13) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. *Nucleic Acids Res.* **2008**, *36*, 157–167.
- (14) Wilhelmsson, L. M.; Holmén, A.; Lincoln, P.; Nielsen, P. E.; Nordén, B. *J. Am. Chem. Soc.* **2001**, *123*, 2434–2435.
- (15) Engman, K. C.; Sandin, P.; Osborne, S.; Brown, T.; Billeter, M.; Lincoln, P.; Nordén, B.; Albinsson, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2004**, *32*, 5087–5095.
- (16) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005**, *33*, 5019–5025.

**Table 2.** Donor and Acceptor 33-mer Oligonucleotides Used

name	sequence <sup>a</sup>
tC <sup>O</sup> 1	5'-CGATCACACAXAAGGACGAGGATAAGGGAGGAGG-3'
tC <sup>O</sup> 2	5'-CGATCACAXACAAGGACGAGGATAAGGGAGGAGG-3'
tC <sup>O</sup> 3	5'-CGATCAACACAAAGGACGAGGATAAGGGAGGAGG-3'
tC <sub>nitro</sub> 1	5'-CCTCCCTCTTATCCTCGTYTTGTGTGATCG-3'
tC <sub>nitro</sub> 2	5'-CCTCCCTCTTATCCTCGTYCTGTGTGATCG-3'
tC <sub>nitro</sub> 3	5'-CCTCCCTCTTATCCTCGTYCTGTGTGATCG-3'
tC <sub>nitro</sub> 4	5'-CCTCCCTCTTATCCTCGTYCTGTGTGATCG-3'

<sup>a</sup> X = tC<sup>O</sup>; Y = tC<sub>nitro</sub>.

(dashed lines) the same overall spectral envelope is found. The general appearance of the CD spectra is that of normal B-form DNA, which is characterized by a positive band centered at 275 nm, a negative band at 250–240 nm, a band that can be either positive or negative at 220 nm, and just below that a narrow negative peak followed by a large positive peak at 190–180 nm.<sup>17</sup> The slight differences that can be seen between the CD of the unmodified and corresponding modified duplex most certainly come as an effect of the differences between the absorption of tC<sub>nitro</sub> compared to cytosine. In conclusion, the CD experiment together with the duplex melting temperature study, as well as the fact that the structurally very similar tC and tC<sup>O</sup> have been shown using CD<sup>13,15</sup> and NMR<sup>15</sup> not to alter the natural form of DNA, indicate that exchanging a cytosine for a tC<sub>nitro</sub> does not perturb the structure of the normal B-form DNA.

The lowest energy absorption maximum of tC<sub>nitro</sub> in dsDNA is centered at approximately 440 nm with an extinction coefficient of 5400 M<sup>-1</sup>cm<sup>-1</sup> (Figure 1). The position of the absorption results in a very good spectral overlap with the emission of tC<sup>O</sup> (Figure 1) which is centered at 465 nm and a smaller overlap with the emission of tC (Emission<sub>max</sub> = 505 nm, not shown). In combination with the quantum yield of tC<sup>O</sup> in dsDNA and a refractive index of 1.4 in DNA,<sup>1</sup> the R<sub>0</sub> of the tC<sup>O</sup>–tC<sub>nitro</sub> FRET-pair is estimated to be 27 Å when using a κ<sup>2</sup> of 2/3 (the use of 2/3 here is only to facilitate comparison with R<sub>0</sub> of common FRET-pairs). The 33-mer oligonucleotides utilized for the measurements in this study are presented in Table 2.

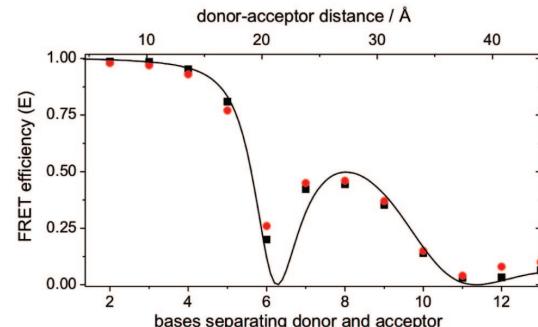
Except for the FRET donor position, the three tC<sup>O</sup> sequences are the same and are complementary to the tC<sub>nitro</sub> sequences, which have the FRET acceptor at four different positions. The positions of the donor and acceptor in the seven sequences are chosen so that every separation between 2 and 13 bases can be monitored combining the strands. Furthermore, the sequences are designed so that tC<sup>O</sup> has the same surrounding bases and the donor and acceptor are situated far from the more dynamic ends of the duplex.

To study the change in the tC<sup>O</sup>–tC<sub>nitro</sub> FRET efficiency for the 12 different separations, both steady-state and time-resolved fluorescence measurements were performed (Table 3). In both cases the results (Figure 3) show an efficiency that is highly dependent on both distance and orientation as the separation and, thus, the direction of the transition dipoles of the base analogs are altered in a stepwise fashion. The data suggest that we have successfully designed an excellent nucleic acid base analog FRET-pair. In the time-resolved measurements tC<sup>O</sup> exhibits single exponential fluorescence decay for most sequences. However, the most quenched sequences need two

**Table 3.** Lifetimes and Steady-State Quenching Data as well as Fitted FRET Efficiency

bases in between	$\tau_1 (\alpha_1)$ /ns	$\tau_2 (\alpha_2)$ /ns	$\langle \tau \rangle^a$ /ns	$\chi^2$	$1 - \langle \tau \rangle / \tau_0$	$1 - II_0$	E <sup>b</sup>
2	0.05 (0.96)	0.48 (0.04)	0.07	1.24	0.99	0.98	1.00
3	0.05 (0.96)	0.54 (0.04)	0.07	1.31	0.99	0.97	0.99
4	0.13 (0.65)	0.38 (0.35)	0.22	1.14	0.95	0.93	0.98
5	1.11 (0.68)	0.36 (0.32)	0.87	1.62	0.81	0.77	0.89
6	3.64 (1)		3.64	1.28	0.20	0.26	0.18
7	2.63 (1)		2.63	1.57	0.42	0.45	0.39
8	2.65 (1)		2.65	1.58	0.44	0.46	0.55
9	3.09 (1)		3.09	1.30	0.35	0.37	0.43
10	3.91 (1)		3.91	1.14	0.14	0.15	0.17
11	4.41 (1)		4.41	1.11	0.03	0.04	0.01
12	4.40 (1)		4.40	1.13	0.03	0.08	0.02
13	4.26 (1)		4.26	1.21	0.06	0.10	0.05

<sup>a</sup>  $\langle \tau \rangle = \alpha_1 \tau_1 + \alpha_2 \tau_2$ . <sup>b</sup> Fitted FRET efficiency.

**Figure 3.** Efficiency of energy transfer for the base analog FRET-pair tC<sup>O</sup>–tC<sub>nitro</sub> estimated using decreases in tC<sup>O</sup>, donor, emission (red circles), and tC<sup>O</sup> average emission lifetimes (black squares) as the two analogs are separated by 2 to 13 bases in a DNA duplex. Curve fitting using eqs 6 and 7 with  $\alpha$  and J<sub>DA</sub> as fit parameters is shown as solid line. Excitation wavelength 370 nm. Measurements performed at 22 °C in 25 mM phosphate buffer (pH 7.5) and [Na<sup>+</sup>] = 100 mM.

exponential components in order to explain the fluorescence decay. There might be several reasons for the nonexponential decay such as: (1) difficulties in measuring highly quenched fluorophores on this short time scale where small amounts of unquenched fluorophores or scattered light might disturb the experiment or (2) distribution of donor–acceptor distances and orientations leading to a distribution of energy transfer efficiencies.

Qualitatively, the data has an appearance exactly as expected for a FRET-pair situated at different positions within a DNA duplex: E decreases sharply with distance while oscillating between local maxima and minima as the transition dipoles of the donor and acceptor change between more parallel and more perpendicular configurations (eq 3). To analyze the measured efficiencies quantitatively we use eqs 6 and 7. These equations take into consideration vector distance and accurate orientations between chromophores and become increasingly similar to the rough model in which the chromophores are placed on top of each other along the DNA helix axis, with increasing base separation. The excellent fit, where J<sub>DA</sub> and  $\alpha$  are varied, to the experimental data and the distinct changes between maxima and minima not only confirm that the  $\phi_D$  and the overlap integral (i.e., donor emission profile and  $\varepsilon_A$ ) are virtually constant, but also gives further evidence that these C-analogs have practically no dynamics on the time-scale of fluorescence; however, faster dynamics cannot be ruled out. The fit is quite insensitive to the magnitude of the overlap integral and the fitted value ( $2 \cdot 10^{14}$

(17) Rodger, A. and Nordén, B. *Circular Dichroism and Linear Dichroism*; Oxford University Press, Oxford, NY, 1997.

$M^{-1} \text{ cm}^{-1} \text{ nm}^4$ ) is close to the one estimated from eq 4 using the spectroscopic properties of the donor and acceptor ( $1.2 \cdot 10^{14} M^{-1} \text{ cm}^{-1} \text{ nm}^4$ ).

From the phase angle parameter,  $\alpha$ , in the curve fitting we also find that the direction of the  $S_1-S_0$  transition of  $tC_{\text{nitro}}$  is rotated  $67^\circ$  compared to that of  $tC^0$  within their three-ring systems. This is in good agreement with the values obtained for  $tC^0$  ( $-33^\circ$ ; anticlockwise from molecule long-axis)<sup>13</sup> and  $tC_{\text{nitro}}$  ( $+25^\circ$ , manuscript in preparation) and proves the high potential of this FRET-pair in detailed structure probing. The slight difference between the curve fitted and the calculated angle of the  $S_1-S_0$  transitions ( $67^\circ$  vs  $58^\circ$ ) likely comes as an effect of the model fitting, in combination with small errors in the experimental determination of the transition dipole orientations, rather than any significant changes in orientations when the base analogs are positioned within the base stack compared to their monomeric forms. We have previously performed circular dichroism experiments on homodimers of  $tC$  as well as  $tC^0$  that are separated by 0–2 bases in DNA duplexes to examine the excitonic effect and found minor effects for the case where the homodimer is separated only by 0 bases (data not shown). The fact that the observed exciton interaction is small most likely comes as an effect of the fairly low extinction coefficients (oscillator strengths) of the analogs. In the current study we have a case where  $tC^0$  and  $tC_{\text{nitro}}$  are separated by at least 2 bases, the extinction coefficients are fairly low and, furthermore, the energy of the  $S_1-S_0$  transitions of  $tC^0$  and  $tC_{\text{nitro}}$  is different (heterodimer) as is the energy of the  $S_1-S_0$  transitions of  $tC^0$  and  $tC_{\text{nitro}}$  compared to the normal bases. This suggests that the direction of the transition dipoles of  $tC^0$  and  $tC_{\text{nitro}}$  are not substantially affected due to interactions between them or between them and the surrounding bases.

## Conclusions

In conclusion, we have designed the first nucleic acid base analog FRET-pair. As a consequence of both the analogs being rigidly located within the base stack, this system enables very high control of the orientation factor. A set of strands containing our FRET-pair at strategically chosen positions, that is, where the slopes are steep in Figure 3, will, thus, make it possible to accurately distinguish distance- from orientation-changes using FRET. In combination with the favorable base pairing properties this will facilitate detailed studies of the inherent dynamics of nucleic acid structures. Moreover, the placement of FRET-pair chromophores inside the base stack will be a great advantage in studies where other (biomacro)molecules interact with the nucleic acid. Lastly, our study gives possibly the first truly solid experimental support to the dependence of  $E$  on orientation of involved transition dipoles as predicted by the Förster theory.

## Experimental Section

**Synthesis of Nucleoside Building Blocks  $tC^0$  and  $tC_{\text{nitro}}$ .** Unless stated otherwise, all reagents were obtained from commercial suppliers and used without further purification: DCM, pyridine and DIPEA were purified by distillation (over calcium hydride). Synthesis of  $tC^0$ , 5-nitro-2-amino-thiophenol and 2-deoxy-3,5-di-*O*-*p*-toluoyl- $\alpha$ -D-erythro-pentofuranosyl was done according to literature procedures.<sup>13,18–20</sup> Deoxygenation of reaction mixtures

was achieved by bubbling nitrogen through the solution for 30 min. Column chromatography was performed using silica gel (Matrex, LC 60Å/35–70  $\mu\text{m}$ ).  $^1\text{H}$  (400 MHz) and  $^{13}\text{C}$  (100.6 MHz) NMR spectra were recorded at room temp. in  $\text{CDCl}_3$  or  $(\text{CD}_3)_2\text{SO}$  using a Jeol Eclipse 400 NMR spectrometer. Chemical shifts are reported relative to residual  $\text{CHCl}_3$  or  $(\text{CH}_3)_2\text{SO}$  ( $\delta = 7.26$  or 2.54 ppm) for  $^1\text{H}$  NMR and ( $\delta = 77.23$  or 40.45 ppm) for  $^{13}\text{C}$  NMR, respectively.  $^{31}\text{P}$  NMR spectrum was recorded on a Bruker AV300 spectrometer at 121 MHz and was externally referenced to 85% phosphoric acid in deuterated water. High-resolution mass spectrum was recorded using the electrospray technique on a Bruker APEX III FT-ICR mass spectrometer. Low-resolution mass spectra were recorded using the electrospray technique on a Fisons VG platform instrument or a Waters ZMD quadrupole mass spectrometer in acetonitrile (HPLC grade). Elemental analyses were performed by H. Kolbe Mikroanalytisches laboratorium, Mülheim an der Ruhr, Germany.

**3:** Degassed NaOH (aq) (0.25 M, 95 mL) was added to a mixture of 5-nitro-2-amino-thiophenol (5.55 g, 32.6 mmol) and 5-bromouracil (6.46 g, 33.8 mmol) under argon and was allowed to reflux for 24 h. The crude product was allowed to cool and was subsequently filtered off. No more purification was done.

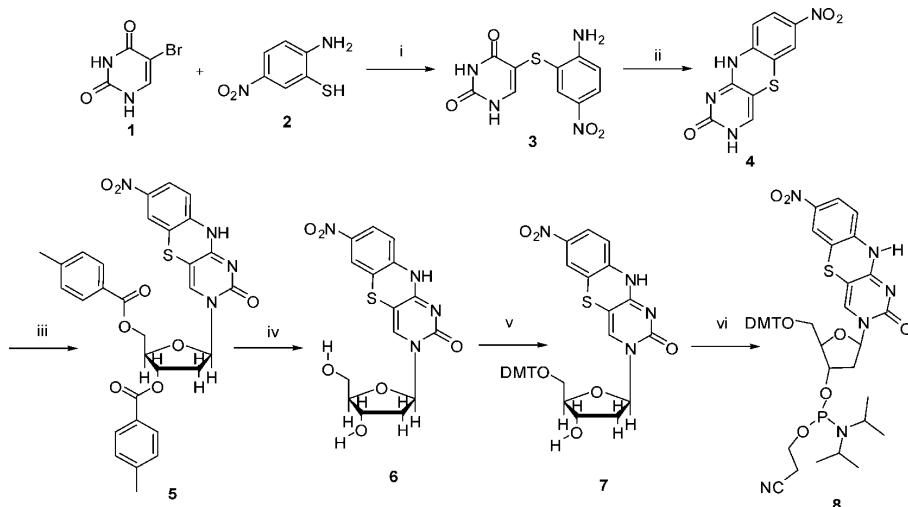
**4:** To the crude product **3** was added EtOH (940 mL) and concentrated HCl (37%, 64 mL). The reaction mixture was refluxed for 24 h whereafter it was allowed to cool down and subsequently filtered off. The filter cake was slurried up in  $\text{NH}_4$  (aq) (8%, 50 mL) at 60 °C for 10 min, cooled down, and filtered off. This was repeated once. The filter cake was washed with water, DMSO, and finally MeOH, resulting in a red insoluble powder (1.3 g, 5 mmol, 15% over two steps). Elemental analysis calculated for  $[\text{C}_{10}\text{H}_6\text{N}_4\text{O}_3\text{S}]$ : C, 45.80; H, 2.31, found: C, 45.97; H, 2.36.

**5:** DMF (15 mL) was added to a mixture of **4** (606 mg, 2.31 mmol) and NaH (60% in mineral oil, 101 mg, 2.54 mmol) under argon. The reaction mixture was left for 1 h where after toluene (15 mL) was added. 2-deoxy-3,5-di-*O*-*p*-toluoyl- $\alpha$ -D-erythro-pentofuranosyl (1.04 g, 2.67 mmol) was added portion wise to the blue reaction mixture for 1 h whereafter it was left overnight. Ethylacetate was added to the mixture whereafter the mixture was filtrated and the filtrate was washed twice with water where upon the solvent was removed in vacuo. Chromatography ( $\text{SiO}_2$ , 1–1.5% MeOH in  $\text{CH}_2\text{Cl}_2$ ) yielded a yellow-red solid (160 mg, 0.26 mmol, 11%);  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta = 7.88$ –8.5 (m, 5H), 7.76 (d, 1H), 7.51 (s, 1H), 7.42 (d, 1H), 7.2–7.3 (m, 4H), 6.32 (dd, 1H), 5.61 (d, 1H), 4.83 (dd, 1H), 4.6–4.7 (m, 2H), 2.94 (dd, 1H), 2.43 (s, 3H), 2.38 (s, 3H), 2.25 (m, 1H) ppm;  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta = 166.2$ , 160.5, 154.1, 144.7, 144.2, 141.6, 134.7, 130.3, 129.9, 129.7, 129.6, 129.4, 129.2, 126.5, 123.5, 121.4, 118.4, 118.3, 96.1, 87.5, 83.9, 75.1, 64.1, 39.5, 21.8 ppm; Elemental analysis calculated for  $[\text{C}_{31}\text{H}_{26}\text{N}_4\text{O}_8\text{S}]$ : C, 60.58; H, 4.26, found: C, 60.52; H, 4.31.

**6:** MeONa (40 mM in MeOH, 25 mL) was added to **5** (137 mg, 0.223 mmol) under argon. The reaction mixture was left overnight whereafter it was neutralized with acetic acid and the solvent was removed in vacuo. Chromatography ( $\text{SiO}_2$ , 10–13% MeOH in  $\text{CH}_2\text{Cl}_2$ ) yielded a yellow-red solid (60 mg, 0.16 mmol, 71%);  $^1\text{H}$  NMR ( $(\text{CD}_3)_2\text{SO}$ )  $\delta = 11$  (br s, 1H), 7.91–8.03 (m, 3H), 7.03 (d, 1H), 6.09 (t, 1H), 5.28 (br s, 1H), 5.15 (br s, 1H), 4.25 (m, 1H), 3.83 (q, 1H), 3.67 (dd, 1H), 3.58 (dd, 1H), 2.21 (m, 1H), 2.07 (m, 1H) ppm;  $^{13}\text{C}$  NMR ( $(\text{CD}_3)_2\text{SO}$ )  $\delta = 159.66$ , 154.41, 143.67, 137.30, 124.33, 122.25, 118.70, 117.56, 94.22, 88.53, 86.62, 70.70, 61.71, 41.50 ppm; Elemental analysis calculated for  $[\text{C}_{15}\text{H}_{14}\text{N}_4\text{O}_6\text{S}]$ : C, 47.62; H, 3.73, found: C, 47.48; H, 3.70.

**7:** Freshly distilled pyridine (5 mL) was added to **6** (222 mg, 0.587 mmol), DMAP (4 mg, 0.03 mmol) and DMT-Cl (240 mg, 0.71 mmol) under argon. The reaction mixture was left overnight whereafter it was quenched with a small amount of  $\text{NaHCO}_3$  (aq, 5%) and the solvent was evaporated in vacuo. The crude product was dissolved in DCM and washed once with 5%  $\text{NaHCO}_3$  (aq) and twice with  $\text{H}_2\text{O}$ . Chromatography ( $\text{SiO}_2$ , 1.5–3% MeOH in  $\text{CH}_2\text{Cl}_2$ ) yielded a yellow-red solid (297 mg, 0.436 mmol, 74%);

- (18) Lin, K. Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995**, *117*, 3873–3874.
- (19) Rolland, V.; Kotera, M.; Lhomme, J. *Synth. Commun.* **1997**, *27*, 3505–3511.
- (20) Chedekel, M. R.; Sharp, D. E.; Jeffery, G. A. *Synth. Commun.* **1980**, *10*, 167–173.

Scheme 1<sup>a</sup>

<sup>a</sup> Reaction conditions: (i)  $\text{NaOH}_{(\text{aq})}$ , 24 h, reflux; (ii)  $\text{EtOH}$ ,  $\text{HCl}$ , 24 h, reflux, 15% over two steps; (iii)  $\text{DMF}$ , toluene, 3,5-di- $O$ -*p*-toloyl- $\alpha$ -D-erythropentofuranosyl,  $\text{NaH}$ , 18 h, rt, 11%; (iv)  $\text{MeONa}$ ,  $\text{MeOH}$ , 18 h, rt, 71%; (v) pyridine, DMAP,  $\text{DMT-Cl}$ , 18 h, rt, 74%; (vi)  $\text{DCM}$ , DIPEA, 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, 1 h, rt, 93%.

<sup>1</sup>H NMR ( $\text{CDCl}_3$ )  $\delta$  = 10.45 (br s, 1H), 7.88 (s, 1H), 7.80 (d, 1H), 7.51 (s, 1H), 7.10–7.45 (m, 10H), 6.83 (m, 4H), 6.37 (t, 1H), 4.68 (m, 1H), 4.17 (m, 2H), 3.74 (s, 3H), 3.71 (s, 3H), 3.41 (d, 1H), 3.33 (d, 1H), 2.88 (m, 1H), 2.34 (m, 1H) ppm; <sup>13</sup>C NMR ( $\text{CDCl}_3$ )  $\delta$  = 160.51, 158.79, 155.20, 144.38, 143.89, 141.82, 136.15, 135.76, 135.39, 130.17, 130.05, 128.21, 128.12, 127.20, 123.21, 121.24, 118.56, 117.78, 113.50, 96.41, 87.43, 87.21, 86.92, 72.10, 63.46, 55.33, 42.32 ppm; Elemental analysis calculated for  $[\text{C}_{36}\text{H}_{32}\text{N}_4\text{O}_8\text{PS}]$ : C, 63.52; H, 4.74, found: C, 63.59; H, 4.67.

**8:** Dry **7** (0.29 mg, 0.43 mmol) was dissolved in  $\text{DCM}$  (10.0 mL) under an atmosphere of argon and DIPEA (0.184 mL, 1.06 mmol) was added. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (0.114 mL, 0.51 mmol) was then added dropwise and after that the reaction mixture was stirred at room temperature for 1 h then transferred under argon into a separating funnel containing degassed  $\text{DCM}$  (20.0 mL). The mixture was washed with degassed saturated aqueous  $\text{KCl}$  (20.0 mL) and the organic layer was separated, dried over sodium sulfate, filtered, and the solvent was removed in vacuo. The phosphoramidite product was dried under vacuum, dissolved in  $\text{DCM}$  (3 mL), and precipitated from hexane (200 mL) at room temperature to give the title compound **8** (the phosphoramidite of 7-nitro-1,3-diaza-2-oxophenothiazine) as an orange precipitate (0.35 g, 93%);  $\delta$  (300 MHz,  $\text{CDCl}_3$ ) 148.21 and 149.14;  $m/z$  LRMS [ES<sup>+</sup>, MeCN] 903 ( $\text{M} + \text{Na}^+$ , 10%); HRMS ( $\text{M} + \text{Na}^+$ ) ( $\text{C}_{45}\text{H}_{49}\text{N}_6\text{NaO}_9\text{PS}$ ) calc., 903.2912, found: 903.2895.

**Oligonucleotide Synthesis.** Standard DNA phosphoramidites, solid supports and additional reagents were purchased from Link Technologies or Applied Biosystems Ltd. Disposable Sephadex NAP columns were purchased from GE Healthcare. All oligonucleotides were synthesized on an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard 0.2  $\mu\text{Mole}$  phosphoramidite cycle of acid-catalyzed detritylation, coupling, capping, and iodine oxidation. Stepwise coupling efficiencies and overall yields were determined by the automated triityl cation conductivity monitoring facility and in all cases were >98.0%. All 2-cyanoethyl phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M. The coupling times were 25 s for normal (A,G,C,T) monomers and 10 min for the modified phosphoramidite monomer. Cleavage of oligonucleotides from the solid support and deprotection was achieved by exposure to concentrated aqueous ammonia solution for 60 min at room temperature followed by heating in a sealed tube for 5 h at 55 °C. For

details in RP-HPLC analysis and purification of oligonucleotides see Supporting Information.

**Photophysical Measurements.** All measurements were made at 22 °C in a phosphate buffer at pH 7.5 in total sodium- and phosphate ion concentrations of 100 mM and 25 mM, respectively. Double stranded concentrations were 2  $\mu\text{M}$  or 9  $\mu\text{M}$  (the higher concentration used for time-resolved measurements of highly quenched sequences). An excess of the acceptor strand were used in all experiments (to ensure complete hybridization). Absorption spectra were recorded from 200 to 600 nm on a Varian Cary 4000 spectrophotometer. The sequences used in the study are 5'-CGA TCA XAX AXA AYY ACG AYY ATA AGG AGG AGG -3', where X is C which can be substituted by a tC<sup>O</sup> and Y is a G where the C on the complementary strand can be substituted by a tC<sub>nitro</sub>. Combination of singly substituted strands results in duplexes with distances ranging from 2 to 13 base pairs separating the tC<sup>O</sup> and tC<sub>nitro</sub> (10 to 48 Å). The extinction coefficient of the tC<sub>nitro</sub> nucleoside was determined by measuring the absorption of samples of known concentration. Samples were prepared by weighing out small amounts of the tC<sub>nitro</sub> nucleoside, typically 1 mg, and dissolving them in known volumes of MQ water (Millipore). The extinction coefficient was determined as an average of three measurements.

Steady state fluorescence was measured on a Spex Fluorolog 3 spectrofluorimeter (JY Horiba). The emission spectra were recorded from 380 to 800 nm with the excitation wavelength fixed at 370 nm.

Fluorescence lifetimes were determined using time-correlated single photon counting. The excitation pulse was provided by a Tsunami Ti:Sapphire laser (Spectra-Physics; 80 MHz) which was pumped by a Millenia Pro X (Spectra-Physics). The Tsunami output at 740 nm was acousto-optically pulse-picked to 4 MHz by a pulse selector (Spectra Physics) when needed and subsequently frequency-doubled yielding an excitation wavelength of 370 nm. The photons were collected by a thermoelectrically cooled micro channel-plate photomultiplier tube (MCP-PMT R3809U-50; Hamamatsu) and fed into a multichannel analyzer with 4096 channels. A minimum of 10 000 counts were recorded in the top channel. The fluorescence decay curves were fitted to exponential expressions by the program FluoFit Pro v.4 (PicoQuant GmbH). The sample response was monitored through a monochromator at 460 ± 16 nm.

UV absorption DNA melting studies were performed on 10-mer oligonucleotides (for sequences see Table 1 or Supporting Informa-

tion) at a concentration of approximately  $4 \mu\text{M}$  using a Varian Cary 4000 spectrophotometer equipped with a programmable multicell temperature block. The samples were heated from 10 to  $80^\circ\text{C}$  at a maximum rate of  $0.2^\circ\text{C min}^{-1}$  whereupon they were cooled to  $10^\circ\text{C}$  at the same rate. The absorption at  $260 \text{ nm}$  was measured with a temperature interval of  $0.5^\circ\text{C}$ . Melting temperatures were determined using the maximum of the derivatives.

Circular dichroism spectra were recorded on 10-mer oligonucleotides (for sequences see Table 1 or Supporting Information) at a concentration of approximately  $3.5 \mu\text{M}$  using a Jasco J-810 spectropolarimeter at  $20^\circ\text{C}$ . Spectra were recorded between 200 and  $500 \text{ nm}$ .

**Theoretical Evaluation.** The energy transfer efficiency for the steady state case as well as for the time-resolved case was determined using eq 4:

$$E = 1 - \frac{F}{F_0} = 1 - \frac{\langle \tau \rangle}{\tau_0} \quad (4)$$

where  $E$  is the energy transfer efficiency,  $F$  and  $F_0$  are the integrated emission intensities of the donor in presence and absence of acceptor, respectively, and  $\tau$  and  $\tau_0$  are the donor lifetimes in presence and absence of acceptor, respectively. The expected energy transfer efficiencies were calculated using eq 1 and the Förster distances,  $R_0$ , were calculated using eq 2 with the refractive index ( $n$ ) and donor quantum yield ( $\phi_D$ ) set to 1.4 and 0.23, respectively.<sup>1,13</sup> The spectral overlap integral ( $J_{DA}$ ) was determined using eq 5:

$$J_{DA} = \int_0^{\infty} F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda \quad (5)$$

where  $F_D$  is the normalized donor emission and  $\epsilon_A$  is the acceptor molar absorptivity. The orientation factor,  $\kappa$ , was calculated using eq 6:<sup>21</sup>

(21) Carlsson, C.; Larsson, A.; Björkman, M.; Jonsson, M.; Albinsson, B. *Biopolymers* **1997**, *41*, 481–494.

$$\kappa_{DA} = \cos(n_{DA}\beta + \alpha) - 3 \left( \frac{a \cdot \sin(n_{DA}\beta + \alpha)}{R_{DA}} \right)^2 \quad (6)$$

where  $n_{DA}$  is the number of base pairs in between the donor and acceptor,  $a$  is the distance between the center of the DNA helix to the center of the chromophore (4 Å),  $R_{DA}$  is the donor acceptor distance,  $\alpha$  is a fitted phase angle, and  $\beta$  is the helical rise angle ( $34.3^\circ$  /base pair). The donor acceptor distance (in Å) is calculated using eq 7:<sup>21</sup>

$$R_{DA} = \sqrt{2a^2(1 - \cos(n_{DA}\beta + \alpha)) + (b(n_{DA} + 1))^2} \quad (7)$$

where  $b$  is the helical rise (3.4 Å /base pair). An in-house made MATLAB program was used to fit the data from the lifetime measurements with respect to the phase angle and the overlap integral. The phase angle is defined as the angle between the transition dipole moments of the donor on one strand and the acceptor on the other, looking along the DNA helix long-axis and when there are no bases separating the donor and the acceptor (i.e., the acceptor is the neighboring base of the guanine that base-pairs with the donor). Considering how the donor and acceptor are oriented in the helix, the phase angle can be translated to an angle describing the difference in the orientation of the transition dipole moments of the donor and acceptor within their three-ring systems.

**Acknowledgment.** Prof. Kristine Kilså at the University of Copenhagen is acknowledged for fruitful discussions. This research is supported by the Swedish Research Council.

**Supporting Information Available:** Details in RP-HPLC analysis and purification of oligonucleotides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA806944W

## TECHNICAL BRIEF: FRETmatrix – A METHOD FOR SIMULATION AND ANALYSIS OF FRET IN OLIGOS

Søren Preus

Department of Chemistry

University of Copenhagen

Copenhagen, DK-2100, Denmark

Marcus Wilhelmsson

Department of Chemical and Biological

Engineering/Physical Chemistry

Chalmers University of Technology

Kemivägen 10, SE-412 96 Göteborg, Sweden

The recent introduction of nucleobase analogues tC, tC<sup>o</sup> and tC<sub>nitro</sub> (Figure 1) into the arsenal of commercial FRET probes has provided researchers with a highly useful alternative to traditional, external dyes when studying the structure, dynamics and function of nucleic acids. Since not only the interpair-distance but also the probe orientation play an important role in the energy transfer process, FRET between rigidly attached nucleobase analogues (base-base FRET) can be exploited to obtain both positional and orientational information about nucleic acid geometries without complications associated with linker flexibility. The design and interpretation of base-base FRET experiments has up till now been challenging to a large majority of researchers.

Addressing this issue, Wilhelmsson and co-workers<sup>1</sup> have developed a freely downloadable software package (FRETmatrix) divided into two parts:

(1) the design and (2) the analysis of base-base FRET experiments.

One part of the software simulates theoretical FRET efficiencies between probes positioned in any 3D nucleic acid structure. To significantly facilitate the design and identify optimal positions of the FRET-probes in oligonucleotides of base-base FRET experiments, the software is used to predict theoretical signal changes between all possible donor-acceptor positions in a nucleic acid sequence, e.g., upon binding of a protein or some other structural change (Figure 2).

To analyse base-base FRET experiments quantitatively multiple time-resolved fluorescence decays can be analysed simultaneously in the second part of the freeware. This novel, rigorous analysis may provide insight into global and/or local 3D structural features, such as the geometry of kinked DNA or the position, orientation and dynamics at specific base positions. The MATLAB-based software is freely available from the following link:

FIGURE 1: TRICYCLIC FLUORESCENT CYTIDINE ANALOGUE PHOSPHORAMIDITES

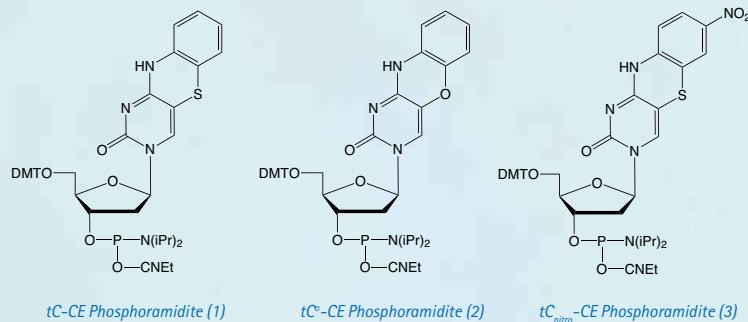
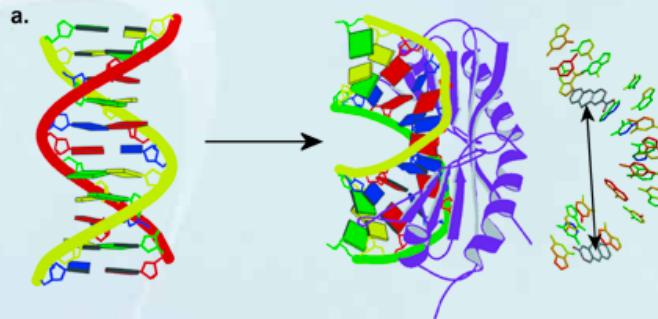


FIGURE 2: EXAMPLE OF A BASE-BASE FRET EXPERIMENT SIMULATED USING FRETmatrix



a. The geometry of DNA upon binding of the TATA-binding protein (TBP)

b. The predicted change in FRET efficiency at selected donor-acceptor positions.

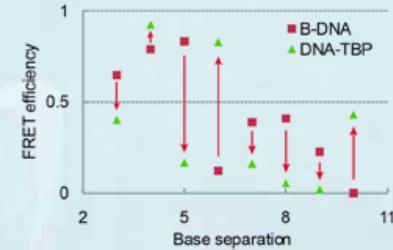
<http://www.chalmers.se/chem/EN/divisions/physical-chemistry/staff/marcus-wilhelmsson/fretmatrix>.

### Reference:

1. S. Preus, K. Kilsa, F.A. Miannay, B. Albinsson, and L.M. Wilhelmsson, *Nucleic Acids Res.*, 2012.

### ORDERING INFORMATION

Item	Catalog No.	Pack	Price(\$)
tC-CE Phosphoramidite	10-1516-95	50 μmole	250.00
	10-1516-90	100 μmole	490.00
	10-1516-02	0.25g	1460.00
tC <sup>o</sup> -CE Phosphoramidite	10-1517-95	50 μmole	250.00
	10-1517-90	100 μmole	490.00
	10-1517-02	0.25g	1460.00
tC <sub>nitro</sub> -CE Phosphoramidite	10-1518-95	50 μmole	265.00
	10-1518-90	100 μmole	520.00
	10-1518-02	0.25g	1460.00

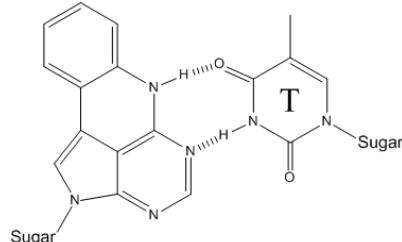


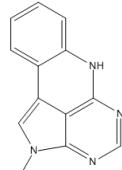
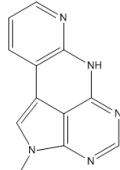
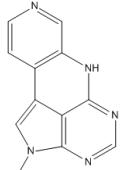
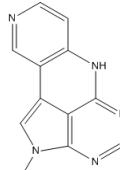
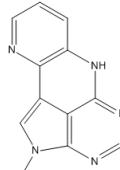
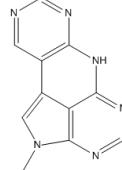
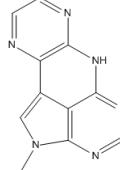
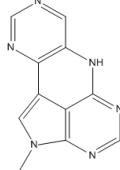
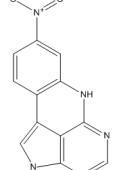
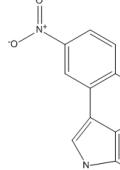
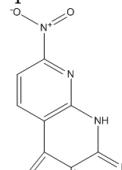
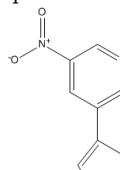
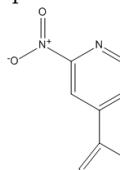
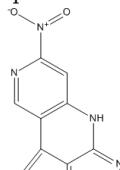
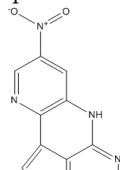
# Library of potential nucleobase analogue structures

---

Three lowest energy electronic transitions were predicted by TDDFT B3LYP/6-311+G(2d) calculations on either the B3LYP/6-31G\*\* or AM1 optimized geometries. All spectra of nitro-aromatics have been predicted *without* solvation, while the transitions of most non-nitro compounds have been predicted *with* a CPCM solvation model for H<sub>2</sub>O unless otherwise stated. *f* denotes oscillator strength.

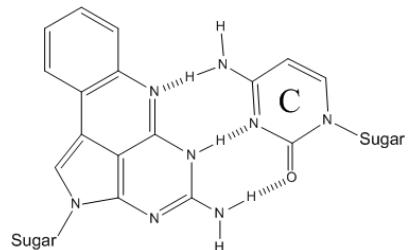
## ADENINE ANALOGUES



qA1		qA2		qA3		qA4		qA5		qA6	
											
350 nm	<i>f</i> =0.060	359 nm	<i>f</i> =0.20	360 nm	<i>f</i> =0.041	359 nm	<i>f</i> =0.035	355 nm	<i>f</i> =0.30	368 nm	<i>f</i> =0.088
330 nm	<i>f</i> =0.20	341 nm	<i>f</i> =0.083	337 nm	<i>f</i> =0.22	324 nm	<i>f</i> =0.18	347 nm	<i>f</i> =0.003	339 nm	<i>f</i> =0.17
283 nm	<i>f</i> =0.10	281 nm	<i>f</i> =0.091	277 nm	<i>f</i> =0.073	279 nm	<i>f</i> =0.082	283 nm	<i>f</i> =0.10	279 nm	<i>f</i> =0.051
											
388 nm	<i>f</i> =0.31	346 nm	<i>f</i> =0.003	346 nm	<i>f</i> =0.14	496 nm	<i>f</i> =0.12	478 nm	<i>f</i> =0.09	339 nm	<i>f</i> =0.17
339 nm	<i>f</i> =0.17	298 nm	<i>f</i> =0.0045	287 nm	<i>f</i> =0.0047	346 nm	<i>f</i> =0.017	340 nm	<i>f</i> =0.07	331 nm	<i>f</i> =0.00
279 nm	<i>f</i> =0.051					334 nm	<i>f</i> =0.00				
											
470 nm	<i>f</i> =0.17	480 nm	<i>f</i> =0.0971	450 nm	<i>f</i> =0.052	459 nm	<i>f</i> =0.086	504 nm	<i>f</i> =0.14	355 nm	<i>f</i> =0.072
355 nm	<i>f</i> =0.072	354 nm	<i>f</i> =0.1501	354 nm	<i>f</i> =0.13	350 nm	<i>f</i> =0.072	353 nm	<i>f</i> =0.088	347 nm	<i>f</i> =0.000
347 nm	<i>f</i> =0.00	330 nm	<i>f</i> =0.0000	342 nm	<i>f</i> =0.000	347 nm	<i>f</i> =0.000	334 nm	<i>f</i> =0.000		

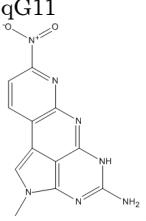
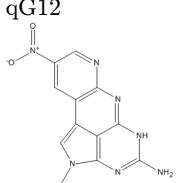
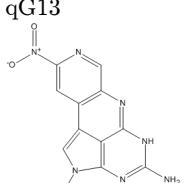
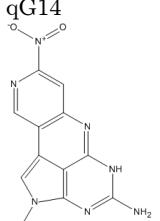
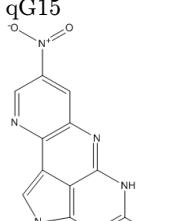
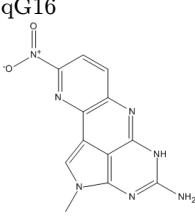
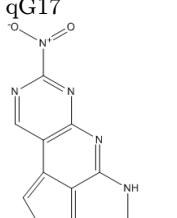
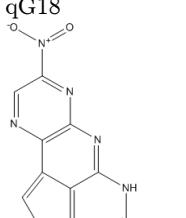
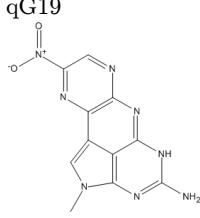
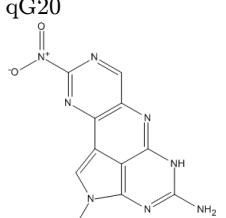
qA16	qA17	qA18	qA19	qA20
460 nm f=0.12	432 nm f=0.11	473 nm f=0.21	456 nm f=0.15	431 nm f=0.064
359 nm f=0.14	364 nm f=0.098	366 nm f=0.077	375 nm f=0.15	376 nm f=0.17
343 nm f=0.000	358 nm f=0.0017	349 nm f=0.0001	342 nm f=0.000	359 nm f=0.0009

## GUANINE ANALOGUES

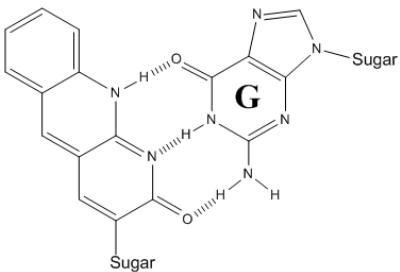


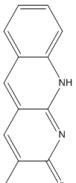
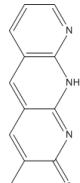
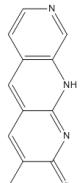
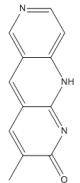
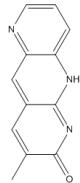
qG1	qG2	qG3	qG4	qG5
347 nm f=0.049	366 nm f=0.20	373 nm f=0.051	369 nm f=0.069	366 nm f=0.18
337 nm f=0.17	351 nm f=0.035	340 nm f=0.17	334 nm f=0.11	351 nm f=0.078
308 nm f=0.18	303 nm f=0.17	302 nm f=0.18	299 nm f=0.18	305 nm f=0.17
qG6	qG7	qG8	qG9	qG10
386 nm f=0.11	408 nm f=0.2	393 nm f=0.099	494 nm f=0.054	474 nm f=0.14
353 nm f=0.085	356 nm f=0.035	356 nm f=0.14	348 nm f=0.005	344 nm f=0.031
295 nm f=0.16	313 nm f=0.004	295 nm f=0.004	334 nm f=0.000	333 nm f=0.000

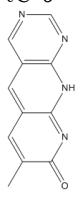
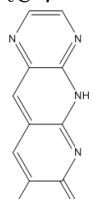
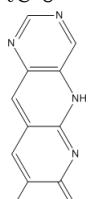
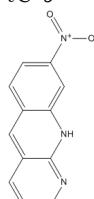
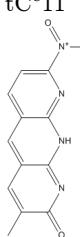
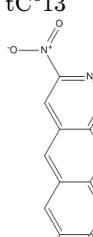
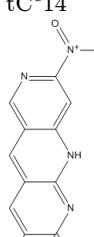
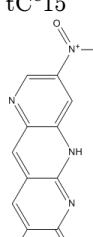
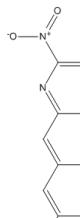
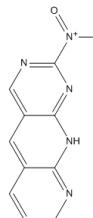
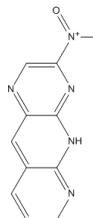
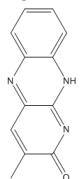
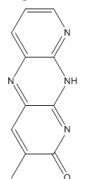
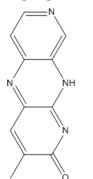
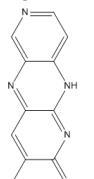
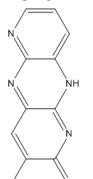
b

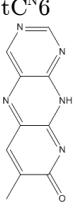
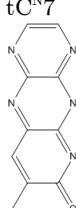
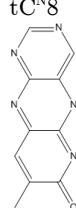
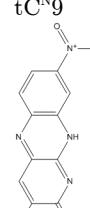
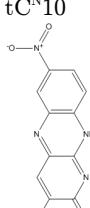
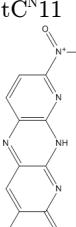
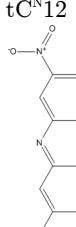
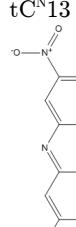
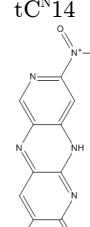
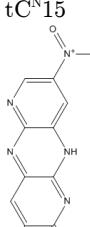
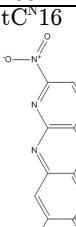
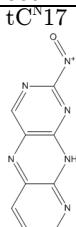
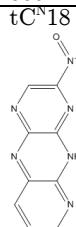
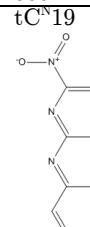
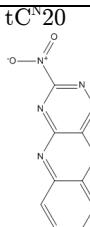
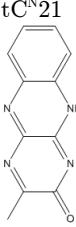
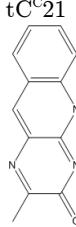
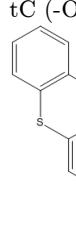
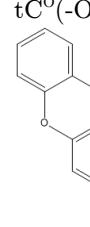
				
465 nm f=0.092 355 nm f=0.074 350 nm f=0.000	478 nm f=0.15 344 nm f=0.055 338 nm f=0.000	450 nm f=0.11 347 nm f=0.056 347 nm f=0.000	451 nm f=0.028 349 nm f=0.053 347 nm f=0.000	510 nm f=0.067 352 nm f=0.069 345 nm f=0.00
				
464 nm f=0.17 352 nm f=0.054 351 nm f=0.000	416 nm f=0.045 367 nm f=0.12 352 nm f=0.007	479 nm f=0.12 372 nm f=0.0008 371 nm f=0.091	465 nm f=0.19 368 nm f=0.001 367 nm f=0.073	438 nm f=0.12 372 nm f=0.058 370 nm f=0.029

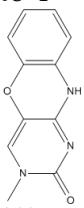
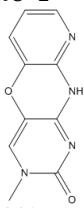
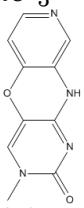
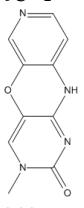
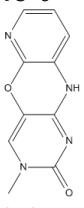
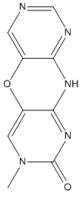
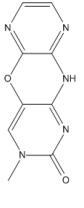
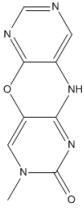
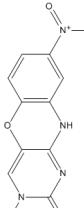
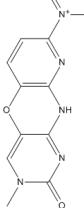
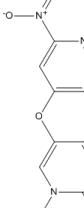
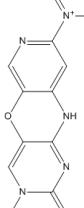
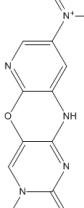
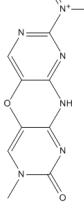
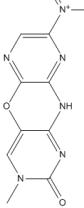
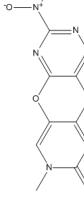
## CYTOSINE ANALOGUES



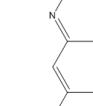
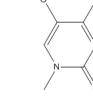
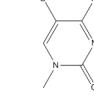
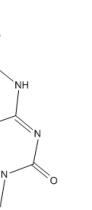
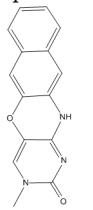
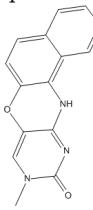
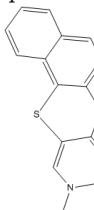
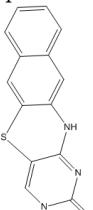
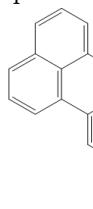
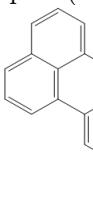
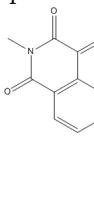
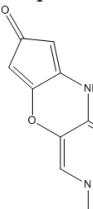
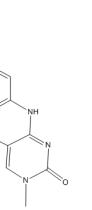
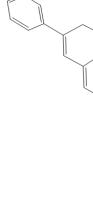
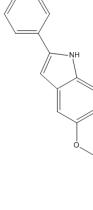
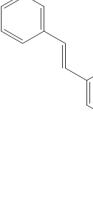
				
370 nm f=0.35 342 nm f=0.36 337 nm f=0.00	370 nm f=0.61 347 nm f=0.00 332 nm f=0.11	382 nm f=0.23 356 nm f=0.00 344 nm f=0.39	353 nm f=0.63 345 nm f=0.00 334 nm f=0.044	376 nm f=0.55 352 nm f=0.00 331 nm f=0.16

				
358 nm f=0.69 357 nm f=0.0004 326 nm f=0.010	385 nm f=0.66 370 nm f=0.0001 348 nm f=0.0020	385 nm f=0.40 374 nm f=0.00 346 nm f=0.0020	464 nm f=0.00 445 nm f=0.24 382 nm f=0.17	426 nm f=0.00 391 nm f=0.21 369 nm f=0.17
				
472 nm f=0.00 429 nm f=0.36 372 nm f=0.085	441 nm f=0.00 394 nm f=0.26 371 nm f=0.21	437 nm f=0.00 385 nm f=0.23 363 nm f=0.0083	449 nm f=0.00 407 nm f=0.33 357 nm f=0.00	480 nm f=0.00 449 nm f=0.30 379 nm f=0.11
				
443 nm f=0.00 390 nm f=0.46 362 nm f=0.037	456 nm f=0.00 395 nm f=0.43 364 nm f=0.0052	497 nm f=0.00 437 nm f=0.41 381 nm f=0.090	467 nm f=0.00 397 nm f=0.52 361 nm f=0.032	460 nm f=0.00 395 nm f=0.29 364 nm f=0.0074
				
406 nm f=0.28 385 nm f=0.0002 365 nm f=0.43	394 nm f=0.0025 394 nm f=0.56 356 nm f=0.0001	415 nm f=0.21 405 nm f=0.0001 352 nm f=0.0001	393 nm f=0.0002 376 nm f=0.53 361 nm f=0.16	403 nm f=0.53 400 nm f=0.0001 354 nm f=0.0000

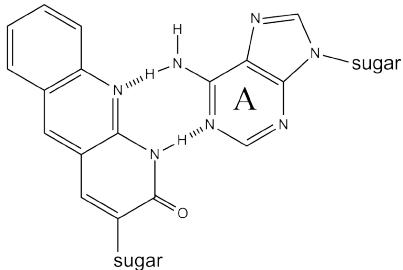
				
404 nm f=0.0001 374 nm f=0.68 373 nm f=0.0000	418 nm f=0.00 404 nm f=0.62 388 nm f=0.0002	423 nm f=0.00 410 nm f=0.37 375 nm f=0.0002	480 nm f=0.00 435 nm f=0.23 393 nm f=0.0002	454 nm f=0.0001 388 nm f=0.34 375 nm f=0.0004
				
487 nm f=0.00 419 nm f=0.36 403 nm f=0.0001	462 nm f=0.00 389 nm f=0.0002 385 nm f=0.47	469 nm f=0.00 396 nm f=0.19 385 nm f=0.0003	471 nm f=0.00 396 nm f=0.32 388 nm f=0.0003	495 nm f=0.00 436 nm f=0.30 412 nm f=0.0001
				
469 nm f=0.0001 392 nm f=0.0002 391 nm f=0.43	476 nm f=0.0001 404 nm f=0.033 386 nm f=0.40	510 nm f=0.00 435 nm f=0.00 426 nm f=0.39	488 nm f=0.00 414 nm f=0.00 393 nm f=0.50	485 nm f=0.00 414 nm f=0.054 403 nm f=0.19
				
454 nm f=0.00 412 nm f=0.29 363 nm f=0.41	467 nm f=0.0001 374 nm f=0.31 321 nm f=0.18	349 nm f=0.014 344 nm f=0.12 306 nm f=0.12	346 nm f=0.21 342 nm f=0.07 309 nm f=0.097	

				
362 nm f=0.19 294 nm f=0.091 290 nm f=0.003	361 nm f=0.21 308 nm f=0.10 288 nm f=0.084	352 nm f=0.12 300 nm f=0.20 281 nm f=0.009	363 nm f=0.12 290 nm f=0.074 286 nm f=0.049	356 nm f=0.21 310 nm f=0.14 288 nm f=0.057
				
365 nm f=0.13 304 nm f=0.072 287 nm f=0.14	358 nm f=0.30 332 nm f=0.037 283 nm f=0.086	350 nm f=0.094 315 nm f=0.26 283 nm f=0.051	438 nm f=0.097 349 nm f=0.037 333 nm f=0.25	452 nm f=0.072 333 nm f=0.21 332 nm f=0.00
				
418 nm f=0.12 346 nm f=0.21 342 nm f=0.00	454 nm f=0.073 331 nm f=0.00 330 nm f=0.13	424 nm f=0.042 342 nm f=0.00 324 nm f=0.10	417 nm f=0.045 342 nm f=0.00 340 nm f=0.18	444 nm f=0.10 347 nm f=0.080 334 nm f=0.18
				
434 nm f=0.097 343 nm f=0.00 332 nm f=0.040	397 nm f=0.043 354 nm f=0.057 349 nm f=0.13	419 nm f=0.14 353 nm f=0.17 340 nm f=0.00	407 nm f=0.058 357 nm f=0.0055 333 nm f=0.0047	407 nm f=0.058 357 nm f=0.0055 333 nm f=0.0047

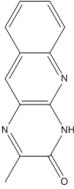
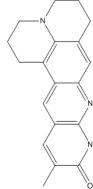
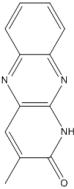
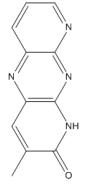
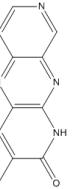
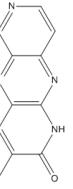
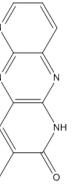
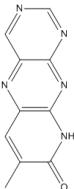
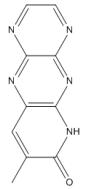
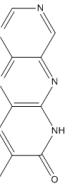
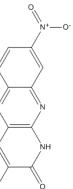
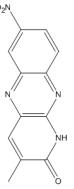
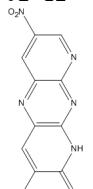
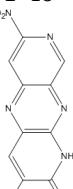
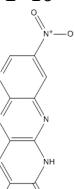
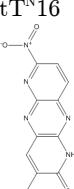
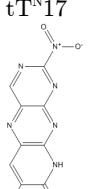
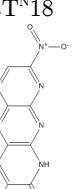
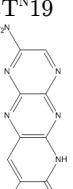
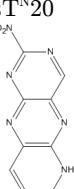
 369 nm f=0.092 309 nm f=0.077 297 nm f=0.029	 370 nm f=0.084 320 nm f=0.11 306 nm f=0.098	 370 nm f=0.044 312 nm f=0.19 298 nm f=0.024	 373 nm f=0.051 308 nm f=0.076 297 nm f=0.052	 372 nm f=0.082 326 nm f=0.17 306 nm f=0.060
 380 nm f=0.049 323 nm f=0.060 304 nm f=0.14	 375 nm f=0.15 363 nm f=0.092 303 nm f=0.097	 380 nm f=0.020 334 nm f=0.25 302 nm f=0.055	 458 nm f=0.080 361 nm f=0.015 345 nm f=0.16	 458 nm f=0.043 346 nm f=0.063 331 nm f=0.0042
 443 nm f=0.095 357 nm f=0.14 344 nm f=0.0007	 466 nm f=0.046 347 nm f=0.016 330 nm f=0.0003	 437 nm f=0.023 343 nm f=0.015 342 nm f=0.0015	 437 nm f=0.042 350 nm f=0.11 344 nm f=0.0098	 478 nm f=0.098 361 nm f=0.082 352 nm f=0.11
 457 nm f=0.066 360 nm f=0.0038 342 nm f=0.0011	 416 nm f=0.038 364 nm f=0.12 351 nm f=0.0062	 458 nm f=0.13 375 nm f=0.13 344 nm f=0.0033	 462 nm f=0.077 371 nm f=0.018 341 nm f=0.0008	 431 nm f=0.038 366 nm f=0.029 357 nm f=0.0038

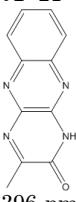
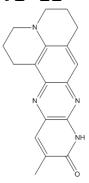
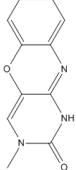
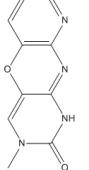
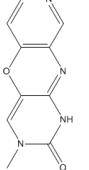
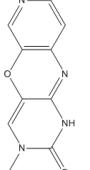
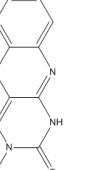
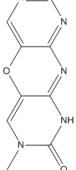
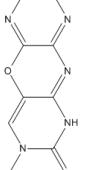
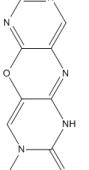
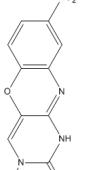
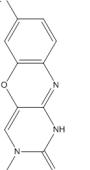
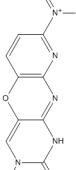
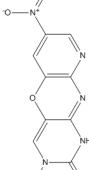
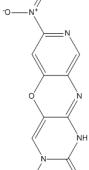
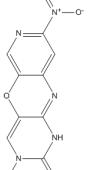
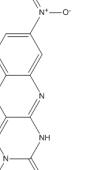
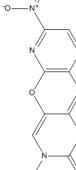
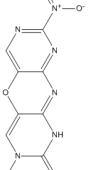
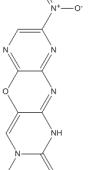
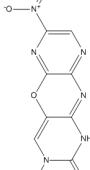
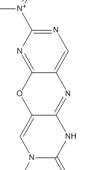
tC <sup>C</sup> 22	tC <sup>N</sup> 22	tC <sup>O</sup> 22	tC <sup>S</sup> 22	
				
491 nm f=0.13 412 nm f=0.00 358 nm f=0.39	603 nm f=0.15 389 nm f=0.54 381 nm f=0.030	391 nm f=0.22 335 nm f=0.084 316 nm f=0.0099	401 nm f=0.14 342 nm f=0.061 332 nm f=0.020	
qC <sup>a</sup> 1	qC <sup>b</sup> 1	qC <sup>c</sup> 1	qC <sup>a</sup> 1	qC <sup>b</sup> 1
				
409 nm f=0.060 351 nm f=0.22 312 nm f=0.048	360 nm f=0.35 333 nm f=0.026 311 nm f=0.22	406 nm f=0.14 348 nm f=0.091 318 nm f=0.094	404 nm f=0.049 344 nm f=0.16 319 nm f=0.025	362 nm f=0.17 346 nm f=0.0097 320 nm f=0.29
qC <sup>c</sup> 1	qC h	qC h (-O)	qC h imide	tC <sup>O</sup> p1
				
407 nm f=0.081 349 nm f=0.062 322 nm f=0.043	420 nm f=0.31 349 nm f=0.04 312 nm f=0.04	443 nm f=0.22 367 nm f=0.042 327 nm f=0.17	450 nm f=0.58 355 nm f=0.01 342 nm f=0.05	457 nm f=0.001 367 nm f=0.34 366 nm f=0.002
Coumarine	Ether	Amine	tCO-st2	tCO-st3
				
439 nm f=0.60 346 nm f=0.29 336 nm f=0.24	426 nm f=0.45 351 nm f=0.033 327 nm f=0.37	394 nm f=0.84 349 nm f=0.054 333 nm f=0.32	414 nm f=0.81 346 nm f=0.030 331 nm f=0.64	420 nm f=0.72 343 nm f=0.053 332 nm f=0.56

## THYMIN ANALOGUES

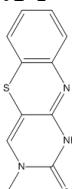
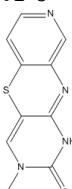
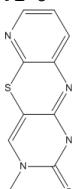
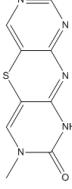
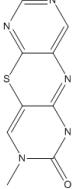
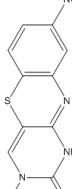
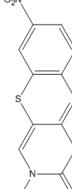
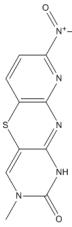
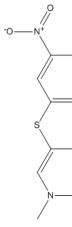
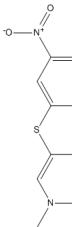
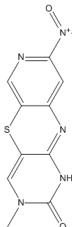
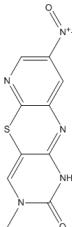
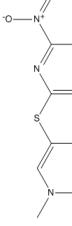
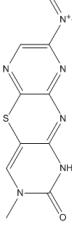
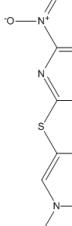
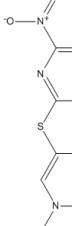


tT <sup>C</sup> 1	tT <sup>C</sup> 2	tT <sup>C</sup> 3	tT <sup>C</sup> 4	tT <sup>C</sup> 5
351 nm f=0.05 331 nm f=0.53	340 nm f=0.47 319 nm f=0.18 311 nm f=0.0002	358 nm f=0.093 322 nm f=0.46 317 nm f=0.0003	333 nm f=0.39 324 nm f=0.13 299 nm f=0.0001	345 nm f=0.45 329 nm f=0.0001 320 nm f=0.20
tT <sup>C</sup> 6	tT <sup>C</sup> 7	tT <sup>C</sup> 8	tT <sup>C</sup> 9	tT <sup>C</sup> 10
335 nm f=0.56 332 nm f=0.0007 310 nm f=0.021	376 nm f=0.0011 355 nm f=0.60 327 nm f=0.058	366 nm f=0.0008 352 nm f=0.32 321 nm f=0.27	392 nm f=0.040 356 nm f=0.42 341 nm f=0.00	363 nm f=0.096 353 nm f=0.11 335 nm f=0.00
tT <sup>C</sup> 11	tT <sup>C</sup> 12	tT <sup>C</sup> 13	tT <sup>C</sup> 14	tT <sup>C</sup> 15
374 nm f=0.25 367 nm f=0.00 338 nm f=0.00	364 nm f=0.26 348 nm f=0.00 334 nm f=0.00	354 nm f=0.14 349 nm f=0.00 324 nm f=0.0001	359 nm f=0.00 349 nm f=0.19 341 nm f=0.25	387 nm f=0.20 374 nm f=0.00 350 nm f=0.25
tT <sup>C</sup> 16	tT <sup>C</sup> 17	tT <sup>C</sup> 18	tT <sup>C</sup> 19	tT <sup>C</sup> 20
360 nm f=0.36 355 nm f=0.0081 337 nm f=0.012	374 nm f=0.050 345 nm f=0.35 332 nm f=0.0078	418 nm f=0.0002 384 nm f=0.35 355 nm f=0.00	405 nm f=0.0003 370 nm f=0.40 348 nm f=0.00	392 nm f=0.014 355 nm f=0.25 346 nm f=0.0095

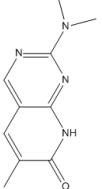
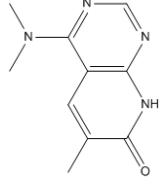
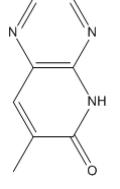
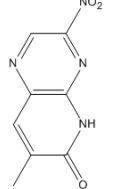
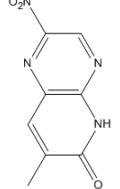
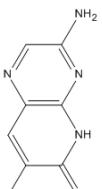
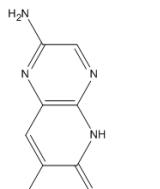
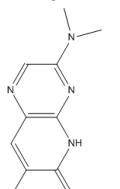
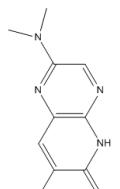
<b>tT<sup>C</sup>21</b>	<b>tT<sup>C</sup>22</b>			
				
358 nm f=0.032	Gas:			
357 nm f=0.0002	460 nm f=0.080			
326 nm f=0.38	342 nm f=0.27			
	303 nm f=0.00			
<b>tT<sup>N</sup>1</b>	<b>tT<sup>N</sup>2</b>	<b>tT<sup>N</sup>3</b>	<b>tT<sup>N</sup>4</b>	<b>tT<sup>N</sup>5</b>
				
373 nm f=0.12	375 nm f=0.0002	379 nm f=0.20	358 nm f=0.0006	378 nm f=0.0002
348 nm f=0.54	360 nm f=0.59	365 nm f=0.0004	350 nm f=0.39	364 nm f=0.58
344 nm f=0.0014	308 nm f=0.00	329 nm f=0.43	343 nm f=0.24	313 nm f=0.00
<b>tT<sup>N</sup>6</b>	<b>tT<sup>N</sup>7</b>	<b>tT<sup>N</sup>8</b>	<b>tT<sup>N</sup>9</b>	<b>tT<sup>N</sup>10</b>
				
399 nm f=0.00	426 nm f=0.0001	408 nm f=0.00	390 nm f=0.12	364 nm f=0.30
350 nm f=0.68	370 nm f=0.68	369 nm f=0.45	369 nm f=0.0007	358 nm f=0.0009
320 nm f=0.00	322 nm f=0.00	324 nm f=0.0001	351 nm f=0.37	351 nm f=0.042
<b>tT<sup>N</sup>11</b>	<b>tT<sup>N</sup>12</b>	<b>tT<sup>N</sup>13</b>	<b>tT<sup>N</sup>14</b>	<b>tT<sup>N</sup>15</b>
				
410 nm f=0.0001	404 nm f=0.0001	378 nm f=0.0003	383 nm f=0.0002	420 nm f=0.0001
375 nm f=0.39	363 nm f=0.48	365 nm f=0.21	353 nm f=0.31	385 nm f=0.33
353 nm f=0.0001	330 nm f=0.00	349 nm f=0.0001	351 nm f=0.0003	347 nm f=0.00
<b>tT<sup>N</sup>16</b>	<b>tT<sup>N</sup>17</b>	<b>tT<sup>N</sup>18</b>	<b>tT<sup>N</sup>19</b>	<b>tT<sup>N</sup>20</b>
				
401 nm f=0.0001	424 nm f=0.0064	470 nm f=0.00	458 nm f=0.00	432 nm f=0.0051
362 nm f=0.50	357 nm f=0.38	383 nm f=0.47	369 nm f=0.58	368 nm f=0.30
350 nm f=0.0002	343 nm f=0.020	362 nm f=0.00	357 nm f=0.00	348 nm f=0.030

				
396 nm f=0.00 371 nm f=0.058 341 nm f=0.0009	516 nm f=0.11 352 nm f=0.23 349 nm f=0.082			
tT^o1	tT^o2	tT^o3	tT^o4	tT^o5
				
374 nm f=0.32 306 nm f=0.078 293 nm f=0.0045	380 nm f=0.34 308 nm f=0.12 301 nm f=0.040	367 nm f=0.247 310 nm f=0.218 289 nm f=0.0011	382 nm f=0.24 300 nm f=0.15 288 nm f=0.0055	371 nm f=0.36 313 nm f=0.14 300 nm f=0.018
tT^o6	tT^o7	tT^o8	tT^o9	tT^o10
				
391 nm f=0.25 302 nm f=0.15 295 nm f=0.054	381 nm f=0.43 331 nm f=0.051 296 nm f=0.076	368 nm f=0.24 322 nm f=0.28 294 nm f=0.021	435 nm f=0.039 357 nm f=0.38 331 nm f=0.00	454 nm f=0.23 336 nm f=0.00 333 nm f=0.15
tT^o11	tT^o12	tT^o13	tT^o14	tT^o15
				
407 nm f=0.049 365 nm f=0.45 344 nm f=0.00	455 nm f=0.24 335 nm f=0.00 334 nm f=0.00	430 nm f=0.20 349 nm f=0.00 321 nm f=0.15	412 nm f=0.0045 352 nm f=0.35 343 nm f=0.00	441 nm f=0.040 356 nm f=0.46 330 nm f=0.00
tT^o16	tT^o17	tT^o18	tT^o19	tT^o20
				
436 nm f=0.28 349 nm f=0.00 327 nm f=0.082	393 nm f=0.052 358 nm f=0.29 345 nm f=0.045	407 nm f=0.050 368 nm f=0.49 342 nm f=0.00	433 nm f=0.32 354 nm f=0.00 341 nm f=0.00	414 nm f=0.23 363 nm f=0.0007 319 nm f=0.0099

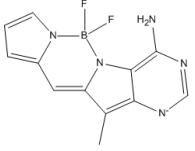
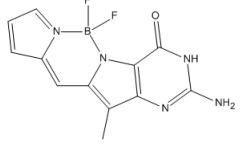
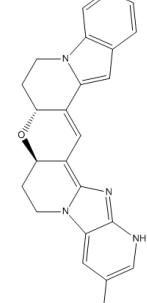
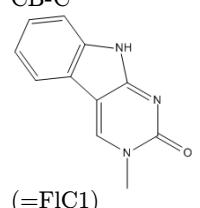
k

				
408 nm f=0.20 329 nm f=0.081 313 nm f=0.020	416 nm f=0.20 333 nm f=0.092 326 nm f=0.11	407 nm f=0.13 331 nm f=0.22 313 nm f=0.018	424 nm f=0.14 326 nm f=0.099 311 nm f=0.019	406 nm f=0.21 337 nm f=0.17 322 nm f=0.042
				
437 nm f=0.14 330 nm f=0.067 323 nm f=0.14	416 nm f=0.26 366 nm f=0.10 318 nm f=0.092	410 nm f=0.11 346 nm f=0.29 316 nm f=0.042	472 nm f=0.038 383 nm f=0.28 333 nm f=0.00	499 nm f=0.15 364 nm f=0.066 336 nm f=0.00
				
442 nm f=0.034 389 nm f=0.34 346 nm f=0.0003	491 nm f=0.16 355 nm f=0.044 342 nm f=0.0011	473 nm f=0.13 351 nm f=0.0001 350 nm f=0.049	449 nm f=0.012 372 nm f=0.23 345 nm f=0.0002	483 nm f=0.041 384 nm f=0.34 331 nm f=0.00
				
480 nm f=0.20 362 nm f=0.0089 350 nm f=0.00	432 nm f=0.041 373 nm f=0.25 350 nm f=0.0012	450 nm f=0.040 398 nm f=0.39 350 nm f=0.0003	482 nm f=0.22 364 nm f=0.00 359 nm f=0.0023	456 nm f=0.15 370 nm f=0.0019 356 nm f=0.012

<b>tT<sup>s</sup>22</b>	<b>qT h</b>	<b>qT h (-O)</b>	<b>qT h2</b>	<b>tT<sup>O</sup>22</b>
430 nm f=0.31 359 nm f=0.0077 342 nm f=0.030	452 nm f=0.38 349 nm f=0.01 319 nm f=0.02	524 nm f=0.28 383 nm f=0.04 339 nm f=0.14	430 nm f=0.39 334 nm f=0.026 312 nm f=0.032	407 nm f=0.39 332 nm f=0.017 327 nm f=0.034
<b>tTS (-O)</b>	<b>tTO (-O)</b>			
423 nm f=0.26 402 nm f=0.032 335 nm f=0.11	390 nm f=0.38 369 nm f=0.033 317 nm f=0.084			
<b>bT1</b>	<b>bT2</b>	<b>bT3</b>	<b>bT4</b>	<b>bT5</b>
310 nm f=0.36 270 nm f=0.00 263 nm f=0.01	350 nm f=0.23 315 nm f=0.00 308 nm f=0.01	333 nm f=0.20 319 nm f=0.00 303 nm f=0.00	393 nm f=0.12 341 nm f=0.003 328 nm f=0.01	327 nm f=0.54 287 nm f=0.029 260 nm f=0.00
<b>bT6</b>	<b>bT7</b>	<b>bT8</b>	<b>bT9</b>	<b>bT10</b>
371 nm f=0.18 281 nm f=0.002 275 nm f=0.099	320 nm f=0.14 304 nm f=0.26 265 nm f=0.00	341 nm f=0.58 294 nm f=0.03 262 nm f=0.00	386 nm f=0.14 287 nm f=0.0077 281 nm f=0.13	342 nm f=0.16 311 nm f=0.23 270 nm f=0.003
<b>bT11</b>	<b>bT12</b>	<b>bT13</b>	<b>bT14</b>	<b>bT15</b>
299 nm f=0.39 288 nm f=0.0014 274 nm f=0.00	302 nm f=0.00 296 nm f=0.00 285 nm f=0.00	333 nm f=0.10 312 nm f=0.00 304 nm f=0.00	321 nm f=0.45 287 nm f=0.12 275 nm f=0.002	316 nm f=0.34 280 nm f=0.043 269 nm f=0.00

bT16	bT17	bT18	bT19	bT20
				
337 nm f=0.45	342 nm f=0.25	330 nm f=0.42	364 nm f=0.28	336 nm f=0.38
297 nm f=0.16	293 nm f=0.13	318 nm f=0.003	343 nm f=0.003	333 nm f=0.0005
285 nm f=0.002	281 nm f=0.0097	286 nm f=0.00	334 nm f=0.001	328 nm f=0.0007
bT21	bT22	bT23	bT24	
				
348 nm f=0.52	379 nm f=0.31	366 nm f=0.55	390 nm f=0.27	
293 nm f=0.078	304 nm f=0.003	301 nm f=0.098	307 nm f=0.034	
289 nm f=0.0034	282 nm f=0.00	294 nm f=0.0028	291 nm f=0.14	

### STRUCTURES BASED ON KNOWN DYES:

BodipyBase BOD-A	BodipyBase BOD-G	MethyleneBase MB-T	CyBase	CarbazoleBase CB-C (=FlC1)
				
645 nm f=0.023	571 nm f=0.071	632 nm f=0.41	Gas: 489 nm f=1.19	335 nm f=0.077
475 nm f=0.0003	412 nm f=0.55	412 nm f=0.30	386 nm f=0.32	302 nm f=0.14
414 nm f=0.59	375 nm f=0.0001	381 nm f=0.00	375 nm f=0.01	262 nm f=0.72
CB-T (=FlT1)	CB-C(-O) (=FlC(-O)1)			
364 nm f=0.01	327 nm f=0.047			
320 nm f=0.33	293 nm f=0.30			
254 nm f=0.66	248 nm f=0.003			

# Part II

APPENDED

PAPERS



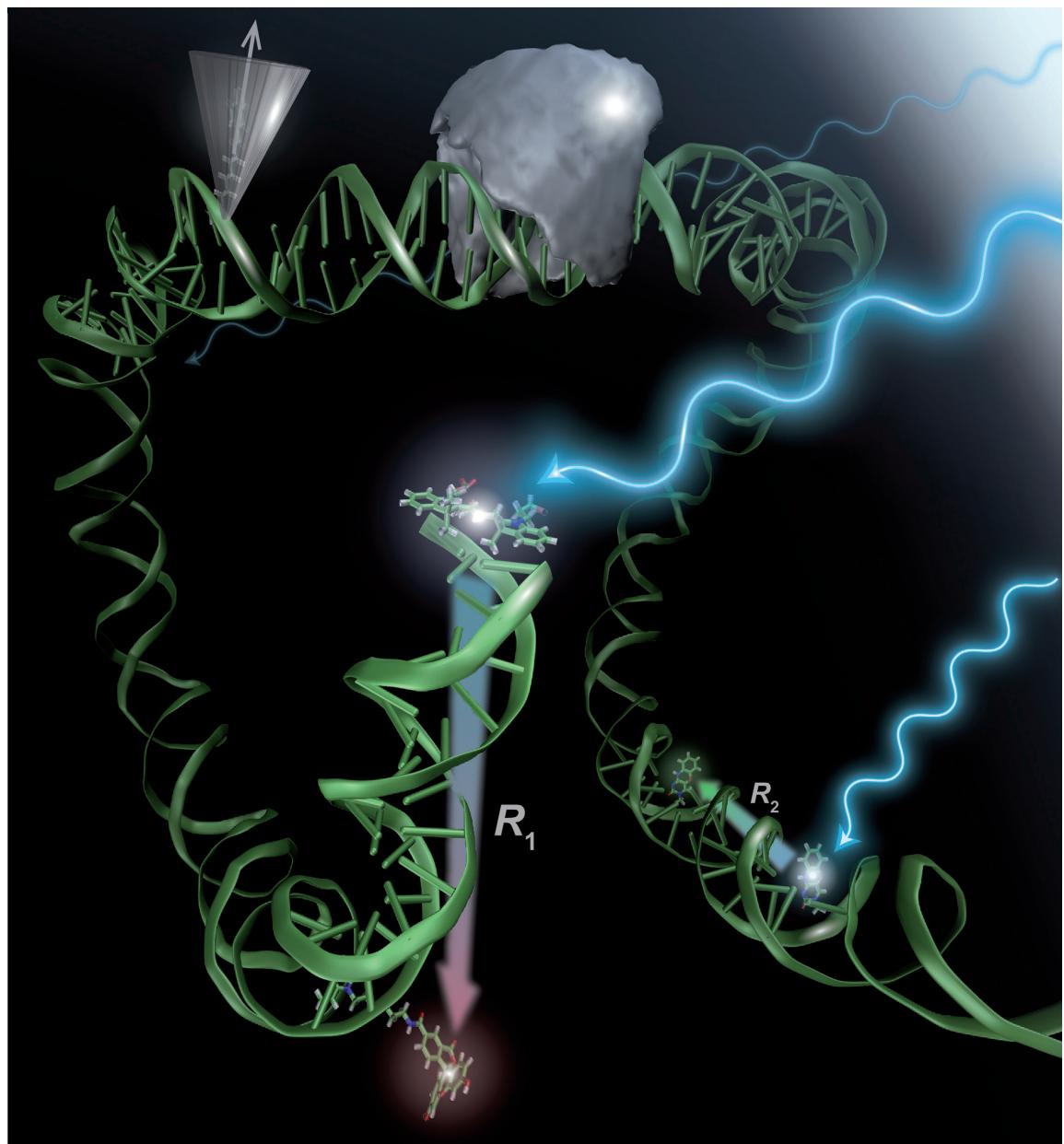
# **Paper I**

*In which we review the field of quantitative FRET...*



# Advances in Quantitative FRET-Based Methods for Studying Nucleic Acids

Søren Preus<sup>[b]</sup> and L. Marcus Wilhelmsson\*<sup>[a]</sup>



Förster resonance energy transfer (FRET) is a powerful tool for monitoring molecular distances and interactions at the nanoscale level. The strong dependence of transfer efficiency on probe separation makes FRET perfectly suited for "on/off" experiments. To use FRET to obtain quantitative distances and

three-dimensional structures, however, is more challenging. This review summarises recent studies and technological advances that have improved FRET as a quantitative molecular ruler in nucleic acid systems, both at the ensemble and at the single-molecule levels.

## 1. Introduction

The three-dimensional structures and conformational dynamics of nucleic acids are directly related to their functions in living organisms. Traditionally, structural information on biomolecules is obtained at atomic resolution by X-ray crystallography<sup>[1]</sup> and NMR spectroscopy.<sup>[2]</sup> Although the importance of both methods is indisputable, the main drawback of crystallography is the need for molecular crystals, whereas NMR primarily suffers from complexity, the need for large amounts of sample and an upper molecular weight limit. Lower-resolution techniques include small-angle scattering,<sup>[3]</sup> cryo-electron microscopy<sup>[4]</sup> and "molecular rulers" such as FRET. Of these, FRET possesses the prominent advantage of being cheap and easy, and provides rapid measurements with just a few nmol of sample even in complex media or large molecular complexes. In addition, FRET is routinely performed at the single-molecule level.<sup>[5]</sup> The main disadvantages of FRET as a quantitative ruler, however, are its dependence on fluorophore photophysical properties, which depend on the interaction between the probe and its microenvironment, and the uncertainty in probe position and orientation relative to the biomolecule. These problems are often associated with the linker connecting the probe to the biomolecule. In single-molecule experiments, FRET suffers from low, fluctuating signals and bleaching. In addition, FRET is limited to distances of <10 nm.

The limitations of FRET have motivated the development of alternative molecular rulers. Site-directed spin labelling offers measures of relative distances and orientations.<sup>[6]</sup> Despite its high potential, spin labelling is not widely used for structural studies of nucleic acids, probably due to limitations in the number of commercially available spin labels.<sup>[7]</sup> In the past six or seven years the interesting electronic properties of metal nanoparticles (NPs) have been exploited in molecular rulers based on NP-NP coupling ("plasmon rulers")<sup>[8]</sup> and small-angle X-ray scattering interference (SAXSI),<sup>[9]</sup> whereas dye-NP coupling has been exploited in rulers based on nanometal surface energy transfer (NSET)<sup>[10]</sup> and surface-enhanced Raman scattering (SERS).<sup>[11]</sup> The main advantages of these techniques are high and robust signals and probe separations exceeding 10 nm. However, each of these methods still requires refinement and more thorough characterization before it can have broad applicability in quantitative structural studies. Most notably, the heterogeneity of colloidal NPs in preparation limits the interpretation of quantitative measurements because the sizes and shapes of NPs are directly related to their optical properties.

As a result, none of these rulers is yet as ubiquitous in the biosciences as FRET. Whereas the theory and applications of

FRET have been reviewed previously,<sup>[12]</sup> this review seeks to provide a short, but comprehensive, overview of recent studies and techniques that have addressed the above issues of FRET and advanced the use of quantitative FRET in nucleic acid studies. We define "quantitative FRET" as when a distance, a distribution of distances, or dynamical parameters are derived from the measured FRET signal. The properties of common FRET probes in nucleic acid environments are reviewed first, followed by an overview of alternative fluorophores. After focusing on the probes we summarise methods developed to measure and to analyse FRET quantitatively, both at the ensemble and at the single-molecule levels, including methods developed to model probe fluctuations. Finally we describe new techniques based on FRET and provide a list of advanced FRET simulation and analysis software. Throughout the article, particular emphasis is given to the advantages and limitations of each strategy.

## 2. Dye Properties in Nucleic Acids

Almost all quantitative FRET experiments are performed by tethering the dyes externally to nucleic acids through linkers (Figure 1A). This highly versatile approach allows practically any probe to be implemented site-specifically by several different methods. The inherent drawback, however, is that quantitative interpretation of experiments becomes complicated by dipole diffusion and reorientation, influencing the donor-acceptor distance ( $R$ ) and the orientation factor ( $\kappa^2$ ), respectively. Only in ideal cases is the almost universal assumption of freely rotating fluorophores—that is,  $\kappa^2 = 2/3$ —valid ( $\kappa^2$  can take values between 0 and 4). In addition, interactions between the dye and the nucleic acid, such as electrostatic interactions with the negatively charged backbone, intercalation and/or end-stacking, influence not only the position and orientation of the probe but also its photophysical properties, thus directly affecting the critical Förster distance ( $R_0$ ; readers interested in the equations describing this relationship are referred to Lakowicz<sup>[12c]</sup>). It is a rule rather than an exception that fluorophore quantum yields and lifetimes depend on dye position and the surrounding DNA sequence.

[a] Assoc. Prof. L. M. Wilhelmsson  
Department of Chemical and Biological Engineering  
Chalmers University of Technology  
41296 Göteborg (Sweden)  
E-mail: marcus.wilhelmsson@chalmers.se

[b] S. Preus  
Department of Chemistry, University of Copenhagen  
Universitetsparken 5, 2100 Copenhagen (Denmark)

Because the energy transfer is dependent on the above parameters, a large number of studies addressing the properties of FRET dyes attached to DNA have been performed (summarised in Table 1). Most of the early studies employing quantitative FRET in nucleic acids combined fluorescein (FAM) as the donor and rhodamine or tetramethylrhodamine (TMR) as the acceptor (Figure 1B).<sup>[14]</sup>

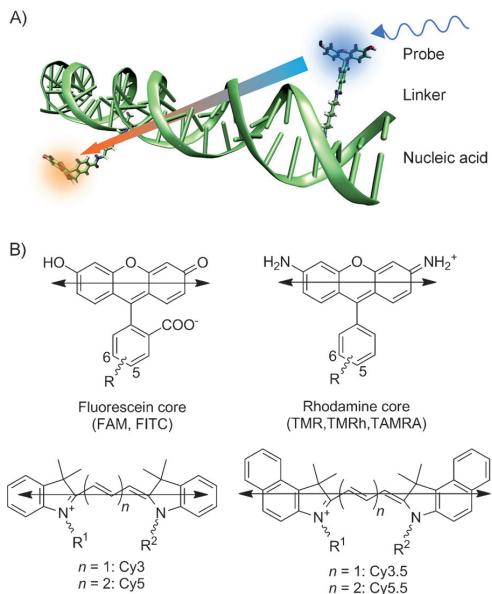
Fluorescein, being negatively charged under biological conditions, is repelled by the DNA backbone, resulting in an approximately random dynamic orientation as observed by the fluorescence anisotropy of the dye coupled to DNA.<sup>[15]</sup> The drawbacks of fluorescein, however, are its relatively poor photostability, pH dependency and a marked quenching effect by guanine.<sup>[16]</sup>

In contrast to fluorescein, the positively charged rhodamine and its derivatives interact strongly with DNA.<sup>[15a–c, 16e, 17]</sup> Neubauer et al. showed that end-labelled rhodamine 6G populates several distinct, end-stacked conformations and suggested that different quenching kinetics, resulting from electron transfer from a nearby guanine, govern the different states.<sup>[17h]</sup> The popular carboxytetramethylrhodamine (TAMRA) also interacts with DNA<sup>[15c]</sup> and displays complex decay kinetics in nucleic acid environments<sup>[16c, 17f]</sup> resulting from quenching by guanine<sup>[16e, 17i,j]</sup> and the presence of Mn<sup>2+</sup>.<sup>[17k]</sup>

Marcus Wilhelmsson received his PhD, mainly focusing on interactions between small ligands and DNA, from Chalmers University of Technology (2003). He has also been a guest researcher at Stanford University in Prof. Richard Zare's lab (2000). After his PhD he switched to DNA nanotechnology and was a supervisor in an EU STREP project involving 25 researchers in four countries. Since 2008 he has had his own group at Chalmers. His main focus is on the design, development and use of fluorescent base analogues and on methodologies to facilitate quantitative FRET. He is also founder of the conference FB<sup>3</sup>.



Søren Preus is currently finishing his PhD in the optical spectroscopy laboratory of Assoc. Prof. Kristine Kilså at Københavns Universitet. His research frequently brings him to the laboratories of Prof. Bo Albinsson and L.M.W. at Chalmers University of Technology, and he was a visiting scholar in Prof. Daniel Herschlag's laboratory at Stanford University. His current research is focused on the application of fluorescent dyes, especially nucleobase analogues to study biological systems at the nanoscale level.



**Figure 1.** A) Illustration of tethered FRET probes. B) Examples of common FRET probes. Arrows denote *emission* transition dipole moments (note that most rhodamines have an additional slightly blue-shifted transition with a transition moment perpendicular to the lowest in energy<sup>[13]</sup>).

Texas Red, a zwitterionic derivative of fluorescein and rhodamine, has been suggested to bind significantly into the minor groove of DNA.<sup>[15c]</sup>

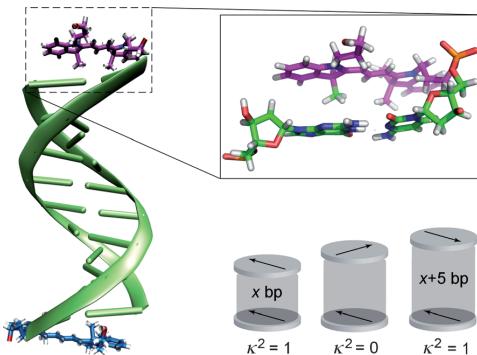
For the last ten years, the most popular FRET probes in nucleic acid studies have been the Cy dyes, primarily owing to their high brightness in single-molecule experiments (Figure 1B). Due to their popularity, the literature on the properties of Cy dyes in nucleic acid environments is comprehensive<sup>[10d, 17a,e, 18]</sup> (summarised in a recent review by Levitus and Ranjit<sup>[18a]</sup>). Levitus and co-workers showed that the photophysical properties of Cy3 in DNA differ substantially from the properties of the free dye and display a strong dependency on dye position,<sup>[18d]</sup> whereas Sabanayagam et al. showed that the fluorescence quantum yield of Cy3 even varies from molecule to molecule.<sup>[18g]</sup> This property is a consequence of a *cis-trans* photoisomerisation decay pathway competing with fluorescence from the first excited singlet state of Cy3 and related polymethine dyes.<sup>[19]</sup> When the microenvironment of the probe hinders rotation of one end of the molecule with respect to the other the fluorescence quantum yield increases. Interestingly, in Cy3B the isomerisation is prevented by rigidification of the trimethine chain, which increases the fluorescence quantum yield and excited state lifetime of the probe,<sup>[18d, 20]</sup> but at the expense of lower photostability.<sup>[20c]</sup>

Lilley and co-workers have shown that end-labelled Cy3 and Cy5 tend to stack on top of the ending base-pair of the helix, resulting in a partially constrained dipole orientation (Figure 2).<sup>[18b,i]</sup> The stacking ability is independent of linker length<sup>[18e]</sup> but the precise orientation of the stacked probe is

**Table 1.** Selected references for commercially available FRET probes characterized in nucleic acid environments.

Dye	Property <sup>[a]</sup> and references	Suppliers <sup>[b]</sup>
Alexa 488	lifetime/QY <sup>[15d, 18k]</sup> position/orientation <sup>[15d, 18k]</sup>	[c]
ATTO 610/647/680	lifetime/QY <sup>[23c, d]</sup> position/orientation <sup>[23c]</sup>	[d]
Cy3	lifetime/QY <sup>[17e, 18d-j]</sup> position/orientation <sup>[17e, 18b-f]</sup>	[e, f, g]
Cy3B	lifetime/QY <sup>[18d, 20]</sup>	[e]
Cy5	lifetime/QY <sup>[10d, 17e, 18j, k]</sup> position/orientation <sup>[10d, 17a, e, 18k, l]</sup>	[e, f, g]
fluorescein/FAM	lifetime/QY <sup>[10d, 15b, d, 16c, d, 24]</sup> position/orientation <sup>[10d, 15, 16d, 17b-d, 24b]</sup>	[c, e, f, g, h, i, j]
rhodamine 6G	lifetime/QY <sup>[17e, h, 25]</sup> position/orientation <sup>[17e, h, 25]</sup>	[i]
rhodamine/TMR/ TAMRA	lifetime/QY <sup>[15b, 16c, 17d-g, 18g, 24c]</sup> position/orientation <sup>[15a-c, 17a-g]</sup>	[c, f, g, h, i, j]
Texas Red	lifetime/QY <sup>[16c]</sup> position/orientation <sup>[15c]</sup>	[c]
tC/tC <sup>O</sup> /tC <sub>nitro</sub>	lifetime/QY <sup>[26]</sup> position/orientation <sup>[26a, 27]</sup>	[f]

[a] QY: Fluorescence quantum yield; "position/orientation" covers studies of the structure and dynamics of the dye in DNA as studied by fluorescence anisotropy, NMR, MD simulations etc. [b] Suppliers of monomeric dyes (phosphoramidites or other reactive forms): [c] Life Technologies, [d] ATTO-TEC, [e] GE Healthcare, [f] Glen Research, [g] Lumiprobe, [h] Sigma-Aldrich, [i] AAT Bioquest, [j] Azco Biotech, Inc. For their commercial incorporation into nucleic acids see reference list.<sup>[28]</sup>



**Figure 2.** Orientations of end-labelled Cy3 (purple) and Cy5 (blue) in B-DNA. Coordinate file courteously provided by Prof. David Lilley. Bottom right: schematic depiction of the effect of duplex length on the transition dipole moment orientations for three extreme cases of  $\kappa^2$  (dyes represented as discs). Adapted with permission from ref. [18e], copyright 2011, Biophysical Society.

linker-dependent, as shown by NMR.<sup>[18c]</sup> Molecular dynamics (MD) simulations, partly supported by lifetime measurements,<sup>[18e]</sup> suggested that the stacking interaction depends on the identity of the ending base pair.<sup>[18f]</sup> The effect of end-stacking on the energy transfer efficiency between terminally attached Cy dyes was demonstrated by Iqbal et al.<sup>[21]</sup> As a result of the constrained transition moments, the FRET efficiency between Cy3 and Cy5 units terminally attached to duplexes of

varying lengths (Figure 2, insert) resulted in a fine-structured FRET efficiency as a function of duplex length with a periodicity in phase with the helical periodicities of the B-form DNA and A-form DNA/RNA helices being probed. Upon internal labelling, Cy dyes are believed to bind into and along the minor or major groove.<sup>[10d, 17a, 22]</sup>

Recently, the ATTO dyes (trademark of ATTO-TEC, Siegen, Germany) have found increased use in single-molecule FRET experiments, due to their high brightnesses and photostabilities. To date, only a few studies have been directed at characterizing this dye series in nucleic acids.<sup>[23]</sup> The oxazine ATTO 655 was shown to interact strongly with double-stranded DNA and is transformed into a reduced, quenched state by guanine.<sup>[23a,b]</sup> Kupstat et al. found that ATTO 610 and ATTO 680 interacted with DNA, however, with the assumption of freely rotating probes giving reasonable agreement between expected and measured FRET efficiencies.<sup>[23c]</sup>

Under the high irradiation conditions of single-molecule setups, most organic dyes undergo blinking or long-time on/off photoswitching.<sup>[29]</sup> Although photoswitchable dyes are useful in, for example, super-resolution imaging for determining dye localizations one probe at a time,<sup>[30]</sup> in FRET experiments blinking can be a serious source of error if interpreted as a FRET state.<sup>[29a]</sup> Blinking can be greatly reduced by use of additives such as a reducing and oxidising system (ROXS)<sup>[29b]</sup> or a triplet-state quencher such as Trolox, either added to the solution<sup>[29c]</sup> or conjugated directly to the dye.<sup>[29d]</sup> In fact, the mechanism of Trolox as an anti-blinking agent was shown by Tinnefeld and co-workers to be similar to that of a ROXS.<sup>[29e]</sup> Trolox can also be used in aerobic solution in combination with an oxygen radical scavenger such as cysteamine.<sup>[29f]</sup> Interestingly, the identities of some long-lived dark states were recently determined for red cyanine dyes<sup>[29g]</sup> and ATTO 655.<sup>[29h]</sup>

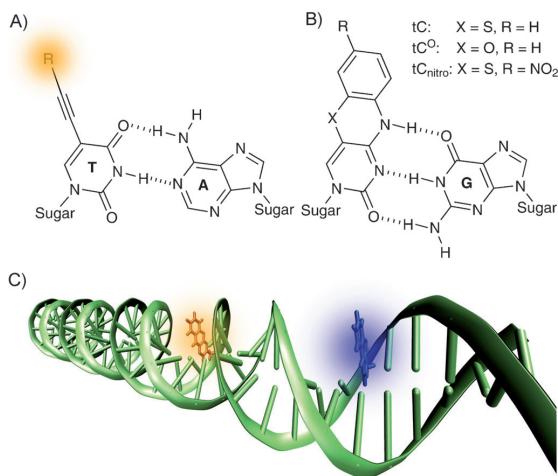
### 3. Alternative Probes

Many alternative FRET probes have been developed for various purposes (for a review see Sapsford et al.<sup>[31]</sup>). FRET between genetically encoded fluorescent proteins is extremely useful for studying protein interactions *in vivo*,<sup>[32]</sup> and fluorescent proteins have been coupled covalently to DNA in the pursuit of light-addressable photonic nanodevices.<sup>[33]</sup> Fluorescent proteins, though, will introduce more unnecessary complexity into quantitative FRET studies than small organic dyes in many *in vitro* applications. Quantum dots (QDs, semiconductor nanoparticles) offer extreme intensities and stabilities relative to organic fluorophores, and for analytical and imaging purposes the potential of QDs is tremendous (for reviews see, for example, ref. [34]). The main disadvantages of QDs in quantitative FRET are their large sizes (30–40 nm in diameter), the difficulty in coupling only one biomolecule per QD, and limitations in donor–acceptor distances due to the thickness of the QD coating.

Background signals resulting from scattering, direct excitation of acceptor and/or autofluorescence can be greatly reduced by use of long-lifetime donors such as lanthanides and transition metal complexes (see ref. [35] and references there-

in). Besides providing the ability to filter out background fluorescence through time-gated detection of acceptor emission, lanthanide emission is unpolarized, resulting in less uncertainty associated with the orientation factor of the energy transfer process (reviewed by Selvin<sup>[35a]</sup>). However, reported studies using long-lifetime probes for quantitative nucleic acid studies have been sparse in relation to those involving conventional labels. This is likely a result of limitations in available probes and the difficulty in introducing lanthanide and transition metal complexes site-specifically and noninvasively into nucleic acids.

Recent efforts have been put into the development of dyes that exhibit less uncertainty in position and orientation relative to the nucleic acid.<sup>[18k,36]</sup> Diffusional mobility is reduced by using short or rigid linkers (Figure 3A),<sup>[18k,36a-e]</sup> but this can



**Figure 3.** FRET probes designed to display limited diffusional and rotational mobility. A) Fluorophore tethered through a rigid ethynyl linker. B) The tricyclic cytosine base analogue FRET probes. C) The base probes positioned in B-DNA.

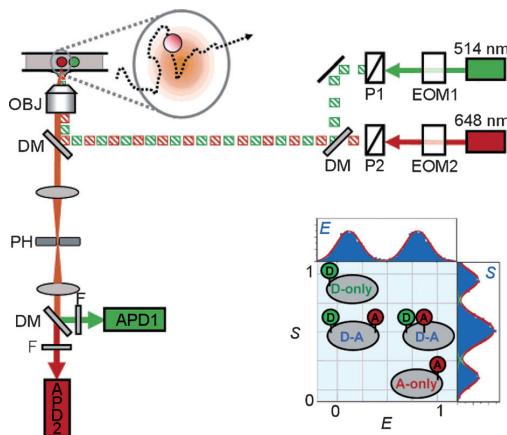
be at the expense of limited rotational freedom of the probe.<sup>[18k,36a]</sup> In some cases, the orientation of the probe can be controlled in a predictable manner.<sup>[36f,g]</sup> Lewis and co-workers demonstrated control of dye orientation in a series of "end-capped" DNA duplexes in which the donor was rigidly attached to both strands of the duplex while the acceptor was allowed to stack on the ending base-pair.<sup>[36f]</sup> Ranjit et al. tested probes consisting of Cy3 and Cy5 rigidly attached to the backbone of DNA and found unusually high FRET efficiencies at large donor–acceptor distances, explained by favourable transition dipole moment orientations.<sup>[36g]</sup>

Our group has developed nucleobase analogue FRET pairs with very high degrees of position and orientation control of both donor and acceptor transition dipole moments (Figure 3B, C).<sup>[26d,27c]</sup> To date, the combination of two donors (tC and tC<sup>O</sup>) and one acceptor (tC<sub>nitro</sub>) has been reported (base-base FRET), but fluorescent base analogues can also be used in

combination with other labels.<sup>[37]</sup> By mimicking the Watson-Crick hydrogen bonding and base-stacking of the natural bases these probes position themselves inside the nucleic acid structure as a replacement for one of the canonical bases, adapting the position and orientation of the substituted base, usually without perturbing the overall nucleic acid structure (reviewed previously<sup>[38]</sup>). The main advantage of this approach is the ability to position the reporters inside the site of interest, where it is impossible to use external fluorophores. In addition, the very high degree of control of both position and orientation of the probes allows more detailed structural information to be retrieved from FRET experiments without complications associated with linker flexibility. Furthermore, the photophysical properties of the tC bases are relatively environment-insensitive, allowing better control of quantum yields and spectral overlaps than with most external dye pairs. Despite their great promise, the base probes cannot compete with external labels on overall brightness and large  $R_0$  distances, and still no fluorescent base analogues sufficiently bright and photostable for single-molecule experiments have been reported. In addition, as a result of their highly constrained orientations, which as mentioned above are unique to these FRET pairs, the value of  $\kappa^2$  cannot be assumed constant (e.g.,  $\kappa^2 = 2/3$ ) but must be included in the analysis. Moreover, theoretical calculations have suggested that the ideal dipole approximation, on which the FRET theory is based, in some cases does not fully describe closely separated, constrained donors and acceptors.<sup>[39]</sup>

#### 4. Measuring FRET Accurately

Aided by the increasing versatility in available probes and technological advances in single-molecule fluorescence spectroscopy (recently reviewed by Hohlbein et al.<sup>[40]</sup>) the FRET toolbox is continuously expanding, allowing FRET to be used quantitatively for more and more complex biological systems. Single-molecule fluorescence detection is generally performed either with TIRF (total internal reflection fluorescence) microscopy on surface-immobilized dyes or with confocal microscopy on freely diffusing dyes (see Selvin and Ha for a practical description of both methods<sup>[5c]</sup>). In TIRF smFRET (single-molecule FRET) microscopy, movies of up to hundreds of individual donor–acceptor pairs are recorded simultaneously, and individual intensity time traces are deduced and analysed from these. In diffusion-based smFRET, intensity time traces are recorded from a small confined excitation volume, detecting the short "bursts" of photons emitted from dyes diffusing into the confocal spot. In both strategies the difficulty in measuring quantitative FRET mainly comes from the requirement for corrections of spectral cross-talk, detection efficiencies and relative donor and acceptor fluorescence quantum yields. These parameters are usually obtained by alternating laser excitation (ALEX) for molecules in solution<sup>[41]</sup> (reviewed by Kapanidis et al.<sup>[41a]</sup>) or by acceptor photobleaching for surface-immobilized fluorophores.<sup>[42]</sup> In ALEX, the donor and acceptor are sequentially excited in a rapid, alternating fashion (Figure 4). A two-dimensional histogram of FRET efficiencies and relative donor–acceptor stoichiometries can be made, enabling molecules to be



**Figure 4.** Schematic illustration of setup used in ALEX. Bottom right: resulting two-dimensional histogram of FRET efficiencies ( $E$ ) and relative donor-acceptor stoichiometries ( $S$ ) used to sort molecules. Adapted with permission from ref. [41a]; copyright 2005, American Chemical Society.

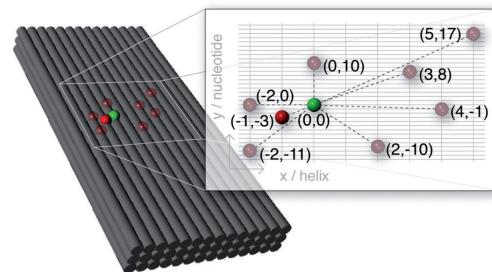
sorted into subpopulations of different physical states (Figure 4, insert). In acceptor photobleaching the corrections are obtained by comparing fluorescence intensities before and after destroying the acceptor by photobleaching.<sup>[42]</sup>

The ability to detect multiple fluorescence parameters at the single-molecule level experimentally offers vast potential in future quantitative FRET.<sup>[43]</sup> ALEX with nanosecond alternation (nsALEX)<sup>[43a]</sup>/pulsed interleaved excitation (PIE)<sup>[43b]</sup> can be used to measure lifetimes, time-resolved fluorescence anisotropies and high-temporal-resolution fluorescence correlation spectroscopy (FCS) at the single-molecule level. Seidel and co-workers have developed single-molecule multiparameter fluorescence detection (MFD)<sup>[43c-h]</sup> in which all possible experimentally ascertainable fluorescence information (intensity, lifetime and anisotropy) is collected simultaneously by time-resolved, polarized detection in both donor and acceptor emission channels (reviewed by Sismanakis et al.<sup>[43d]</sup>). In time-resolved measurements, FRET efficiencies can be calculated directly from the donor decay rates without the need for many of the correction factors required in intensity-based measurements.<sup>[43g]</sup> Recently, MFD was combined with the alternating excitation scheme of PIE, paving the way for accurate smFRET of freely diffusing molecules in a single experiment.<sup>[43h]</sup> Although single-molecule multiparameter techniques are highly promising, the main disadvantage is their complexity, requiring equipment not available in most microscopy facilities.

The determination of FRET efficiencies from ensemble measurements has been described<sup>[44]</sup> well both for ensemble spectroscopy<sup>[44a,b]</sup> and for FRET microscopy (refs. [44c-e] and references therein). Small signal changes are particularly difficult to measure accurately through conventional intensity or lifetime-based measurements, thus decreasing the accuracy when measuring long donor-acceptor distances. Widengren and co-workers recently showed that by monitoring the acceptor triplet state even very low FRET efficiencies (<5%) can be mea-

sured accurately, potentially allowing longer distances to be probed with conventional microscopy setups.<sup>[45]</sup> At very short donor-acceptor separations, on the other hand, measured FRET efficiencies are often systematically lower than expected.<sup>[41b,46]</sup> Di Fiori and Meller showed that at short separations direct dye-dye interactions can result in a mixture of quenched and unquenched dye states, leading to lower apparent measured FRET efficiencies.<sup>[46a]</sup> In single-molecule experiments, direct dye-dye interactions can be distinguished from FRET states, leading to more accurate measurements.

For quantitative measurements, a “calibration ruler” with known structure and donor-acceptor separations is valuable as a reference for testing of new FRET-based methods or for studying the influence of photophysical, structural and instrumental factors.<sup>[47]</sup> The standards employed are usually polyprolines<sup>[47a,b]</sup> or oligonucleotides<sup>[41b,46a,47d]</sup> (to reference but a few), both possessing the advantages of being simple and cheap to fabricate and allowing donor-acceptor separations to be varied in a systematic manner. The main disadvantage of polyprolines as model systems, however, is their deviation from ideal rigidity,<sup>[47a,b]</sup> whereas the quantitative analysis of FRET in oligonucleotides can be complicated by the helicity of dsDNA. Recently, a rigid DNA origami block was demonstrated as a quantitative FRET ruler (Figure 5).<sup>[47c]</sup> This model system is more complex in some aspects and increases the expense of fabrication relative to regular dsDNA, but it provides an intriguing versatility in structure design and the ability to position the dyes on the same side of the DNA surface.



**Figure 5.** DNA origami calibration ruler. Reprinted with permission from ref. [47c]; copyright 2011, Wiley-VCH.

## 5. Quantitative Analysis of smFRET

The quantitative analysis of FRET measurements is particularly challenging in single-molecule experiments as a result of low signal-to-noise ratios. The theory and methods used to analyse smFRET data are therefore under continuous development, in studies both of immobilized<sup>[48]</sup> and of freely diffusing dyes.<sup>[48i,49]</sup>

The main advantage of using immobilized dyes in smFRET is long observation times of each individual molecule (up to minutes), allowing kinetic rates between discrete states to be extracted from histograms of dwell times in the various states. However, the interpretation of intensity traces into quantitative

distances is difficult, due to shot-noise and the low numbers of photons detected in each time-bin. Addressing this issue, Watkins et al. used smFRET to derive distance distributions by combining a maximum information analysis of the photon traces and a model-free maximum-entropy-based deconvolution of the raw probability distribution histograms.<sup>[48a]</sup> Landes and co-workers have reported a denoising algorithm for smFRET trajectories of immobilized dyes based on wavelet shrinkage.<sup>[48b,c]</sup> Backović et al. recently reported a distribution-based method for generating smFRET efficiency histograms from intensity trajectories by use of Bayesian updating.<sup>[48d]</sup>

Usually, smFRET efficiency histograms are analysed by fitting a sum of weighted Gaussians, from which the average FRET efficiency is extracted. In order also to extract quantitative information, such as structural heterogeneities and molecular dynamics, from the widths and shapes of smFRET histograms, the contribution from shot-noise must be accounted for. Here one must distinguish between the real FRET efficiency and the often used proximity ratio (PR), a convenient but uncorrected measure of FRET calculated as the ratio of the acceptor intensity ( $I_A$ ) to the total intensity ( $I_A + I_D$ ) upon donor excitation:

$$\text{PR} = \frac{I_A}{I_A + I_D} \quad (1)$$

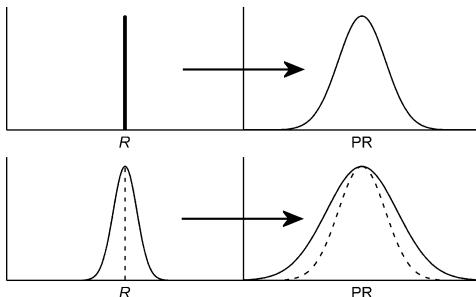
Although the PR provides a relative measure of the degree of FRET it cannot be translated directly into quantitative distances. In diffusion-based measurements, the shot-noise limited PR histogram (PRH)—that is, the histogram broadened only by shot-noise—can be calculated by using the measured burst size distribution and Poisson statistics. Any additional broadening of the PRH beyond shot-noise is then assigned to static and/or dynamic heterogeneities of the system under investigation (Figure 6). This strategy is used in PRH analysis, described by Nir et al.,<sup>[49a]</sup> and probability distribution analysis (PDA), developed by Seidel and co-workers.<sup>[49b-d]</sup> PDA has been used to model both multiple static states<sup>[49d,e]</sup> and dynamics between interconverting states<sup>[49c,f]</sup> (for a review covering PDA see Sismanakis et al.,<sup>[43c]</sup> although the model in a given study must be chosen on the basis of prior assumptions. Kapanidis and co-workers showed that dynamics can be distinguished from mul-

tiple static states by burst variance analysis (BVA), in which the standard deviation of FRET in each burst is calculated and compared to that expected from shot-noise only.<sup>[49g]</sup> Nir and co-workers also reported a statistical method based on photon arrival times, in which the internal photon distribution is analysed and used to classify bursts according to various properties of interests, thus acting as a filter to identify subpopulations in FRET and ALEX measurements.<sup>[49h]</sup> The derivation of distance distributions from diffusion-based smFRET measurements by the method of classic maximum entropy, having the advantage of analysing the data in a model-free manner, has also been demonstrated.<sup>[49i]</sup>

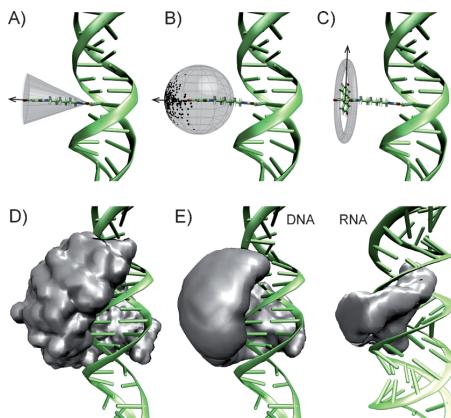
## 6. Modelling Probe Dynamics

In order to interpret measured FRET efficiencies into quantitative structural information about the nucleic acid, both the positions and the orientations (including diffusion and reorientation) of the tethered probes relative to the biomolecule must be known or somehow included in the analysis. Although  $R$  and  $\kappa^2$  might in some situations be directly correlated,<sup>[50]</sup> dye diffusion and reorientation are usually modelled separately for purposes of simplicity. The simplest way of modelling probe diffusion in ensemble measurements is by fitting the donor intensity decay with use of a Gaussian distribution of donor–acceptor distances (referred to as time-resolved FRET).<sup>[17d,51]</sup> In this case, the width and centre of the fitted distribution provide quantitative measures of dye diffusion and average donor–acceptor distance, respectively. Because the broadening of the distribution results both from tether flexibility and from global nucleic acid dynamics, time-resolved FRET can be used to derive information on the flexibility of structural elements in nucleic acids, in particular if the contribution from the linker is known (a detailed review is provided by Klostermeier and Millar<sup>[51a]</sup>). The often overlooked power of time-resolved FRET is its ability to reveal trends in the donor decay reflecting molecular dynamics and to extract more than one quantitative molecular parameter from experiments, facilitated by the large number of data points analysed in such experiments.<sup>[46b]</sup>

Although a distance distribution does not itself give information on rotational dynamics, orientation factor distributions can be implemented in time-resolved FRET in a similar manner for the description of dye reorientation (for a useful discussion see Parkhurst et al.<sup>[51b]</sup>). Traditionally, the distribution of dye orientations during the energy transfer is described by a simple wobbling-in-cone model, with semi-angles obtained from fluorescence anisotropy measurements (or directly from microscope images<sup>[52]</sup>), from which upper and lower limits of  $\langle \kappa^2 \rangle$  can be estimated (Figure 7A).<sup>[51b,53]</sup> This model was recently explored for quantitative interpretation of FRET efficiency histograms in single-molecule FRET.<sup>[54]</sup> In the cone model, the transition moment of the dye is assumed to be equally distributed within a cone. Rindermann et al. recently modelled dye orientation and reorientation by using the Mises–Fisher directional probability distribution (Figure 7B).<sup>[46b]</sup> This model describes the direction of the transition moment vector by an axial Gaussian probability density function weighted most heavily at the



**Figure 6.** Illustration of the proximity ratio histogram resulting from shot-noise only (top) and shot-noise including additional broadening due to a distribution of D–A distances (bottom).



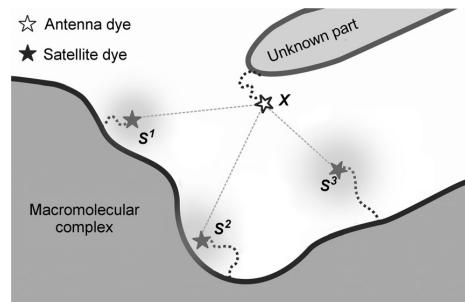
**Figure 7.** Modelling probe dynamics. A–C) Representing dye reorientation by: A) a wobbling-in-cone model, B) the Mises–Fisher distribution, and C) a distribution-in-disc model. D, E) Illustration of tethered dye diffusion modelled by: D) MD simulation, and E) an accessible volume model. Coordinate files used in (D) and (E) were courteously provided by Prof. Claus A. M. Seidel and Simon Sindbert.<sup>[18k]</sup>

mean orientation of the dye. Through the use of the direction and distribution width of the transition dipole vectors as fitting parameters, this strategy provided estimates of the three-dimensional positions, orientations and rotational dynamics of the dyes in DNA.<sup>[46b]</sup> For dyes with transition moments perpendicular to the linker axis—FAM and TAMRA, for example—a distribution-in-disc model might be more physically accurate (Figure 7C).<sup>[18k]</sup>

The downside of using theoretical distance or orientation distribution models in describing probe diffusion and reorientation is that the inherent assumptions imposed by such models—that the distributions are isotropic and uncorrelated, for example—are nearly impossible to validate in practice. If the local structure of the nucleic acid around the tether is known, the exact positions and orientations of the probes can be predicted by use of MD simulations (Figure 7D).<sup>[17a, 18k, 50, 55]</sup> MD simulations provide the most detailed insight into probe position, orientation and dynamics by offering an atomic-level view of distance and  $\kappa^2$  distributions, as well as any correlation between the two. MD simulations, however, are still rather cumbersome and time-consuming for everyday use. Recently, very simple and fast geometric search algorithms have been implemented for the prediction of dye<sup>[18k]</sup> and spin label<sup>[6b, 56]</sup> positions in DNA. The accessible volume (AV) of the dye can be determined simply by calculating the volume within the extended linker-dye radius and the point of tether, excluding all positions resulting in a steric clash between the tethered dye and the nucleic acid surface (Figure 7E).<sup>[18k]</sup> Assuming that all dye positions within the AV are equally populated provides a reasonable representation of the dye, as shown by Sindbert et al. (this recent paper presents a very thorough investigation of fluorophore-linker fluctuations in DNA and RNA<sup>[18k]</sup>). Alternatively, more customized Metropolis-Monte-Carlo simulations have also been used to model tether flexibility in DNA.<sup>[57]</sup>

## 7. Emerging FRET-Based Techniques

An interesting development in quantitative FRET is the use of multiple FRET-based distance measurements to obtain three-dimensional structural information.<sup>[55a, 58]</sup> By exploitation of partially known geometrical elements of the biomolecular structure, the three-dimensional position of an unknown structural element can be determined accurately by means of triangulation/trilateration.<sup>[55a, 58b–j]</sup> Triangulation-based FRET techniques exploit the concept of three or more “satellite dyes”, positioned in the known part of the structure, in determining the position of one or more “antenna dyes” (Figure 8). Using FRET

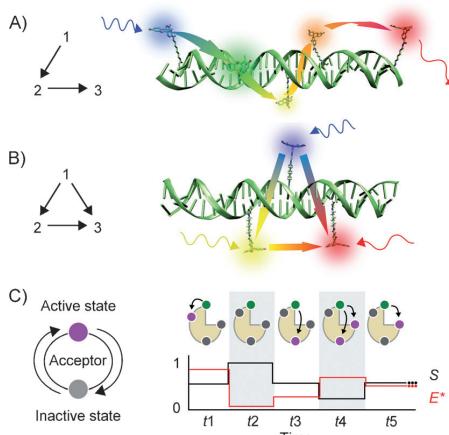


**Figure 8.** Schematic illustration of antenna and satellite dyes used in FRET triangulation. Adapted with permission from ref. [58b]; copyright 2008, Macmillan Publishers Ltd.

triangulation, Wozniak et al. determined the three-dimensional geometries of adenine bulges of various sizes by combining several donor and acceptor positions across the B-DNA helical arms.<sup>[55a]</sup> Muschielok et al. have developed a “nano-positioning system” (NPS) for the analysis of data from FRET triangulation measurements with the aid of Bayesian parameter estimation.<sup>[58b, c]</sup> By accounting for many of the uncertainties in FRET experiments, the NPS provides an informative picture of the information obtainable from the experiment by producing probability distributions of fluorophore positions. Using FRET triangulation and the NPS, Andrecka et al. determined the position of nascent RNA exiting RNA polymerase II,<sup>[58d]</sup> as well as the positions of the non-template and upstream DNA in the same complex.<sup>[58e]</sup> More recently, Balci et al. used FRET triangulation and the NPS to probe the position of the ss/dsDNA junction in a helicase–DNA complex,<sup>[58f]</sup> and Treutlein et al. determined the three-dimensional architecture of an RNA polymerase II open promoter complex.<sup>[58g]</sup> Using a slightly different approach, Sabir et al. combined several smFRET measurements with MD simulations in determining the global structure of a four-stranded DNA fork<sup>[58h]</sup> and, more recently, a three-way DNA junction.<sup>[58i]</sup>

The simultaneous combination of more than one donor–acceptor pair can be exploited for multistep and multicolour FRET (put in a general theoretical framework by Watrob et al.<sup>[59]</sup>). Through the introduction of mediator dyes in between the donor and acceptor, energy has been transferred

over distances longer than 20 nm by multistep FRET (Figure 9A).<sup>[60]</sup> The use of such systems as photonic wires,<sup>[60a-c]</sup> photonic switches<sup>[60d-f]</sup> or as a trick for long-range FRET mea-



**Figure 9.** Illustration of: A) multistep FRET, B) multicolor FRET (represented by three-color ALEX<sup>[61f]</sup>), and C) switchable FRET.  $S$  is the donor/acceptor stoichiometry and  $E^*$  is the apparent FRET efficiency. Part (C) is adapted with permission from ref. [62a]; copyright 2010, Macmillan Publishers Ltd.

urements<sup>[60g]</sup> has been suggested. Labelling with more than one donor–acceptor pair (multicolour FRET) can be used to monitor multiple distances simultaneously (Figure 9B).<sup>[61]</sup> Multicolour FRET can be used in ensemble measurements to limit the number of samples needed<sup>[61a-c]</sup> but is particularly powerful in single-molecule experiments for the simultaneous monitoring of multiple structural elements in real time without hiding molecular heterogeneity and dynamics behind an ensemble average.<sup>[61d,f-i]</sup>

The limitation of multicolour FRET is its complexity. After excitation a cascade of events occurs, complicating quantitative data analysis. In addition, multicolour FRET requires several fluorophores that are spectrally separated but have sufficiently large spectral overlaps for FRET. To address these issues, Kapandidis and co-workers reported switchable FRET in which photo-switchable dyes are used to confine individual donor–acceptor pairs temporarily within a multilabelled system.<sup>[62]</sup> In switchable FRET, single FRET pairs are sequentially activated and deactivated by stochastically switching the acceptor dyes between active and inactive states (Figure 9C). Monitoring the stoichiometry between active donor and acceptor dyes in each biomolecule by ALEX allows individual FRET states to be resolved in real time.<sup>[62a]</sup>

## 8. FRET Simulation and Analysis Software

The customized software developed in individual laboratories for the simulation and analysis of FRET is occasionally made publicly available.<sup>[63]</sup> Table 2 lists a selection of such software.

**Table 2.** Selected software packages for the simulation and analysis of FRET data.

Name	Capabilities	Free	Link
TwoTone	automated ALEX and TIRF-FRET image analysis <sup>[63a]</sup>	MATLAB-based	[63b]
CPLC package	construction of smFRET traces and histograms from TIRF-FRET measurements <sup>[5d]</sup>	MATLAB-based	[63c]
vbFRET	identification of states and kinetic rates from sm-FRET trajectories by HMM and ME <sup>[48e]</sup>	MATLAB-based	[63f]
HaMMY	determination of kinetic rates from smFRET trajectories by HMM and ML <sup>[48f]</sup>	yes	[63d]
QuB	determination of kinetic rates from smFRET trajectories by HMM and ML	yes	[63e]
CSM	identification of states from smFRET trajectories by causal state modelling <sup>[48g]</sup>	MATLAB-based	[63i]
FRETshrink	denoise smFRET trajectories by wavelet shrinkage <sup>[48b]</sup>	MATLAB-based	[63j]
FretTrace	reconstruction of distance trajectories from smFRET burst trajectories by ML <sup>[49k]</sup>	yes	[63k]
Tatiana	analysis of smFRET histograms by PDA; part of software for MFD analysis <sup>[49b]</sup>	no	[63l]
exiFRET	simulation of FRET in complex geometries <sup>[63m]</sup>	yes	[63n]
FRETmatrix	simulation and analysis of FRET in three-dimensional nucleic acid geometries	MATLAB-based	[63o]
FRETlab	simulation of time-resolved FRET by use of the Mises–Fisher distribution <sup>[46b]</sup>	IGOR Pro-based	[63p]
FRETsg	translation of multiple FRET distances into three-dimensional positions <sup>[58g]</sup>	yes	[63q]
FRETnps	simulation and analysis of FRET by the NPS triangulation method <sup>[58b]</sup>	MATLAB runtime	[63r]

Some freeware packages offer smFRET trajectories and histograms from raw TIRF image data, such as TwoTone from the Kapandidis lab<sup>[63b]</sup> and the data acquisition and analysis software package from the Ha lab.<sup>[5d, 63c]</sup> The analysis of smFRET trajectories into kinetic rates between different conformational states can be accomplished by hidden Markov modelling (HMM) with a maximum likelihood (ML) optimisation as in HaMMY<sup>[48f, 63d]</sup> and QuB<sup>[63e]</sup> or a maximum evidence (ME) optimisation as in vbFRET<sup>[48e, 63f]</sup> (smFRET analysis by HMM is reviewed by Blanco and Walter<sup>[64j]</sup>). Software packages for analysing bursts of diffusing molecules are available from a number of commercial suppliers: SymPhoTime from PicoQuant<sup>[63g]</sup> and Burst Analyzer from Becker & Hickl, for example.<sup>[63h]</sup> Analysis of MFD data, including the analysis of smFRET histograms by PDA,<sup>[49b]</sup> can be performed with a package from the Seidel lab.<sup>[63l]</sup> Corry and co-workers have developed exiFRET, a web-based platform for simulating FRET between probes distributed in various complex geometries.<sup>[63m,n]</sup> Our own software package, FRETmatrix, provides a general framework for the simulation and analysis of FRET in nucleic acids.<sup>[63o]</sup> FRETmatrix directly correlates three-dimensional nucleic acid structures with FRET efficiencies and can be used to analyse both nucleic acid structure and dynamics from multiple time-resolved FRET measurements. FRETmatrix is particularly powerful when the positions and orientations of the probes are highly or partly constrained, such as in base-base FRET.

## 9. Summary and Outlook

We have summarised recent methodological advances in the use of FRET as a quantitative molecular ruler, with particular focus on studies of nucleic acid structure and dynamics. New knowledge on the relative positions, orientations and dynamics of several commercially available dyes attached to DNA should aid the design and interpretation of FRET experiments (Section 2). Alternative FRET probes designed with well-defined positions and orientations within DNA now offer interesting alternatives for more specialized studies such as retrieval of orientation information and site-specific probing of nucleobases (Section 3). FRET can now be measured accurately at the single-molecule level even for diffusing molecules in solution, and this field is pushed further by the development of multi-parameter fluorescence techniques capable of monitoring several dye properties simultaneously (Section 4). The quantitative analysis of smFRET signals into donor–acceptor distance distributions has the potential to provide quantitative insight into structural and dynamic heterogeneities of DNA and RNA systems (Section 5), further aided by new methods optimised to model local dipole fluctuations (Section 6). The systematic probing of multiple FRET pair positions in DNA/RNA structures can be used to triangulate the three-dimensional positions of unknown structural elements, whereas multicolour FRET can now be used to monitor two or more distances simultaneously at the single-molecule level (Section 7).

We believe that continuing progress in quantitative FRET-based methods will be greatly catalysed by sharing of the knowledge and specialized software developed in individual laboratories (Section 8). As such, it is our hope that the overview provided here will inspire further development and use of quantitative FRET as a tool complementing more complex, higher-resolution techniques in the study of nucleic acids.

## Acknowledgements

S.P. gratefully acknowledges Assoc. Prof. Kristine Kilså and Prof. Bo Albinsson for their support and supervision. This work is funded by the Swedish Research Council (VR) and the Danish Council for Independent Research and Natural Sciences (FNU).

**Keywords:** DNA structures • fluorescence spectroscopy • fluorescent probes • FRET • nucleic acids

- [1] S. R. Holbrook, *Annu. Rev. Biophys.* **2008**, *37*, 445–464.
- [2] M. P. Foster, C. A. McElroy, C. D. Amero, *Biochemistry* **2007**, *46*, 331–340.
- [3] R. P. Rambo, J. A. Tainer, *Curr. Opin. Struct. Biol.* **2010**, *20*, 128–137.
- [4] J. Frank, *Q. Rev. Biophys.* **2009**, *42*, 139–158.
- [5] a) E. T. Mollova, *Curr. Opin. Chem. Biol.* **2002**, *6*, 823–828; b) S. A. McKinney, E. Tan, T. J. Wilson, M. K. Nahas, A. C. Declais, R. M. Clegg, D. M. Lilley, T. Ha, *Biochem. Soc. Trans.* **2004**, *32*, 41–45; c) P. R. Selvin, T. Ha, *Single-Molecule Techniques: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, **2008**; d) R. Roy, S. Hohng, T. Ha, *Nat. Methods* **2008**, *5*, 507–516; e) C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, T. Ha, *Annu. Rev. Biochem.* **2008**, *77*, 51–76.
- [6] a) Q. Cai, A. K. Kusnetzow, W. L. Hubbell, I. S. Haworth, G. P. Gacho, N. Van Eps, K. Hideg, E. J. Chambers, P. Z. Qin, *Nucleic Acids Res.* **2006**, *34*, 4722–4730; b) Q. Cai, A. K. Kusnetzow, K. Hideg, E. A. Price, I. S. Ha-

worth, P. Z. Qin, *Biophys. J.* **2007**, *93*, 2110–2117; c) A. Marko, V. Deny-senkov, D. Margraf, P. Cekan, O. Schiemann, S. T. Sigurdsson, T. F. Prisner, *J. Am. Chem. Soc.* **2011**, *133*, 13375–13379; d) O. Schiemann, N. Piton, Y. Mu, G. Stock, J. W. Engels, T. F. Prisner, *J. Am. Chem. Soc.* **2004**, *126*, 5722–5729; e) O. Schiemann, T. F. Prisner, *Q. Rev. Biophys.* **2007**, *40*, 1–53.

- [7] S. T. Sigurdsson, *Pure Appl. Chem.* **2011**, *83*, 677–686.
- [8] a) B. M. Reinhard, S. Sheikholeslami, A. Mastroianni, A. P. Alivisatos, J. Liphardt, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2667–2672; b) B. M. Reinhard, M. Siu, H. Agarwal, A. P. Alivisatos, J. Liphardt, *Nano Lett.* **2005**, *5*, 2246–2252; c) C. Sönnichsen, B. M. Reinhard, J. Liphardt, A. P. Alivisatos, *Nat. Biotechnol.* **2005**, *23*, 741–745.
- [9] a) A. J. Mastroianni, D. A. Sivak, P. L. Geissler, A. P. Alivisatos, *Biophys. J.* **2009**, *97*, 1408–1417; b) R. S. Mathew-Fenn, R. Das, J. A. Silverman, P. A. Walker, P. A. Harbury, *PLoS one* **2008**, *3*, e3229.
- [10] a) J. Griffin, A. K. Singh, D. Senapati, P. Rhodes, K. Mitchell, B. Robinson, E. Yu, P. C. Ray, *Chemistry* **2009**, *15*, 342–351; b) T. L. Jennings, M. P. Singh, G. F. Strouse, *J. Am. Chem. Soc.* **2006**, *128*, 5462–5467; c) J. Seelig, K. Leslie, A. Renn, S. Kuhn, V. Jacobsen, M. van de Corput, C. Wyman, V. Sandoghdar, *Nano Lett.* **2007**, *7*, 685–689; d) M. P. Singh, T. L. Jennings, G. F. Strouse, *J. Phys. Chem. B* **2009**, *113*, 552–558.
- [11] S. Lal, N. K. Grady, G. P. Goodrich, N. J. Halas, *Nano Lett.* **2006**, *6*, 2338–2343.
- [12] a) P. Wu, L. Brand, *Anal. Biochem.* **1994**, *218*, 1–13; b) D. M. J. Lilley, T. J. Wilson, *Curr. Opin. Chem. Biol.* **2000**, *4*, 507–517; c) J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Vol. 3, 3rd ed., Springer, New York, **2006**; d) D. M. J. Lilley, *Methods Enzymol.* **2009**, *469*, 159–187; e) H. E. Grecco, P. J. Verveer, *ChemPhysChem* **2011**, *12*, 484–490; f) H. Sahoo, *J. Photochem. Photobiol. C* **2011**, *12*, 20–30.
- [13] A. Penzkofer, J. Wiedmann, *Opt. Commun.* **1980**, *35*, 81–86.
- [14] a) A. I. Murchie, R. M. Clegg, E. von Kitzing, D. R. Duckett, S. Diekmann, D. M. Lilley, *Nature* **1989**, *341*, 763–766; b) R. M. Clegg, A. I. Murchie, A. Zechel, D. M. Lilley, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2994–2998; c) R. M. Clegg, A. I. Murchie, D. M. Lilley, *Biophys. J.* **1994**, *66*, 99–109; d) C. Gohlke, A. I. Murchie, D. M. Lilley, R. M. Clegg, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11660–11664; e) F. Stühmeier, J. B. Welch, A. I. Murchie, D. M. Lilley, R. M. Clegg, *Biochemistry* **1997**, *36*, 13530–13538.
- [15] a) R. M. Clegg, A. I. Murchie, A. Zechel, C. Carlberg, S. Diekmann, D. M. Lilley, *Biochemistry* **1992**, *31*, 4846–4856; b) L. Wang, A. K. Gaigalas, J. Blasic, M. J. Holden, *Spectrochim. Acta Part A* **2004**, *60*, 2741–2750; c) J. R. Unruh, G. Gokulrangan, G. H. Lushington, C. K. Johnson, G. S. Wilson, *Biophys. J.* **2005**, *88*, 3455–3465; d) J. E. Noble, L. Wang, K. D. Cole, A. K. Gaigalas, *Biophys. Chem.* **2005**, *113*, 255–263; e) Q. H. Xu, S. Wang, D. Korystov, A. Mikhailovsky, G. C. Bazan, D. Moses, A. J. Heeger, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 530–535.
- [16] a) L. Song, E. J. Hennink, I. T. Young, H. J. Tanke, *Biophys. J.* **1995**, *68*, 2588–2600; b) R. Sjöback, J. Nygren, M. Kubista, *Spectrochim. Acta Part A* **1995**, *51*, L7–L21; c) J. R. Unruh, G. Gokulrangan, G. S. Wilson, C. K. Johnson, *Photochem. Photobiol.* **2005**, *81*, 682–690; d) I. Nazareno, R. Pires, B. Lowe, M. Obaidy, A. Rashtchian, *Nucleic Acids Res.* **2002**, *30*, 2089–2195; e) M. Torimura, S. Kurata, K. Yamada, T. Yokomaku, Y. Kamagata, T. Kanagawa, R. Kurane, *Anal. Sci.* **2001**, *17*, 155–160.
- [17] a) E. Dolghih, A. E. Roitberg, J. L. Krause, *J. Photochem. Photobiol. A* **2007**, *190*, 321–327; b) M. Lorenz, A. Hillisch, S. D. Goodman, S. Diekmann, *Nucleic Acids Res.* **1999**, *27*, 4619–4625; c) M. Lorenz, A. Hillisch, D. Payet, M. Buttinielli, A. Travers, S. Diekmann, *Biochemistry* **1999**, *38*, 12150–12158; d) R. A. Hochstrasser, S. M. Chen, D. P. Millar, *Biophys. Chem.* **1992**, *45*, 133–141; e) A. Dietrich, V. Buschmann, C. Müller, M. Sauer, *J. Biotechnol.* **2002**, *82*, 211–231; f) G. Vámosi, C. Gohlke, R. M. Clegg, *Biophys. J.* **1996**, *71*, 972–994; g) C. Eggeling, J. R. Fries, L. Brand, R. Günther, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1556–1561; h) H. Neubauer, N. Gaiko, S. Berger, J. Schaffer, C. Eggeling, J. Tuma, L. Verdier, C. A. Seidel, C. Griesinger, A. Volkmer, *J. Am. Chem. Soc.* **2007**, *129*, 12746–12755; i) T. Heinlein, J.-P. Klemeyer, O. Piester, M. Sauer, *J. Phys. Chem. B* **2003**, *107*, 7957–7964; j) Y. Ohya, K. Yabuki, T. Ouchi, *Supramol. Chem.* **2003**, *15*, 149–154; k) E. M. Stennett, G. Kodis, M. Levitus, *ChemPhysChem* **2012**, *13*, 909–913.
- [18] a) M. Levitus, S. Ranjit, *Q. Rev. Biophys.* **2011**, *44*, 123–151; b) D. G. Norman, R. J. Grainger, D. Uhrin, D. M. Lilley, *Biochemistry* **2000**, *39*, 6317–6324; c) L. Urnavicius, S. A. McPhee, D. M. Lilley, D. G. Norman, *Biophys. J.* **2012**, *102*, 561–568; d) M. E. Sanborn, B. K. Connolly, K. Guru-

- nathan, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 11064–11074; e) J. Ouellet, S. Schorr, A. Iqbal, T. J. Wilson, D. M. Lilley, *Biophys. J.* **2011**, *101*, 1148–1154; f) J. Spiriti, J. K. Binder, M. Levitus, A. van der Vaart, *Biophys. J.* **2011**, *100*, 1049–1057; g) M. Massey, W. R. Algar, U. J. Krull, *Anal. Chim. Acta* **2006**, *568*, 181–189; h) B. J. Harvey, C. Perez, M. Levitus, *Photochem. Photobiol. Sci.* **2009**, *8*, 1105–1110; i) C. R. Sabanayagam, J. S. Eid, A. Meller, *J. Chem. Phys.* **2005**, *122*, 061103; j) X. Li, Y. Yin, X. Yang, Z. Zhi, X. S. Zhao, *Chem. Phys. Lett.* **2011**, *513*, 271–275; k) S. Sindbert, S. Kalinini, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller, C. A. Seidel, *J. Am. Chem. Soc.* **2011**, *133*, 2463–2480; l) A. Iqbal, L. Wang, K. C. Thompson, D. M. Lilley, D. G. Norman, *Biochemistry* **2008**, *47*, 7857–7862; m) B. J. Harvey, M. Levitus, *J. Fluoresc.* **2009**, *19*, 443–448.
- [19] a) P. F. Aramendia, R. M. Negri, E. San Roman, *J. Phys. Chem.* **1994**, *98*, 3165–3173; b) A. K. Chibisov, G. V. Zakharova, H. Goerner, Yu. A. Sogolyaev, I. L. Muskhali, A. I. Tolmachev, *J. Phys. Chem.* **1995**, *99*, 886–893.
- [20] a) M. Cooper, A. Ebner, M. Briggs, M. Burrows, N. Gardner, R. Richardson, R. West, *J. Fluoresc.* **2004**, *14*, 145–150; b) L. Linck, P. Kapusta, U. Resch-Genger, *Photochem. Photobiol.* **2012**, *88*, 867–875; c) L. Linck, U. Resch-Genger, *Eur. J. Med. Chem.* **2010**, *45*, 5561–5566.
- [21] A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, D. M. Lilley, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11176–11181.
- [22] B. A. Armitage in *DNA Binders and Related Subjects*, Vol. 253 (Eds.: J. B. Chaires, M. J. Waring), **2005**, Springer, Berlin, pp. 55–76.
- [23] a) J. Vogelsang, T. Cordes, P. Tinnefeld, *Photochem. Photobiol. Sci.* **2009**, *8*, 486–496; b) R. Zhu, X. Li, X. S. Zhao, A. Yu, *J. Phys. Chem. B* **2011**, *115*, 5001–5007; c) A. Kupstat, T. Ritschel, M. U. Kumke, *Bioconjugate Chem.* **2011**, *22*, 2546–2557; d) V. Buschmann, K. D. Weston, M. Sauer, *Bioconjugate Chem.* **2003**, *14*, 195–204.
- [24] a) R. Sjöback, J. Nygren, M. Kubista, *Biopolymers* **1998**, *46*, 445–453; b) T. Kaji, S. Ito, S. Iwai, H. Miyasaka, *J. Phys. Chem. B* **2009**, *113*, 13917–13925; c) R. F. Delgadillo, L. J. Parkhurst, *Photochem. Photobiol.* **2010**, *86*, 261–272.
- [25] a) A. Ivanova, G. Jezierski, E. Vladimirov, N. Rosch, *Biomacromolecules* **2007**, *8*, 3429–3438; b) M. Kabeláč, F. Zimandl, T. Fessl, Z. Chval, F. Lánkaš, *Phys. Chem. Chem. Phys.* **2010**, *12*, 9677–9684.
- [26] a) P. Sandin, K. Borjesson, H. Li, J. Martensson, T. Brown, L. M. Wilhelmsson, B. Albinsson, *Nucleic Acids Res.* **2008**, *36*, 157–167; b) P. Sandin, L. M. Wilhelmsson, P. Lincoln, V. E. Powers, T. Brown, B. Albinsson, *Nucleic Acids Res.* **2005**, *33*, 5019–5025; c) L. M. Wilhelmsson, P. Sandin, A. Holmén, B. Albinsson, P. Lincoln, B. Nordén, *J. Phys. Chem. B* **2003**, *107*, 9094–9101; d) S. Preus, K. Borjesson, K. Kilsa, B. Albinsson, L. M. Wilhelmsson, *J. Phys. Chem. B* **2010**, *114*, 1050–1056.
- [27] a) K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeret, P. Lincoln, B. Norden, B. Albinsson, L. M. Wilhelmsson, *Nucleic Acids Res.* **2004**, *32*, 5087–5095; b) S. Preus, K. Kilsa, L. M. Wilhelmsson, B. Albinsson, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8881–8892; c) K. Borjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson, L. M. Wilhelmsson, *J. Am. Chem. Soc.* **2009**, *131*, 4288–4293.
- [28] List: <http://www.glenresearch.com/Reference/oligohouses.html>.
- [29] a) C. R. Sabanayagam, J. S. Eid, A. Meller, *J. Chem. Phys.* **2005**, *123*, 224708; b) J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, P. Tinnefeld, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8107–8112; c) I. Rasnik, S. A. McKinney, T. Ha, *Nat. Methods* **2006**, *3*, 891–893; d) R. B. Altman, D. S. Terry, Z. Zhou, Q. Zheng, P. Geggier, R. A. Kolster, Y. Zhao, J. A. Javitch, J. D. Warren, S. C. Blanchard, *Nat. Methods* **2012**, *9*, 68–71; e) T. Cordes, J. Vogelsang, P. Tinnefeld, *J. Am. Chem. Soc.* **2009**, *131*, 5018–5019; f) L. A. Campos, J. Liu, X. Wang, R. Ramanathan, D. S. English, V. Munoz, *Nat. Methods* **2011**, *8*, 143–146; g) G. T. Dempsey, M. Bates, W. E. Kowtoniuk, D. R. Liu, R. Y. Tsien, X. Zhuang, *J. Am. Chem. Soc.* **2009**, *131*, 18192–18193; h) T. Kottke, S. van de Linde, M. Sauer, S. Kakorin, M. Heilemann, *J. Phys. Chem. Lett.* **2010**, *1*, 3156–3159; i) J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, P. Tinnefeld, *Angew. Chem.* **2008**, *120*, 5545–5550; *Angew. Chem. Int. Ed.* **2008**, *47*, 5465–5469.
- [30] M. Bates, B. Huang, X. Zhuang, *Curr. Opin. Chem. Biol.* **2008**, *12*, 505–514.
- [31] K. E. Sapsford, L. Berti, I. L. Medintz, *Angew. Chem.* **2006**, *118*, 4676–4704; *Angew. Chem. Int. Ed.* **2006**, *45*, 4562–4589.
- [32] B. N. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science* **2006**, *312*, 217–224.
- [33] a) G. Heiss, V. Lapiene, F. Kukolka, C. M. Niemeyer, C. Brauchle, D. C. Lamb, *Small* **2009**, *5*, 1169–1175; b) F. Kukolka, C. M. Niemeyer, *Org. Biomol. Chem.* **2004**, *2*, 2203–2206; c) F. Kukolka, B. K. Muller, S. Paternoster, A. Arndt, C. M. Niemeyer, C. Brauchle, D. C. Lamb, *Small* **2006**, *2*, 1083–1089; d) F. Kukolka, O. Schoeps, U. Woggon, C. M. Niemeyer, *Bioconjugate Chem.* **2007**, *18*, 621–627; e) V. Lapiene, F. Kukolka, K. Kiko, A. Arndt, C. M. Niemeyer, *Bioconjugate Chem.* **2010**, *21*, 921–927.
- [34] a) W. R. Algar, U. J. Krull, *Anal. Bioanal. Chem.* **2008**, *391*, 1609–1618; b) U. Resch-Genger, M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke, T. Nann, *Nat. Methods* **2008**, *5*, 763–775; c) I. L. Medintz, H. Mattoussi, *Phys. Chem. Chem. Phys.* **2009**, *11*, 17–45.
- [35] a) P. R. Selvin, *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 275–302; b) P. R. Selvin, J. E. Hearst, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10024–10028; c) A. N. Kapanidis, Y. W. Ebright, R. D. Ludescher, S. Chan, R. H. Ebright, *J. Mol. Biol.* **2001**, *312*, 453–468; d) F. Yuan, L. Griffin, L. Phelps, V. Buschmann, K. Weston, N. L. Greenbaum, *Nucleic Acids Res.* **2007**, *35*, 2833–2845; e) F. Yuan, N. L. Greenbaum, *Methods* **2010**, *52*, 173–179.
- [36] a) D. J. Hurley, Y. Tor, *J. Am. Chem. Soc.* **2002**, *124*, 13231–13241; b) A. Fegan, P. S. Shirude, S. Balasubramanian, *Chem. Commun.* **2008**, 2004–2006; c) T. J. Bandy, A. Brewer, J. R. Burns, G. Marth, T. Nguyen, E. Stulz, *Chem. Soc. Rev.* **2011**, *40*, 138–148; d) R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579–2619; e) L. M. Hall, M. Gerowska, T. Brown, *Nucleic Acids Res.* **2012**; DOI: 10.1093/nar/gks303. f) F. D. Lewis, L. Zhang, X. Zuo, *J. Am. Chem. Soc.* **2005**, *127*, 10002–10003; g) S. Ranjit, K. Gurunathan, M. Levitus, *J. Phys. Chem. B* **2009**, *113*, 7861–7866.
- [37] a) G. Stengel, J. P. Gill, P. Sandin, L. M. Wilhelmsson, B. Albinsson, B. Norden, D. Millar, *Biochemistry* **2007**, *46*, 12289–12297; b) L. M. Wilhelmsson, A. Holmen, P. Lincoln, P. E. Nielsen, B. Norden, *J. Am. Chem. Soc.* **2001**, *123*, 2434–2435; c) Y. Xie, T. Maxson, Y. Tor, *J. Am. Chem. Soc.* **2010**, *132*, 11896–11897.
- [38] L. M. Wilhelmsson, *Rev. Biophys.* **2010**, *43*, 159–183.
- [39] A. Muñoz-Losa, C. Curutchet, B. P. Krueger, L. R. Hartsell, B. Mennucci, *Biophys. J.* **2009**, *96*, 4779–4788.
- [40] J. Hohlein, K. Gryte, M. Heilemann, A. N. Kapanidis, *Phys. Biol.* **2010**, *7*, 031001.
- [41] a) A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, S. Weiss, *Acc. Chem. Res.* **2005**, *38*, 523–533; b) N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright, S. Weiss, *Biophys. J.* **2005**, *88*, 2939–2953; c) A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8936–8941.
- [42] T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, S. Weiss, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 893–898.
- [43] a) T. A. Laurence, X. Kong, M. Jager, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17348–17353; b) B. K. Müller, E. Zaychikov, C. Brauchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522; c) E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, C. A. Seidel, *Methods Enzymol.* **2010**, *475*, 455–514; d) R. Kühnemuth, C. A. M. Seidel, *Single Mol.* **2001**, *2*, 251–254; e) C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, C. A. Seidel, *J. Phys. Chem. A* **2006**, *110*, 2979–2995; f) J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken, C. A. Seidel, *Anal. Chem.* **2006**, *78*, 2039–2050; g) P. J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B. M. Wohrl, T. Restle, R. S. Goody, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1655–1660; h) V. Kudryavtsev, M. Sikor, S. Kalinin, D. Mokranjac, C. A. Seidel, D. C. Lamb, *ChemPhysChem* **2012**, *13*, 1060–1078.
- [44] a) R. M. Clegg, *Methods Enzymol.* **1992**, *211*, 353–388; b) Z. K. Majumdar, R. Hickerson, H. F. Noller, R. M. Clegg, *J. Mol. Biol.* **2005**, *351*, 1123–1145; c) E. A. Jares-Erijman, T. M. Jovin, *Nat. Biotechnol.* **2003**, *21*, 1387–1395; d) E. A. Jares-Erijman, T. M. Jovin, *Curr. Opin. Chem. Biol.* **2006**, *10*, 409–416; e) Y. Sun, H. Wallrabe, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B. M. Wohrl, T. Restle, R. S. Goody, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 462–474.
- [45] H. Hevekerl, T. Spielmann, A. Chmyrov, J. Widengren, *J. Phys. Chem. B* **2011**, *115*, 13360–13370.
- [46] a) N. Di Fiori, A. Meller, *Biophys. J.* **2010**, *98*, 2265–2272; b) J. J. Riedermann, Y. Akhtman, J. Richardson, T. Brown, P. G. Lagoudakis, *J. Am. Chem. Soc.* **2010**, *132*, 279–285; c) H. Sahoo, D. Roccatano, A. Hennig, W. M. Nau, *J. Am. Chem. Soc.* **2007**, *129*, 9762–9772.
- [47] a) B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2754–2759; b) R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1060–1065.

- 2007**, **104**, 18964–18969; c) I. H. Stein, V. Schuller, P. Bohm, P. Tinnefeld, T. Liedl, *ChemPhysChem* **2011**, **12**, 689–695; d) J. J. McCann, U. B. Choi, L. Zheng, K. Weninger, M. E. Bowen, *Biophys. J.* **2010**, **99**, 961–970.
- [48] a) L. P. Watkins, H. Chang, H. Yang, *J. Phys. Chem. A* **2006**, **110**, 5191–5203; b) J. N. Taylor, C. F. Landes, *J. Phys. Chem. B* **2011**, **115**, 1105–1114; c) J. N. Taylor, D. E. Makarov, C. F. Landes, *Biophys. J.* **2010**, **98**, 164–173; d) M. Backovic, E. S. Price, C. K. Johnson, J. P. Ralston, *J. Chem. Phys.* **2011**, **134**, 145101; e) J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez, Jr., C. H. Wiggins, *Biophys. J.* **2009**, **97**, 3196–3205; f) S. A. McKinney, C. Joo, T. Ha, *Biophys. J.* **2006**, **91**, 1941–1951; g) D. Kelly, M. Dillingham, A. Hudson, K. Wiesner, *PLoS one* **2012**, **7**, e29703; h) I. V. Gopich, A. Szabo, *J. Phys. Chem. B* **2009**, **113**, 10965–10973; i) I. V. Gopich, A. Szabo, *J. Chem. Phys.* **2005**, **122**, 014707.
- [49] a) E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, S. Weiss, *J. Phys. Chem. B* **2006**, **110**, 22103–22124; b) M. Antonik, S. Felekyan, A. Gaiduk, C. A. Seidel, *J. Phys. Chem. B* **2006**, **110**, 6970–6978; c) S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, C. A. Seidel, *J. Phys. Chem. B* **2010**, **114**, 7983–7995; d) S. Kalinin, S. Felekyan, A. Valeri, C. A. Seidel, *J. Phys. Chem. B* **2008**, **112**, 8361–8374; e) A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Toth, J. Langowski, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **2009**, **106**, 15308–15313; f) Y. Santoso, J. P. Torella, A. N. Kapanidis, *ChemPhysChem* **2010**, **11**, 2209–2219; g) J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, *Biophys. J.* **2011**, **100**, 1568–1577; h) T. E. Tomov, R. Tsukanov, R. Masoud, M. Liber, N. Plavner, E. Nir, *Biophys. J.* **2012**, **102**, 1163–1173; i) M. S. Devore, S. F. Gull, C. K. Johnson, *J. Phys. Chem. B* **2012**, **116**, 4006–4015; j) A. Hoffmann, D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, B. Schuler, *Phys. Chem. Chem. Phys.* **2011**, **13**, 1857–1871; k) G. F. Schröder, H. Grubmüller, *J. Chem. Phys.* **2003**, **119**, 9920; l) I. V. Gopich, A. Szabo, *J. Phys. Chem. B* **2007**, **111**, 12925–12932.
- [50] D. B. VanBeek, M. C. Zwier, J. M. Shorb, B. P. Krueger, *Biophys. J.* **2007**, **92**, 4168–4178.
- [51] a) D. Klostermeier, D. P. Millar, *Biopolymers* **2001**, **61**, 159–179; b) L. J. Parkhurst, K. M. Parkhurst, R. Powell, J. Wu, S. Williams, *Biopolymers* **2001**, **61**, 180–200; c) L. J. Parkhurst, *Methods Enzymol.* **2004**, **379**, 235–262; d) P. Sandin, P. Lincoln, B. Albinsson, *J. Phys. Chem. C* **2008**, **112**, 13089–13094; e) R. Lamichhane, G. M. Daubner, J. Thomas-Crusells, S. D. Auweter, C. Manatschal, K. S. Austin, O. Valniuk, F. H. Allain, D. Rueda, *Proc. Natl. Acad. Sci. USA* **2010**, **107**, 4105–4110.
- [52] B. Corry, D. Jayatilaka, B. Martinac, P. Rigby, *Biophys. J.* **2006**, **91**, 1032–1045.
- [53] a) K. Kinoshita, Jr., S. Kawato, A. Ikegami, *Biophys. J.* **1977**, **20**, 289–305; b) G. Lipari, A. Szabo, *Biophys. J.* **1980**, **30**, 489–506; c) R. E. Dale, J. Eisinger, W. E. Blumberg, *Biophys. J.* **1979**, **26**, 161–193; d) V. Ivanov, M. Li, K. Mizuuchi, *Biophys. J.* **2009**, **97**, 922–929.
- [54] D. Badali, C. C. Grdinariu, *J. Chem. Phys.* **2011**, **134**, 225102.
- [55] a) A. K. Wozniak, G. F. Schröder, H. Grubmüller, C. A. Seidel, F. Oesterhelt, *Proc. Natl. Acad. Sci. USA* **2008**, **105**, 18337–18342; b) E. Deplazes, D. Jayatilaka, B. Corry, *Phys. Chem. Chem. Phys.* **2011**, **13**, 11045–11054; c) M. Hoefling, N. Lima, D. Haenni, C. A. Seidel, B. Schuler, H. Grubmüller, *PLoS one* **2011**, **6**, e19791; d) E. Dolghih, W. Ortiz, S. Kim, B. P. Krueger, J. L. Krause, A. E. Roitberg, *J. Phys. Chem. A* **2009**, **113**, 4639–4646; e) A. L. Speelman, A. Muñoz-Losa, K. L. Hinkle, D. B. VanBeek, B. Mennucci, B. P. Krueger, *J. Phys. Chem. A* **2011**, **115**, 3997–4008.
- [56] E. A. Price, B. T. Sutcliffe, Q. Cai, P. Z. Qin, I. S. Haworth, *Biopolymers* **2007**, **87**, 40–50.
- [57] G. Zheng, L. Czapla, A. R. Srinivasan, W. K. Olson, *Phys. Chem. Chem. Phys.* **2010**, **12**, 1399–1406.
- [58] a) A. T. Brunger, P. Strop, M. Vrljic, S. Chu, K. R. Weninger, *J. Struct. Biol.* **2011**, **173**, 497–505; b) A. Muschielok, J. Andrecka, A. Jawhari, F. Bruckner, P. Cramer, J. Michaelis, *Nat. Methods* **2008**, **5**, 965–971; c) A. Muschielok, J. Michaelis, *J. Phys. Chem. B* **2011**, **115**, 11927–11937; d) J. Andrecka, R. Lewis, F. Bruckner, E. Lehmann, P. Cramer, J. Michaelis, *Proc. Natl. Acad. Sci. USA* **2008**, **105**, 135–140; e) J. Andrecka, B. Treutlein, M. A. Arcusa, A. Muschielok, R. Lewis, A. C. Cheung, P. Cramer, J. Michaelis, *Nucleic Acids Res.* **2009**, **37**, 5803–5809; f) H. Balci, S. Arslan, S. Myong, T. M. Lohman, T. Ha, *Biophys. J.* **2011**, **101**, 976–984; g) G. F. Schröder, H. Grubmüller, *Comput. Phys. Commun.* **2004**, **158**, 150–157; h) I. Rasnik, S. Myong, W. Cheng, T. M. Lohman, T. Ha, *J. Mol. Biol.* **2004**, **336**, 395–408; i) X. Sun, D. F. Mierke, T. Biswas, S. Y. Lee, A. Landy, M. Radman-Livaja, *Mol. Cell* **2006**, **24**, 569–580; j) B. Treutlein, A. Muschielok, J. Andrecka, A. Jawhari, C. Buchen, D. Kostrewa, F. Hog, P. Cramer, J. Michaelis, *Mol. Cell* **2012**, **46**, 136–146; k) T. Sabir, G. F. Schröder, A. Toumlin, P. McGlynn, S. W. Magennis, *J. Am. Chem. Soc.* **2011**, **133**, 1188–1191; l) T. Sabir, A. Toumlin, L. Ma, A. C. Jones, P. McGlynn, G. F. Schröder, S. W. Magennis, *J. Am. Chem. Soc.* **2012**, **134**, 6280–6285; m) R. Hickerison, S. K. Majumdar, A. Baucom, R. M. Clegg, H. F. Noller, *J. Mol. Biol.* **2005**, **354**, 459–472.
- [59] H. M. Watrob, C. P. Pan, M. D. Barkley, *J. Am. Chem. Soc.* **2003**, **125**, 7336–7343.
- [60] a) M. Heilemann, P. Tinnefeld, G. Sanchez Mosteiro, M. Garcia Parajo, N. F. Van Hulst, M. Sauer, *J. Am. Chem. Soc.* **2004**, **126**, 6514–6515; b) P. Tinnefeld, M. Heilemann, M. Sauer, *ChemPhysChem* **2005**, **6**, 217–222; c) J. K. Hannestad, P. Sandin, B. Albinsson, *J. Am. Chem. Soc.* **2008**, **130**, 15889–15895; d) J. K. Hannestad, S. R. Gerrard, T. Brown, B. Albinsson, *Small* **2011**, **7**, 3178–3185; e) I. H. Stein, C. Steinhauer, P. Tinnefeld, *J. Am. Chem. Soc.* **2011**, **133**, 4193–4195; f) E. Graugnard, D. L. Kellis, H. Bui, S. Barnes, W. Kuang, J. Lee, W. L. Hughes, W. B. Knowlton, B. Yurke, *Nano Lett.* **2012**, **12**, 2117–2122; g) E. Haustein, M. Jahnz, P. Schwille, *ChemPhysChem* **2003**, **4**, 745–748.
- [61] a) J. Liu, Y. Lu, *J. Am. Chem. Soc.* **2002**, **124**, 15208–15216; b) D. Klostermeier, P. Sears, C. H. Wong, D. P. Millar, J. R. Williamson, *Nucleic Acids Res.* **2004**, **32**, 2707–2715; c) J. C. Lam, Y. Li, *ChemBioChem* **2010**, **11**, 1710–1719; d) S. Hohng, C. Joo, T. Ha, *Biophys. J.* **2004**, **87**, 1328–1337; e) J. P. Clamme, A. Deniz, *ChemPhysChem* **2005**, **6**, 74–77; f) N. K. Lee, A. N. Kapanidis, H. R. Koh, Y. Korlann, S. O. Ho, Y. Kim, N. Gassman, S. K. Kim, S. Weiss, *Biophys. J.* **2007**, **92**, 303–312; g) J. Ross, P. Buschkamp, D. Fetting, A. Donnermeyer, C. M. Roth, P. Tinnefeld, *J. Phys. Chem. B* **2007**, **111**, 321–326; h) J. Lee, S. Lee, K. Ragunathan, C. Joo, T. Ha, S. Hohng, *Angew. Chem.* **2010**, **122**, 10118–10121; *Angew. Chem. Int. Ed.* **2010**, **49**, 9922–9925; i) S. Lee, J. Lee, S. Hohng, *PLoS one* **2010**, **5**, e12270.
- [62] a) S. Uphoff, S. J. Holden, L. Le Reste, J. Periz, S. van de Linde, M. Heilemann, A. N. Kapanidis, *Nat. Methods* **2010**, **7**, 831–836; b) S. Uphoff, K. Gryte, G. Evans, A. N. Kapanidis, *ChemPhysChem* **2011**, **12**, 571–579.
- [63] a) S. J. Holden, S. Uphoff, J. Hohlbein, D. Yadin, L. Le Reste, O. J. Britton, A. N. Kapanidis, *Biophys. J.* **2010**, **99**, 3102–3111; b) TwoTone: <http://www.physics.ox.ac.uk/Users/kapanidis/Group/Main.Software.html>; c) CPLC: <https://physics.illinois.edu/cplc/software/>; d) HaMMY: <http://bio.physics.illinois.edu/HaMMY.html>; e) QuB: <http://www.qub.buffalo.edu/>; f) vBFRET: <http://vbfret.sourceforge.net/>; g) PicoQuant: <http://www.picoquant.com/>; h) Becker+Hickl: <http://www.becker-hickl.de/>; i) CSM: <http://www.mathworks.com/matlabcentral/fileexchange/33217>; j) FRETshrink: <http://www.lrg.rice.edu/>; k) FretTrace: <http://www.mpiibpc.mpg.de/home/grubmueller/downloads/FretTrace/index.html>; l) MFD: <http://www.mpc.uni-duesseldorf.de/seidel/index.htm>; m) E. Deplazes, D. Jayatilaka, B. Corry, *J. Biomed. Opt.* **2012**, **17**, 011005; n) exiFRET: <http://exifret.com/>; o) FRETmatrix: <http://www.chalmers.se/chem/EN/divisions/physical-chemistry/staff/marcus-wilhelmsen/fretmatrix/>; p) FRETLab: <http://www.hybrid.soton.ac.uk/research/fret-lab/>; q) RETsg: <http://www.mpiibpc.mpg.de/home/grubmueller/downloads/FRETsg/index.html>; r) FRETnps: <http://www.cup.uni-muenchen.de/pc/michaelis/software.html>.
- [64] M. Blanco, N. G. Walter, *Methods Enzymol.* **2010**, **472**, 153–178.

Received: June 13, 2012

Published online on August 31, 2012

## **Paper II**

*In which tC<sub>nitro</sub> is characterized as a FRET acceptor...*



## Characterization of Nucleobase Analogue FRET Acceptor tC<sub>nitro</sub>

Søren Preus,<sup>†</sup> Karl Börjesson,<sup>‡</sup> Kristine Kilså,<sup>†</sup> Bo Albinsson,<sup>‡</sup> and L. Marcus Wilhelmsson\*,<sup>‡</sup>

Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, and Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden

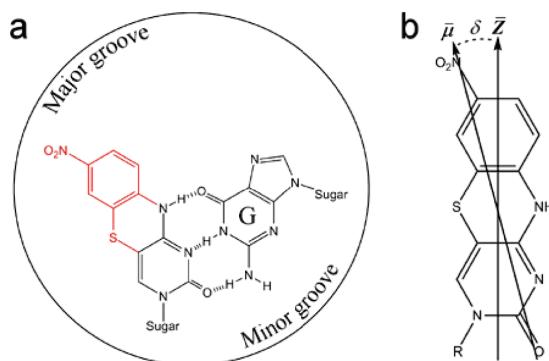
Received: October 2, 2009; Revised Manuscript Received: November 23, 2009

The fluorescent nucleobase analogues of the tricyclic cytosine (tC) family, tC and tC<sup>O</sup>, possess high fluorescence quantum yields and single fluorescence lifetimes, even after incorporation into double-stranded DNA, which make these base analogues particularly useful as fluorescence resonance energy transfer (FRET) probes. Recently, we reported the first all-nucleobase FRET pair consisting of tC<sup>O</sup> as the donor and the novel tC<sub>nitro</sub> as the acceptor. The rigid and well-defined position of this FRET pair inside the DNA double helix, and consequently excellent control of the orientation factor in the FRET efficiency, are very promising features for future studies of nucleic acid structures. Here, we provide the necessary spectroscopic and photophysical characterization of tC<sub>nitro</sub> needed in order to utilize this probe as a FRET acceptor in nucleic acids. The lowest energy absorption band from 375 to 525 nm is shown to be the result of a single in-plane polarized electronic transition oriented ~27° from the molecular long axis. This band overlaps the emission bands of both tC and tC<sup>O</sup>, and the Förster characteristics of these donor–acceptor pairs are calculated for double-stranded DNA scenarios. In addition, the UV–vis absorption of tC<sub>nitro</sub> is monitored in a broad pH range and the neutral form is found to be totally predominant under physiological conditions with a pK<sub>a</sub> of 11.1. The structure and electronic spectrum of tC<sub>nitro</sub> is further characterized by density functional theory calculations.

### Introduction

The development of synthetic nucleobase analogues continues to provide new tools for researchers working in the field of nucleic acids. Over the past 15 years, artificial nucleobases have been developed to enhance duplex and triplex stability,<sup>1,2</sup> to detect single nucleotide polymorphisms (SNPs),<sup>3–5</sup> as electron spin labels,<sup>6,7</sup> and to gain insight into the function of DNA polymerases,<sup>8</sup> to name but a few. In particular, fluorescent nucleobase analogues have attracted wide attention as probes for nucleic acid structure, dynamics, and interactions (for reviews, see refs 9–13). This specific class of probes includes 2-AP,<sup>14,15</sup> the pteridines 3MI and 6MI,<sup>16,17</sup> pyrrolo-C (PC),<sup>18</sup> the base discriminating fluorescent bases (BDFs) of Saito and co-workers,<sup>11,19–21</sup> the size-expanded (xDNA) nucleobases of Kool and co-workers,<sup>22,23</sup> and other more recently reported fluorescent base analogues.<sup>24–33</sup> Common for all of the mentioned fluorescent base analogues are their sensitivity to the surrounding environment and most often a significant quenching of fluorescence in the base-stacking environment provided by double-stranded DNA. These features have been exploited in applications such as quencher free molecular beacons,<sup>5</sup> in monitoring or characterizing protein activity,<sup>34–50</sup> and in probing the local structure and dynamics of nucleic acids.<sup>51–58</sup> However, the low and unpredictable fluorescence quantum yields of these fluorophores in double-stranded DNA make them highly unsuitable for studies involving fluorescence resonance energy transfer (FRET).

The two fluorescent nucleobase analogues of the tricyclic cytosine family, 1,3-diaza-2-oxophenothiazine (tC) and 1,3-



**Figure 1.** (a) Chemical structure of the tricyclic cytosine analogue tC<sub>nitro</sub> and its base-pairing with guanine. The tricyclic expansion of natural cytosine is highlighted in red. Also shown is the direction of the major and minor groove when looking down the long axis of double-stranded DNA. (b) Direction of the transition dipole moment,  $\bar{\mu}$ , associated with the lowest energy electronic transition of tC<sub>nitro</sub> ( $\delta = 27^\circ$ ). For KtC<sub>nitro</sub>, R = CH<sub>2</sub>COO<sup>-</sup>K<sup>+</sup>.

diaza-2-oxophenoxazine (tC<sup>O</sup>), have previously been characterized and shown to work excellently as fluorescent probes in both single- and double-stranded DNA.<sup>59–64</sup> The high fluorescence quantum yields and single fluorescence lifetimes, combined with a rigid and well-defined position inside the DNA double helix, make these base analogues particularly well suited for studies involving fluorescence anisotropy and FRET in nucleic acid containing systems.<sup>65</sup> Following this track, we recently demonstrated a novel FRET pair consisting of tC<sup>O</sup> as the donor and the nitro-substituted 7-nitro-1,3-diaza-2-oxophenothiazine (tC<sub>nitro</sub>, Figure 1) as the acceptor.<sup>66</sup> The well-defined position and orientation of this FRET pair inside the DNA

\* To whom correspondence should be addressed. E-mail: marcus.wilhelmsson@chalmers.se. Phone: +46 31 7723051. Fax: +46 31 7723858.

<sup>†</sup> University of Copenhagen.

<sup>‡</sup> Chalmers University of Technology.

double helix is extraordinary and facilitates an unprecedented high control of the orientation factor in the FRET efficiency compared to other labeling strategies.<sup>67–71</sup> However, in order to utilize this FRET pair for structural studies of more complex systems, the spectroscopic and photophysical properties of tC<sub>nitro</sub> must be thoroughly characterized in terms of the number of transitions constituting the lowest energy absorption band, as well as the direction and magnitude of absorbing and emitting transition dipole moments.

Here, we characterize the chromophoric properties of tC<sub>nitro</sub> with special focus put on its use as an energy acceptor in FRET experiments. In particular, we show that the lowest energy absorption band of tC<sub>nitro</sub> is the result of a single electronic transition with an in-plane polarized transition dipole moment oriented ~27° toward the NO<sub>2</sub> group from the molecular long axis (Figure 1b). In addition, the UV-vis absorption spectrum of tC<sub>nitro</sub> is monitored and shown to be preserved in a wide pH range. Furthermore, the molecular geometry of tC<sub>nitro</sub> is optimized and electronic transitions are predicted using quantum chemical calculations.

## Experimental Methods

**Synthesis.** The tC<sub>nitro</sub> nucleoside was prepared as described previously.<sup>66</sup> The potassium salt of 7-nitro-1,3-diaza-2-oxophenothiazine-3-yl acetic acid (KtC<sub>nitro</sub>) (Figure 1b) was synthesized by saponification of the tertbutyl ester, which was obtained by alkylation of the anion of 7-nitro-1,3-diaza-2-oxophenothiazine<sup>66</sup> with tertbutylbromoacetate, following the general procedure of Eldrup et al.<sup>72</sup>

**Calculations.** DFT geometry optimizations were performed in the ground state of the molecule using the B3LYP functional<sup>73–75</sup> and a 6-31G(d,p) basis set as implemented in the Gaussian 03 program package.<sup>76</sup> Restricted Hartree–Fock (RHF) wavefunctions were used. TDDFT<sup>77,78</sup> B3LYP/6-311+G(2d) calculations of the 10 lowest energy excitations were performed using Gaussian 03. The amount of HOMO → LUMO character of the lowest energy electronic transition of the investigated compound was determined from the calculated CI coefficients.

**Linear Dichroism.** Linear dichroism (LD) is the difference in absorption of two mutually perpendicular planes of linearly polarized light, and was in this work exploited to estimate the direction of the transition dipole moment of tC<sub>nitro</sub> aligned in a stretched polyvinyl alcohol (PVA) film.<sup>79</sup> The reduced LD of a uniaxial sample is defined as

$$\text{LD}^r = \frac{A_{\parallel} - A_{\perp}}{A_{\text{iso}}} = 3 \frac{A_{\parallel} - A_{\perp}}{A_{\parallel} + 2A_{\perp}}$$

where  $A_{\parallel}$  and  $A_{\perp}$  are the absorption of light oriented parallel ( $\parallel$ ) and perpendicular ( $\perp$ ) to the macroscopic orientation axis (the direction of stretching) and  $A_{\text{iso}}$  is the absorbance of the corresponding isotropic sample. The LD<sup>r</sup> of a molecule with rod-like orientation is related to the angle,  $\delta_i$ , between the  $i$ th transition moment and the orientation axis by

$$\text{LD}^r = \frac{3}{2} S(\cos^2 \delta_i - 1)$$

where  $S$  is the Saupe orientation factor for the orientation axis. The LD measurements of tC<sub>nitro</sub> in stretched PVA film were performed using a Varian Cary 4B spectrophotometer equipped with Glan air-space calcite polarizers in both sample and reference beam. The film was made from a 12.5% (w/w) aqueous solution of PVA that was prepared by dissolving PVA in water under heating to 85–95 °C and continuous stirring. A

5 mL portion of the transparent PVA solution was mixed with 3 mL of an aqueous solution of the potassium salt of tC<sub>nitro</sub> (~0.2 mg of substance). The mixture was poured onto rinsed horizontal glass plates and left to dry in a dust-free environment for a week. The film was then removed from the plates and mechanically stretched 4 times the original length under hot air from a hairdryer using a manually operated, in-house built device.

**Fluorescence Anisotropy.** Fluorescence anisotropy,  $r$ , measurements were performed by exciting the sample using vertically polarized light, and it was calculated as the intensity ratio between the polarized and total emission emanating from the sample<sup>80</sup>

$$r = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + 2GI_{\perp}}$$

Here,  $I_{\parallel}$  and  $I_{\perp}$  are the emission intensities measured through polarizers oriented parallel and perpendicular to the incident wave, respectively, and the  $G$ -factor is the ratio of the instrumental sensitivities for vertically and horizontally polarized light. For an immobile fluorophore, the angle between absorbing and emitting transition moments,  $\alpha$ , is related to the fundamental anisotropy by

$$r_0 = \frac{1}{5}(3 \cos^2 \alpha - 1)$$

The fundamental fluorescence anisotropy of tC<sub>nitro</sub> was measured by immobilizing the molecule in a propylene glycol (PG) glass at 200 K using an Oxford optistat DN cryostat and measuring the excitation spectrum through Glan polarizers in both excitation and emission beam. Spectra were recorded on a Spex Fluorolog 3 spectrofluorimeter (JY Horiba) using excitations in the range 290–510 nm in 4 nm intervals with the emission monitored at 530 nm.

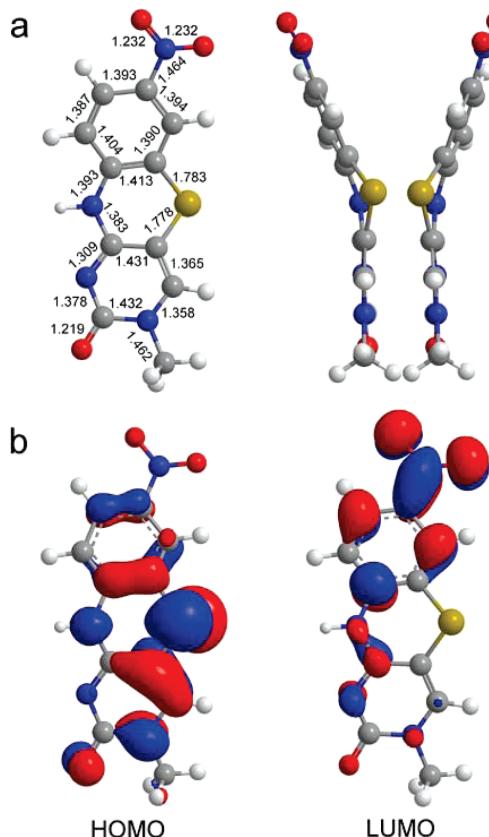
**Magnetic Circular Dichroism.** In magnetic circular dichroism (MCD), two CD spectra are recorded of the sample in the presence of a magnetic field oriented from north to south and south to north. The MCD spectrum is then calculated by subtracting the two<sup>79,81</sup>

$$2\text{MCD}(\lambda) = \text{NS}(\lambda) - \text{SN}(\lambda)$$

where NS and SN represent the measured CD with the magnetic field oriented north–south and south–north, respectively. In the presence of an external magnetic field, two electronic transitions close in energy but with different polarizations will split into two MCD signals of opposite sign, resulting in a bisignate signal over the absorption band. Thus, MCD is a useful method to determine whether an absorption band contains one or several transitions. The MCD of tC<sub>nitro</sub> was measured on a Jasco-J-720 CD spectropolarimeter equipped with a permanent horseshoe magnet. The MCD signal was recorded in the range 300–550 nm at a rate of 20 nm/min, and the final spectrum was determined as an average of eight acquisitions. A baseline of the solvent, a water/methanol (1:2) mixture, was used for both the NS and SN acquisitions. The concentration of the tC<sub>nitro</sub> nucleoside was 240 μM.

**UV-vis Absorption.** UV-vis absorption spectra were recorded on a Varian Cary 4000 spectrophotometer in 1 cm quartz cuvettes using pure solvent as the baseline.

The pH titration of tC<sub>nitro</sub> was performed by mixing a set of samples from aqueous stock solutions of tC<sub>nitro</sub>, 2 M KOH, and 1 M HCl producing 10–12 samples of equal tC<sub>nitro</sub> concentrations but different pH values. The pH range was tested from pH 0.2 to 12.2 and measured on a standard calibrated pH meter.



**Figure 2.** (a) B3LYP/6-31G(d,p) optimized ground state structures of the isolated tC<sub>nitro</sub> base. Two local energy minima were identified on the PES of tC<sub>nitro</sub> corresponding to geometries folded along the middle sulfur–nitrogen axis. Left: front view. Right: side view. Bond lengths are given in Ångströms. (b) Frontier KS orbitals of tC<sub>nitro</sub> optimized at the B3LYP/6-311+G(2d) level.

Singular value decomposition (SVD) is a mathematical tool with application in the deconvolution of spectra resulting from several absorbing species.<sup>82</sup> SVD analysis was performed using an in-house built Matlab script. The pH was rewritten into a vector of [H<sup>+</sup>] and used with a matrix containing the experimentally recorded spectra with each column representing a pH value and rows representing wavelengths. The pK<sub>a</sub> was determined from the [H<sup>+</sup>] at which the ratio between the protonated and deprotonated components in the SVD analysis was 1.

## Results and Discussion

**Molecular Geometry and Electron Distribution.** The strict conditions set by the DNA double helix on a nucleobase analogue in terms of H-bonding, base-stacking, and steric hindrances make it highly relevant to determine the molecular geometry of tC<sub>nitro</sub>. Two local energy minima on the potential energy surface of tC<sub>nitro</sub> were identified from a B3LYP/6-31G(d,p) conformational search (Figure 2a). Calculations of the vibrational spectra confirmed that the optimized structures correspond to minima on the potential energy surface. As seen in Figure 2a, the two different geometries are the result of a distortion of the tricyclic framework symmetry due to the large sulfur atom present in the middle ring. The two local minima

are mirror images corresponding to two geometries folded along the middle sulfur–nitrogen axis. This is also the result of AM1 calculations performed on the structurally similar tC base.<sup>64</sup> However, a slight distortion of the sulfur atom from the molecular plane is predicted by the DFT calculations compared to the AM1 optimized geometry. This is accompanied by a larger degree of bending (an angle of 26° for the DFT optimized geometry compared to 18° for AM1). The result that tC<sub>nitro</sub> adopts a bent geometry is supported by the X-ray structure of the parent compound phenothiazine.<sup>83,84</sup>

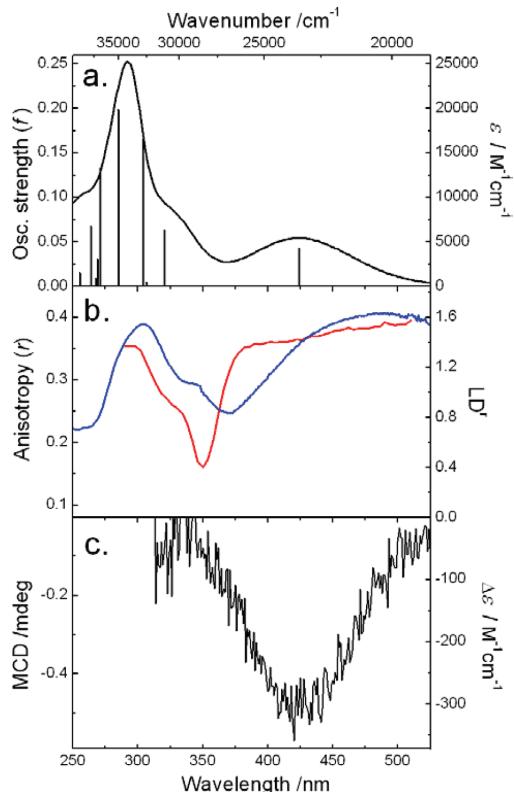
The bent geometries of tC<sub>nitro</sub> (and tC) do not affect the B-DNA secondary structure sterically, since the expansion of these base analogues, compared to natural cytosine, is directed into the major groove of B-DNA. In addition, given the structural similarities between tC, tC<sup>O</sup>, and tC<sub>nitro</sub>, as well as the fact that tC<sub>nitro</sub> does not decrease duplex stability nor perturb the B-DNA structure in any significance,<sup>66</sup> we find it very likely that the base-flipping rate of tC<sub>nitro</sub> is as slow as that of tC and tC<sup>O</sup>.<sup>60–62</sup> Furthermore, a very pronounced relationship between the DNA double helical structure and previously measured FRET efficiencies between tC<sup>O</sup> and tC<sub>nitro</sub> positioned inside double-stranded DNA supports the claim that tC<sub>nitro</sub> is firmly stacked inside the DNA double helix.<sup>66</sup>

The two optimized frontier Kohn–Sham (KS) orbitals of tC<sub>nitro</sub> are shown in Figure 2b. The HOMO orbital is distributed almost throughout the entire molecular plane (although with low density at the nitro group), while the LUMO orbital is mainly centered at the electron withdrawing nitro group. As a result of the small spatial overlap of the HOMO and LUMO orbitals of tC<sub>nitro</sub>, the lowest energy electronic transition has a certain degree of charge-transfer character, which is a characteristic often seen for polycyclic nitroaromatics.<sup>85–88</sup>

**Calculations of Electronic Transitions and Frontier Orbitals.** The UV–vis absorption spectrum of the nucleoside of tC<sub>nitro</sub> in H<sub>2</sub>O at 295 K is shown in Figure 3a. The lowest energy absorption band is centered at 424 nm ( $\epsilon_{424\text{ nm}} = 5400 \text{ M}^{-1}\text{cm}^{-1}$ ) but becomes slightly red-shifted (10–15 nm) upon incorporation into DNA.<sup>66</sup> This Gaussian-shaped band is energetically well separated from higher energy absorption bands ( $\Delta E \sim 1 \text{ eV}$ ).

The TDDFT B3LYP/6-311+G(2d) calculated electronic spectrum of the B3LYP/6-31G(d,p) optimized molecular geometry of tC<sub>nitro</sub> is also depicted in Figure 3a (vertical lines). In general, the overall appearance of the calculated spectrum agrees well with the experimentally recorded UV–vis absorption spectrum in H<sub>2</sub>O. The lowest energy absorption band of tC<sub>nitro</sub> is found to be the result of a single electronic transition having 92% HOMO → LUMO character which is energetically well separated from higher energy transitions ( $\Delta E \sim 1 \text{ eV}$ ). The direction of the calculated transition dipole moment associated with the S<sub>0</sub> → S<sub>1</sub> transition was predicted to be oriented in-plane and tilted 9° toward the nitro group from the molecular long axis. The direction of the transition dipole moment of tC<sub>nitro</sub> does not change when interconverting between the two bent geometries.

It is noted here that the CT character of most of the excited states of tC<sub>nitro</sub> (data not shown for transitions above the lowest in energy) results in slightly underestimated excitation energies, which is a well-known problem when using local functionals, such as B3LYP, in the prediction of excitation energies of transitions involving KS orbitals of small spatial overlap.<sup>89–91</sup> The excitation energies shown in Figure 3a have therefore been multiplied by a factor 1.08 to facilitate comparison with the overall spectral shape of the UV–vis absorption spectrum of tC<sub>nitro</sub> in H<sub>2</sub>O.



**Figure 3.** (a) Isotropic UV-vis absorption spectrum (full drawn line) and calculated electronic transitions (vertical lines). The absorption spectrum was measured in H<sub>2</sub>O at 295 K. Electronic transitions were predicted in the gas phase at the TDDFT B3LYP/6-311+G(2d) level. Note that the calculated excitation energies have been multiplied by a factor of 1.08 to facilitate comparison with the UV-vis absorption spectrum in H<sub>2</sub>O. (b) Excitation anisotropy spectrum (red) of the tC<sub>nitro</sub> nucleoside and reduced linear dichroism spectrum (blue) of the potassium salt of tC<sub>nitro</sub>. The anisotropy was measured in a propylene glycol glass at 200 K, and the linear dichroism was recorded using a stretched PVA film. (c) MCD spectrum of tC<sub>nitro</sub> nucleoside in a H<sub>2</sub>O/methanol mixture (1:2).

**Experimental Characterization of the Lowest Energy Absorption Band.** The electronic transition responsible for the lowest energy absorption band of tC<sub>nitro</sub> was experimentally characterized using three different polarized optical spectroscopy techniques: fluorescence anisotropy, MCD, and LD. Figure 3b shows the fluorescence anisotropy of the nucleoside of tC<sub>nitro</sub> immobilized in a PG glass at 200 K (red). The change from H<sub>2</sub>O to PG does not change the shape of the absorption spectrum of the chromophore and only slightly shifts the absorption maximum of the lowest energy band to higher energy (5 nm; data not shown). As can be seen, the excitation anisotropy of tC<sub>nitro</sub> almost reaches the theoretical maximum value of  $r_A = 0.4$  over the lowest energy absorption band. First of all, this shows that the fluorophore is practically immobilized on the time scale of the excited state decay. Second, this value of  $r_A$  corresponds to completely parallel absorbing and emitting transition dipole moments. Combined with the fact that  $r_A$  is essentially constant over the entire lowest energy absorption band, this indicates that the low energy absorption band of tC<sub>nitro</sub> only contains a single electronic transition which, in turn, is the S<sub>0</sub> → S<sub>1</sub> transition.

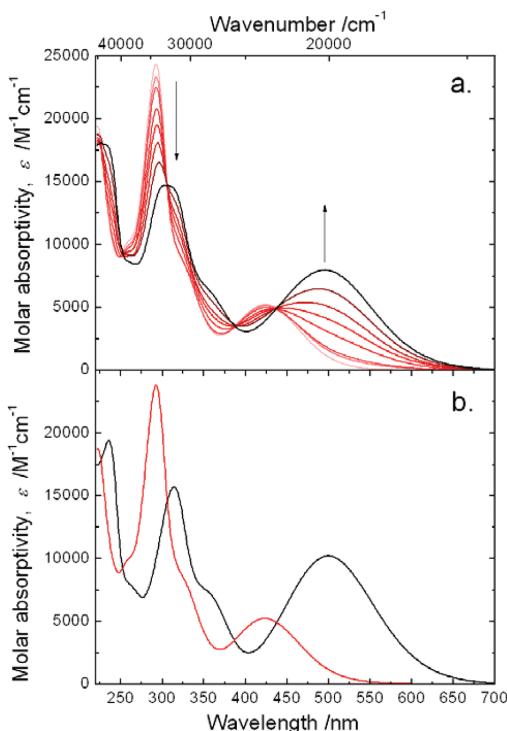
The MCD of the nucleoside of tC<sub>nitro</sub> in a H<sub>2</sub>O/methanol mixture (1:2) is shown in Figure 3c. The use of a H<sub>2</sub>O/methanol mixture as solvent was necessary in order to dissolve a sufficient amount of sample to detect the MCD signal. The change from pure H<sub>2</sub>O to a H<sub>2</sub>O/methanol mixture does not change the absorption spectrum of tC<sub>nitro</sub> significantly (data not shown). As seen from Figure 3c, the MCD signal over the region of the lowest absorption band of tC<sub>nitro</sub> bears the same sign and appears as the mirror image of the absorption spectrum. Together with the fluorescence anisotropy measurement and the high level quantum chemical calculations (*vide supra*), this result strongly indicates that only a single electronic transition is responsible for the lowest energy absorption band of tC<sub>nitro</sub>.

The direction of the S<sub>0</sub> → S<sub>1</sub> transition dipole moment of tC<sub>nitro</sub> was estimated on the basis of the reduced linear dichroism of the potassium salt of tC<sub>nitro</sub> (Figure 1b where R = CH<sub>2</sub>COO<sup>-</sup>K<sup>+</sup>) aligned in a stretched polyvinyl alcohol (PVA) film (Figure 3b, blue line). As can be seen, the LD<sup>r</sup> of tC<sub>nitro</sub> reaches a value of LD<sup>r</sup> = 1.62 over the lowest energy absorption band. Assuming that tC<sub>nitro</sub> orients similarly to the structurally comparable compound methylene blue within the PVA film (rod-like with a Saupe orientation factor of 0.78)<sup>92</sup> as previously assumed for tC and tC<sup>O</sup> aligned in stretched PVA films,<sup>61,64</sup> the direction of the absorbing transition moment of tC<sub>nitro</sub> is calculated to be tilted 27° from the molecular long axis. The TDDFT calculations (*vide supra*) together with a FRET study in double-stranded DNA<sup>66</sup> have shown that the tilt is toward the nitro group. The fact that the LD<sup>r</sup> is not constant over the entire absorption band of tC<sub>nitro</sub> was also observed for the LD<sup>r</sup> of tC and tC<sup>O</sup>.<sup>61,64</sup> This feature may be the result of several absorbing species oriented differently in the film, such as aggregates of molecules formed due to the high sample concentration used in the film.

**Emissive Properties.** The fluorescence quantum yield of tC<sub>nitro</sub> is virtually zero, both in its monomeric form and after incorporation into single- and double-stranded DNA, but increases at temperatures below 225 K with an emission band centered at 530 nm in PG (data not shown). The weak or complete lack of fluorescence is a general property of nitroaromatic compounds and is usually either due to a fast intersystem crossing<sup>88,93-95</sup> or internal conversion<sup>85,96</sup> process. For tC<sub>nitro</sub>, we suggest that the quenching is due to an efficient S<sub>1</sub>–S<sub>0</sub> internal conversion related to the charge transfer character of the first excited state. Studies addressing this are currently performed in our lab.

**pH Dependency.** The pH range of the neutral form of tC<sub>nitro</sub> was investigated by monitoring the UV-vis absorption of tC<sub>nitro</sub> at various pH values. The experiment shows that the neutral form of tC<sub>nitro</sub> is predominant in a pH range from pH 11 down to at least pH 0.2 (the most acidic pH tested). The result of a pH 8–12 titration of the nucleoside of tC<sub>nitro</sub> in H<sub>2</sub>O is shown in Figure 4a in which the arrows denote the spectral evolution as the pH increases. When increasing the pH, the low energy absorption band of tC<sub>nitro</sub> at  $\lambda_{\text{max}} = 424$  nm is replaced by a more intense band at  $\lambda_{\text{max}} = 500$  nm corresponding to the deprotonated compound (deprotonation of the central enamine). The color of the sample correspondingly changes from a clear yellow to a deep red color. The presence of several isosbestic points on the titration curves (at 227, 253, 306, 387, and 436 nm) strongly indicates that only the protonated and deprotonated forms of tC<sub>nitro</sub> are present in the solution.

Since the measured absorption spectrum of tC<sub>nitro</sub>, at each of the different pH values, is a linear combination of the absorption spectra of the protonated and deprotonated species, the isolated



**Figure 4.** (a) Isotropic absorption spectra of tC<sub>nitro</sub> at various pH values. Arrows indicate an increase in pH. pH values used: 8.00, 9.88, 10.22, 10.74, 10.89, 11.13, 11.34, 11.74, and 12.18. (b) Isolated spectra of the protonated (red) and deprotonated (black) of tC<sub>nitro</sub> in H<sub>2</sub>O obtained from the singular value decomposition (SVD).

absorption spectrum of each of the two components can successfully be deconvoluted using singular value decomposition. The isolated spectra obtained from the SVD analysis are shown in Figure 4b. The overall isolated absorption spectrum of the deprotonated component is red-shifted compared to the neutral compound. This is most likely due to electrostatic interactions with the solvent resulting from the net negative charge of the deprotonated compound. The red-shift in wavenumbers is 3500 cm<sup>-1</sup> for the lowest energy absorption band and 2400 cm<sup>-1</sup> for all of the higher energy absorption maxima and the shoulder at 350 nm. This suggests that the energy of S<sub>0</sub> of tC<sub>nitro</sub> is increased by 0.30 eV (2400 cm<sup>-1</sup>) for the deprotonated anion, while the S<sub>1</sub> energy is stabilized by 0.14 eV (3500–2400 cm<sup>-1</sup> = 1100 cm<sup>-1</sup>), compared to the neutral protonated species in H<sub>2</sub>O.

The pK<sub>a</sub> of tC<sub>nitro</sub> is calculated from the SVD analysis to be 11.1. This pK<sub>a</sub> value of tC<sub>nitro</sub> is 2 units lower than that found for tC (pK<sub>a</sub> = 13.2),<sup>64</sup> most likely due to the electron-withdrawing NO<sub>2</sub> group that stabilizes the negative charge of the deprotonated compound.

**Förster Characteristics.** The red-shifted absorption of tC<sub>nitro</sub> has a considerable spectral overlap with the fluorescence spectra of both tC and tC<sup>O</sup>. In Table 1, we have summarized the calculated overlap integrals<sup>80</sup> between the nucleobase analogue FRET pairs of the tricyclic cytosine family. Because the absorption and emission spectra, as well as the fluorescence quantum yields, of the nucleobase analogues change upon incorporation into DNA, the overlap integrals are also shown for the FRET pairs positioned in double-stranded DNA.

**TABLE 1:** Calculated Overlap Integrals and Critical Förster Distances (for  $\kappa^2 = 2/3$ ) of the tC–tC<sub>nitro</sub> and tC<sup>O</sup>–tC<sub>nitro</sub> Nucleobase Analogue Donor–Acceptor Pairs

donor	$\Phi_f$	$n^a$	$J/M^{-1} \text{cm}^{-1} \text{nm}^4$	$R_0/\text{\AA}$
Monomer <sup>b</sup>				
tC	0.13	1.35	$4.7 \times 10^{13}$	21.8
tC <sup>O</sup>	0.30	1.35	$1.0 \times 10^{14}$	28.5
In dsDNA <sup>c</sup>				
tC	0.20	1.40	$5.4 \times 10^{13}$	23.4
tC <sup>O</sup>	0.22	1.40	$1.2 \times 10^{14}$	27.2

<sup>a</sup> Since the refractive index is a macroscopic property, the chosen value of 1.40 is an approximate guess based on the chemical structure of the DNA molecule. <sup>b</sup> Quantum yield values are for the nucleosides. <sup>c</sup> Since the quantum yields of tC and tC<sup>O</sup> are slightly dependent on neighboring bases,<sup>61,62</sup> the values used here are averages.

Furthermore, representative values of the critical Förster distances are shown in Table 1 assuming  $\kappa^2 = 2/3$  (freely rotating chromophores) and averaged donor quantum yields.<sup>80</sup> The overlap integrals between the absorption of monomeric tC<sub>nitro</sub> and the emission of monomeric tC and tC<sup>O</sup> are  $4.7 \times 10^{13}$  and  $1.0 \times 10^{14} \text{ M}^{-1} \text{cm}^{-1} \text{nm}^4$ , respectively. The corresponding critical distances are 21.8 Å between monomeric tC–tC<sub>nitro</sub> and 28.5 Å between tC<sup>O</sup>–tC<sub>nitro</sub>. Due to the red-shifted absorption of tC<sub>nitro</sub> upon incorporation into DNA, the overlap integrals for FRET pairs positioned in DNA are slightly larger compared to the monomeric FRET pairs ( $5.4 \times 10^{13}$  and  $1.2 \times 10^{14} \text{ M}^{-1} \text{cm}^{-1} \text{nm}^4$  for tC and tC<sup>O</sup> as donors, respectively). Using the average quantum yields of tC and tC<sup>O</sup> positioned in DNA, the corresponding critical distances become 23.4 Å for tC as the donor and 27.2 Å for tC<sup>O</sup> as the donor. These values may be compared to the dimensions of the B-DNA double helix which has a length of approximately 34 Å per turn.

Since the Förster formulation is based on a point dipole approximation, it is noted here that the FRET theory is not valid for close lying chromophores in which electronic couplings of other origins can be anticipated.<sup>97</sup> We have previously performed circular dichroism experiments on homodimers of tC as well as tC<sup>O</sup> separated by 0–2 bases in DNA duplexes to examine the excitonic effect and found minor effects for the case where the homodimer is separated only by 0 bases (data not shown). The dipole–dipole resonance mechanism is therefore expected to be the totally predominant electronic interaction between these base analogues in the distance range in which they operate (appr. 15–45 Å).

## Conclusions and Outlook

We have characterized the nucleobase analogue tC<sub>nitro</sub> with special focus put on its use as a FRET acceptor in nucleic acid studies. With a pK<sub>a</sub> of 11.1, the neutral form of tC<sub>nitro</sub> is totally predominant under physiological conditions. The lowest energy absorption band (375–525 nm) of this compound is the result of a single electronic transition well separated (~1 eV) from higher energy excitations and associated with an in-plane polarized transition dipole moment oriented ~27° toward the nitro group from the molecular long axis. A DFT conformational search performed at the B3LYP/6-31G(d,p) level identified two local minima on the potential energy surface of tC<sub>nitro</sub> corresponding to two geometries folded along the middle sulfur–nitrogen axis; however, the direction of the transition dipole moment of tC<sub>nitro</sub> is virtually the same within its two geometries.

Currently available FRET pairs for use in nucleic acid containing systems normally have longer Förster distances (>50

Å at  $\kappa^2 = 2/3$ ) than the system presented here but lack high control of geometry and/or a stable quantum yield for the donor, resulting in inaccuracies in estimating the Förster distance. This work, on the other hand, lays the foundation for future detailed structural studies of nucleic acid containing systems implementing tC<sub>nitro</sub> as a FRET acceptor and tC or tC<sup>O</sup> as the donor. The fact that these nucleobase donor–acceptor pairs can be rigidly positioned close to or inside the actual site of interest when studying nucleic acids opens up a number of new possibilities in the structural investigation of nucleic acids in solution. Since the energy transfer process is highly dependent on both orientation and distance between the nucleobases, very detailed studies of nucleic acid structures and dynamics are in principle possible. Furthermore, through carefully designed experiments, the nucleobase analogue FRET pairs have the potential to thoroughly probe even subtle structural changes occurring in nucleic acids, e.g., resulting from the sequence specific interaction with other (bio)molecules or a change of physical chemical conditions.

**Acknowledgment.** This research is supported by the Swedish Research Council (VR) and the Danish Council for Independent Research | Natural Sciences (FNU).

## References and Notes

- (1) Herdewijn, P. *Antisense Nucleic Acid Drug Dev.* **2000**, *10*, 297.
- (2) Wojciechowski, F.; Hudson, R. H. E. *Curr. Top. Med. Chem.* **2007**, *7*, 667.
- (3) Wagenknecht, H. A. *Ann. N.Y. Acad. Sci.* **2008**, *1130*, 122.
- (4) Okamoto, A.; Tainaka, K.; Ochi, Y.; Kanatani, K.; Saito, I. *Mol. Biosyst.* **2006**, *2*, 122.
- (5) Venkatesan, N.; Seo, Y. J.; Kim, B. H. *Chem. Soc. Rev.* **2008**, *37*, 648.
- (6) Barhate, N.; Cekan, P.; Massey, A. P.; Sigurdsson, S. T. *Angew. Chem., Int. Ed.* **2007**, *46*, 2655.
- (7) Cekan, P.; Smith, A. L.; Barhate, N.; Robinson, B. H.; Sigurdsson, S. T. *Nucleic Acids Res.* **2008**, *36*, 5946.
- (8) Jung, K. H.; Marx, A. *Cell. Mol. Life Sci.* **2005**, *62*, 2080.
- (9) Asseline, U. *Curr. Org. Chem.* **2006**, *10*, 491.
- (10) Hawkins, M. E. In *Topics in Fluorescence Spectroscopy: DNA Technology*; Lakowicz, J. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2003.
- (11) Okamoto, A.; Saito, Y.; Saito, I. *J. Photochem. Photobiol., C* **2005**, *6*, 108.
- (12) Rist, M. J.; Marino, J. P. *Curr. Org. Chem.* **2002**, *6*, 775.
- (13) Wilson, J. N.; Kool, E. T. *Org. Biomol. Chem.* **2006**, *4*, 4265.
- (14) Freese, E. *J. Mol. Biol.* **1959**, *1*, 87.
- (15) Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228.
- (16) Hawkins, M. E. *Cell Biochem. Biophys.* **2001**, *34*, 257.
- (17) Hawkins, M. E. In *Methods in Enzymology - Fluorescence Spectroscopy*; Elsevier Academic Press Inc: San Diego, CA, 2008; Vol. 450, p 201.
- (18) Berry, D. A.; Jung, K. Y.; Wise, D. S.; Sercel, A. D.; Pearson, W. H.; Mackie, H.; Randolph, J. B.; Somers, R. L. *Tetrahedron Lett.* **2004**, *45*, 2457.
- (19) Okamoto, A.; Tainaka, K.; Saito, I. *Tetrahedron Lett.* **2003**, *44*, 6871.
- (20) Okamoto, A.; Tainaka, K.; Saito, I. *J. Am. Chem. Soc.* **2003**, *125*, 4972.
- (21) Okamoto, A.; Tanaka, K.; Fukuta, T.; Saito, I. *J. Am. Chem. Soc.* **2003**, *125*, 9296.
- (22) Krueger, A. T.; Lu, H. G.; Lee, A. H. F.; Kool, E. T. *Acc. Chem. Res.* **2007**, *40*, 141.
- (23) Krueger, A. T.; Kool, E. T. *J. Am. Chem. Soc.* **2008**, *130*, 3989.
- (24) Greco, N. J.; Sinkeldam, R. W.; Tor, Y. *Org. Lett.* **2009**, *11*, 1115.
- (25) Mizuta, M.; Seio, K.; Miyata, K.; Sekine, M. *J. Org. Chem.* **2007**, *72*, 5046.
- (26) Srivatsan, S. G.; Weizman, H.; Tor, Y. *Org. Biomol. Chem.* **2008**, *6*, 1334.
- (27) Butler, R. S.; Cohn, P.; Tenzel, P.; Abboud, K. A.; Castellano, R. K. *J. Am. Chem. Soc.* **2009**, *131*, 623.
- (28) Dyrager, C.; Börjesson, K.; Diner, P.; Elf, A.; Albinsson, B.; Wilhelmsson, L. M.; Grøtli, M. *Eur. J. Org. Chem.* **2009**, *1515*.
- (29) Mizuta, M.; Seio, K.; Ohkubo, A.; Sekine, M. *J. Phys. Chem. B* **2009**, *113*, 9562.
- (30) Miyata, K.; Mineo, R.; Tamamushi, R.; Mizuta, M.; Ohkubo, A.; Taguchi, H.; Seio, K.; Santa, T.; Sekine, M. *J. Org. Chem.* **2007**, *72*, 102.
- (31) Miyata, K.; Tamamushi, R.; Ohkubo, A.; Taguchi, H.; Seio, K.; Santa, T.; Sekine, M. *Org. Lett.* **2006**, *8*, 1545.
- (32) Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784.
- (33) Wojciechowski, F.; Hudson, R. H. E. *J. Am. Chem. Soc.* **2008**, *130*, 12574.
- (34) Harirharan, C.; Reha-Krantz, L. *J. Biochemistry* **2005**, *44*, 15674.
- (35) Wojtuszezki, K.; Hawkins, M. E.; Cole, J. L.; Mukerji, I. *Biochemistry* **2001**, *40*, 2588.
- (36) Allan, B. W.; Beechem, J. M.; Lindstrom, W. M.; Reich, N. O. *J. Biol. Chem.* **1998**, *273*, 2368.
- (37) Frey, M. W.; Sowers, L. C.; Millar, D. P.; Benkovic, S. J. *Biochemistry* **1995**, *34*, 9185.
- (38) Srivatsan, S. G.; Greco, N. J.; Tor, Y. *Angew. Chem., Int. Ed.* **2008**, *47*, 6661.
- (39) Raney, K. D.; Sowers, L. C.; Millar, D. P.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6644.
- (40) Hochstrasser, R. A.; Carver, T. E.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1994**, *33*, 11971.
- (41) Jia, Y. P.; Kumar, A.; Patel, S. S. *J. Biol. Chem.* **1996**, *271*, 30451.
- (42) Ujvari, A.; Martin, C. T. *Biochemistry* **1996**, *35*, 14574.
- (43) Zhong, X. J.; Patel, S. S.; Werneburg, B. G.; Tsai, M. D. *Biochemistry* **1997**, *36*, 11891.
- (44) Holz, B.; Klimasauskas, S.; Serva, S.; Weinhold, E. *Nucleic Acids Res.* **1998**, *26*, 1076.
- (45) Allan, B. W.; Reich, N. O.; Beechem, J. M. *Biochemistry* **1999**, *38*, 5308.
- (46) Allan, B. W.; Reich, N. O. *Biochemistry* **1996**, *35*, 14757.
- (47) Stivers, J. T.; Pankiewicz, K. W.; Watanabe, K. A. *Biochemistry* **1999**, *38*, 952.
- (48) Liu, C. H.; Martin, C. T. *J. Biol. Chem.* **2002**, *277*, 2725.
- (49) Harirharan, C.; Bloom, L. B.; Helquist, S. A.; Kool, E. T.; Reha-Krantz, L. *J. Biochemistry* **2006**, *45*, 2836.
- (50) Liu, C. H.; Martin, C. T. *Mol. Biol.* **2001**, *308*, 465.
- (51) Menger, M.; Eckstein, F.; Porschke, D. *Biochemistry* **2000**, *39*, 4500.
- (52) Guest, C. R.; Hochstrasser, R. A.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1991**, *30*, 3271.
- (53) Xu, D. G.; Evans, K. O.; Nordlund, T. M. *Biochemistry* **1994**, *33*, 9592.
- (54) Ramreddy, T.; Kombrabail, M.; Krishnamoorthy, G.; Rao, B. J. *J. Phys. Chem. B* **2009**, *113*, 6840.
- (55) Jean, J. M.; Hall, K. B. *Biochemistry* **2004**, *43*, 10277.
- (56) Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G. *J. Phys. Chem. B* **2007**, *111*, 5757.
- (57) Larsen, O. F. A.; van Stokkum, I. H. M.; Gobets, B.; van Grondelle, R.; van Amerongen, H. *Biophys. J.* **2001**, *81*, 1115.
- (58) Dash, C.; Rausch, J. W.; Le Grice, S. F. J. *Nucleic Acids Res.* **2004**, *32*, 1539.
- (59) Börjesson, K.; Sandin, P.; Wilhelmsson, L. M. *Biophys. Chem.* **2009**, *139*, 24.
- (60) Engman, K. C.; Sandin, P.; Osborne, S.; Brown, T.; Billeter, M.; Lincoln, P.; Nordén, B.; Albinsson, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2004**, *32*, 5087.
- (61) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. *Nucleic Acids Res.* **2008**, *36*, 157.
- (62) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005**, *33*, 5019.
- (63) Wilhelmsson, L. M.; Holmén, A.; Lincoln, P.; Nielson, P. E.; Nordén, B. *J. Am. Chem. Soc.* **2001**, *123*, 2434.
- (64) Wilhelmsson, L. M.; Sandin, P.; Holmén, A.; Albinsson, B.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2003**, *107*, 9094.
- (65) Stengel, G.; Gill, J. P.; Sandin, P.; Wilhelmsson, L. M.; Albinsson, B.; Nordén, B.; Millar, D. *Biochemistry* **2007**, *46*, 12289.
- (66) Börjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. *J. Am. Chem. Soc.* **2009**, *131*, 4288.
- (67) Clegg, R. M.; Murchie, A. I. H.; Zechel, A.; Lilley, D. M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2994.
- (68) Hurley, D. J.; Tor, Y. *J. Am. Chem. Soc.* **2002**, *124*, 13231.
- (69) 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.
- (70) Lewis, F. D.; Zhang, L. G.; Zuo, X. B. *J. Am. Chem. Soc.* **2005**, *127*, 10002.
- (71) Sapsford, K. E.; Berti, L.; Medintz, I. L. *Angew. Chem., Int. Ed.* **2006**, *45*, 4562.
- (72) Eldrup, A. B.; Nielsen, B. B.; Haaima, G.; Rasmussen, H.; Kastrup, J. S.; Christensen, C.; Nielsen, P. E. *Eur. J. Org. Chem.* **2001**, *1781*.
- (73) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (74) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (75) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (76) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.;

- Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.
- (77) Marques, M. A. L.; Gross, E. K. U. *Annu. Rev. Phys. Chem.* **2004**, 55, 427.
- (78) Burke, K.; Werschnik, J.; Gross, E. K. U. *J. Chem. Phys.* **2005**, 123.
- (79) Rodger, A.; Nordén, B. *Circular Dichroism and Linear Dichroism*; Oxford University Press: Oxford, UK, 1997.
- (80) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: New York, 2006; Vol. 3.
- (81) Mason, W. M. *Magnetic Circular Dichroism*; John Wiley & Sons: Hoboken, NJ, 2007.
- (82) Handler, R. W.; Shrager, R. I. *J. Biochem. Biophys. Methods* **1994**, 28, 1.
- (83) Bell, J. D.; Blount, J. F.; Briscoe, O. V.; Freeman, H. C. *Chem. Commun.* **1968**, 1656.
- (84) McDowell, J. J. H. *Acta Crystallogr., Sect. B* **1976**, 32, 5.
- (85) Mohammed, O. F.; Vauthey, E. *J. Phys. Chem. A* **2008**, 112, 3823.
- (86) Khalil, O. S.; Bach, H. G.; McGlynn, S. P. *J. Mol. Spectrosc.* **1970**, 35, 455.
- (87) Mikula, J. J.; Stuebing, E. W.; Anderson, R. W.; Harris, L. E. *J. Mol. Spectrosc.* **1972**, 42, 350.
- (88) Zugazagoitia, J. S.; Almora-Diaz, C. X.; Peon, J. *J. Phys. Chem. A* **2008**, 112, 358.
- (89) Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, 126, 4007.
- (90) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, 119, 2943.
- (91) Tozer, D. J. *J. Chem. Phys.* **2003**, 119, 12697.
- (92) Nordén, B. *J. Chem. Phys.* **1980**, 72, 5032.
- (93) Ohtani, H.; Kobayashi, T.; Suzuki, K.; Nagakura, S. *Bull. Chem. Soc. Jpn.* **1980**, 53, 43.
- (94) Morales-Cueto, R.; Esquivelzeta-Rabell, M.; Saucedo-Zugazagoitia, J.; Peon, J. *J. Phys. Chem. A* **2007**, 111, 552.
- (95) Takezaki, M.; Hirota, N.; Terazima, M. *J. Phys. Chem. A* **1997**, 101, 3443.
- (96) Kovalenko, S. A.; Schanz, R.; Farztdinov, V. M.; Hennig, H.; Ernsting, N. P. *Chem. Phys. Lett.* **2000**, 323, 312.
- (97) Beljonne, D.; Curutchet, C.; Scholes, G. D.; Silbey, R. J. *J. Phys. Chem. B* **2009**, 113, 6583.

JP909471B



## **Paper III**

*In which FRETmatrix is presented...*



# FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids

Søren Preus<sup>1</sup>, Kristine Kilså<sup>1</sup>, Francois-Alexandre Miannay<sup>2</sup>, Bo Albinsson<sup>2</sup> and L. Marcus Wilhelmsson<sup>2,\*</sup>

<sup>1</sup>Department of Chemistry, University of Copenhagen, Copenhagen, DK-2100, Denmark and <sup>2</sup>Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden

Received July 11, 2012; Revised August 17, 2012; Accepted August 20, 2012

## ABSTRACT

Förster resonance energy transfer (FRET) is a technique commonly used to unravel the structure and conformational changes of biomolecules being vital for all living organisms. Typically, FRET is performed using dyes attached externally to nucleic acids through a linker that complicates quantitative interpretation of experiments because of dye diffusion and reorientation. Here, we report a versatile, general methodology for the simulation and analysis of FRET in nucleic acids, and demonstrate its particular power for modelling FRET between probes possessing limited diffusional and rotational freedom, such as our recently developed nucleobase analogue FRET pairs (base–base FRET). These probes are positioned inside the DNA/RNA structures as a replacement for one of the natural bases, thus, providing unique control of their position and orientation and the advantage of reporting from inside sites of interest. In demonstration studies, not requiring molecular dynamics modelling, we obtain previously inaccessible insight into the orientation and nanosecond dynamics of the bases inside double-stranded DNA, and we reconstruct high resolution 3D structures of kinked DNA. The reported methodology is accompanied by a freely available software package, FRETmatrix, for the design and analysis of FRET in nucleic acid containing systems.

## INTRODUCTION

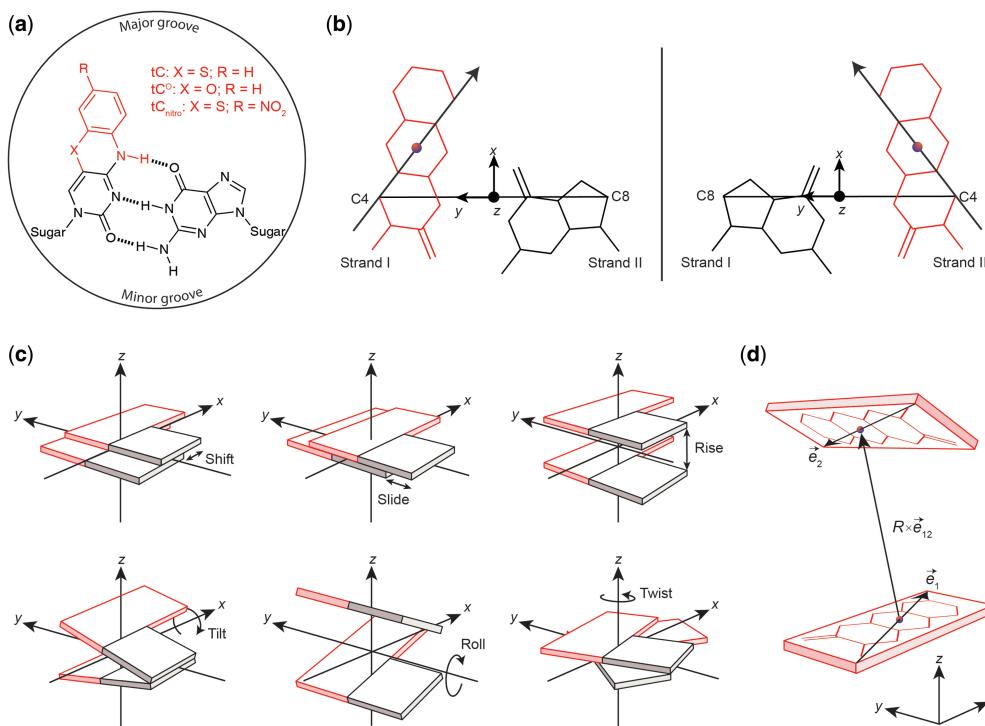
As cornerstones of the central dogma and fundamental players in gene regulation, nucleic acids and their structures, dynamics, conformational changes and interactions

with other biomolecules is key to the understanding of living organisms. Traditionally, high-resolution structural insight into nucleic acids is accomplished using nuclear magnetic resonance (NMR) spectroscopy (1) or X-ray crystallography (2), often being complemented by lower-resolution techniques, such as Förster resonance energy transfer (FRET) (3,4). Of these, FRET possesses the prominent advantage of providing rapid measures of just a few nanomoles of sample in solution even in complex media or large molecular complexes. The method relies on the ability of a donor fluorophore to transfer its excitation energy to an acceptor chromophore through an oscillating transition dipole–dipole resonance mechanism. An inter-pair distance is obtained from the measured FRET efficiency provided that a reasonable value of the orientation factor,  $\kappa^2$ , can be estimated based on previous knowledge of the system (equations explaining the relationship between transfer efficiency and distance as well as orientation can be found in the ‘Materials and Methods’ section) (5). Because of limitations in available dyes (6), by far the most FRET experiments use external fluorophores being tethered to the nucleic acid through a linker, thus, introducing dye diffusion and reorientation hampering the interpretation of quantitative experiments (7–9). Although recent advances have progressed the modelling of linker flexibility in quantitative FRET measurements (10–13), external labelling will always be accompanied by an inherent limitation in the information obtainable from the technique.

We previously reported an all-nucleobase FRET pair system consisting of two, now commercially available, base analogues, tC<sup>O</sup> and tC<sub>nitro</sub> (Figure 1a) (14). These probes possess relatively stable photophysical properties in nucleic acid environments and are rigidly positioned inside the DNA/RNA structure mimicking the hydrogen bonding and base-stacking of natural cytosine (14–16) (manuscript in preparation regarding tC-family in RNA). The well-defined transition dipole position and

\*To whom correspondence should be addressed. Tel: +46 31 772 30 51; Fax: +46 31 772 38 58; Email: marcus.wilhelmsson@chalmers.se  
Present address:

Francois-Alexandre Miannay, Physical Chemistry Department, University of Geneva, Geneva, CH-1211, Switzerland.



**Figure 1.** Using base probe building blocks for constructing 3D nucleic acid models and base–base FRET simulations. (a) Chemical structures of the tC base probes in their base pairing environment with guanine. tC and tC' may act as FRET donor with tC<sub>nitro</sub> as acceptor. (b) Definition of local tC–G base pair coordinate frame with tC in strand I (left) and strand II (right). In the base pair coordinate frame the y-axis is parallel to the line connecting the C4 atom of the tC base and the C8 atom of the complementary G base pointing from strand II to strand I. The z-axis points in the 5'-→3' direction of strand I, and the x-axis completes a right-handed set. The origin of the base pair frame is the midpoint of the C4–C8 line. (c) Definition of base pair step parameters used to construct 3D nucleic acid geometries. Shift, slide and rise are translational parameters, whereas tilt, roll and twist are rotational parameters. (d) Definition of 3D unit vectors used to simulate FRET in the constructed nucleic acid geometries.

orientation at the timescale of energy transfer is a considerable advantage of these FRET probes offering the potential to retrieve distance and orientational information from FRET measurements. Furthermore, the ability to position the reporters inside the very site of interest is an attractive, if not vital, feature in a majority of studies (17,18). However, the technical and theoretical challenges involved in the simulation and quantitative analysis of these probes have up till now posed serious limitations in the design and analysis of base–base FRET experiments. This calls for a diverse general methodology to simulate and analyse base–base FRET in any kind of nucleic acid structure.

Here, we report a general extendable methodology for the (i) simulation and (ii) quantitative analysis of FRET in nucleic acids. However, the method is particularly powerful for modelling constrained probes, including not only modified bases but also rigidly bound dyes (19–25). In the simulation part (i), an all-atom 3D nucleic acid model is constructed followed by a FRET simulation between static donor and acceptor dipole vectors. The generality of the implemented model building scheme makes the simulations completely independent of the

geometrical shape of the nucleic acid allowing FRET in any structure to be modelled fast. To facilitate quantitative evaluation of FRET experiments in the analysis part (ii), we introduce user-defined directional probability distributions representing nanosecond rotational dynamics of the probes. The analysis routine performs a direct global fit of multiple time-resolved donor intensity decays from FRET experiments that increases the level of information obtainable from the measured data compared with analysing FRET efficiencies only. Depending on the objective of a given study, the analysis routine can easily be modified to search the data for local or global structural or dynamical features and provides a direct correlation between 3D nucleic acid structures and measured FRET signals.

The method is demonstrated experimentally by a combinatorial base–base FRET pairing approach, in which multiple probe positions are combined to gain quantitative information about the structure and the dynamics of the nucleic acid. In demonstration Study 1, the method is used to probe the local orientation and rotational fluctuations of the bases inside double-stranded DNA in solution. The second experimental study demonstrates

how base-base FRET in combination with FRETmatrix can be used to reconstruct the 3D structure of local kinks, such as protein/ligand binding sites, bulges, junctions and DNA lesions. Here, we study two model structures allowing us to gauge our method against previous results and find excellent agreement. The information gained using the presented method is otherwise obtainable only hypothetically by means of much more complex and time-consuming NMR or single-molecule methods; thus, the method shows great promise for future innovative investigations. All of the reported methodology is accompanied by a MATLAB-based software package, FRETmatrix, equipped with a user-interface and freely available from <http://www.chalmers.se/chem/EN/divisions/physical-chemistry/staff/marcus-wilhelmsen/fretmatrix>.

## MATERIALS AND METHODS

### Building nucleic acid geometrical models

FRETmatrix implements a matrix-based base-centred calculation scheme [the Cambridge University Engineering Department Helix Computation Scheme (CEHS)] to build nucleic acid geometries (26). The scheme uses a standardized reference frame for the description of nucleic acid structures, in which the geometry of all base pairs and base pair steps in the structure are characterized by a set of rigid body parameters as described in detail previously (27–29). The reference frame definitions associate each base and each base pair with a local coordinate frame in which the Cartesian coordinates of all  $n$  atoms are predefined and stored in coordinate matrices:

$$\text{XYZ} = \begin{bmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \quad (1)$$

Building a 3D geometrical model is then accomplished by a set of matrix operations serving to position each local coordinate frame within the global coordinate frame of the structure. Such rotations are accomplished using the three rotation matrices:

$$\text{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos a & -\sin a \\ 0 & \sin a & \cos a \end{bmatrix} \quad (2)$$

$$\text{R}_y = \begin{bmatrix} \cos a & 0 & \sin a \\ 0 & 1 & 0 \\ -\sin a & 0 & \cos a \end{bmatrix} \quad (3)$$

$$\text{R}_z = \begin{bmatrix} \cos a & -\sin a & 0 \\ \sin a & \cos a & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

For example, if we wish to rotate a set of coordinates arranged in an **XYZ** matrix (equation 1) through an angle  $a$  about the  $x$ -axis, this is analogous to the matrix product:

$$\text{xyz} = \text{XYZ}^* \text{R}_x^T(a) \quad (5)$$

where  $\text{R}_x^T$  is the transpose of the rotation matrix  $\text{R}_x$ . Translating the transformed coordinates by a vector  $\text{v}_{xyz}$ , arranged in an  $n \times 3$  matrix  $\text{v}$ , is then accomplished by

$$\text{xyz} = \text{xyz} + \text{v} \quad (6)$$

to yield the final set of 3D coordinates  $\text{xyz}$  defining the new position and orientation of the body. Using equations (1–6), one can construct and orient any geometrical body defined in three dimensions. In particular, building nucleic acid geometries is accomplished by rotating and translating the local base and/or base pair coordinate frames into the global nucleic acid coordinate frame. For example, going from one base pair to the next is accomplished by three consecutive rotations (tilt, roll and twist) and a three-coordinate translation (shift, slide and rise). See El Hassan and Calladine (26) and Lu *et al.* (27) for additional descriptions of the matrix-based equations used in CEHS.

### Building geometries containing FRET probes

Based on the detailed information we have previously obtained regarding the structural and electronic properties of the tC bases (15,30,31), we can implement these base probes within the standardized nucleic acid reference frame and, thus, the CEHS model building scheme. We define the local tC base coordinate frame directly analogously to the pyrimidine base frame, and the local base pair coordinate frame of a synthetic tC-G base pair directly analogously to the corresponding C-G pair (Figure 1b and Supplementary Figure S1). The vectorial representations of the transition dipole moments of the probes are similarly defined within the local coordinate frames allowing the dipole positions and orientations to be unambiguously specified relative to the atomic coordinates of the base pair (Supplementary Tables S1 and S2).

Defining the local coordinate frame of a synthetic base allows us to simulate the exact position and orientation of probes positioned in any 3D nucleic acid structure by building up a geometrical model using the standard structural parameters to describe each dinucleotide step in the structure (Figure 1c). The result of this model building routine is a global coordinate matrix describing the position of all atoms and the positions and directions of the 3D transition dipole vectors of FRET pairs positioned in the structure.

### Building two structural units separated by a kink

Building geometrical models consisting of multiple joined structural modules, such as two B-DNA helices separated by a kink, is achieved by FRETmatrix in a similar manner as constructing a base pair step namely using three Euler angles and a translation vector to describe the exact relative orientation and position of the two units. In FRETmatrix, the first base pair of unit 1 is defined as the global coordinate frame of the structure. The second unit, i.e. the second helix, is initially built within the coordinate frame of the first base pair of unit 2 and is then subsequently rotated and translated according to the specified kink parameters and is finally aligned with the last base pair of the previous unit using equations (1–6). Using this approach, an unlimited number of structural units can be joined.

### Simulating FRET in 3D nucleic acid structures

In short, the FRET efficiency between a FRET pair positioned in the geometrical model is given by

$$E = \frac{R^6}{R_0^6 + R^6} \quad (7)$$

where the critical Förster distance (the distance at which  $E = 0.5$ ) is calculated as

$$R_0 = 0.211 \left( \frac{\kappa^2 \Phi_D J}{\eta^4} \right)^{\frac{1}{6}} \quad (8)$$

Here,  $\Phi_D$  is the donor quantum yield in absence of acceptor,  $J$  is the spectral overlap integral between the donor emission and acceptor absorption spectrum and  $\eta$  is the refractive index, all being parameters exploited as previous knowledge. With the global coordinate matrix at hand, the two geometrical parameters,  $\kappa$  and  $R$ , are simulated using the extracted donor and acceptor dipole vectors: the orientation factor in the energy transfer process is calculated as a sum of vector products

$$\kappa = \vec{e}_1 \cdot \vec{e}_2 - 3(\vec{e}_1 \cdot \vec{e}_{12})(\vec{e}_{12} \cdot \vec{e}_2) \quad (9)$$

where  $\vec{e}_1$  and  $\vec{e}_2$  are the dipole unit vectors, and  $\vec{e}_{12}$  is the unit vector connecting the two dipole centres (Figure 1d), whereas the distance between two dipole centres,  $(x_1, y_1, z_1)$  and  $(x_2, y_2, z_2)$ , is calculated as

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (10)$$

More detailed information on the calculation of FRET is provided in Supplementary Note S1.

### Modelling base dynamics

The simulations described above assume rigid nucleic acid geometries and static probe dipoles. The results of such first principle simulations are particularly useful for qualitative guiding purposes. In solution, nucleic acid dynamics occurring on the same timescales as the energy transfer process, i.e. from the sub picosecond range to the nanosecond regime, result in dipole orientational fluctuations during the excited state lifetimes of the probes. In ensemble measurements, the result is a distribution of donor–acceptor orientations, which should be implemented in the simulation to evaluate FRET experiments more quantitatively.

To model rotational dynamics of the probes, we extend the structure building scheme by including dipole vector directional distributions in the FRET simulation (Supplementary Note S2). First, it is recognized that the anisotropic microenvironment of the base probe when base pairing to its complementary base leads to the definition of two spherical coordinates representing two different modes of rotational movement of the base (Figure 2a). Here, the angle  $\theta$  describes in-plane movement primarily being influenced by the potential energy imposed by the H-bonding to the complementary base. The angle  $\phi$ , on the other hand, describes out-of-plane movement influenced by the base-stacking with neighbouring bases

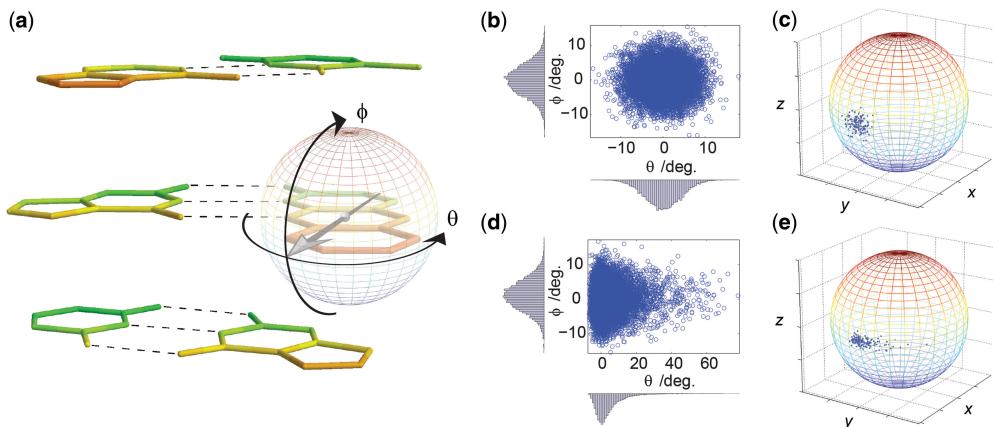
and internal bending modes of the structural framework of the probe. If the simulation is to be as physically accurate as possible, the directional probability distribution representing dipole reorientation should describe these fundamentally different characteristics along the  $\theta$ - and  $\phi$ -coordinates of the potential energy surface in which the base moves.

We model base fluctuations by assigning a 1D potential energy function to each of the two modes of dipole rotation. The corresponding Boltzmann distribution at room temperature then provides us with two 1D probability density functions,  $P_i(\theta)$  and  $P_o(\phi)$ , describing the probability of finding the dipole vector along each of the two coordinates (Supplementary Figure S2). Here, the probabilities of  $\theta$  and  $\phi$  are independent, and the joint probability distribution over  $\theta$  and  $\phi$  is, therefore, given by

$$P(\theta, \phi) = P_i(\theta) \times P_o(\phi) \quad (11)$$

Although this equation accurately describes the probability distribution of the orientation of the base probe within a base stack, we have not been able to identify a closed analytical expression relating the energy transfer rate constant with expression (equation 11). The desired vector probability distribution is therefore implemented by sampling the joint probability distribution of  $P_i$  and  $P_o$  over each of the spherical coordinates (Figure 2b and d) and then subsequently transform the samples into Cartesian space (Figure 2c and e). To circumvent samples being piled up at the poles (Supplementary Figure S3), samples are initially drawn with a maximum likelihood of  $(\theta, \phi) = (0, 0)$ , i.e. in Cartesian space, the unit vector pointing in the  $(x, y, z) = (1, 0, 0)$  direction. The drawn samples are then subjected to a set of Euler rotations aligning the maximum likelihood of the distribution, with the direction of the transition moment vector of the probe positioned in its local coordinate frame. The distribution is put into place within the global nucleic acid structure in the same way as described earlier in the text for positioning a base pair. The resulting vector population, now representing the dipole orientational distribution traced during an energy transfer event, is used to calculate the corresponding energy transfer rate constant and simulate the donor decay. This stepwise construction of the dipole unit vector distribution ensures that the sampled vector distribution represents the potentials defined for the microenvironment of the probe, provided that the width of  $P_o(\phi)$  does not allow samples to be drawn near  $\phi = \pm 90^\circ$ . The form and width of the sampled dipole vector distribution directly reflect the base dynamics occurring on the timescale of the energy transfer, usually being in the range of 0.5–5 ns.

Assigning independent functional forms to the in-plane and out-of-plane rotational modes of the base probes makes the form and width of the simulated dipole vector distribution highly versatile (Supplementary Figure S4). For example, assigning harmonic potentials to  $\theta$  and  $\phi$  produces dipole vector distributions with appearances similar to the Kent or the Mises–Fisher directional distributions (Figure 2b and c), the analogues of the bivariate



**Figure 2.** Simulating dipole vector directional distributions. (a) Definition of two spherical angles describing in-plane ( $\theta$ ) and out-of-plane rotation ( $\phi$ ) of a base probe in double-stranded DNA (only the nucleobases are shown). The two neighbouring base pairs are shown above and below the tC–G base pair. The strand in which the base probe is positioned has 5' directed downwards and 3' upwards. (b–e) Sampling dipole vectors from two marginal Boltzmann probability distributions. First, samples are drawn in spherical coordinates (b, d) and then subsequently transformed into Cartesian space where samples are located on the unit sphere representing dipole unit vectors (c, e). The form and width of the distribution in (b) and (d) is the same as in (c) and (e), respectively. In (b, c), a harmonic potential is used to describe in-plane and out-of-plane motion, corresponding to sampling a bivariate Gaussian distribution. In (d, e), a harmonic potential is used to describe out-of-plane motion, and a Lennard-Jones potential is used to describe the potential imposed by the in-plane hydrogen-bonding to the complementary base.

Gaussian distribution for directional data in three dimensions (10). To mimic the base pairing environment of a base probe, we model the in-plane movement using a Lennard-Jones potential (Figure 2d and e). This function first of all possesses the correct appearance expected for a hydrogen-bond and, secondly, is straight forward to model (Supplementary Figure S2). As shown below, the out-of-plane movement is almost negligible at the time-scale of the energy transfer, making the choice of this functional form less important. For this reason, we use a harmonic potential to represent the out-of-plane movement of the base.

### Simulating donor intensity decays

As the entire time-resolved donor intensity decay contains several hundred more data points than the corresponding steady-state FRET efficiency, the amount of information obtainable from a FRET experiment is greatly increased by introducing the structural and dynamical parameters directly in the donor intensity decay fitting. We simulate the donor decays based on an exponential decay model that includes the Förster characteristics derived from the constructed nucleic acid geometry and associated dipole vector distributions. In the dynamic averaging regime the donor intensity decay in presence of FRET is

$$I_{\text{dyn}}(t) = e^{-\frac{t}{\tau_D} \left( 1 + C \times \frac{\langle \kappa^2 \rangle}{R^6} \right)} \quad (12)$$

where  $C = 0.211^6 \left( \frac{\Phi_D J(\lambda)}{\eta^{-4}} \right)$ ,  $\tau_D$  is the donor lifetime in absence of acceptor, and  $\kappa$  and  $R$  are defined earlier in the text. Taking into account the potential presence of other fluorophores but the donor, the total decay is

simulated as

$$I_{\text{sim}}(t) = I_0 \left( b \left[ a e^{-\frac{t}{\tau_D} \left( 1 + C \times \frac{\langle \kappa^2 \rangle}{R^6} \right)} + (1-a) e^{-\frac{t}{\tau_D}} \right] + (1-b) e^{-\frac{t}{\tau_2}} \right) \quad (13)$$

where  $I_0$  is a pre-exponential factor,  $(1-a)$  is the fraction of donors not coupled to an acceptor (Supplementary Figure S5),  $(1-b)$  is the fraction of emission from any other fluorophore but the donor, with a lifetime of  $\tau_2$  (in our experiments being <5%). The simulated decay is convolved with the instrument response function of the time-correlated single photon counting measurement:

$$I_c(t) = \int_0^t \text{IRF}(t') \times I_{\text{sim}}(t-t') dt' \quad (14)$$

executed in FRETmatrix using the efficient fast Fourier transform-based method of overlap-add. The reduced  $\chi^2$  of the fit is then used to evaluate the goodness of the fit and is calculated as

$$\chi_r^2 = \left[ \sum_{k=1}^n \frac{(I_m(t_k) - I_c(t_k))^2}{I_m(t_k)} \right] \times \frac{1}{n} \quad (15)$$

where  $I_m(t_k)$  is the number of measured counts at time  $t_k$ ,  $I_c(t_k)$  is the calculated number of counts at time  $t_k$ , and  $n$  is the total number of datapoints (number of channels  $\times$  number of samples) (32).

### Experimental details

All experiments were performed in phosphate buffer ( $[\text{Na}^+] = 0.1 \text{ M}$ ) at  $T = 295 \text{ K}$ . Donor concentrations were  $0.5\text{--}2 \mu\text{M}$ , with 30% excess of the acceptor strand.

Annealing was performed by heating up to 368 K followed by slow cooling to room temperature (RT) (12 h). Fluorescence lifetimes measurements were performed using time-correlated single photon counting. The excitation wavelength at 377 nm was delivered by a 10 MHz pulsed laser diode (PicoQuant). The fluorescence of the samples was spectrally filtered at 460 nm by a monochromator and detected by a thermoelectrically cooled microchannel plate photomultiplier tube Hamamatsu R3809U-66. The counts were sent to a multichannel analyser (Lifespec, Edinburgh Analytical Instruments) adjusted to 2048 channels, where a minimum of 10 000 counts were recorded in the top channel. The spectral resolution of the monochromator was fixed at 5 nm (emission slit width) and the time window was 100 ns. The instrument response function has a full width at half time maximum (FWHM) of 60 ps. For all FRET-pair positions, the lifetime measured of the donor-no-acceptor sample was used as donor reference.

#### Demonstration Study 1

One sequence motif was used to design three donor strands, each being complementary to four acceptor strands as described previously (Supplementary Table S3) (14). All probes were associated with identical directional vector distributions in the analysis. The in-plane movement was described by a Lennard-Jones potential defined by a dissociation energy of 0.1 eV, an experimentally determined average value of the three C–G hydrogen bonds (33,34). For the out-of-plane movement, we applied a harmonic potential. To obtain a low standard deviation in the calculation of  $\chi_r^2$ , we sampled  $N = 1000$  vectors when constructing the dipole distributions of each probe corresponding to  $10^6$  dipole–dipole combinations of each FRET-pair resulting in a standard deviation of the calculated  $\chi_r^2$  of  $s(\chi_r^2) < 0.1$  (Supplementary Figure S6). Increasing  $N$  for higher statistical significance is at the expense of increased computational time. All decays were analysed using equation (13) with experimentally determined input parameters (Supplementary Table S4).

#### Demonstration Study 2

Monomeric probes were purchased from Glen Research (Sterling, Virginia) and incorporated into DNA oligonucleotides and purified by reversed phase high-performance liquid chromatography (14) by ATDBio (Southampton, UK). DNA samples were designed to yield a large number of donor–acceptor combinations across the kinks with one global sequence motif for all samples (Supplementary Table S5). Donor decays from 18 FRET combinations of the 0A bulge and 16 combinations of the 3A bulge were combined and analysed using the six kink parameters as the only variables. To increase the number of FRET combinations, the design was made with the kink at two different positions in the sequence (only one kink per sample). In the analysis, the same set of kink parameters were used for both kink positions. Before beginning the FRET analysis, all decays were fitted to an exponential decay model providing decay parameters needed for the analysis (Supplementary Tables S6

and S7). All decays were then analysed using FRETmatrix and equation (13), using the experimentally determined input parameters (Supplementary Table S8). All donor–acceptor pairs were associated with the same set of simulation parameters as used in demonstration Study 1 (see earlier in the text) and, in addition  $\theta\text{-FWHM} = 13^\circ$ ,  $\phi\text{-FWHM} = 2^\circ$ ,  $\phi_A = 25^\circ$ ,  $\phi_D = 8^\circ$ .

Two geometrical constraints were included in the DNA kink analysis algorithm (Supplementary Note S3):

- Kink parameter values resulting in a sterical clash between two or more atoms were automatically discarded. Here, the atomic van der Waals radius was set to 1.3 Å.
- A maximum distance of 11 Å between two neighbouring bases positioned in the same strand was allowed. This structural constraint is imposed physically by the covalent bonds connecting the two neighbouring bases.

For the 0A structure parameter, values optimized with and without geometrical constraints were identical. For the 3A structure parameter, values optimized with and without geometrical constraints were close to being identical (Supplementary Table S9).

#### Construction of $\chi^2$ surfaces

The calculation of  $\chi_r^2$  surfaces was built from 50–1600 point calculations depending on the parameter range. Interpolation was used to gain smooth surfaces. At each coordinate on the surface the parameters not being constrained in the analysis was optimized. Different algorithms were needed to optimize the parameters in the model. For parameters not associated with the random sampling process, a gradient-based algorithm with upper and lower boundary conditions was used. However, for  $\phi\text{-FWHM}$  and  $\theta\text{-FWHM}$ , the non-linear pattern search algorithm was chosen, as the random sampling step makes the object function non-differentiable. Optimizing kink parameters, including geometrical constraints, was best achieved using a two-step algorithm: an initial rough parameter search using the simulated annealing algorithm followed by a gradient-based optimization for determining the minimum more accurately. The confidence intervals of each fit parameter in the models were calculated based on the *F*-statistic using  $P = 0.05$  ( $>95\%$  chance that parameter value is consistent with data) (32). Running the  $\chi_r^2$  surfaces shown requires anything from a few minutes to several hours on a modern day laptop, depending on the size of the surface and the range of parameter values investigated.

#### Software availability

The reported methodology is implemented in a MATLAB-based software package, which can be downloaded for free at <http://www.chalmers.se/chem/EN/divisions/physical-chemistry/staff/marcus-wilhelmsson/fretmatrix>. The program is designed for the preliminary design and subsequent quantitative analysis of experiments involving constrained FRET probes. A user guide is provided (Supplementary Note S3).

### Implementing a new probe in FRETmatrix

To implement a new FRET probe in the methodology, the atomic coordinates of the base relative to the base reference frame must first be known and the direction of the transition dipole vector within the structural framework of the probe. A base pair building block is then constructed using a set of rigid body base pair parameters (stagger, stretch, shear, propeller, opening and buckle), the analogues of base pair step parameters (29). More detailed information, including a script constructing base pair building blocks from the atomic coordinates of a new base, is supplied with FRETmatrix (*Supplementary Note S3*). Examples of other dyes currently implemented in FRETmatrix are shown in *Supplementary Figure S7*.

## RESULTS

### Simulations of model geometries

#### Building model geometries

The versatility of the model building approach is demonstrated by three representative output structures from the FRETmatrix software (*Figure 3a–c*). FRETmatrix can build and gather regular A-form and B-form helices containing the FRET probes directly from an input sequence using base pair step parameters previously derived from experimentally determined structures (*Figure 3a* and *b* and *Supplementary Table S10*) (28,29). The program produces a Protein Data Bank (PDB) molecule file of the simulated structure for visual inspection purposes only. More complex nucleic acid structures, predefined in a PDB file, can be simulated through the use of a structural analysis routine that extracts all the structural parameters necessary to rebuild the geometrical arrangement of all bases in the structure (*Supplementary Note S3*) (28,35). The extracted structural parameters are then used as input for FRETmatrix that rebuilds a geometrical model of the structure inserting base probes at any desired positions. This rebuilding routine is demonstrated by PDB entry 1TGH, the complex formed between the TATA-binding protein (TBP) and its DNA target (*Figure 3c* and *Supplementary Data S1*) (36). In this form of simulation, the inserted FRET probe will possess the same geometrical position and orientation as the substituted base.

#### Simulating FRET in model geometries

The simulated FRET efficiencies of 16 base–base FRET combinations systematically positioned in standard A-form and B-form helices illustrate some of the advantages provided by the methodology presented here (*Figure 3d* and *e*). Firstly, the characteristic helical periodicity of A-form and B-form DNA are automatically seen as structural fingerprints without the need of imposing structure-dependent geometrical definitions and mathematical expressions relating the transition dipole vectors to the nucleic acid helical framework, such as in the still widely used model of Clegg *et al.* (37). Secondly, the method automatically takes into account any 5'/3' effects, thus showing some insightful differences between the energy transfer efficiencies calculated for donor–acceptor pairs separated

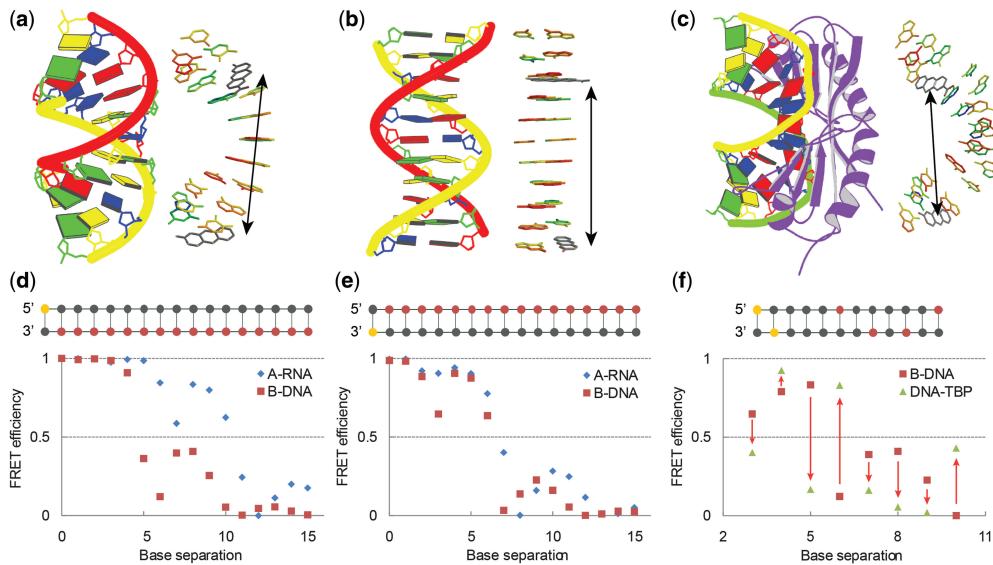
by the same distance but positioned 5' (*Figure 3d*) versus 3' ends (*Figure 3e*) on separate strands. These FRET differences are because of the difference in the relative orientation between donor and acceptor dipole vectors as can be concluded from the calculated FRET characteristics (*Supplementary Data S2*).

The high potential of base–base FRET to report on detailed structural changes of nucleic acids, e.g. on binding of a protein or other DNA ligands, is seen by comparing the simulated FRET signals at chosen donor–acceptor positions in regular B-DNA with the same set of donor–acceptor positions in the complex formed on binding of TBP (*Figure 3f*). These calculations predict the signal change that would occur when TBP binds to DNA, assuming the structural model is correct. The donor–acceptor positions chosen for this demonstrational simulation were selected from a screening of all possible FRET combinations, performed using an automated feature in FRETmatrix, with the criteria of yielding high signal changes on binding of substrate (*Supplementary Note S3*). Such simulations are particularly useful in studies requiring high throughput, e.g. for complementing higher resolution structural techniques when comparing different protein homologues bound to DNA. The high change in FRET signal at these positions is a combination of a change in donor–acceptor distance and, importantly, a change in the relative orientation between donor and acceptor (*Supplementary Data S2*).

#### Demonstration Study 1

To demonstrate experimentally how FRETmatrix and base–base FRET can be used to extract quantitative information on local nucleobase dynamics, we performed a global analysis of nine combinations of the tC<sup>O</sup>–tC<sub>nitro</sub> pair positioned in B-DNA with distances varying from 5–13 base pairs (*Figure 4b* insert and *Supplementary Table S3*). As the overall helical structure of this model system is already well known, this allows us to analyse the data in terms of the geometry and directional fluctuations of the base probes in their base pairing environment with guanine. Besides demonstrating the power of the method, this study additionally provides information that can be exploited as previous knowledge in studies where other structural features of the nucleic acid are being probed as shown later in the text.

For this study, we defined four unknown parameters to be analysed while leaving all other variables constrained based on previous knowledge (*Supplementary Tables S4* and *S10*). Two parameters describe the nanosecond dynamics of the probes, namely the FWHM of the directional distributions representing in-plane and out-of-plane base fluctuations. We additionally recognized the out-of-plane dipole bending angle,  $\phi_b$ , of tC<sup>O</sup> and tC<sub>nitro</sub> as two unknown variables of the system (*Figure 4c* insert). This geometrical parameter was introduced based on previous theoretical calculations suggesting that the tricyclic framework of the tC probes is highly flexible in terms of out-of-plane bending along the middle S-N axis of tC/tC<sub>nitro</sub> or the O-N axis of tC<sup>O</sup> (38). In there, we proposed a model in which the structure of the tC bases



**Figure 3.** Demonstration of geometrical model building combined with FRET simulations in three model structures. (a–c) Representative examples of output geometries produced by FRETmatrix (right) along with the block representation of the corresponding structures produced by 3DNA (left). Inserted FRET pairs are shown in grey and marked with arrows. The simulated structures are A-form RNA (a), B-form DNA (b), and PDB entry 1TGH: the complex between the TBP and DNA (c). (d, e) Simulated FRET efficiencies between tC<sup>O</sup> and tC<sub>nitro</sub> at selected positions within model structures (Supplementary Data S2). The positions of tC<sup>O</sup> and tC<sub>nitro</sub> in the strands are illustrated in top with the position of tC<sup>O</sup> marked in yellow and the position of tC<sub>nitro</sub> marked in red. Base separation denotes number of base pairs in between the FRET pair. In (f), red arrows denote change in FRET signal that would occur on binding of TBP to double-stranded DNA.

is energetically guided into a conformer characterized by bending into the major groove as a result of steric interactions with the 5' neighbour. Confirming/disconfirming this model not only demonstrates the power of the method but also provides valuable insight into the properties of the tC probes in confined biological environments.

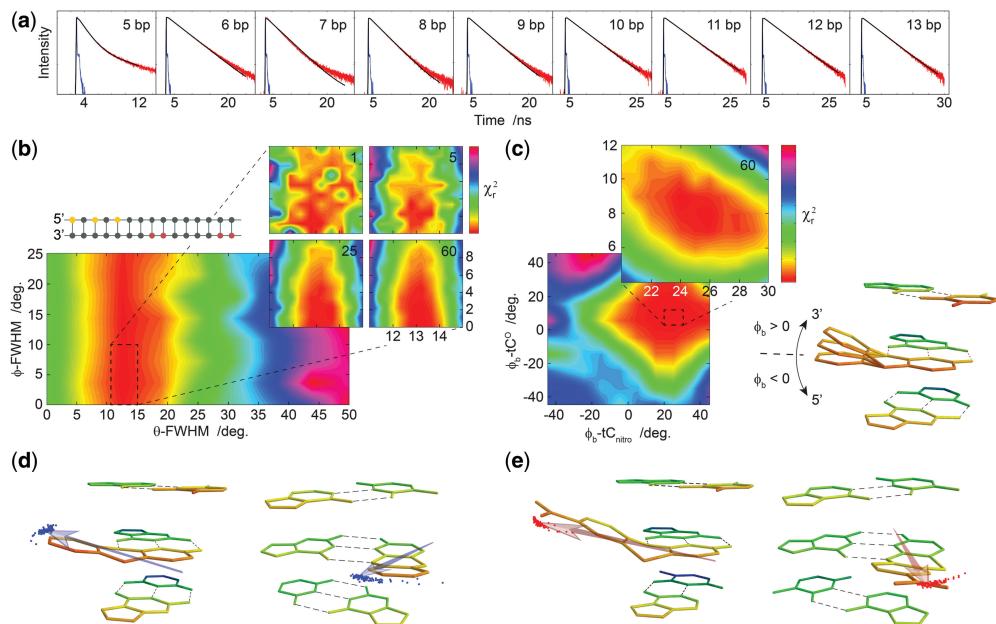
As the object function of our simulation, being the fit between measured and theoretical intensity decays, is dependent on random sampling, optimizing the parameters of the model is not achievable using a traditional gradient-based algorithm. Instead of performing an automated parameter optimization, we systematically investigate the entire parameter space relating all parameter values with the goodness of the global fit. The resulting  $\chi^2_r$  surfaces not only reveal the optimal parameter values but also provide a direct visualization of the uniqueness of the best fit and an estimation of the associated confidence intervals (Figure 4b and c). To overcome the inherent variations of the calculated  $\chi^2_r$  as a result of the implemented sampling procedure (Supplementary Figure S6), the simulation is performed several times averaging the calculated  $\chi^2_r$  values in the final plot (Figure 4b and c inserts).

From this  $\chi^2_r$  analysis routine, a single deep minimum is revealed on the  $\chi^2_r$  surface (Figure 4b and c) with optimal parameter values of  $\theta$ -FWHM =  $13.1^\circ \pm 0.3^\circ$ ,  $\phi$ -FWHM <  $2.5^\circ$ ,  $\phi_b$ -tC<sup>O</sup> =  $8.2 \pm 1.7^\circ$  and  $\phi_b$ -tC<sub>nitro</sub> =  $25.0 \pm 2.5^\circ$ . Visualizing the dipole vector distribution corresponding to these parameter values reveals a highly constrained orientational freedom of the base probes

inside double-stranded DNA (Figure 4d and e). The negligible out-of-plane movement of the bases at the timescale of the energy transfer provides a direct insight into the influence of base-stacking on the dynamics of nucleobases. Strikingly, the out-of-plane bending angles correspond to bending towards the 3' end of the strand in which the probes are positioned, thus confirming the previously suggested model in which the tricyclic framework of the tC bases is guided into the major groove (38). The fitted dipole bending values are close to the folding angles of the tC bases being < $10^\circ$  and  $26^\circ$  for tC<sup>O</sup> and tC<sub>nitro</sub>, respectively, predicted using density functional theory (38).

### Demonstration Study 2

Using the measured emission decay data from 16–18 different positions of the tC<sup>O</sup>–tC<sub>nitro</sub> FRET-pair, we demonstrate how base–base FRET together with FRETmatrix can be used experimentally to reconstruct 3D nucleic acid structures. The following two model systems are studied: a regular base pair step (0A bulge, Figure 5a) and a three adenine bulge (3A bulge, Figure 5b). A translation vector,  $\vec{v}$ , and three Euler angles with a ZXZ convention describe the relative position and orientation of the two structural units separated by the kink (Figure 5c), whereas the two helices themselves are modelled using base pair step parameters for regular B-DNA (Supplementary Table S10). Here, the 0A bulge constitutes the only model system that allows us to compare all six fitted kink parameters with known values, whereas mainly the DNA bending



**Figure 4.** Using quantitative base–base FRET to obtain information about the orientation and nanosecond dynamics of the base probes in DNA. **(a)** Globally fit decays using the four optimized parameters. **(b)** Calculated  $\chi^2$  surface of the width of the directional distribution in the in-plane direction ( $\phi$ -FWHM) and the out-of-plane direction ( $\phi$ -FWHM). Insert (left) illustrates the positions of donor and acceptor in the samples with  $tC^O$  marked in yellow and  $tC_{nitro}$  in red. In the experiment, only one donor–acceptor pair is positioned in each sample corresponding to nine different donor decays. Colour bar for large surface: 3.5:42. Insert (right) shows calculated  $\chi^2_r$  surface averaged from 1, 5, 25 and 60 simulations (colour bar: 3.5:4.3). **(c)** Calculated  $\chi^2$  surface of the dipole bending angle of  $tC^O$  ( $\phi_b-tC^O$ ) and  $tC_{nitro}$  ( $\phi_b-tC_{nitro}$ ). Colour bar: 3.5:170. Insert (upper) shows averaged  $\chi^2$  surface after 60 simulations (colour bar: 3.5:5.5). Insert (right) illustrates the definition of  $\phi_b$ . The model shows a base probe in B-DNA where the 3' end of the strand in which the probe is positioned points upward and 5' is pointing downwards. Only the nucleobases are shown. **(d, e)** Illustration of the distribution made with the optimized parameters viewed from the side (left) and from the front (right) of  $tC^O$  (d) and  $tC_{nitro}$  (e).

angle,  $\beta$ , of the 3A bulge can be compared with results from previous studies (12,39,40). In the analysis, the base probe bending angles and dipole distributions obtained from demonstration Study 1 were exploited as previous knowledge, allowing the DNA geometries to be probed with greater accuracy.

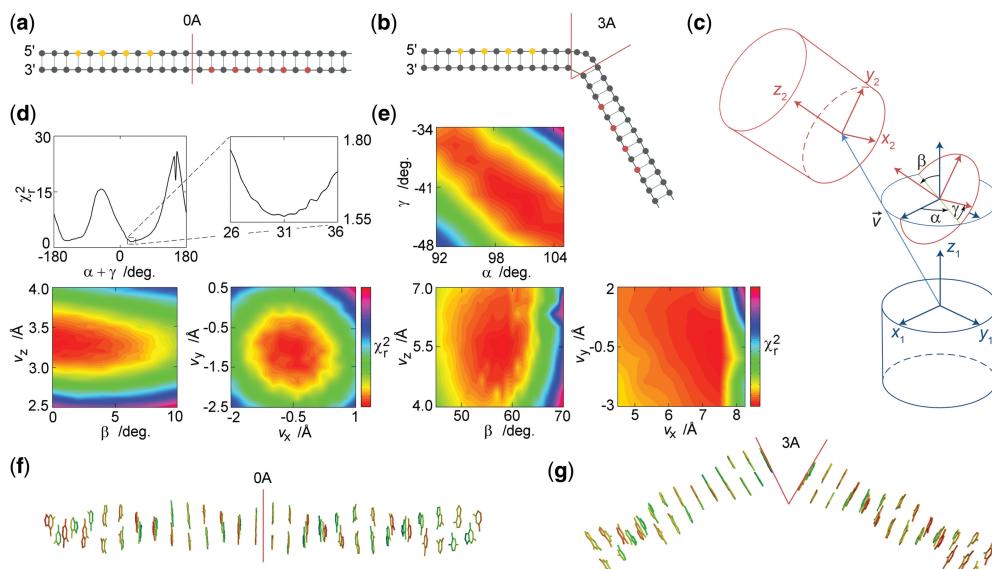
For both model structures, the results are in good agreement with expected geometries (Figure 5d–g). The optimized 0A kink parameter values are  $v_x = -0.55 \pm 0.3 \text{ \AA}$ ,  $v_y = -1.1 \pm 0.3 \text{ \AA}$ ,  $v_z = 3.3 \pm 0.1 \text{ \AA}$ ,  $\alpha + \gamma = 31 \pm 1^\circ$ ,  $\beta < 2^\circ$  (Figure 5 d and f). All six 0A parameter values are thus in close agreement with the values for a regular base pair step with an accuracy of  $< 1 \text{ \AA}$  for the translational shift and a few degrees for the angles of rotation. Notice that for  $\beta = 0$ ,  $\alpha$  and  $\gamma$  are directly correlated and display a sinusoidal influence on  $\chi^2$  with a period of  $180^\circ$  (Figure 5d), as twisting the two helical units by an angle  $\alpha + \gamma = 180^\circ$  from an initial to a second state results in the dipole vectors being oppositely oriented but practically parallel as in the initial state.

The optimized 3A kink parameter values are  $v_x = 7.4 \pm 0.1 \text{ \AA}$ ,  $v_y = -0.55 \pm 0.6 \text{ \AA}$ ,  $v_z = 5.5 \pm 0.2 \text{ \AA}$ ,  $\alpha = 101 \pm 2^\circ$ ,  $\beta = 57 \pm 1^\circ$ ,  $\gamma = -43 \pm 2^\circ$  (Figure 5e and g). The 3A DNA bending angle of  $\beta = 57^\circ \pm 1^\circ$  is in good agreement with previous studies using ensemble

FRET ( $50^\circ$ – $70^\circ$ ) (39), transient electric birefringence ( $58^\circ \pm 4^\circ$ ) (40), and single-molecule FRET ( $56^\circ \pm 4^\circ$ ) (12). The all-atom 3A structure shows good overlap with the structure of another 3A bulge geometry obtained using single-molecule FRET with multiparameter fluorescence detection (Supplementary Figure S8) (12). The global fit between measured and theoretical intensity decays of all D-A separations is good for the 0A and 3A model systems (Supplementary Figures S9 and S10, respectively). It is noted that the full  $\chi^2$  surface oscillates with a period of  $\alpha + \gamma = 180^\circ$  (Supplementary Figure S11) as observed for the 0A bulge model system. The true global minimum was identified based on the assumption that DNA bending follows the right handed helical twist of the two helices ( $\alpha + \gamma \approx 55^\circ$ ).

## DISCUSSION

Compared with external labelling, base–base FRET offers a unique possibility to position the reporters inside the site of interest probing the local orientation and dynamics at specific base positions within nucleic acids. Global nucleic acid dynamics occurring at timescales exceeding the probe lifetimes are in principle also obtainable using the methodological framework presented here, although this was



**Figure 5.** Using quantitative base–base FRET to reconstruct the 3D structure of nucleic acids. (a) Model system 1: a regular base pair step simulated as a local site in B-DNA. Donor positions are in yellow and acceptor positions in red. (b) Model system 2: a three adenine bulge. (c) Definition of kink parameters. The two helical coordinate systems are the base pair coordinate frames of the two base pairs neighbouring the kink. (d)  $\chi^2$  surfaces of the 0A system based on the global analysis of 18 donor decays (Supplementary Figure S9). Colour bars: 1.55:1.97 (left) and 1.55:1.78 (right). (e) Global analysis of the 3A bulge system based on the global analysis of 16 donor decays (Supplementary Figure S10). Figures show only the global minimum on the  $\chi^2$  surface (Supplementary Figure S11). Colour bars: 1.85:2.12 (top), 1.85:2.39 (bottom left), 1.85:2.90 (bottom right). (f) Optimized 0A structure. (g) Optimized 3A structure.

not demonstrated herein. The combined signal from several donor–acceptor positions can provide high-resolution distance and orientational information of nucleic acid structures without complications associated with fluorophore linker flexibility or DNA–dye interactions.

Various approaches have previously been developed to model probe dynamics during the energy transfer process (10–12,19). Importantly, our method, which uses user-defined directional distributions, provides a one-step analysis without the need to include force field molecular dynamics simulations in the analysis. The highly versatile form of the vector distribution is a particular advantage when modelling the orientationally constrained base probes in various nucleic acid environments. Using actual energy potentials to describe the nucleic acid dynamics paves the way for novel experimental studies of the fundamental physical properties of DNA and RNA structures.

It is recognized that there are two modes of dipole dynamics, reorientation and diffusion, being reflected in the energy transfer efficiency through the values of  $\kappa^2$  and  $R$ , respectively (12). Although dipole diffusion is pronounced when measuring FRET between external fluorophores, the tC base probes are rigidly positioned at relatively close distances inside the DNA structures. For this reason, we only modelled the orientational fluctuations of the probes. However, the method is expandable to include dipole diffusion when studying more dynamic structures.

In the demonstration studies, we used dynamic averaging of  $\kappa^2$ , which assumes that the rotational correlation time of the probes is much faster than the energy transfer (5). This assumption was shown to be valid for external fluorophores (12) and is supported here by an internal correlation time of the base probes of  $\tau_{int} = 350$  ps estimated from the time-resolved fluorescence anisotropy decay of tC<sup>O</sup> in high-viscosity solution (Supplementary Figure S12). In addition, the donor decays used in our studies are all well-fit using a single lifetime to represent FRET, which is a strong indication of dynamic averaging (Supplementary Table S6).

## SUMMARY AND OUTLOOK

We have developed a general methodological platform for simulating FRET in nucleic acids and demonstrated its particular power in modelling probes possessing limited degree of diffusional and rotational freedom. The method is based on the ability to rapidly construct any 3D nucleic acid geometry and simulate FRET between probes positioned anywhere within the structure. Directional vector distributions are implemented to model rotational dynamics of the probes, which, in combination with direct global intensity decay fitting of multiple donor and acceptor pairs, may provide quantitative information about structural and dynamical properties of nucleic acids. The method was used in combination with base–base FRET to obtain insight into base

dynamics occurring on the timescale of energy transfer and to probe the exact 3D structure of kinked DNA in solution. Importantly, the method is versatile and expandable.

As a result of the rapidly progressing field of fluorescent nucleobase analogues (16–18) and other rigidly attached probes (19–25), including the popular Cy3–Cy5 pair shown to be partly constrained when tethered to the ends of nucleic acids (7,9,19,41,42), we anticipate that many fluorescent markers will be modelled in the future using the methodology presented here. Given the versatility of base–base FRET combined with the ready-to-use methodological platform reported here, we believe that new possibilities for experimental studies of nucleic acid structure and dynamics have opened up.

## SUPPLEMENTARY DATA

**Supplementary Data** are available at NAR Online: Supplementary Tables 1–10, Supplementary Figures 1–12, Supplementary Notes 1–3, Supplementary Data sets 1 and 2 and Supplementary References [1–13].

## ACKNOWLEDGEMENTS

Prof. P. G. Lagoudakis and J. J. Rindermann from Southampton University and K. S. Cordua from DTU Informatics are gratefully acknowledged for fruitful discussions. Dr J. Fock from University of Copenhagen and A. Dierckx from Chalmers University are thankfully acknowledged for helping with or testing the software. The authors thank P. Sandin from Chalmers for providing valuable feedback on the manuscript.

## FUNDING

Swedish Research Council (VR); Stiftelsen Olle Engkvist Byggmästare; Danish Council for Independent Research | Natural Sciences (FNU). Funding for open access charge: The Swedish Research Council (VR).

*Conflict of interest statement.* None declared.

## REFERENCES

1. Foster,M.P., McElroy,C.A. and Amero,C.D. (2007) Solution NMR of large molecules and assemblies. *Biochemistry*, **46**, 331–340.
2. Holbrook,S.R. (2008) Structural principles from large RNAs. *Annu. Rev. Biophys.*, **37**, 445–464.
3. Clegg,R.M. (1992) Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.*, **211**, 353–388.
4. Selvin,P.R. (2000) The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.*, **7**, 730–734.
5. Dale,R.E. and Eisinger,J. (1974) Intramolecular distances determined by energy transfer. Dependence on orientational freedom of donor and acceptor. *Biopolymers*, **13**, 1573–1605.
6. Sapsford,K.E., Berti,L. and Medintz,I.L. (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor–acceptor combinations. *Angew. Chem. Int. Ed. Engl.*, **45**, 4562–4589.
7. Iqbal,A., Wang,L., Thompson,K.C., Lilley,D.M. and Norman,D.G. (2008) The structure of cyanine 5 terminally attached to double-stranded DNA: implications for FRET studies. *Biochemistry*, **47**, 7857–7862.
8. Neubauer,H., Gaiko,N., Berger,S., Schaffer,J., Eggeling,C., Tuma,J., Verdier,L., Seidel,C.A., Griesinger,C. and Volkmer,A. (2007) Orientational and dynamical heterogeneity of rhodamine 6G terminally attached to a DNA helix revealed by NMR and single-molecule fluorescence spectroscopy. *J. Am. Chem. Soc.*, **129**, 12746–12755.
9. Norman,D.G., Grainger,R.J., Uhrin,D. and Lilley,D.M. (2000) Location of cyanine-3 on double-stranded DNA: importance for fluorescence resonance energy transfer studies. *Biochemistry*, **39**, 6317–6324.
10. Rindermann,J.J., Akhtman,Y., Richardson,J., Brown,T. and Lagoudakis,P.G. (2011) Gauging the flexibility of fluorescent markers for the interpretation of fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, **133**, 279–285.
11. Sindbert,S., Kalinin,S., Nguyen,H., Kienzler,A., Clima,L., Bannwarth,W., Appel,B., Muller,S. and Seidel,C.A. (2011) 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.
12. Wozniak,A.K., Schroder,G.F., Grubmuller,H., Seidel,C.A. and Oesterhelt,F. (2008) Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl Acad. Sci. USA*, **105**, 18337–18342.
13. Muschielok,A., Andrecka,J., Jawhari,A., Bruckner,F., Cramer,P. and Michaelis,J. (2008) A nano-positioning system for macromolecular structural analysis. *Nat. Methods*, **5**, 965–971.
14. Börjesson,K., Preus,S., El-Sagheer,A.H., Brown,T., Albinsson,B. and Wilhelmsson,L.M. (2009) Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems. *J. Am. Chem. Soc.*, **131**, 4288–4293.
15. Sandin,P., Börjesson,K., Li,H., Mårtensson,J., Brown,T., Wilhelmsson,L.M. and Albinsson,B. (2008) Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res.*, **36**, 157–167.
16. Wilhelmsson,L.M. (2010) Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.*, **43**, 159–183.
17. Sinkel�,R.W., Greco,N.J. and Tor,Y. (2010) Fluorescent analogs of biomolecular building blocks: design, properties, and applications. *Chem. Rev.*, **110**, 2579–2619.
18. Srivatsan,S.G. and Sawant,A.A. (2011) Fluorescent ribonucleoside analogues as probes for investigating RNA structure and function. *Pure Appl. Chem.*, **83**, 213–232.
19. Iqbal,A., Arslan,S., Okumus,B., Wilson,T.J., Giraud,G., Norman,D.G., Ha,T. and Lilley,D.M. (2008) Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl Acad. Sci. USA*, **105**, 11176–11181.
20. Lewis,F.D., Zhang,L. and Zuo,X. (2005) Orientation control of fluorescence resonance energy transfer using DNA as a helical scaffold. *J. Am. Chem. Soc.*, **127**, 10002–10003.
21. Ranjit,S., Gurunathan,K. and Levitus,M. (2009) Photophysics of backbone fluorescent DNA modifications: reducing uncertainties in FRET. *J. Phys. Chem. B*, **113**, 7861–7866.
22. Hall,L.M., Gerowska,M. and Brown,T. (2012) A highly fluorescent DNA toolkit: synthesis and properties of oligonucleotides containing new Cy3, Cy5 and Cy3B monomers. *Nucleic Acids Res.*, **40**, e108.
23. Bandy,T.J., Brewer,A., Burns,J.R., Marth,G., Nguyen,T. and Stulz,E. (2011) DNA as supramolecular scaffold for functional molecules: progress in DNA nanotechnology. *Chem. Soc. Rev.*, **40**, 138–148.
24. Wilson,J.N. and Kool,E.T. (2006) Fluorescent DNA base replacements: reporters and sensors for biological systems. *Org. Biomol. Chem.*, **4**, 4265–4274.
25. Varghese,R. and Wagenknecht,H.A. (2009) DNA as a supramolecular framework for the helical arrangements of chromophores: towards photoactive DNA-based nanomaterials. *Chem. Commun.*, 2615–2624.
26. El Hassan,M.A. and Calladine,C.R. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.*, **251**, 648–664.

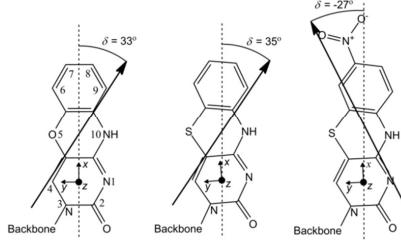
27. Lu,X.J., El Hassan,M.A. and Hunter,C.A. (1997) Structure and conformation of helical nucleic acids: rebuilding program (SCHNArP). *J. Mol. Biol.*, **273**, 681–691.
28. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
29. Olson,W.K., Bansal,M., Burley,S.K., Dickerson,R.E., Gerstein,M., Harvey,S.C., Heinemann,U., Lu,X.J., Neidle,S., Shakhed,Z. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.
30. Preus,S., Börjesson,K., Kilså,K., Albinsson,B. and Wilhelmsson,L.M. (2010) Characterization of nucleobase analogue FRET acceptor tCnitro. *J. Phys. Chem. B*, **114**, 1050–1056.
31. Wilhelmsson,L.M., Sandin,P., Holmén,A., Albinsson,B., Lincoln,P. and Nordén,B. (2003) Photophysical characterization of fluorescent DNA base analogue, tC. *J. Phys. Chem. B*, **107**, 9094–9101.
32. Lakowicz,J.R. (2006) *Principles of Fluorescence Spectroscopy*, 3rd edn. Springer, New York.
33. Freier,S.M., Sugimoto,N., Sinclair,A., Alkema,D., Neilson,T., Kierzek,R., Caruthers,M.H. and Turner,D.H. (1986) Stability of XGCGCp, GCGCyp, and XGCGCYp helices: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry*, **25**, 3214–3219.
34. Turner,D.H., Sugimoto,N., Kierzek,R. and Dreiker,S.D. (1987) Free-energy increments for hydrogen-bonds in nucleic-acid base-pairs. *J. Am. Chem. Soc.*, **109**, 3783–3785.
35. Zheng,G., Lu,X.J. and Olson,W.K. (2009) Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, **37**, W240–W246.
36. Juo,Z.S., Chiu,T.K., Leiberman,P.M., Baikalov,I., Berk,A.J. and Dickerson,R.E. (1996) How proteins recognize the TATA box. *J. Mol. Biol.*, **261**, 239–254.
37. Clegg,R.M., Murchie,A.I., Zechel,A. and Lilley,D.M. (1993) Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA*, **90**, 2994–2998.
38. Preus,S., Kilså,K., Wilhelmsson,L.M. and Albinsson,B. (2010) Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family. *Phys. Chem. Chem. Phys.*, **12**, 8881–8892.
39. Gohlke,C., Murchie,A.I., Lilley,D.M. and Clegg,R.M. (1994) Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA*, **91**, 11660–11664.
40. Zacharias,M. and Hagerman,P.J. (1995) Bulge-induced bends in RNA: quantification by transient electric birefringence. *J. Mol. Biol.*, **247**, 486–500.
41. Ouellet,J., Schorr,S., Iqbal,A., Wilson,T.J. and Lilley,D.M. (2011) Orientation of cyanine fluorophores terminally attached to DNA via long, flexible tethers. *Biophys. J.*, **101**, 1148–1154.
42. Urnavicius,L., McPhee,S.A., Lilley,D.M. and Norman,D.G. (2012) The structure of sulfoindocarbocyanine 3 terminally attached to dsDNA via a long, flexible tether. *Biophys. J.*, **102**, 561–568.

# FRETmatrix: A general methodology for the simulation and analysis of FRET in nucleic acids

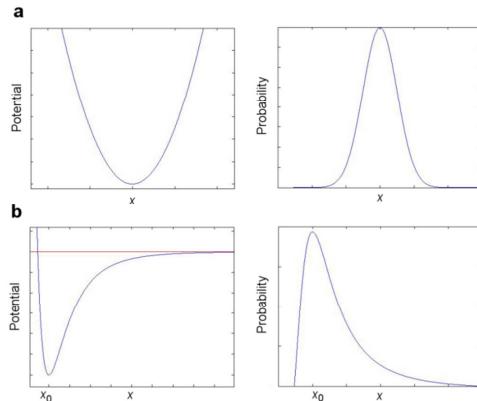
Søren Preus, Kristine Kilså, Francois-Alexandre Miannay, Bo Albinsson, and L. Marcus Wilhelmsson

<b>Supplementary Figure 1</b>	The position and direction of the transition dipole moments of tC, tC <sup>O</sup> and tC <sub>nitro</sub> in their base reference frames.
<b>Supplementary Figure 2</b>	Modelled potentials and corresponding Boltzmann distributions.
<b>Supplementary Figure 3</b>	Illustration of the pole pile up effect.
<b>Supplementary Figure 4</b>	Sampled vector distributions characterized by different widths.
<b>Supplementary Figure 5</b>	Effect of $\alpha$ on the simulated decay.
<b>Supplementary Figure 6</b>	Standard deviation of the reduced global chi-square as a function of the number of dipoles drawn in the simulation.
<b>Supplementary Figure 7</b>	Examples of other dyes but the tC bases implemented in FRETmatrix
<b>Supplementary Figure 8</b>	Optimized 3A structure overlay.
<b>Supplementary Figure 9</b>	Global fits of demonstration study 2 (0A bulge).
<b>Supplementary Figure 10</b>	Global fits of demonstration study 2 (3A bulge).
<b>Supplementary Figure 11</b>	Full $\alpha, \gamma$ chi-square surface of 3A bulge model system.
<b>Supplementary Figure 12</b>	Time-resolved fluorescence anisotropy of tC <sup>O</sup> in double-stranded DNA.
<b>Supplementary Table 1</b>	Atomic coordinates and transition dipole coordinates of tC, tC <sup>O</sup> and tC <sub>nitro</sub> in the nucleobase coordinate frames.
<b>Supplementary Table 2</b>	Atomic coordinates of tC, tC <sup>O</sup> and tC <sub>nitro</sub> in the base-pair coordinate frame.
<b>Supplementary Table 3</b>	DNA sequences in demonstration study 1. Base-pair step parameters of regular A-form and B-form helices used by FRETmatrix.
<b>Supplementary Table 4</b>	Parameter values used in demonstration study 1.
<b>Supplementary Table 5</b>	DNA sequences in demonstration study 2.
<b>Supplementary Table 6</b>	Double stranded donor reference decays fit of demonstration study 2.
<b>Supplementary Table 7</b>	Donor decays in demonstration study 2 fit in presence of FRET.
<b>Supplementary Table 8</b>	Parameter values used in demonstration study 2.
<b>Supplementary Table 9</b>	Fitted kink parameter values of the 3A bulge.
<b>Supplementary Table 10</b>	Base-pair step parameters of regular A-form and B-form helices.
<b>Supplementary Note 1</b>	Simulating FRET using three dimensional dipole vector representations.
<b>Supplementary Note 2</b>	Flow chart for simulating dipole dynamics.
<b>Supplementary Note 3</b>	FRETmatrix v1.0 user guide.
<b>Supplementary Data 1</b>	Base-pair step parameter matrix of model DNA-TBP complex.
<b>Supplementary Data 2</b>	Simulated FRET characteristics of model systems.

## SUPPLEMENTARY FIGURES



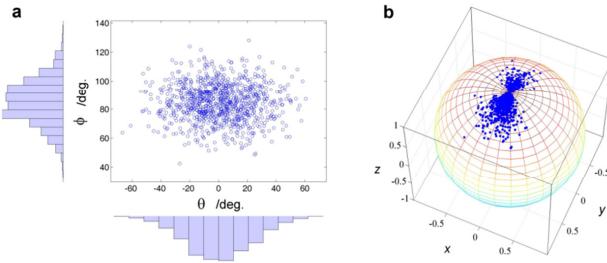
**Supplementary Figure 1.** The position and direction of the transition dipole moments of tC, tC<sup>O</sup> and tC<sub>nitro</sub> and their base reference frames being centred in the bottom ring system for the base positioned in strand I. The corresponding atomic coordinates and transition dipole coordinates are shown in Supplementary Table 1. The y-axis of this frame is pointing towards the C4 atom, the x-axis is pointing in the direction of the outer ring and the z-axis is pointing perpendicularly to the base plane. For the base in strand II the coordinate frame is rotated 180° about the positive x-axis. Since the sulphur atom perturbs the symmetry of the tricyclic frameworks of tC and tC<sub>nitro</sub>, a small angular correction was made for the direction of the molecular long axis. The angle between the molecular long axis (dashed line) and the base reference frame x-axes are 0°, 5°, and 3° for tC<sup>O</sup>, tC, and tC<sub>nitro</sub>, respectively. The dipole centers were set to be in the center of the mid ring of the tricyclic framework. Values of  $\delta$  have been estimated experimentally for tC (1), tC<sup>O</sup> (2), and tC<sub>nitro</sub> (3).



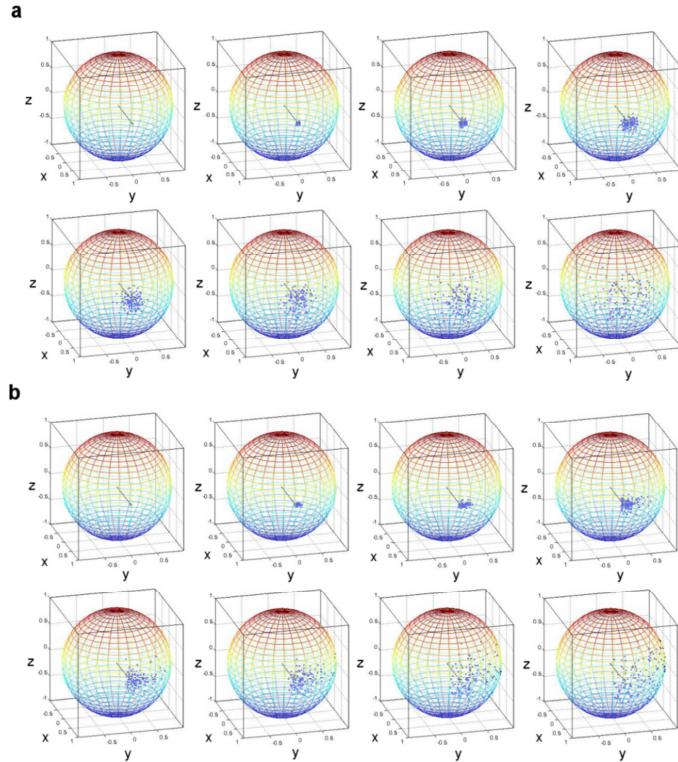
**Supplementary Figure 2.** Modeled potentials (left) and corresponding Boltzmann distributions (right). (a) Harmonic potential. (b) The Lennard Jones potential we use for the in-plane movement is given by

with a dissociation energy of  $D = 0.03$  eV and a temperature of  $T = 296$  K.  $V$  has a minimum at  $V(x_{\min})$ . The Boltzmann probability distribution of a species in the energy potential  $V$  is then

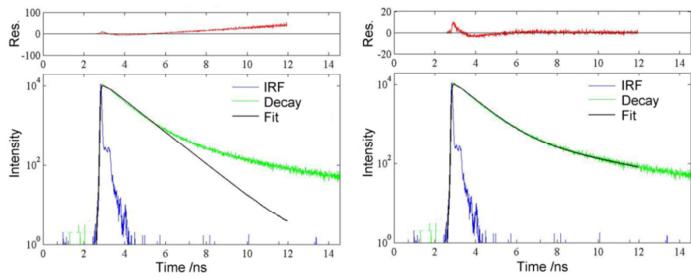
where  $A$  is a normalization factor  $k_B$  is the Boltzmann constant and  $T$  is the temperature. Only points below the defined red threshold line are used when constructing the probability distribution.



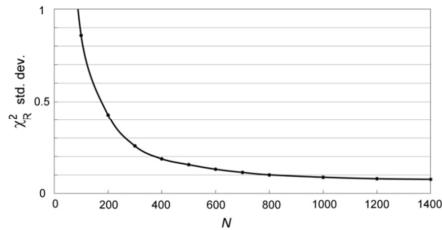
**Supplementary Figure 3.** When transforming sampled spherical coordinates into Cartesian coordinates, samples drawn near the poles of the unit sphere (i.e. near  $\phi = \pm 90^\circ$ ) will pile up in Cartesian space compared to points sampled near the equator. (A) Scatter and histograms of samples drawn from a bivariate Gaussian distribution centered at  $\theta = 0^\circ$  and  $\phi = 85^\circ$ . (B) Cartesian coordinates of the same samples as in (A) showing the pole pile up effect. The piling up of samples near the poles arise because of the smaller distance covered by a walk  $\theta = 360^\circ$  around the sphere when walking near the pole ( $\phi = 90^\circ$ ) compared to near the equator ( $\phi = 0^\circ$ ).



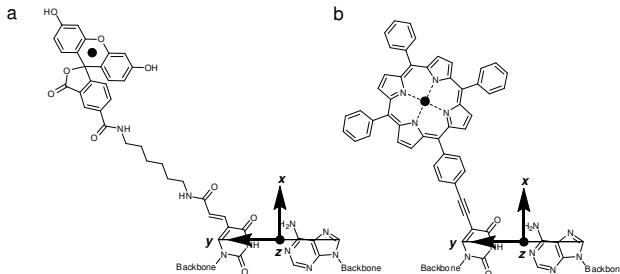
**Supplementary Figure 4.** Sampled vector distributions in Cartesian space characterized by different distribution widths. (a) Using a harmonic potential to describe both the  $\phi$  and the  $\theta$  coordinate which corresponds to sampling a bivariate Gaussian distribution. The same FWHM is used in both directions. Top row from left to right: FWHM =  $0.1^\circ$ ,  $2^\circ$ ,  $5^\circ$ ,  $10^\circ$ . Bottom from left to right: FWHM =  $15^\circ$ ,  $20^\circ$ ,  $30^\circ$ ,  $40^\circ$ . (b) Using a harmonic potential for the  $\phi$  direction and a Lennard Jones potential for the  $\theta$  direction. The FWHM for the two coordinates in each subfigure is the same as in (a).



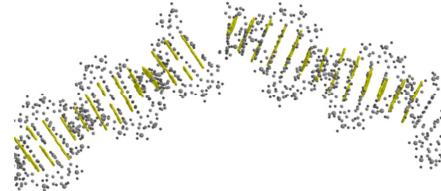
**Supplementary Figure 5.** Effect of  $\alpha$  on the simulated decay. Left:  $\alpha = 0.0000$ . Right:  $\alpha = 0.0063$  corresponding to 0.63% of donor fluorophores not coupled to an acceptor. Blue is the instrument response function (IRF), green is the measured decay, and black is the simulated decay convolved with the IRF. The example is a sample with a 5 base-pair separation between donor and acceptor. In both cases  $b = 0$ .



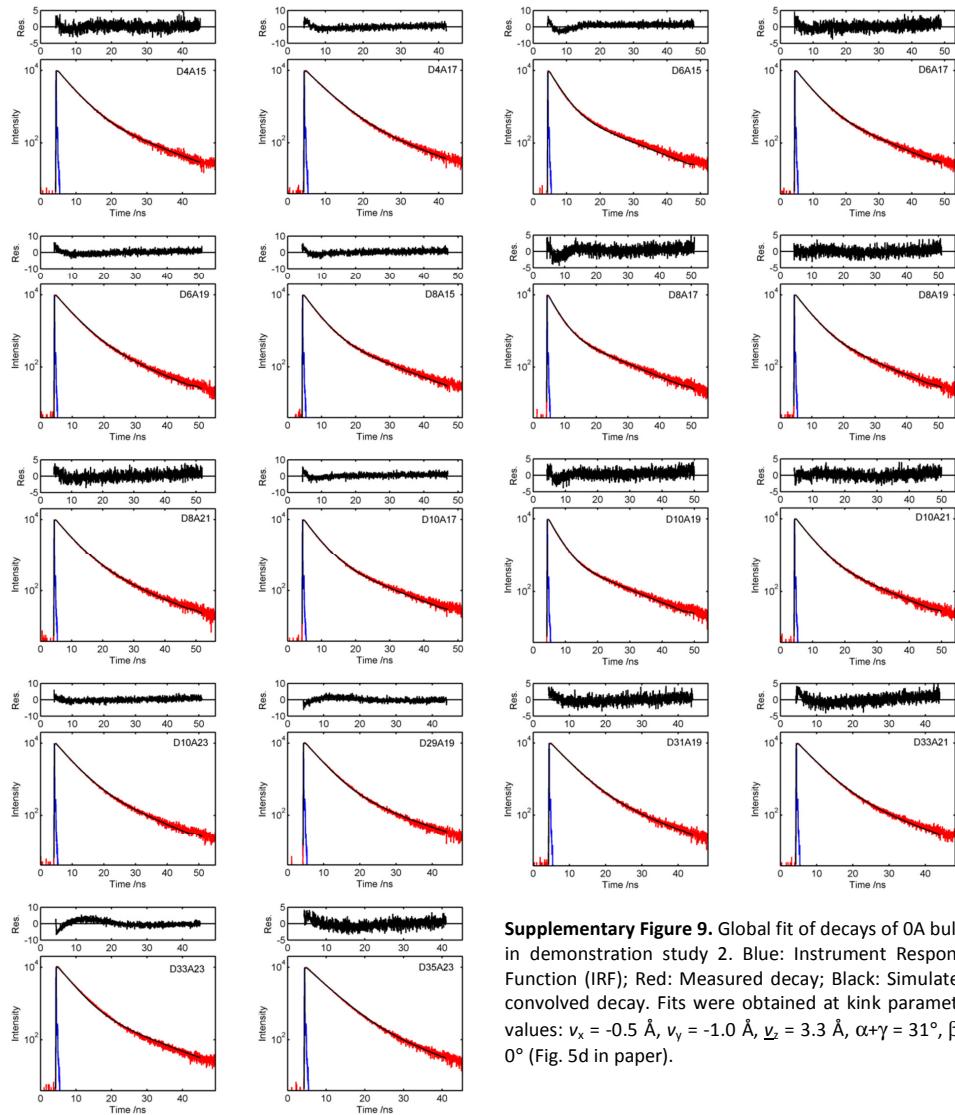
**Supplementary Figure 6.** Standard deviation of the reduced global chi-square as a function of the number of dipoles,  $N$ , drawn in the simulation. Since the object function is based on random sampling of the directional dipole probability distribution the accuracy of the simulation depends on the number of dipoles,  $N$ , drawn. The calculated  $\chi^2$  values are based on >100 repeated simulations of the data from demonstration study 1 when sampling from a distribution with  $\theta\text{FWHM} = 12^\circ$  and  $\phi\text{FWHM} = 4^\circ$ .



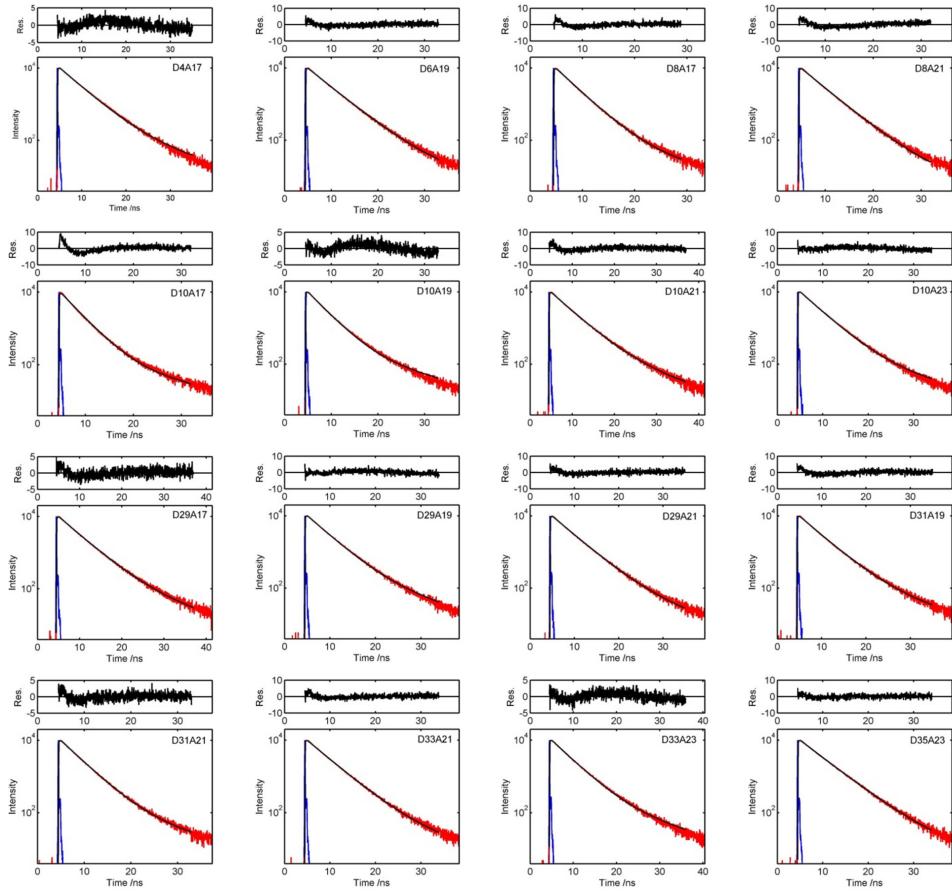
**Supplementary Figure 7.** Examples of other dyes but the tC bases implemented in FRETmatrix (in the figure dyes are positioned in strand I). A) Fluorescein/rhodamine-based dyes tethered to thymine by a C6 linker. In FRETmatrix the dyes are positioned in their mean positions relative to the B-DNA helix as determined for the structurally similar Alexa 488 fluorophore by Sindbert *et al.* using MD simulations.(4) B) A Zn/2H porphyrin dye rigidly tethered to thymine through an acetylenic linker. The origin of the coordinate frame is the mid-point of the line connecting the C8 of adenine and C6 for thymine. The y-axis is parallel to the C8-C6 line pointing from strand II to strand I. The x-axis points into the major groove while the z-axis completes a right-handed set. The dipole centers are marked as black dots and are  $(x,y,z) = (0.86, 12.69, 5.06)$  and  $(x,y,z) = (11.94, 8.37, 0.0)$ , respectively (in Å).



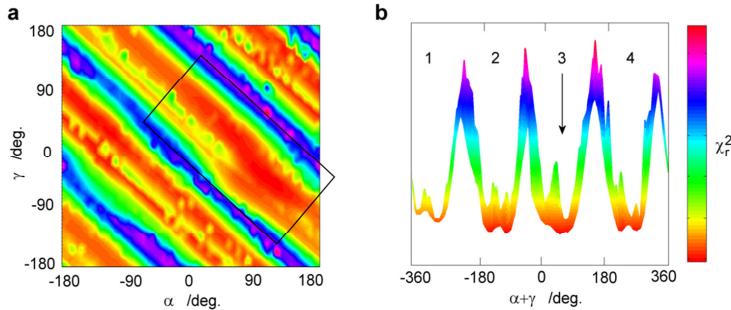
**Supplementary Figure 8.** Optimized 3A bulge structure (yellow) overlayed on the atomic framework of a 3A bulge structure obtained by Wozniak *et al.* using Multiparameter Fluorescence Detection Single-Molecule FRET (grey) (5).



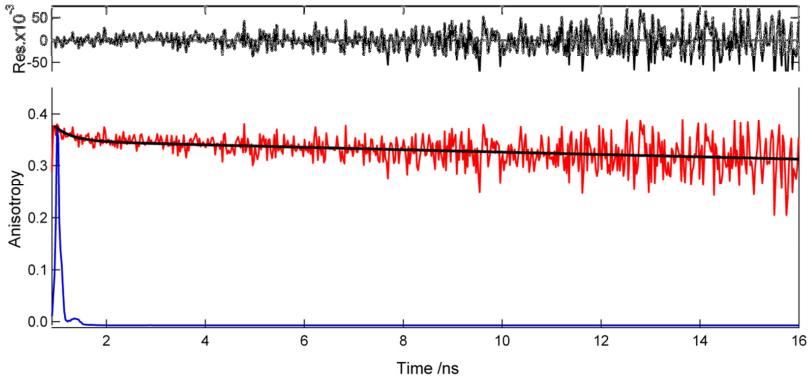
**Supplementary Figure 9.** Global fit of decays of OA bulge in demonstration study 2. Blue: Instrument Response Function (IRF); Red: Measured decay; Black: Simulated, convolved decay. Fits were obtained at kink parameter values:  $v_x = -0.5 \text{ \AA}$ ,  $v_y = -1.0 \text{ \AA}$ ,  $v_z = 3.3 \text{ \AA}$ ,  $\alpha + \gamma = 31^\circ$ ,  $\beta = 0^\circ$  (Fig. 5d in paper).



**Supplementary Figure 10.** Global fit of 3A bulge decays in demonstration study 2. Blue: Instrument Response Function (IRF); Red: Measured decay; Black: Simulated, convolved decay. Fits were obtained with kink parameter values:  $v_x = 7.4 \text{ \AA}$ ,  $v_y = -0.5 \text{ \AA}$ ,  $v_z = 5.5 \text{ \AA}$ ,  $\alpha = 101^\circ$ ,  $\beta = 57^\circ$ ,  $\gamma = -43^\circ$ .



**Supplementary Figure 11.** (a) Full  $\alpha, \gamma$  chi-square surface of 3A bulge model system. (b) Sideview of surface in (a) looking from lower right corner to upper left corner. The chi-square oscillates with a period of  $\alpha + \gamma = 180^\circ$  as is also observed for the OA bulge model system (Fig. 5d in paper). These oscillations are caused by the symmetric properties of the two DNA helices separated by the kink. We interpret trough 3 ( $\alpha + \gamma \approx 55^\circ$ ) as the true global minimum based on the assumption that the DNA bending follows the right handed helical twist of the two helices. Constructing all other kink parameter chi-square surfaces (Fig. 5e in paper) were done with the values of  $\alpha$  and  $\gamma$  optimized within the intervals marked by the black rectangle in (a). Color bar: 1.85:13 for both (a) and (b).



**Supplementary Figure 12.** Time-resolved fluorescence anisotropy of  $tC^0$  in double-stranded DNA in high viscosity phosphate buffer solution (55% sucrose,  $0.1\text{ M Na}^+$ ,  $T = 296\text{ K}$ ). In high viscosity solution, the rotational correlation times affecting the anisotropy decay in the measurable timescales are those due to DNA torsional motions with a long rotational correlation of  $\tau_2$ , and the fast internal fluorophore motions with a rotational correlation time of  $\tau_f$  while all other rotational motions can be assumed as infinite. In this regime the decay is fitted to

Here,  $r_0$  is the fundamental anisotropy of the probe, 0.375, (2)  $\alpha$  is a parameter in between 0 and 1 representing the internal rotational flexibility of the probe. The measured time-resolved anisotropy decay (red) was fitted using  $\alpha$ ,  $\tau_f$  and  $\tau_2$  as fitting parameters:

Sample	$r_0$	$\alpha$	$\tau_f$ /ns	$\tau_2$ /ns	$R^2$
180bp, high visc.	0.375	0.075	0.35	143	1.39

## SUPPLEMENTARY TABLES

**Supplementary Table 1.** Atomic coordinates of tC, tC<sup>O</sup> and tC<sub>nitro</sub> in the nucleobase coordinate frames defined in Supplementary Figure 1 (the bases positioned in strand I). “Dipole center” and “dipole end” are used to define the position and orientation of the transition dipole vectors. The dipole end is the end point of the unit vector going from the dipole center and tilted an angle  $\delta$  relative to the molecular long axis and is calculated as:  $(x,y,z) = (\text{Center}(x,y,z) + (\cos(\delta \cdot L), -\sin(\delta \cdot L), 0))$ , where  $L = 0^\circ$ ,  $L = 5^\circ$  and  $L = 3^\circ$  for tC<sup>O</sup>, tC and tC<sub>nitro</sub>, respectively (Supplementary Figure 1). Atomic coordinates are based on the DFT B3LYP/6-31G\*\* optimized geometries. The ground-state structures of tC and tC<sub>nitro</sub> are degenerate, corresponding to two geometries folded along the middle sulphur/oxygen-nitrogen axis(6), however, only the coordinates of the average, planar geometry of tC and tC<sub>nitro</sub> are shown. Because of the flexibility of the tricyclic frameworks, the out-of-plane bending was set as a variable in the analysis as described in the paper.

Atom	tC <sup>O</sup>			tC			tC <sub>nitro</sub>		
	x	y	z	x	y	z	x	y	z
N1	0.000	-1.387	0.000	0.000	-1.371	0.000	0.000	-1.360	0.000
C2	-1.206	-0.726	0.000	-1.202	-0.736	0.000	-1.220	-0.715	0.000
N3	-1.179	0.697	0.000	-1.181	0.691	0.000	-1.186	0.711	0.000
C4	0.000	1.387	0.000	0.000	1.383	0.000	0.000	1.395	0.000
O5/S5	2.371	1.438	0.000	2.716	1.683	0.000	2.718	1.671	0.000
C6	4.737	1.416	0.000	5.283	0.774	0.000	5.266	0.736	0.000
C7	5.947	0.729	0.000	6.303	-0.175	0.000	6.258	-0.241	0.000
C8	5.956	-0.659	0.000	5.980	-1.531	0.000	5.946	-1.598	0.000
C9	4.758	-1.363	0.000	4.646	-1.927	0.000	4.610	-1.970	0.000
N10	2.316	-1.357	0.000	2.289	-1.429	0.000	2.268	-1.440	0.000
C	1.128	-0.690	0.000	1.138	-0.687	0.000	1.118	-0.682	0.000
C	1.180	0.737	0.000	1.196	0.747	0.000	1.190	0.747	0.000
C	3.545	0.713	0.000	3.940	0.387	0.000	3.926	0.362	0.000
C	3.543	-0.686	0.000	3.613	-0.980	0.000	3.586	-1.007	0.000
O	-2.272	-1.314	0.000	-2.283	-1.305	0.000	-2.302	-1.278	0.000
N							7.661	0.173	0.000
O							7.901	1.382	0.000
O							8.518	-0.710	0.000
Dipole:									
Center	2.360	0.000	0.000	2.450	-0.176	0.000	2.450	-0.176	0.000
End <sup>†</sup>	3.199	-0.545	0.000	3.216	-0.819	0.000	3.364	0.231	0.000

**Supplementary Table 2.** Atomic coordinates of tC,  $tC^0$  and  $tC_{\text{nitro}}$  in the base-pair coordinate frame defined in **Fig. 1b left** of the paper. Dipole center and dipole end define the position and orientation of the transition dipoles. All base-pair coordinates in the base-pair reference frames were constructed from the individual base coordinate frames using the base-pair parameter values corresponding to flat base-pairs(7):

propeller =  $0^\circ$ , opening =  $0^\circ$ , buckle =  $0^\circ$ , stagger =  $0 \text{ \AA}$ , stretch =  $5.45 \text{ \AA}$ , shear =  $0 \text{ \AA}$

The above definitions are all for the base analogues positioned in strand I and the complementary base in strand II as defined in the CEHS scheme.(8) Base analogues positioned in strand II have identical coordinates as for strand I except all y values are multiplied by -1.

Atom	$tC^0$			tC			$tC_{\text{nitro}}$		
	x	y	z	x	y	z	x	y	z
N1	-0.465	2.154	0.000	-0.465	2.152	0.000	-0.465	2.157	0.000
C2	-1.545	2.990	0.000	-1.542	2.981	0.000	-1.558	2.999	0.000
N3	-1.279	4.388	0.000	-1.281	4.384	0.000	-1.284	4.399	0.000
C4	0.000	4.869	0.000	0.000	4.867	0.000	0.000	4.873	0.000
O5/S5	2.346	4.519	0.000	2.728	4.704	0.000	2.726	4.687	0.000
C6	4.674	4.098	0.000	5.104	3.375	0.000	5.080	3.336	0.000
C7	5.751	3.217	0.000	5.950	2.267	0.000	5.893	2.205	0.000
C8	5.525	1.847	0.000	5.402	0.985	0.000	5.356	0.920	0.000
C9	4.226	1.356	0.000	4.021	0.820	0.000	3.977	0.779	0.000
N10	1.820	1.774	0.000	1.782	1.709	0.000	1.758	1.696	0.000
C	0.761	2.631	0.000	0.772	2.635	0.000	0.752	2.637	0.000
C	1.053	4.029	0.000	1.071	4.038	0.000	1.064	4.034	0.000
C	3.380	3.606	0.000	3.715	3.220	0.000	3.696	3.193	0.000
C	3.142	2.228	0.000	3.162	1.928	0.000	3.130	1.901	0.000
O	-2.695	2.590	0.000	-2.704	2.603	0.000	-2.720	2.626	0.000
N							7.345	2.377	0.000
O							7.786	3.528	0.000
O							8.041	1.362	0.000
Dipole:									
Center	2.092	3.104	0.000	2.150	2.911	0.000	2.150	2.911	0.000
End	2.827	2.425	0.000	2.798	2.154	0.000	3.119	3.158	0.000

**Supplementary Table 3.** DNA sequences in demonstration study 1. X =  $tC^0$ . Y =  $tC_{\text{nitro}}$ .

Name	Sequence
D7	5'-CGA TCA XAC ACA AGG ACG AGG ATA AGG AGG AGG
D9	5'-CGA TCA CAX ACA AGG ACG AGG ATA AGG AGG AGG
D11	5'-CGA TCA CAC AXA AGG ACG AGG ATA AGG AGG AGG
A14	5'-CCT CCT CCT TAT CCT CGT CYT TGT GTG TGA TCG
A15	5'-CCT CCT CCT TAT CCT CGT YCT TGT GTG TGA TCG
A20	5'-CCT CCT CCT TAT CYT CGT CCT TGT GTG TGA TCG
A21	5'-CCT CCT CCT TAT YCT CGT CCT TGT GTG TGA TCG

**Supplementary Table 4.** Parameter values in demonstration study 1.

Parameter	Value
$J / M^{-1}cm^{-1}nm^4$	1.3e14
$\Phi_D$	0.23 (reference (2))
$\tau_D /ns$	4.32, 4.35, 4.55
$a$	0.992–1
$b$	0
$\eta$	1.4

**Supplementary Table 5.** DNA sequences in demonstration study 2.

Name	Donor sequences ( $X = tC^0$ ): 0A-bulge
D4	5'-CCA XAC ACA CAC GTG AGA GAG AGA CGT ACA CAC ACT CC
D6	5'-CCA CAX ACA CAC GTG AGA GAG AGA CGT ACA CAC ACT CC
D8	5'-CCA CAC AXA CAC GTG AGA GAG AGA CGT ACA CAC ACT CC
D10	5'-CCA CAC ACA XAC GTG AGA GAG AGA CGT ACA CAC ACT CC
D29	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGT AXA CAC ACT CC
D31	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGT ACA XAC ACT CC
D33	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGT ACA CAX ACT CC
D35	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGT ACA CAC AXT CC
	Donor sequences ( $X = tC^0$ ): 3A-bulge
D4	5'-CCA XAC ACA CAC GAAATG AGA GAG AGA CGT ACA CAC ACT CC
D6	5'-CCA CAX ACA CAC GAAATG AGA GAG AGA CGT ACA CAC ACT CC
D8	5'-CCA CAC AXA CAC GAAATG AGA GAG AGA CGT ACA CAC ACT CC
D10	5'-CCA CAC ACA XAC GAAATG AGA GAG AGA CGT ACA CAC ACT CC
D29	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGAAAT AXA CAC ACT CC
D31	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGAAAT ACA XAC ACT CC
D33	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGAAAT ACA CAX ACT CC
D35	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGAAAT ACA CAC AXT CC
	Non-modified strands
	5'-CCA CAC ACA CAC GTG AGA GAG AGA CGT ACA CAC ACT CC
	5'-GGA GTG TGT GTA CGT CTC TCT ACG TGT GTG TGT GG
	Acceptor sequences ( $Y = tC_{nitro}$ )
A15	5'-GGA GTG TGT GTA CGT CTC TCT CTY ACG TGT GTG TGT GG
A17	5'-GGA GTG TGT GTA CGT CTC TCT YTC ACG TGT GTG TGT GG
A19	5'-GGA GTG TGT GTA CGT CTC TYT CTC ACG TGT GTG TGT GG
A21	5'-GGA GTG TGT GTA CGT CTY TCT CTC ACG TGT GTG TGT GG
A23	5'-GGA GTG TGT GTA CGT YTC TCT CTC ACG TGT GTG TGT GG

**Supplementary Table 6.** Double stranded donor reference decays of demonstration study 2 fitted to a double-exponential decay:

$$I(t) = I_0(\alpha_1 e^{-t/\tau_1} + \alpha_2 e^{-t/\tau_2})$$

All decays were fitted in one global step with the value of  $\tau_2$  set to be equal for all samples,  $\tau_1$  set to be equal for the same position in the strand (e.g. 0A-D35 and 3A-D35) while  $\alpha_1$  and  $\alpha_2 = 1 - \alpha_1$  were allowed to vary in between samples. The long lifetime of  $\tau_2 = 12.01$  ns is the result of an unidentified impurity present in all samples of demonstration study 2, while  $\tau_1$  is the tC<sup>0</sup> donor reference lifetime.

0A	$\alpha_1$	$\alpha_2$	$\tau_1$	$\tau_2$	$\chi^2$
D4	0.96	0.045	4.21	12.01	1.11
D6	0.95	0.054	4.24	12.01	1.13
D8	0.94	0.056	4.17	12.01	1.04
D10	0.95	0.054	4.16	12.01	1.03
D29	0.96	0.041	4.28	12.01	1.10
D31	0.96	0.036	4.25	12.01	1.03
D33	0.96	0.040	4.32	12.01	1.10
D35	0.97	0.034	4.60	12.01	0.99

3A	$\alpha_1$	$\alpha_2$	$\tau_1$	$\tau_2$	$\chi^2$
D4	0.99	0.013	4.21	12.01	1.15
D6	0.99	0.013	4.24	12.01	1.08
D8	0.99	0.011	4.17	12.01	1.15
D10	0.98	0.016	4.16	12.01	1.16
D29	0.98	0.017	4.28	12.01	1.06
D31	0.99	0.015	4.25	12.01	1.02
D33	0.98	0.017	4.32	12.01	1.10
D35	0.99	0.012	4.60	12.01	1.05

Global	1.08
--------	------

**Supplementary Table 7.** Fit of  $tC^0$  donor decays of demonstration study 2 in presence of FRET using a double-exponential decay. The value of  $\tau_2$  were constrained to 12.01 ns for all samples while  $\tau_1$ ,  $\alpha_1$ , and  $\alpha_2 = 1 - \alpha_1$  were optimized for each sample. The values of  $\alpha_1$  from these fits were used as the parameter  $b$  in the decay model when analyzing the decays during the FRET analysis.

0A	$\alpha_1$	$\alpha_2$	$\tau_1$	$\tau_2$	$\chi^2$
D4A15	0.966	0.034	3.61	12.01	1.15
D4A17	0.962	0.038	3.98	12.01	1.20
D6A15	0.975	0.025	2.53	12.01	1.20
D6A17	0.955	0.045	3.60	12.01	1.11
D6A19	0.950	0.050	3.98	12.01	1.08
D8A15	0.964	0.036	3.12	12.01	1.20
D8A17	0.973	0.027	2.51	12.01	1.29
D8A19	0.954	0.046	3.60	12.01	1.12
D8A21	0.950	0.050	3.98	12.01	1.05
D10A17	0.965	0.035	3.17	12.01	1.13
D10A19	0.976	0.024	2.50	12.01	1.23
D10A21	0.957	0.043	3.61	12.01	1.11
D10A23	0.950	0.050	3.99	12.01	1.10
D29A19	0.967	0.033	3.83	12.01	1.02
D31A19	0.966	0.034	3.93	12.01	1.04
D33A21	0.965	0.035	3.89	12.01	1.13
D33A23	0.967	0.033	3.78	12.01	1.06
D35A23	0.971	0.029	4.17	12.01	1.09
Global					1.13
3A	$\alpha_1$	$\alpha_2$	$\tau_1$	$\tau_2$	$\chi^2$
D4A17	0.99	0.014	4.18	12.01	1.27
D6A19	0.99	0.011	4.02	12.01	1.19
D8A17	0.99	0.007	3.12	12.01	1.31
D8A21	0.99	0.009	3.79	12.01	1.13
D10A17	0.99	0.009	2.90	12.01	2.18
D10A19	0.99	0.013	3.29	12.01	1.74
D10A21	0.98	0.019	4.04	12.01	1.63
D10A23	0.99	0.014	3.88	12.01	1.32
D29A17	0.99	0.018	4.15	12.01	1.14
D29A19	0.99	0.015	3.97	12.01	1.19
D29A21	0.99	0.013	4.05	12.01	1.13
D31A19	0.98	0.017	3.94	12.01	1.38
D31A21	0.99	0.010	3.51	12.01	1.27
D33A21	0.99	0.012	4.01	12.01	1.17
D33A23	0.98	0.016	3.86	12.01	1.57
D35A23	0.99	0.010	4.43	12.01	1.17
Global					1.36

**Supplementary Table 8.** Parameter values in demonstration study 2 for both the OA and 3A structures.

Parameter	Value
$J / M^{-1} \text{cm}^{-1} \text{nm}^4$	1.3e14
$\Phi_D$	0.23 for AA neighbours(2) 0.27 for AT neighbours(2)
$\tau_D / \text{ns}$	$\tau_1$ from Supplementary Table 6
A	0.996
B	$\alpha_1$ from Supplementary Table 7
$\eta$	1.4

**Supplementary Table 9.** Fitted kink parameter values of the 3A bulge in demonstration study 2 with (+) and without (-) geometrical constraints. The minimum without constraints is in violence with constraint #2: maximum distance between neighbouring bases in the same strand.

Parameter	+	-
$v_x$	7.4 Å	8.5 Å
$v_y$	-0.5 Å	-1.8 Å
$v_z$	5.5 Å	5.01 Å
$\alpha$	101°	101°
$\beta$	57°	64°
$\gamma$	-43°	-43°
$\chi^2_r$	1.85	1.83

**Supplementary Table 10.** Base-pair step parameters of regular A-form and B-form helices used by FRETmatrix. The sequence dependent parameters shown in the top are from the calf thymus fiber models of Arnott & co-workers(9,10) as implemented in w3DNA (11). Average A-DNA and B-DNA parameters are from Olson *et al.* (12).

Step (5'-3')	Shift /Å	Slide /Å	Rise /Å	Tilt /deg.	Roll /deg.	Twist /deg.
AA	0	0.45	3.36	0	1.71	35.96
AT	0	0.44	3.35	0	1.71	35.67
AC	0	0.49	3.36	0	1.71	36.83
AG	0	0.42	3.35	0	1.71	34.79
TA	0	0.47	3.36	0	1.71	36.26
TT	0	0.45	3.36	0	1.71	35.96
TC	0	0.50	3.37	0	1.71	37.13
TG	0	0.44	3.35	0	1.71	35.09
CA	0	0.44	3.35	0	1.71	35.10
CT	0	0.42	3.35	0	1.71	34.79
CC	0	0.47	3.36	0	1.71	35.96
CG	0	0.41	3.34	0	1.71	33.92
GA	0	0.50	3.37	0	1.71	37.13
GT	0	0.49	3.36	0	1.71	36.83
GC	0	0.54	3.38	0	1.71	38.00
GG	0	0.47	3.36	0	1.71	35.96
Average:						
B-DNA	-0.02	0.23	3.32	-0.1	0.60	36.00
A-DNA	0	-1.53	3.32	0.1	8.00	31.10

## Supplementary Note 1. Simulating FRET using three dimensional dipole vector representations

When the global coordinate matrix of a structure has been constructed the coordinates defining the transition dipoles of donor and acceptor in the global coordinate frame are extracted. We denote these by

$$\text{Donor center} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \quad \text{Donor end} = \begin{pmatrix} x_{1,\text{end}} \\ y_{1,\text{end}} \\ z_{1,\text{end}} \end{pmatrix}$$

$$\text{Acceptor center} = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}, \quad \text{Acceptor end} = \begin{pmatrix} x_{2,\text{end}} \\ y_{2,\text{end}} \\ z_{2,\text{end}} \end{pmatrix}$$

The normalized dipole unit vectors in the global coordinate frame of the nucleic acid are:

$$\vec{e}_1 = \begin{pmatrix} x_{1,\text{end}} - x_1 \\ y_{1,\text{end}} - y_1 \\ z_{1,\text{end}} - z_1 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} x_{2,\text{end}} - x_2 \\ y_{2,\text{end}} - y_2 \\ z_{2,\text{end}} - z_2 \end{pmatrix}$$

The FRET efficiency is then calculated as

$$E = \frac{R_0^6}{R_0^6 + R^6}$$

where  $R_0$  is the critical Förster distance and  $R$  is the distance between the dipoles:

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

The critical Förster distance is given by

$$R_0 = 0.211 \left( \frac{\kappa^2 \Phi_D J(\lambda)}{\eta^{-4}} \right)^{\frac{1}{6}}$$

where  $\kappa^2$  is the orientation factor,  $\Phi_D$  is the donor fluorescence quantum yield in absence of acceptor,  $\eta$  is the refractive index,  $J$  is the spectral overlap integral between donor emission and acceptor absorption calculated as

$$J \equiv \int_0^\infty \varepsilon_A(\lambda) \lambda^4 F_D(\lambda) d\lambda$$

Here  $\varepsilon_A$  is the acceptor molar absorptivity spectrum,  $\lambda$  is the wavelength, and  $F_D$  is the normalized donor emission spectrum. The orientation factor is given by (see Fig. 1d in paper)

$$\kappa = \vec{e}_1 \cdot \vec{e}_2 - 3(\vec{e}_1 \cdot \vec{e}_{12})(\vec{e}_{12} \cdot \vec{e}_2)$$

where the unit vector between the two dipole centres is calculated as

$$\vec{e}_{12} = \frac{1}{R} \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix}$$

### Simulating FRET between dipole vector distributions

If  $N$  donor and  $N$  acceptor dipoles are sampled in the simulation, the  $\kappa^2$  is calculated between every pair combination in the two distributions ( $N^2$  pairs in total) and the average  $\kappa^2$  is then evaluated as

$$\langle \kappa^2 \rangle = \frac{\sum_i N^2 \kappa_i^2}{N^2}$$

In the dynamic averaging regime (dipole reorientation much faster than energy transfer rate) the FRET efficiency is calculated as

$$E = \frac{1}{1 + C^{-1} \times \frac{R^6}{\langle \kappa^2 \rangle}}$$

where  $C = 0.211^6 \left( \frac{\Phi_D J(\lambda)}{\eta^{-4}} \right)$ .

## Supplementary Note 2. Flow chart for simulating dipole dynamics

The simulation of base-base FRET including dipole distributions is divided into three steps:

1. Construct atomic coordinate matrix of structure and extract dipole vectors
2. Construct dipole vector distributions
3. Simulate FRET between dipole distributions

Step 1 is described in the paper. Step 2 is divided into five steps:

- 2A. Define a potential energy function
- 2B. Construct Boltzmann probability density function,  $P(x)$ , from potential
- 2C. Sample  $P(\theta)$  and  $P(\phi)$  in spherical coordinates
- 2D. Transform sampled spherical coordinates into Cartesian coordinates
- 2E. Align sampled distribution with dipole vector in local coordinate frame
- 2F. Position vector distribution into the global coordinate frame

Step 3 is further divided into four steps:

- 3A. Calculate  $\langle \kappa^2 \rangle$  between sampled distributions of donor and acceptor
- 3B. Simulate intensity decays (or a steady-state FRET efficiency)
- 3C. Convolve simulated decay with measured IRF
- 3D. Evaluate simulated decays with measured: calculate  $\chi^2$

When constructing chi-square surface the above procedure is repeated for a new set of parameters. All the resulting chi-square values are stored in a matrix for subsequent analysis and chi-square surface plots.

**Supplementary Note 3. FRETmatrix v1.0 user guide**

# FRETmatrix user guide

v1.0, Jan. 2012

Go to <https://sites.google.com/site/fretmatrix/> for updates and corrections.

Contact: [spreus@nano.ku.dk](mailto:spreus@nano.ku.dk) or [marcus.wilhelmsson@chalmers.se](mailto:marcus.wilhelmsson@chalmers.se)

## Requirements:

MATLAB 2009 or newer with the Bioinformatics Toolbox installed.

Some features in the analysis part require the MATLAB Signal Processing Toolbox, the Global Optimization Toolbox and the Statistics Toolbox 2010 or newer.

## Installation

The software does not need to be installed but is run directly from folder. Download the FRETmatrix zip file and unzip it to a local directory.

## Folders

'Calls': Contains the scripts and functions being called by FRETmatrix.

'Inputs': Contains input files for FRETmatrix supplied by the user.

'Outputs': Contains output files produced by FRETmatrix during and after a simulation or analysis.

## Part I: Simulating base-base FRET

## Open the FRETmatrix user interface

- 1) Open MATLAB
- 2) Make the FRETmatrix folder your current directory
- 3) In the command window type: FRETmatrix
- 4) Press enter. This will open the FRETmatrix user interface (Figure 13). The user interface can be used for the simulation of base-base FRET (steady state) in any nucleic acid geometry.

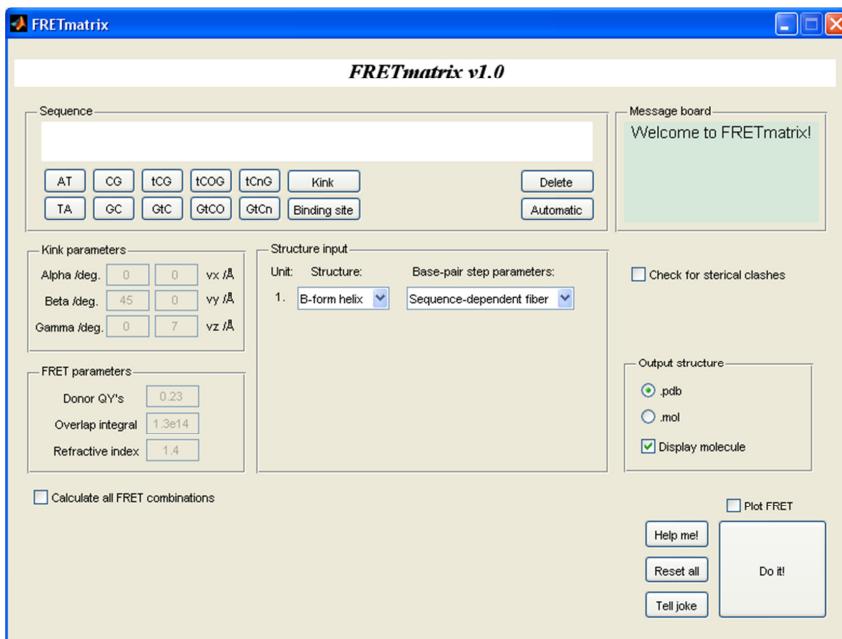


Figure 13. Start window of the FRETmatrix user interface.

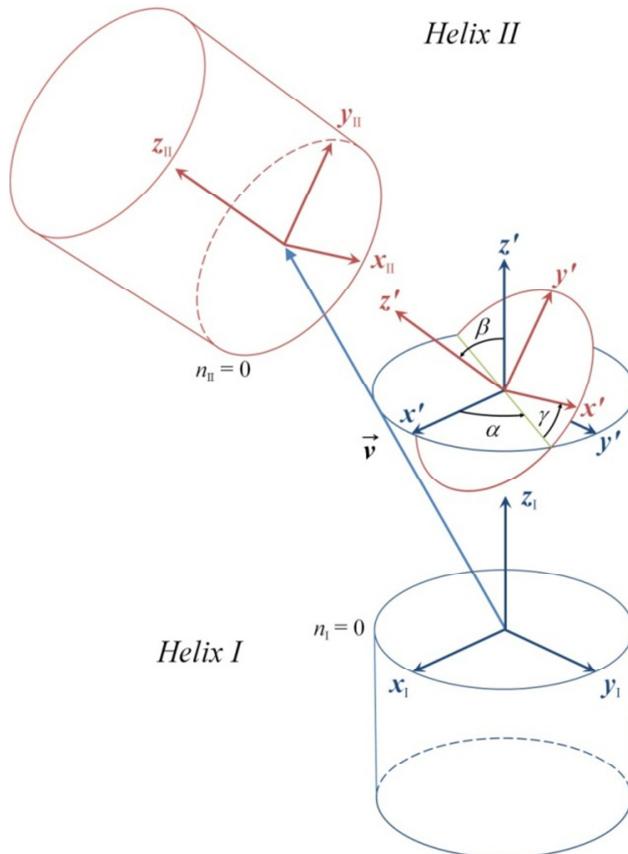
## Panels and buttons

**Sequence panel** Here the sequence of the modeled structure is specified. Type in a sequence by pushing the buttons representing different base-pairs. ‘AT’ inserts an A positioned in strand I and a T and strand II, while ‘TA’ inserts a T positioned in strand I and an A in strand II, etc..

<b>Kink button</b>	Inserts a kink after the last base-pair. A kink separates two structural units. The relative orientation and position of two neighbouring units are specified by the kink parameters (Figure 14).
<b>Binding site button</b>	The ‘Binding site’ button has the exact same function as ‘Kink’ in the program.
<b>Automatic button</b>	Generates an automated sequence of the active unit of the sequence (the last unit). Automatic has different meanings

	<p>depending on the structure being simulated:</p> <ol style="list-style-type: none"> <li>1) For A-form and B-form helices automatic puts in a default 10 bp sequence of pure CG base pairs.</li> <li>2) For structures reconstructed from an input file the sequence is read directly from the file.</li> </ol> <p>It is possible to have the sequence of one structural unit being automatic while other units are manually input.</p>
<b>Delete button</b>	Deletes the last base-pair or kink of the sequence.
<b>Kink parameters panel</b>	Three Euler angles and three translation coordinates defining the relative orientation and position of two structural units. The angles are defined with respect to the coordinate frames of the last base-pair of the first unit and the first base-pair of the second unit (Figure 14).
<b>FRET parameters panel</b>	Specifies the parameters needed to calculate the FRET characteristics between base probes in the structure. Different donor quantum yields can be specified for more than one donor. The overlap integral is in units of $M^{-1}cm^{-1}nm^4$ .
<b>Structure input panel</b>	Here, the structural geometry of each unit in the sequence is specified. A new structural unit appears in this panel when a kink + the next base-pair is entered in the sequence (Figure 15).
<b>Structure</b>	<p>When choosing A-form or B-form helix the base-pair step parameters must be set to “sequence-dependent”, being the base-pair step parameters of Arnott’s fiber models as implemented in w3DNA, or “average”, being the average base-pair step parameters derived by Olson <i>et al.</i><sup>8,9,11</sup></p> <p>Choosing ‘Imported’ is used for reconstructing geometries defined in a pdb file. FRETmatrix requires an input file specifying the structural parameters necessary to rebuild the structure. Three different input files can be used: .par, .out, or .dat, all of which are output files produced by the structural analysis routine of the 3DNA software<sup>12</sup>. The .out file is also achievable using w3DNA<sup>10</sup>. Examples are provided in the inputs folder.</p>
<b>Calculate all FRET combinations checkbox</b>	<p>Checking this box activates the associated panel (Figure 16) and automatically calculates FRET between all donor and acceptor positions in the sequence. This feature is for experiment design purposes. The number of FRET combinations can be limited by introducing constraints in the form of no. of base-pairs separating the donor and acceptor. A constrained site can also be inserted. Clicking this checkbox will only calculate FRET combinations in which the donor and acceptor are positioned on opposite sides of the constrained site. The results are exported to a txt file and an Excel file by pushing the checkbox, ‘Export to Excel’.</p>
<b>Check for Sterical Clashes</b>	This will display a warning if there are any sterical clashes between two bases in the structure. The program does this by calculating the distance between all atoms in the different base-pairs. A sterical clash is registered when two

<b>checkbox</b>	atoms are closer than two times the Van der Waal radius, i.e.: $r < 2 \cdot v_{\text{dw}}$ .
<b>Output structure panel</b>	The program generates a molecule file of the built structure for visual inspection purposes. Choose pdb or mol format as output. The pdb format is optimal for most purposes as there is no upper limit on the total number of atoms in the structure, which is the case for the mol format (>1000 atoms).
<b>Plot FRET checkbox</b>	Checking this box will plot the resulting FRET efficiencies against the donor-acceptor distance.
<b>Help me button</b>	Opens the html file called <code>FRETmatrix_help.html</code> located in the 'Calls' folder.
<b>Reset all button</b>	Restarts the FRETmatrix user interface.



**Figure 14.** Parameters describing the relative position and orientation of two structural units separated by a kink. Alpha, beta, and gamma are three Euler angles describing the relative rotations according to the convention

$R_x(\alpha)R_y(\beta)R_z(\gamma)$ . The translation vector,  $v$ , has three components:  $v_x$ ,  $v_y$ ,  $v_z$  defined with respect to the coordinate system of unit 1 (blue).

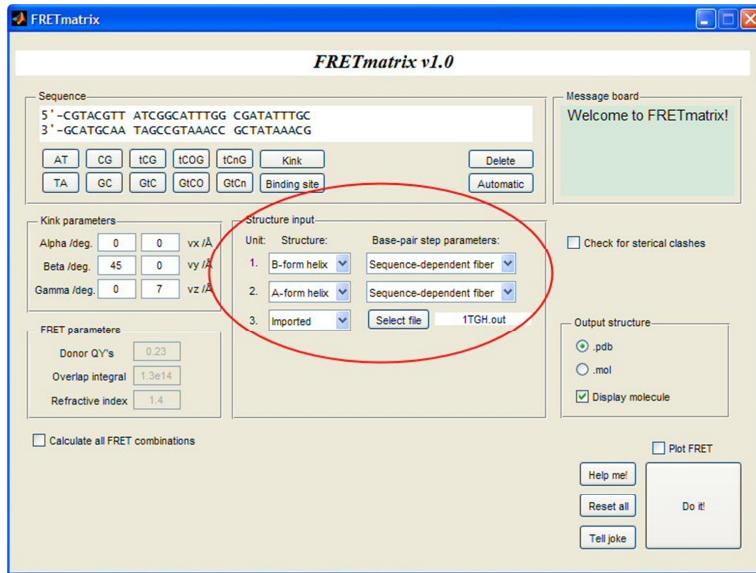


Figure 15. The structure input panel defines what structure is built from the input sequence. Here, two kinks and thus three structural units has been entered in the sequence.

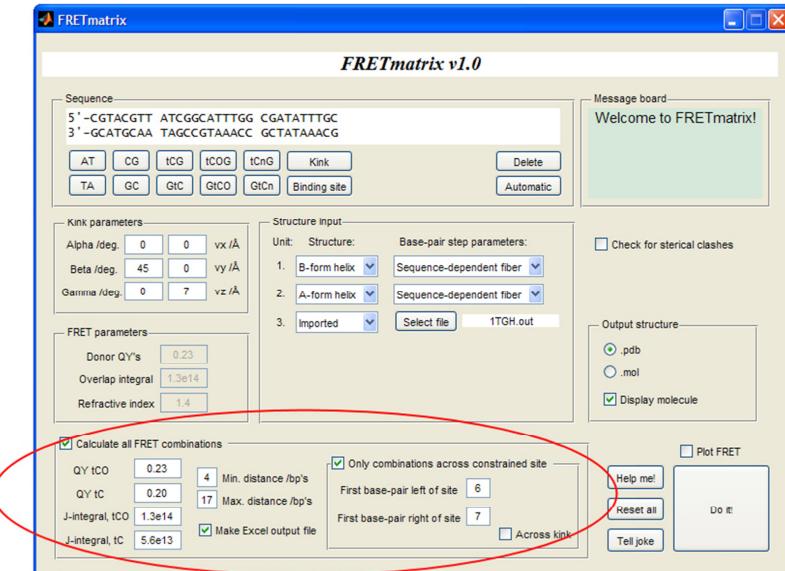


Figure 16. Activating the 'Calculate all FRET combinations' panel inserts FRET pairs at all possible combinations in the sequence and calculate the corresponding FRET characteristics. To narrow down the results list a number of

constraints can be chosen, such as the D-A distance (in bps) and a constrained site. The results are exported to an Excel file located in the outputs folder.

### Quick example 1: Construct two B-DNA helices separated by a kink and simulate FRET between base probes positioned on opposite sides of the kink

- 1) Type in a sequence by pushing the base-pair buttons in sequential order. Insert donor base probes, e.g. 'tCOG', at any positions in the sequence.
- 2) After typing in the sequence of helix 1 (referred to as unit 1) press the 'Kink' button and start typing in the sequence of helix II. An extra structural unit appears in the 'Structure input' panel and the 'Kink parameters' panel becomes activated. In the sequence the two structural units are separated by a space
- 3) Insert acceptor base probes, e.g. 'GtCn', at any positions in the sequence of unit II. This activates the 'FRET parameters' panel.
- 4) In the 'Kink parameters' panel type in the six rigid body parameters describing the relative orientation and position of the two units (Figure 14). Using the default parameters,  $\alpha = 0^\circ$ ,  $\beta = 45^\circ$ ,  $\gamma = 0^\circ$ , will make a  $45^\circ$  bending angle between the two helices (unit 1 and unit 2).
- 5) In the 'FRET parameters' panel type in the parameters describing the FRET pair. See reference values in Sandin *et al.* and Preus *et al.*<sup>2,3,13</sup>.
- 6) Press 'Do it'.

The program will construct a molecule file located in the 'Outputs' folder called mol.pdb. If the 'Display molecule' is checked the produced molecule file will be opened directly by the molviewer in MATLAB (.). FRET will be simulated between all FRET pairs inserted in the sequence with the results being displayed in the MATLAB command window.

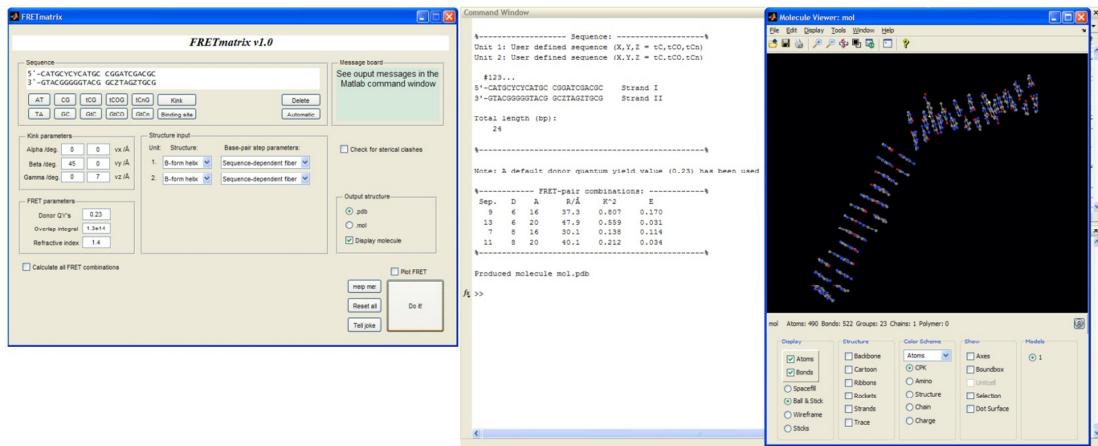


Figure 17. Screenshots from working example 1.

## Quick example 2: Reconstruct structure from pdb file and simulate FRET between all possible positions in the structure.

This example is useful for experiment design purposes.

- 1) Use w3DNA to analyze the pdb structure (<http://w3dna.rutgers.edu/>). Download the resulting output .out file. Examples can be found in the FRETmatrix/inputs folder.
- 2) Select ‘Automatic’ in the ‘Sequence’ panel.
- 3) In the ‘Structure input’ panel choose ‘Imported’ for unit 1. This will activate the ‘Select file’ button next to the unit in the ‘Structure input’ panel.
- 4) Press the ‘Select file’ button and localize the .out file and press ok. The ‘Inputs’ folder of FRETmatrix contains examples that can be used as well.
- 5) Press the ‘Calculate all’ checkbox. This activates the ‘Calculate all’ panel and will automatically combine all possible FRET combinations in the structure (also positions not occupied by a C in the sequence).
- 6) Press ‘Do it’

If the ‘Export to Excel’ is checked an Excel file is generated with the resulting FRET characteristics calculated for all FRET positions in the Structure. The Excel file contains the simulations for both tC and tC<sup>O</sup> but in separate sheets. The file is saved in the FRETmatrix/outputs folder.

Note: For some structures, the output file of w3DNA cannot be used. In this case 3DNA must be used to make either a .par file or a ref\_frames.dat file of the structure. Both of these files can be used as input for FRETmatrix. See the input folder for examples.

The screenshot displays three windows related to FRETmatrix v1.0:

- Main Window:** Shows the 'FRETmatrix v1.0' interface with various input fields for sequences, FRET parameters, and calculation options. A 'Do it!' button is at the bottom right.
- Command Window:** Displays MATLAB command history, including the command to import a sequence from a file named '123.out'. It also shows the total length of the sequence as 12.
- Molecule Viewer:** A 3D visualization of a nucleic acid structure with atoms colored by element (blue for nitrogen, red for oxygen, grey for carbon).
- Lower Insert:** A Microsoft Excel spreadsheet titled 'FRETmatrix\_LTGUDS [Compatibility Mode] - Microsoft Excel'. It contains two sheets: 'NCO' and 'NCO in strand I'. The 'NCO' sheet lists positions for NCO and NCO in strand II across multiple strands (A through Z). The 'NCO in strand I' sheet provides detailed FRET parameters for each position, including donor and acceptor A, E, k<sub>d</sub>, k<sub>a</sub>, and k<sub>T</sub>. The data spans from row 3 to 127.

Figure 18. Screenshots from a working example 2. Lower insert shows output excel file.

## Part II: Analysis of base-base FRET

## Introductory comment

The FRETmatrix analysis program performs a global analysis of the time-resolved donor intensity decays from multiple FRET pairs in a base-base FRET experiment. To allow for flexibility the analysis part of FRETmatrix is not accompanied by a user interface but is run via the `RunProgram.m` file. Preparation of data for the analysis is automated by running the `PrepareDecays.m` script. If large chi-square surfaces will be run the decays can be pre-fit using `MakeStoredFits.m` prior analysis to save computational time.

## 1 Prepare decays for analysis

- 1) Name the decay files according to the positions of the donor (D) and acceptor (A) in the sequence. For example:

D10A19.txt	Decay for donor at position 10 and acceptor at position 19.
D10.txt	Reference decay for donor at position 10.

If in doubt of the numbering, run a single calculation of the sequence using the user interface in Part I above and deduce the D and A positions from the results displayed in the MATLAB command window.

- 2) The decay files must be set up in the following format:

Some lines of text			
t_decay	I_decay	t_IRF	I_IRF
...			
2.7052	14	2.7674	2194
2.7126	25	2.7747	3487
2.7199	27	2.7821	4801
2.7272	33	2.7894	6457
...			

Here, `t_decay` contains the times associated with the decay intensity counts in `I_decay`, and `t_IRF` contains the times associated with the IRF intensity counts in `I_IRF`. Columns are separated by a tab. If the decay files are not automatically set up like this, they must be done so manually, e.g. using Excel. Examples are provided in the `inputs/decays` folder.

- 3) Open `PrepareDecays.m`. This script optimizes the decay parameters of each decay needed for the later analysis. The aim is to automatically determine values of time intervals and shifts of each decay. This is done by fitting the decays to an exponential decay model. The results are exported to a text file which is used by FRETmatrix in the FRET analysis part (below). The `PrepareDecays.m` will look something like this:

```
% Specify files:  
filenames = 'D7A14' 'D7A15' 'D7A20' 'D7A21' 'D9A15';  
extension = '.dat';  
folder = 'C:\FRETmatrix\inputs\decays\B-DNA';  
Studyname = 'B-DNA';  
  
% Settings:  
I_tstart = 10; % Start the time when intensity reaches this count
```

```

I_tend = 20; % End the time when intensity reaches this count
skiplines = 13; % Skip these first lines in each datafile (being
% comments)
plotfit = 'N'; % Plot final fits? 'Y' or 'N'
Save_results = 'Y'; % Saved as .txt in same folder as the decays
L = 10; % Interpolation factor. If L=10, then shift=10 is 1 channel
shift.
shiftrange = [0 150]; % The shift range investigated

% Start guesses:
alpha123 = [1; 0.05; 10]; % Pre-exponential factors
tau123 = [3.0; 5; 0.1]; % Lifetimes

% Set upper and lower boundaries [min max]
a1 = [0 100]; % Pre-exponential factor 1
a2 = [0 100];
a3 = [0 0];
t1 = [0 inf]; % Lifetime 1
t2 = [0 inf];
t3 = [1 1];

% Optimization options (can be increased for higher accuracy):
MaxFunEvals = 5000; % Maximum no. of function evaluations
MaxIter = 1000; % Maximum no. of iterations

```

Choose the filenames to be analysed and run the script (press F5 or type `PrepareDecays` from the MATLAB command window). The fitting will take some time depending on the number of files and the size of the shift range (Figure 19). A progressbar will show the estimated calculation time and results are continuously updated in the MATLAB command window.

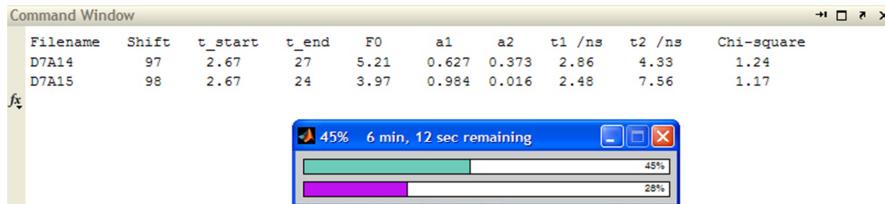


Figure 19. Running `PrepareDecays.m`

The output file, named 'Studyname' \_decaypar.txt, will look something like this:

Filename	Shift	t_start	t_end	F0	a1	a2	t1/ns	t2/ns	Chi-
<b>square</b>									
D7A14	97	2.67	27	5.21	0.63	0.37	2.86	4.33	1.24
D7A15	98	2.67	24	3.97	0.98	0.02	2.48	7.56	1.17
D7A20	11	3.55	30	4.80	0.22	0.78	3.44	4.37	1.13
D7A21	32	3.56	29	4.64	0.68	0.32	3.78	4.61	

1.10								
D9A15	86	2.70	17	0.61	0.99	0.01	0.91	4.13
1.95								

Here, Shift is a channel shift between the decay and the IRF (actually channelshift\*L), t\_start and t\_end is the start and end times, respectively, F0, a1, and a2 are pre-exponential factors, t1 and t2 are the fitted lifetimes, and the Chi-square value is for the fit obtained with the optimized parameter values. Note that in some cases the time-intervals chosen by the fitting routine is not optimal and must be manually set instead. *In general, the fits obtained using the optimized parameters should always be thoroughly checked before continuing with the analysis, as the decay parameters will affect the results of the analysis if not properly set.* Check the resulting fits by setting plotfit = 'Y'.

## 2 Speeding up analysis: Store pre-fit chi squares

This step is optional but can be used to save computational time during the chi-square FRET analysis if large chi-square surfaces will be calculated. The purpose here is to pre-fit all decays for different values of the  $\kappa^2/R^6$  ratio prior the actual FRET analysis. The fitted chi-squares from this step are then called during the FRET analysis when a given  $\kappa^2/R^6$  is encountered for each D-A pair. This eliminates the need to fit the decays over and over again when running the chi square surfaces during a FRET analysis, however, running all the pre-fits requires some computational time invested. If this step is not chosen, FRETmatrix will fit all the decays during a chi-square surface FRET analysis run.

- 1) Open MakeStoredFits.m. The input for the script will look something like this:

```
% Specify study:
Studyname = 'B-DNA'; % Position decays:
FRETmatrix/inputs/decays/Studyname
Samples = 'D7A14 D7A15 D7A20 D7A21 D9A15'; % D-A combinations for
analysis.
file_extension = '.dat';

% Chi square parameters:
M = 2500;           % No. of calculated chi-squares for each decay
E_lower = 0;         % Lower efficiency cut off
E_upper = 0.8;       % Upper efficiency cut off

% Donor reference lifetimes in same order as in sequence:
tau_d = [4.32 4.35 4.55];           % /ns

% Fraction of donor components in decays. Same donor order as in
Sequence
frac_Ds = [ 1 1 1 1 1 1 1 1 1];    % Fraction of donor component in
decay
tauIMP = 12.01;                     % Lifetime of possible impurity
frac_FRET = [0.996 0.996];          % Fraction of donors coupled to
acceptor

% FRET parameters:
Jint = 1.3e14;                      % Overlap integral M.1cm.1nm4
```

```

ny = 1.4; % Refractive index
Qd = 0.23; % Donor quantum yield

% Optimization settings:
decayoptim_Tol = [0.01 0.001]; % [TolFun TolX] Fun: ChiSq, X: F0 & A
Decay_MaxFunEvals = 10000; % Maximum no. of function evaluations
Decay_MaxIter = 5000; % Maximum no. of iterations in decay
fitting

% Export results to .txt file?
save_storedchis = 'Y'; % Will be saved in the decay folder

```

- 2) Specify samples and decay parameters and run the script by pressing F5. The script will call the shift and time-intervals from the `decaypar.txt` file produced as described above. Here, all decays will be fit at M different FRET efficiencies ranging from `E_lower` to `E_upper`. An output file called '`Studyname'_storedchis.txt` containing all the fit chi-square values and the parameters used to produce the fits will be saved to the decays folder. This output file is then called when running the FRETmatrix chi-square surface analysis (below). If the stored chis are used, the FRET parameters used to calculate the pre-fit chi-squares, specified in the `storedchis.txt` file, will automatically be imported in the FRET analysis. The resulting `storedchis.txt` file will look something like this:

```

22
Studyname: B-DNA
Samples:
D7A14 D7A15 D7A20 D7A21 D9A15
M = 2500
tau_d:
4.32 4.35 4.55
tauIMP:
12.01
frac_Ds:
1.000 1.000 1.000 1.000 1.000
frac_FRET:
0.996
Jint:
1.30e+014
ny:
1.40
Qd:
0.230 0.230

Stored chi-square values:
k^2/R^6 D7A14 D7A15 D7A20 D7A21 D9A15
0.000000e+000 49.874270 266.638946 2.008859 5.506675 561.976776
4.662394e-013 49.699960 266.301566 1.992618 5.450659 561.840096
9.327776e-013 49.508359 265.946423 1.966164 5.400735 561.700064
1.399615e-012 49.331958 265.602516 1.941512 5.352094 561.560488
1.866751e-012 49.157133 265.260176 1.922117 5.292344 561.421369
2.334187e-012 48.979170 264.913367 1.900920 5.242812 561.287494
2.801923e-012 48.796988 264.563507 1.876595 5.192448 561.152260
...

```

### 3 Run base-base FRET analysis

Open RunProgram.m. This file contains all variables and choices that must be specified when running the FRETmatrix analysis. This file is thus equivalent to the user interface in part I, and can also be used instead of the user interface in part I. The RunProgram.m will look something like this:

```
%% Structure and FRET pairing input parameters

% Sequence in sequential base-pairs: (Write: kink, site or space for
separating structural units). Tip: You can always write Automatic;
% Example:
% Sequence = [CG GC AT TA] corresponds to:
% 5'-CGAT Strand I
% 3'-GCTA Strand II

Sequence = [tCOG AT tCOG AT tCOG AT AT GtCn GtCn AT CG GC AT GtCn GtCn
AT];

% Structure information:
Structure = 'B'; % 'A', 'B'-form or 'I' for imported
filename = 'None'; % For structure import: .dat, .par, .out

% Display structure in Matlab? 'Y' or 'N'.
display_mol = 'N'; % Displays optimized molecular structure after
analysis

% Kink parameters. If no. of parameters specified < no. of kinks, then
all
% kinks in the structure are defined by the same set of parameters:
% Kink # [1 2...; ]. Won't be used for analysis part. See below.

KinkPar = [0;... % Alpha /deg
           45;... % Beta /deg
           0;... % Gamma /deg
           0;... % vx /Å
           0;... % vy /Å
           6]; % vz /Å

% FRET parameters:
Qd = [0.23 0.23 0.23]; % Donor quantum yields
Jint = 1.3e14; % Overlap integral

% Automatically calculate all FRET combinations?
Calc_All = 'N'; % Y or N
SepCutOff = [4 15]; % FRET-pair separation range (in bp)
ConstrainSite = [10 11]; % Left and right bp next to the site

% Additional input parameters:
moloutfile = 'pdb'; % Structure output file: 'pdb' or 'mol'
checkclashes = 'N'; % Check for sterical clashes? For analysis, see
below.
vdwradius = 1.3; % Van der Waals radii for steric clash detection
/Å
bspar = 'seq'; % Base-pair step parameters: seq, avg
ny = 1.4; % Refractive index

% For calculating all FRET combinations:
```

```

MakeExcel = 'Y'; % Make excel file with output data? 'Y' or 'N'
QY_tCO = 0.23; % Quantum yield of tCO
QY_tC = 0.20; % Quantum yield of tC
J_tCO = 1.3e14; % Overlap integral with tCO as donor
J_tC = 5.6e13; % Overlap integral with tC as donor

% Chi-Square analysis? 'Y' or 'N'.
Analyse = 'Y'; % Specify input parameters below

%% Input parameters for fitting the decays:

% Specify study:
Studyname = 'B-DNA'; % Decay folder:
FRETmatrix/inputs/decays/Studyname
Samples = 'D7A14 D7A15 D7A20 D7A21 D9A15'; % D-A combinations for
analysis.
file_extension = '.dat'; % Suffix of decay files

% Donor reference lifetimes in same order as in sequence:
tau_d = [4.32 4.35 4.55];

% Fraction of donor components in decays. Same donor order as in
Sequence
frac_FRET = [0.99 1]; % Fraction donors w. FRET
frac_Ds = [1 1 1 1 1 1 1 1 1]; % Fraction of donor component in
decay
frac_don = [0.99 1]; % Fraction of donor component in
decay
tauIMP = 12.01; % Lifetime of possible impurity
use_frac_Ds = 'N'; % Use specific values of frac_don? defined in
frac_Ds

% Decay parameter file:
decayparfile = 'B-DNA_decaypar.txt'; % Positioned in inputs folder.

%% Input parameters for chi-square analysis:

% Run several chi-square surfaces and average?
chiruns = 1; % No. of chi-square surface runs
Save_ChiSqSumMatrix = 'N'; % Save the calculated chi-square
surfaces?
PlotAvg_ChiSqSumMatrix = 'N'; % Plot the averaged chi-square surface?

% Sampling parameters and chi-square analysis
N = 1000; % No. of samples drawn from distribution

% Lower and upper bounds in optimization:
% Insert values as: [start end #steps]. If #steps = 1, the parameter
will % be optimized within the specified start-end interval at each
coordinate % on the calculated chi-square surface.

% Base probe parameters:
FWHMtheta_lu = [0 20 4]; % Theta-FWHM /deg.
FWHMphi_lu = [0 20 4]; % Phi-FWHM /deg.
bendA_lu = [25 25 1]; % Acceptor bending angle /deg.
bendD_lu = [8 8 1]; % Donor bending angle /deg.
J_lu = [1.3e14 1.3e14 1]; % Overlap integral /M-lcm-1nm4

```

```

% Kink parameters:
alpha_lu = [0 0 1]; % /deg.
beta_lu = [0 0 1]; % /deg.
gamma_lu = [0 0 1]; % /deg.
vx_lu = [0 0 1]; % /Å
vy_lu = [0 0 1]; % /Å
vz_lu = [0 0 1]; % /Å

% Type of distributions: 'LJ'/'H' is Len. Jones or Harmonic,
respectively.
thetaType = 'LJ'; % Potential in theta direction (in
plane).
phiType = 'H'; % Potential in phi direction (out of
plane)

% Use pre-fitted chi-square values from make_storedchis.m
use_storedchis = 'Y'; % 'Y' or 'N'
storedchis_filename = 'B-DNA_storedchis_ex.txt'; % Located in decays
folder

% Geometrical constraints in optimization? 'Y' or 'N'.
clash_constraint = 'N'; % Use sterical clash constraint in
analysis
neighbdist_constraint = 'N'; % Use distance constraint in analysis
longstrand = 1; % The strand with extra bases (1 or 2)
vdwradius = 1.3; % van der Waals radii
neighbdistmax = 11; % Max. neighbour-distances in same strand
/Å
ChiSqPenalty = 60; % Chi-square penalty for violated
constraint
sa_constraints = 'Y'; % Use simulated annealing for first optimization
step

% Plots? 'Y' or 'N'
plot_chisqsurf = 'Y'; % Plot all calculated chi-square
surfaces?
plot_decay = 'N'; % Plot fitted decays?
plot_FRET = 'N'; % Plot resulting FRET efficiencies?
plot_sampling = 'N'; % Plot sampled distribution?
savefits = 'N'; % Save fitted decays?

% Matlab Pooling? 'Y' or 'N'
pooling = 'N'; % If possible, activate for faster distribution
optimization

% Optimization settings:
Joptim_Tol = [0.05 0.01e14]; % [TolFun TolX] Fun: ChiSqSum, X: Jint
ChiOpt_TolX = 0.001; % TolX: FWHM, bend
ChiOpt_TolFun = 0.001; % TolFun: ChiSqSum
KinkOpt_Tol = [0.01 0.01]; % [TolFun TolX] Fun: ChiSqSum, X:
KinkPar
decayoptim_Tol = [0.01 0.001]; % [TolFun TolX] Fun: ChiSq, X: F0 &
a,b
Decay_MaxFunEvals = 5000; % Max function evaluation in decay
fitting
Decay_MaxIter = 1000; % Max iterations in decay fitting

% Simulated annealing algorithm settings (see also Matlab saoptimset
help):
sim_anneal_optim = 'N'; % Optimize kink using simulated annealing?
saLimit = 30; % Stop optimization when the chisq reaches this value

```

```

saTolFun = 0.001; % Stop optimization when change in chisq < saTolFun
saTimeLimit = 3*60; % Stop optimization after saTimeLimit seconds
saStallTimeLimit = 500; % Stop iteration after saStallTimeLimit
seconds
saDisplay = 'iter'; % Display data from optimization in Matlab
workspace.
saPlot = 'N'; % Plot data computed by the simulated annealing
algorithm?
saTempFcn = @temperatureexp; % Decrease T according to saTempFcn
function
saInitialTemp = 200; % Initial temperature

% Print results to file? 'Y' or 'N'.
Resultsfile = 'N'; % Will be written to
Studyname_analysis.txt

```

Here, the red section is analogues to the user interface above, while the blue sections contain the parameters used to run the chi-square analysis. The input parameters are described in the comments shown above. Some additional notes of clarification:

<b>Running analysis</b>	To run a chi-square analysis Analyse must be set to 'Y', else the program will only perform a single point calculation using the parameters specified in the red section.
<b>Study name</b>	The study name is used as an identifier by FRETmatrix to keep track of more than one studies. It has an impact in a number of different ways: 1) When running an analysis, FRETmatrix will look in the folder: FRETmatrix/inputs/decays/'Studyname' for the decays. 2) The decaypar.txt file will/must be named 'Studyname'_decaypar.txt and be placed in the decays/'Studyname' folder. 3) The pre-fit chi-squares file will be called 'Studyname'_storedchis.txt and must be placed in the decays/'Studyname' folder.
<b>Specifying decays</b>	Using the filename convention described above, the program automatically match a decay file specified in the Samples input with a donor-acceptor position in the Sequence. The names do not need to be specified in any specific order.
<b>Choosing D-A combinations</b>	Any sub-group of donor-acceptor pairs out of the total number of possible combinations can be specified for analysis in the Samples input. In case not all FRET combinations were measured, or not all decays are considered to be well resolved for the analysis, the program will only use the combinations specified in Samples and automatically pair the chosen D-A combinations with the corresponding parameters specified in the decaypar.txt file.
<b>Presence of impurity or non-coupled donors</b>	The presence of even very small amounts of uncoupled donors or an emitting impurity in the sample will have an impact on the fitted decays when the FRET efficiency is high (short donor lifetimes). The fraction of donors in the sample (1-'fraction of impurity') is set by the frac_Ds (for choosing specific values) or the frac_don (for optimizing the value). The lower and upper boundary of fraction of

	donors coupled to an acceptor is set by the <code>frac_FRET</code> .
<b>Running several chi-square surfaces</b>	The program will run the chi-square surface using the specified parameters “chiruns” times. If <code>Save_ChiSqSumMatrix = 'Y'</code> all the calculated surfaces will be saved in the outputs folder. If <code>PlotAvg_ChiSqSumMatrix = 'Y'</code> the calculated surfaces will be saved to the outputs folder and the surface averaged from all runs will be made, saved and plotted.
<b>Setting parameters for chi-square analysis</b>	The parameter intervals are set under “Lower and upper bounds in optimization” in the blue section above. Parameters are set according to <code>*_lu = [lower upper #steps]</code> . E.g. the setting <code>alpha_lu = [0 30 4]</code> will make 4 alpha steps from 0° to 30° (i.e. 0°, 10°, 20°, and 30°), whereas <code>alpha_lu = [0 30 1]</code> will optimize the value of $\alpha$ within the lower and upper boundaries (here 0°-30°). The program automatically identifies what parameters are chosen as the x and y of the $\chi^2$ surface being two parameters where <code>#steps &gt; 1</code> . At each coordinate on the surface, the parameters not included in the chi-square surface are optimized within the chosen lower and upper conditions. Also one dimensional chi-square surfaces can be run if only one parameter is set at <code>#steps &gt; 1</code> . If <code>#steps = 1</code> for all parameters the program will begin a global optimization of all parameters.
<b>Type of distribution</b>	Defines the appearance of the probability distributions representing in-plane and out-of-plane base motion. In the program a harmonic potential is simulated by sampling a Gaussian distribution.
<b>Imposing geometrical constraints in analysis</b>	When having geometrical constraints activated in the analysis all parameter values resulting in a nucleic acid structure in violence with the chosen constraints are automatically discarded in the analysis and the $\chi^2$ is set to a large value (the <code>ChiSqPenalty</code> ). In the final chi-square plot, those parameter values that resulted in a violence of one of the imposed constraints will automatically be associated with the highest chi-square value on the surface.
	The <code>clash_constraint</code> discards structures in which there is a sterical clash between two atoms in different bases. A sterical clash is when the interatomic distance is $r < 2*vdw$ , where <code>vdw</code> is the atomic van der Waals radius. The <code>neighbdist_constraint</code> imposes a maximum distance between two neighbouring bases in a strand, set by the <code>neighbdistmax</code> value.
	In some cases the gradient based algorithm may get stuck at the <code>ChiSqPenalty</code> if the parameter value start guesses themselves are in violence with the constraints. To avoid this, an initial simulated annealing algorithm <sup>14</sup> optimization step can be chosen by setting <code>sa_constraints = 'Y'</code> . This first step will find a set of suitable starting conditions for the gradient based parameter optimization algorithm.
<b>Plotting</b>	<code>plot_chisqsurf</code> : Will plot the calculated chi-square surface. If <code>chiruns &gt; 1</code> each calculated surface will be plotted if this setting is activated. If only the averaged surface is wanted, set

	<pre>PlotAvg_ChiSqSumMatrix = 'Y' and plot_chisqsurf = 'N'.</pre> <p><code>plot_decay</code>: Will plot all the fitted decays. This setting can only be chosen using one single specific set of parameters (i.e. not when an analysis is run).</p> <p><code>plot_FRET</code>: Will plot the FRET efficiency graph. Can only be chosen when using one single specific set of parameters.</p> <p><code>plot_sampling</code>: Will plot the sampled dipole distributions of the first donor and last acceptor in the sequence in Cartesian space. The coordinates of directions can be directly compared within the coordinate frame of the global structure. To see the coordinate axes of the structure, set 'Axes' on in the molviewer when visualizing the molecule.</p> <p><code>savefits</code>: Will save the measured and fitted decays in two matrices in the outputs folder for later plotting: <code>allmeasdecays.mat</code> and <code>decayfitresults.mat</code>. The saved decays and fits can then be plotted one by one using the <code>plotsaveddecays.m</code> script located in the 'Calls' folder. This setting should only be chosen when using one single specific set of parameters.</p>
<b>MATLAB pool</b>	See the MATLAB instructions for a description. This setting activates parallel computing on a pool of workers (e.g. more than one processor). In some cases it can speed up the optimization of the parameters: $\theta$ -FWHM, $\phi$ -FWHM and the bending angles.
<b>Balancing speed and accuracy of fit</b>	The "Optimization settings" provide a means to balance the speed and accuracy of the optimization. In general, the <code>TolFun</code> and <code>TolX</code> values provide more accurate fits at lower values while the <code>MaxFunEvals</code> and <code>MaxIter</code> provide more accurate fits at larger values.
<b>Choosing optimization algorithm</b>	When optimizing all kink parameters at once choosing the simulated annealing optimization algorithm <sup>14</sup> may be advantageous if the chi-square surface is complex. This algorithm is chosen by setting <code>sim_anneal_optim = 'Y'</code> . However, whereas this algorithm may find the global minimum it does so very slowly. If in doubt, disable simulated annealing. See the Matlab manual for a description of simulated annealing optimization settings.
<b>Resultsfile</b>	This will export all details of the fits from each step in the chi-square surface analysis to a .txt file located in the 'outputs' folder named ' <code>Studyname'_analysis.txt</code> .

**Working example 1: Run chi-square surface with  $\theta$ -FWHM and  $\phi$ -FWHM as parameters.**

- 1) Open RunProgram\_Example1.m.
- 2) Run a chi-square surface using the predefined settings by pressing F5.

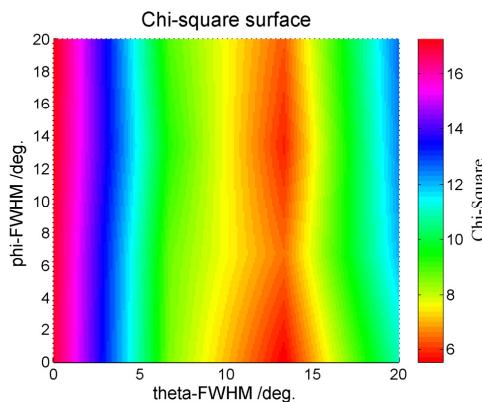
This will run a chi-square surface using the  $\theta$ -FWHM and  $\phi$ -FWHM as the two surface parameters. The samples are a set of donor-acceptor combinations in B-DNA located in the inputs/decays/B-DNA folder. The input sequence is:

```
%-----%
(X, Y, Z = tC, tCO, tCn)

#123...
5'-CGATCAYAYAAGGACGAGGATAAGGAGGAGG      Strand I
3'-GCTAGTGTGTGTTZTGCTZTATTCCCTCCTCC      Strand II

Total length (bp):
33
%-----%
```

In this example, four steps of each surface parameter are calculated in the interval from  $0^\circ$  to  $20^\circ$  (i.e. at values  $0^\circ$ ,  $6.67^\circ$ ,  $13.3^\circ$ , and  $20^\circ$ ) yielding sixteen coordinates on the resulting chi-square surface (Figure 20). Points in between the sixteen steps are interpolated. In this example, the other parameter values are not optimized at each coordinate on the surface since the upper and lower boundaries are set to be equal. Change the parameter intervals by setting the \*\_LU parameters. This example uses pre-fit decays and thus requires the B-DNA\_storedchis\_ex.txt (made by MakeStoredFits.m) to be located in the decays folder. The resulting chi-square surface is an average of 3 runs.



**Figure 20. Result from a run of Working example 1.**

## Working example 2: Optimize kink parameters without constructing chi-square surface

Although it is possible to run a simple optimization, this example is not advised for obtaining the parameter values since there is a high risk that the optimization algorithm gets stuck in a local minimum. In addition, the simulation is based on random sampling and the found parameters will therefore differ from one optimization to the next. In some cases it may be a useful first hint of where to look on the chi-square surface (Working example 1 and 3).

- 1) Open RunProgram\_Example2.m.
- 2) Run a parameter optimization using the predefined settings by pressing F5.

In this example the same samples as in Working example 1 are analysed. However, a kink has now been defined in the sequence in between bp 12 and bp 13. The sequence is now:

```
#123...
5'-CGATCAYAYA AGGACGAGGATAAGGAGGAGG      Strand I
3'-GCTAGTGTGT TZGTGCTZTATTCCCTCCTCC      Strand II
```

The structure is thus modelled by the program as two structural B-DNA units separated by a kink with the kink being a regular base-pair. The intrinsic probe parameters are set to:  $\theta$ -FWHM = 13°,  $\phi$ -FWHM = 2°,  $tC_{nitro}$ - $\phi_b$  = 25°,  $tC^0$ - $\phi_b$  = 8°. The sterical clash constraint has been enabled in the optimization.

A resulting output in the MATLAB command window will look like this:

```
...
No parameters have been specified for a chi-square analysis
%-----%
Global Chi-square: 3.415

Base probe parameters:
thetaFWHM = 13.00;
phiFWHM = 2.00;
bendA = 25.00;
bendD = 8.00;
J = 130000000000000.00;

Kink parameters:
alpha = 40.80;
beta = 1.81;
gamma = 0.00;
vx = 1.41;
vy = -2.31;
vz = 2.86;
%-----%
```

Note: The simulated annealing algorithm<sup>14</sup> can be chosen for this example if the chi-square surface is complex. Set `sim_anneal_optim = 'Y'` to activate simulated annealing.

### Working example 3: Run chi-square surface of kink parameters.

- 1) Open RunProgram\_Example3.m.
- 2) Run a chi-square surface using the predefined settings by pressing F5.

This will run a chi-square surface of the same setup as in Working example 2. The  $x$ - and  $y$ -values of the surface are the DNA-bending angle,  $\beta$ , and the helical rise,  $v_z$ , respectively, each calculated at four steps (Figure 21). The other parameters are optimized within the specified intervals which is why this example requires considerably longer computational time than Working example 1. The sterical clash constraint is enabled which is why the chi-square value is observed to drastically increase at  $v_z$  values below 3 Å. Setting `display_mol = 'Y'` will show the structure corresponding to the minimum found on the calculated chi-square surface which in this case is a regular B-DNA helix (Figure 22).

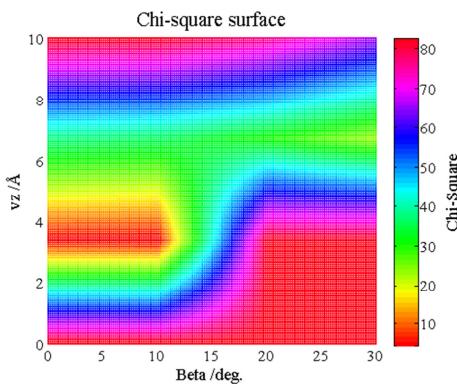


Figure 21. Resulting chi-square surface of one run of Working Example 3.

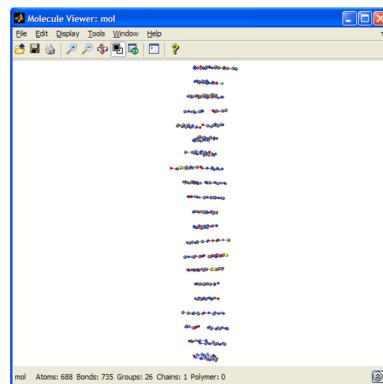


Figure 22. Resulting output structure from Working example 3.

## Implementing a new base probe in the program

This is done using the `MakeBP.m` script located in the calls folder. A punctuated protocol is supplied in the help intro of the script.

## User Guide References

- (1) Arnott, S.; Hukins, D. W. L. *Biochemical and Biophysical Research Communications* **1972**, *47*, 1504.
- (2) Arnott, S.; Hukins, D. W. L. *J. Mol. Biol.* **1973**, *81*, 93.
- (3 )Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakhed, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229.
- (4) Lu, X. J.; Olson, W. K. *Nucleic Acids Research* **2003**, *31*, 5108.
- (5) Zheng, G. H.; Lu, X. J.; Olson, W. K. *Nucleic Acids Research* **2009**, *37*, W240.
- (6) Preus, S.; Börjesson, K.; Kilså, K.; Albinsson, B.; Wilhelmsson, L. M. *J. Phys. Chem. B* **2010**, *114*, 1050.
- (7) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. *Nucleic Acids Research* **2008**, *36*, 157.
- (8) Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Research* **2005**, *33*, 5019.
- (9) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.

End of user guide

## SUPPLEMENTARY DATA

**Supplementary Data 1.** Base-pair step parameter matrix of pdb entry 1TGH, the complex formed between the TATA binding protein and its DNA target (13). The data was obtained using w3DNA (11).

### Local base-pair step parameters

step	Shift	Slide	Rise	Tilt	Roll	Twist
1 CG/CG	0.21	0.65	2.95	0.30	8.52	35.43
2 GT/AC	0.16	-0.53	3.58	-3.24	-1.37	41.06
3 TA/TA	0.46	-0.23	5.08	5.75	45.10	18.59
4 AT/AT	-1.87	-0.24	3.69	1.58	24.83	16.06
5 TA/TA	0.45	2.37	3.17	0.06	7.94	25.61
6 AT/AT	0.80	0.95	3.14	-0.16	21.26	1.46
7 TA/TA	0.02	2.26	3.97	0.51	31.05	27.83
8 AT/AT	0.58	0.04	3.23	-1.79	13.98	20.50
9 TA/TA	-0.08	-0.81	5.49	0.89	50.70	16.68
10 AC/GT	-0.74	0.09	2.78	4.55	13.80	15.74
11 CG/CG	-0.59	1.36	3.27	-2.18	3.07	32.12

**Supplementary Data 2.** Simulated FRET characteristics of model systems (Figure 3 in paper). Sep. denotes donor-acceptor separation in base-pairs. D and A is donor and acceptor positions, respectively, in the sequence. R is the dipole center-center distance, K<sup>2</sup> is  $\chi^2$  and E is the FRET efficiency.

B-DNA. Donor positioned in strand I, acceptor in strand II.

FRET-pair combinations:				
Sep.	D	A	R/Å	K <sup>2</sup>
0	1	2	5.2	0.311
1	1	3	6.7	0.003
2	1	4	9.8	0.483
3	1	5	13.5	0.760
4	1	6	17.4	0.487
5	1	7	21.1	0.086
6	1	8	24.6	0.049
7	1	9	27.9	0.484
8	1	10	31.0	0.921
9	1	11	34.0	0.784
10	1	12	37.1	0.219
11	1	13	40.3	0.021
12	1	14	43.6	0.489
13	1	15	47.1	0.950
14	1	16	50.6	0.754
15	1	17	54.2	0.187

%-----%

B-DNA. Donor positioned in strand II, acceptor in strand I:

FRET-pair combinations:				
Sep.	D	A	R/Å	K <sup>2</sup>
0	1	2	8.3	0.156
1	1	3	10.7	0.247
2	1	4	13.0	0.086
3	1	5	15.4	0.047
4	1	6	18.0	0.554
5	1	7	20.8	0.898
6	1	8	24.0	0.489
7	1	9	27.4	0.020
8	1	10	31.0	0.213
9	1	11	34.6	0.755
10	1	12	38.2	0.893
11	1	13	41.6	0.452
12	1	14	44.9	0.021
13	1	15	48.1	0.199
14	1	16	51.2	0.777

15	1	17	54.4	0.964	0.025
----	---	----	------	-------	-------

%-----%

**A-RNA. Donor in strand I, acceptor in strand II:**

FRET-pair combinations: -----%					
Sep.	D	A	R/Å	K^2	E
0	1	2	6.8	0.395	1.000
1	1	3	7.8	0.647	0.999
2	1	4	9.0	0.267	0.997
3	1	5	10.3	0.011	0.861
4	1	6	11.7	0.414	0.991
5	1	7	13.5	0.547	0.984
6	1	8	15.6	0.128	0.860
7	1	9	18.0	0.057	0.536
8	1	10	20.7	0.556	0.831
9	1	11	23.4	0.953	0.798
10	1	12	26.2	0.781	0.621
11	1	13	29.0	0.272	0.240
12	1	14	31.6	0.000	0.000
13	1	15	34.1	0.277	0.108
14	1	16	36.4	0.824	0.195
15	1	17	38.8	1.036	0.173

%-----%

**A-RNA. Donor in strand II, acceptor in strand I:**

FRET-pair combinations: -----%					
Sep.	D	A	R/Å	K^2	E
0	1	2	6.4	0.170	0.999
1	1	3	7.8	0.223	0.999
2	1	4	10.0	0.042	0.966
3	1	5	12.7	0.036	0.853
4	1	6	15.6	0.256	0.925
5	1	7	18.5	0.431	0.881
6	1	8	21.3	0.371	0.734
7	1	9	23.9	0.139	0.339
8	1	10	26.3	0.000	0.000
9	1	11	28.7	0.175	0.178
10	1	12	30.9	0.540	0.298
11	1	13	33.2	0.687	0.261
12	1	14	35.6	0.414	0.123
13	1	15	38.0	0.053	0.012
14	1	16	40.6	0.075	0.011
15	1	17	43.3	0.490	0.049

%-----%

**DNA-TBP: No TBP**

FRET-pair combinations: -----%					
Sep.	D	A	R/Å	K^2	E
4	1	6	18.7	0.279	0.819
6	1	8	24.6	0.050	0.135
8	1	10	30.9	0.923	0.419
10	1	12	37.2	0.003	0.001
3	2	6	15.4	0.047	0.708
5	2	8	21.6	0.801	0.843
7	2	10	27.4	0.400	0.395
9	2	12	34.6	0.752	0.231

%-----%

**DNA-TBP: With TBP bound**

FRET-pair combinations: -----%					
Sep.	D	A	R/Å	K^2	E
4	1	6	18.3	0.658	0.923

6	1	8	19.0	0.393	0.853
8	1	10	22.9	0.019	0.084
10	1	12	33.1	1.433	0.430
3	2	6	17.9	0.032	0.398
5	2	8	19.8	0.017	0.164
7	2	10	24.0	0.049	0.150
9	2	12	34.2	0.042	0.017

## REFERENCES

1. Wilhelmsson, L.M., Sandin, P., Holmen, A., Albinsson, B., Lincoln, P. and Norden, B. (2003) Photophysical characterization of fluorescent DNA base analogue, tC. *J Phys Chem B*, **107**, 9094-9101.
2. Sandin, P., Borjesson, K., Li, H., Martensson, J., Brown, T., Wilhelmsson, L.M. and Albinsson, B. (2008) Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res*, **36**, 157-167.
3. Preus, S., Borjesson, K., Kilsa, K., Albinsson, B. and Wilhelmsson, L.M. (2010) Characterization of nucleobase analogue FRET acceptor tCnitro. *J Phys Chem B*, **114**, 1050-1056.
4. Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Muller, S. and Seidel, C.A. (2011) Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity. *J Am Chem Soc*, **133**, 2463-2480.
5. Woźniak, A.K., Schroder, G.F., Grubmuller, H., Seidel, C.A. and Oesterhelt, F. (2008) Single-molecule FRET measures bends and kinks in DNA. *Proc Natl Acad Sci U S A*, **105**, 18337-18342.
6. Preus, S., Kilsa, K., Wilhelmsson, L.M. and Albinsson, B. (2010) Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family. *Phys Chem Chem Phys*, **12**, 8881-8892.
7. Lu, X.J., El Hassan, M.A. and Hunter, C.A. (1997) Structure and conformation of helical nucleic acids: rebuilding program (SCHNArP). *J Mol Biol*, **273**, 681-691.
8. El Hassan, M.A. and Calladine, C.R. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J Mol Biol*, **251**, 648-664.
9. Arnott, S. and Hukins, D.W.L. (1972) Optimized parameters for A-DNA and B-DNA. *Biochem Biophys Res Co*, **47**, 1504-&.
10. Arnott, S. and Hukins, D.W. (1973) Refinement of the structure of B-DNA and implications for the analysis of x-ray diffraction data from fibers of biopolymers. *J Mol Biol*, **81**, 93-105.
11. Zheng, G., Lu, X.J. and Olson, W.K. (2009) Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res*, **37**, W240-246.
12. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakkeb, Z. et al. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol*, **313**, 229-237.
13. Juo, Z.S., Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. and Dickerson, R.E. (1996) How proteins recognize the TATA box. *J Mol Biol*, **261**, 239-254.



## **Paper IV**

*In which we present a multistate, reversible switch  
based on a functionalized DNA hairpin loop...*



# A DNA based five-state switch with programmed reversibility††

**Jonathan R. Burns,<sup>a</sup> Søren Preus,<sup>b</sup> Daniel G. Singleton<sup>a</sup> and Eugen Stulz<sup>\*a</sup>**

*Received 10th August 2012, Accepted 21st September 2012*

DOI: 10.1039/c2cc35799b

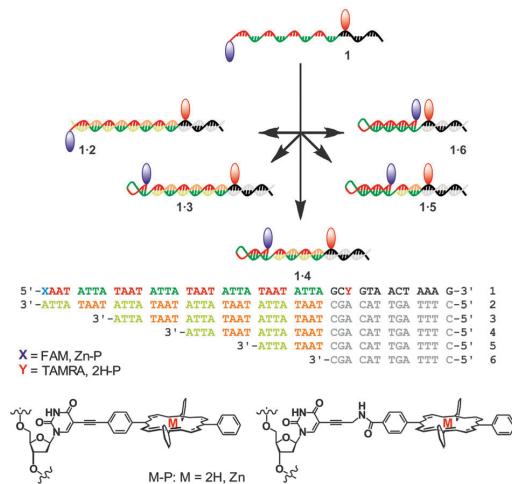
A programmable switch based on a DNA hairpin loop is functionalised with a rigid or flexible porphyrin or FAM and TAMRA FRET pair, which provides insight into the restructuring of the hairpin as well as porphyrin-porphyrin coupling. The switch contains five discrete states which can be accessed independently and followed by real-time spectroscopy, opening the way to a quinary computing code.

The unique self-recognition properties of DNA have been explored extensively for the formation of new self-assembled nano-architectures over the past decades. By taking the DNA out of its biological context, the emerging field of DNA nanotechnology is becoming increasingly attractive to advanced research in drug delivery, autonomous machines or computing<sup>1</sup>. Additional functionalities such as redox active metal complexes<sup>2</sup> or organic chromophores<sup>3</sup> are increasingly being incorporated into DNA, resulting in operational DNA based nano-devices. Still, the incorporation of chemically modified DNA into these nano-architectures is in its infancy, and the exploration of modified DNA as building blocks remains an important aspect to understand their behaviour and suitability for DNA nanotechnology. In particular, organic chromophores<sup>4</sup> such as pyrenes<sup>5</sup> and porphyrins<sup>6</sup> are gaining increasing attention as DNA modifiers e.g. for the creation of photonic wires, or as diagnostic tools since their optical properties (absorption, emission) can vary with the DNA sequence or a change in the environment such as pH, temperature or secondary structure.

Particularly intriguing structures arise when partially self-complementary DNA strands assemble to form intramolecular hairpin loops. The loops can be opened and closed through sequential addition of suitable complementary DNA strands, thus creating moving parts within a DNA nanostructure. The concept has been used in molecular beacons for DNA analysis,<sup>7</sup> in switchable DNA nanostructures with optical responses,<sup>8</sup> and in autonomous DNA walkers.<sup>9</sup> Here, we report a programmable switch based on a molecular beacon, where the DNA is partially self-complementary with repeating ATTA-TAAT sequences (Fig. 1). An additional 13

base sequence allows for specific recognition of complementary strands including various repeats of the complementary ATTA-TAAT box, thus enabling controlled elongation or contraction of the stem region. To demonstrate functionality, FRET pairs were attached to the extremes of the repeat region, giving access to a tuneable energy transfer system with well-defined chromophore distances from the same DNA strand by simply adding the appropriate complementary strand. The porphyrin based FRET system (denoted **1P**) comprises of a zinc porphyrin (donor, D) and a free-base porphyrin (acceptor, A), where the porphyrins are attached either *via* a rigid alkynyl linker or a more flexible propargyl-amide linker.<sup>10</sup> The rigid alkynyl linker ensures that the chromophores have a low diffusional mobility and remain in a well oriented environment, particularly in terms of the transition dipole moment, introduced in order to study the influence of the porphyrin distance and angle on the electronic coupling compared to the flexible linker. As a control system, we have also synthesised the analogous FAM and TAMRA labelled switch strand (denoted **1F**). Because this FRET pair is tethered *via* a longer and more flexible linker, the angular dependence on the FRET efficiency should in ideal cases be eliminated ( $\kappa^2 = 2/3$ ).

The  $T_m$  values of the two extreme complexes (the full length duplex **1P-2** and the hairpin with the longest possible stem of



**Fig. 1** Schematic of the adjustable hairpin loops with DNA sequences. The red and blue markers indicate attachment of FRET pairs.

<sup>a</sup> School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK. E-mail: est@soton.ac.uk;

<sup>b</sup> Department of Chemistry, University of Copenhagen,

† This article is part of the 'Nucleic acids: new life, new materials' web-themed issue.

† Electronic supplementary information (ESI) available: Theoretical FRET calculations, synthetic methods and spectroscopic analysis of the DNA strands. See DOI: [10.1039/c2cc35799b](https://doi.org/10.1039/c2cc35799b)

**Table 1** Melting temperatures of the DNA system and FRET efficiencies (experimental and theoretical)

DNA system	$T_m$ <b>1P</b> <sup>a</sup> /°C	$T_m$ <b>1F</b> <sup>b</sup> /°C	FRET FAM-TAMRA Exp. (calcd)
<b>1</b>	61.7	60.0	0.8 (0.98)
<b>1·2</b>	54.1	59.0	0.00 (0.01)
<b>1·3</b>	50.8	59.1	0.01 (0.03)
<b>1·4</b>	54.4	58.6	0.17 (0.13)
<b>1·5</b>	55.2	59.7	0.52 (0.57)
<b>1·6</b>	61.2	60.2	0.81 (0.98)

<sup>a</sup>  $T_m$  values obtained from UV-melting (260 nm). <sup>b</sup>  $T_m$  values obtained from fluorescence melting ( $\lambda_{ex} = 495$  nm,  $\lambda_{em} = 510$  nm).

DNA **1P-6**) show a clear difference, where **1P-2** has a  $T_m$  value of 54.1 °C, and for **1P-6** the  $T_m$  was determined to be 61.2 °C (Table 1). The molecular beacon **1P** on its own displays a  $T_m$  of 61.7 °C, overall confirming the intramolecular stabilisation effect on the duplex. In the presence of the complementary strands, the other hairpin-duplex combinations show intermediate  $T_m$  values in the expected order.

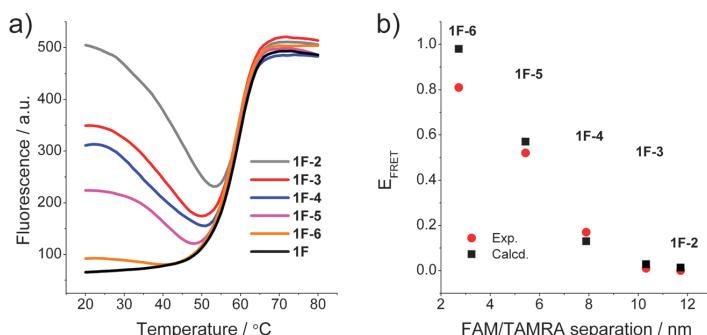
To determine changes in the direct environment of the modifications, we attempted to record the melting profiles monitoring the porphyrin region. However, the thermal lability of the zinc porphyrin complex interferes with the measurement (loss of metal at higher temperatures),<sup>10a</sup> thus fluorescence melting was conducted on the FAM and TAMRA system (Fig. 2a). This system displayed clear transitions when monitoring the donor emission of FAM. The initial donor emission intensity at 20 °C corresponds to the programmed donor–acceptor distance, and decreases in the expected order from **1F-2** to **1F-6**, resulting from the variable length of the stem. The denaturing profiles, however, are very different to the UV-melting curves,<sup>‡</sup> and show an initial decrease in donor emission, indicating a decrease in distance of the chromophores and increase in FRET between the donor and acceptor. Thus the dissociation of the complementary DNA strand is accompanied by a restructuring of the self-complementary DNA to form a longer and more stable stem-loop system as an intermediate structure, which then denatures to give the random coil single strand DNA, though the intermediate full length stem may not be formed in all cases. This is also corroborated by the fact that the fluorescence melting shows the same  $T_m$  of around 60 °C for all systems. The final donor fluorescence intensity is equal in all

systems and corresponds to the initial intensity of the system **1F-2**. This confirms both the complete denaturing at higher temperature as well as formation of the full length duplex in **1·2**.

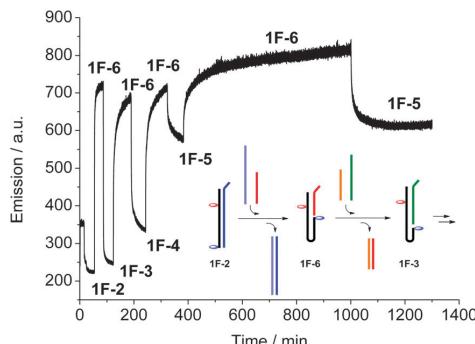
The FRET efficiencies for both the Zn- and 2H-porphyrin and the FAM and TAMRA system were measured and compared, in order to evaluate the difference between the FRET pairs where the chromophores are attached *via* a rigid or flexible linker. Since energy transfer efficiencies can vary greatly with probe diffusion and reorientation, the linker moiety plays an important role. The FRET parameters of the two FRET pairs (quantum yield, extinction coefficient and spectral overlap) were identified by steady state fluorescence spectroscopy.<sup>‡</sup> The FRET efficiencies ( $E_{FRET}$ ) were determined by spectral decomposition of the combined donor and acceptor emissions at an excitation wavelength of 426 nm. The quantum yield of **Zn-P** was determined to be  $\Phi = 0.12$  (using quinine sulphate as the standard) which in combination with a spectral overlap integral of  $J = 2.3 \times 10^{14} \text{ M}^{-1} \text{ cm}^{-1} \text{ nm}^4$  yields a Förster distance of  $R_0 = 28.4 \text{ Å}$ . The Förster distance of the FAM and TAMRA pair was determined to be  $R_0 = 57 \text{ Å}$ . In addition, the FRET efficiencies were predicted theoretically for all systems using a custom made FRET simulation program (FRET matrix).<sup>11</sup> Here, all DNA conformations were simulated as rigid B-form geometries and a full atomistic description of the tethered dyes was used to calculate the theoretical donor–acceptor distances.<sup>‡</sup>

The experimental FRET values of the **1F-x** combinations are in excellent agreement with the theoretically predicted values (Fig. 2b) and show a decreasing FRET efficiency upon increasing the fluorophore distance. Exceptions are the states **1F** and **1F-6** showing a smaller measured FRET efficiency compared to that expected from the calculated donor–acceptor distance. This deviation is often observed for closely spaced dyes and is likely to be a result of direct dye–dye interactions interfering with the FRET process.<sup>12</sup> Overall, the good correlation between expected and measured FRET efficiencies demonstrates the well-defined, discrete states of the switch.

Compared to the FAM and TAMRA control system, the porphyrin FRET system shows a very different behaviour. Despite that porphyrins attached to duplex DNA show FRET efficiencies which are dependent on the distance (including a control sequence used to determine the quantum yields of the porphyrin–DNA, see ESI<sup>‡</sup>), the switch sequences all show a



**Fig. 2** (a) Fluorescence melting curves of the DNA switch by monitoring FAM emission ( $\lambda_{ex} = 495$  nm,  $\lambda_{em} = 510$  nm); (b) calculated vs. experimental FRET of the modified system in the FAM and TAMRA system **1F**; (c) putative intra- and intermolecular stacking of the porphyrins in the DNA beacon systems.



**Fig. 3** Real-time monitoring of the switching between the five different switching states, recording acceptor emission ( $\lambda_{\text{ex}} = 495 \text{ nm}$ ,  $\lambda_{\text{em}} = 583 \text{ nm}$ ), and schematic drawing of the first steps. The final switches between **1F-6** and **1F-5** show the reversibility and stability of the system.

FRET efficiency close to 100%. No variation could be detected in either the different conformations or upon changing the linker to the porphyrin. Based on the  $R_0$  value, the calculations would predict FRET efficiencies of 43% for **1P-6** and <10% for all other combinations. This indicates that the porphyrins are in close contact at a maximum distance of about 3–10 Å (10–30% of  $R_0$  where 100% FRET is expected). The system is therefore disrupted by the hydrophobic interactions of the porphyrins which are probably stacked to give a very efficient energy transfer between the Zn- and 2H-porphyrin (Fig. 2c). At this point we cannot rule out intermolecular interactions as we<sup>10a</sup> and others<sup>6c</sup> have found efficient intermolecular stacking of porphyrins and small organic molecules covalently attached to DNA, which can act as a molecular glue between DNA strands.<sup>13</sup> Since the association constant for these intermolecular assemblies is unknown, it cannot be estimated to which extent they contribute to the FRET if they do so.

DNA has recently been used to generate molecular machines, where addition of specific complementary strands is used to trigger switching between different structures and states, and up to eleven discrete states have been achieved using a DNA tile actuator.<sup>14</sup> Our simple switch equally responds to a change in the system: by extending the complementary strand by eight bases to form an overhang (toehold), the strand can be peeled off, and the system switches to a different state by subsequent addition of a longer or shorter complementary strand. The switching can be monitored using real-time fluorescence spectroscopy (Fig. 3). Starting with the full-length duplex **1F-2**, subsequent switching between the various states leads to a system where five different states can easily and reversibly be addressed. The concept is schematically drawn in Fig. 3 for the switching between the states **1F-2** → **1F-6** → **1F-3**. In this way, not only a binary open-closed code can be made, but this system could form the basis for a quinary code that could significantly enhance the information density in DNA based computing.<sup>15</sup>

In summary, a DNA based adjustable strap was obtained by synthesising DNA strands with partially self-complementary boxes, and addition of complementary strands allows for a programmable adjustment of the stem-loop size. Such a system may be used as part of a molecular motor or computer, where a DNA input results in a change in structure thus encouraging motion and yielding a corresponding change in

output (e.g. change in energy transfer). The presented programmable switch, containing fluorophores attached by both rigid and flexible linkers, reveals that the formation of a particular DNA system can strongly be influenced by the nature of the modification. The hydrophobic nature of the porphyrins disrupts the stem structure in the beacon, and close contacts between the chromophores are evident, which shows the system's limitations. However, by tuning the properties of the modification specific structures and motion are retained.

We acknowledge generous support by ATDBio (Southampton, UK) and sponsorship by the EPSRC (CASE-DTA to JRB); MS analysis by the EPSRC National Mass Spectrometry Service, Swansea, is greatly acknowledged.

## Notes and references

- 1 (a) E. Stulz, *Chem.-Eur. J.*, 2012, **18**, 4456–4469; (b) E. Stulz, G. Clever, M. Shionoya and C. Mao, *Chem. Soc. Rev.*, 2011, **40**, 5633–5635; (c) T. J. Bandy, A. Brewer, J. R. Burns, G. Marth, T. Nguyen and E. Stulz, *Chem. Soc. Rev.*, 2011, **40**, 138–148.
- 2 (a) J. R. Burns, J. Zekonyte, G. Siligardi, R. Hussain and E. Stulz, *Molecules*, 2011, **16**, 4912–4922; (b) S. Bezer, S. Rapireddy, Y. A. Skorik, D. H. Ly and C. Achim, *Inorg. Chem.*, 2011, **50**, 11929–11937; (c) S. I. Khan, A. E. Beilstein and M. W. Grinstaff, *Inorg. Chem.*, 1999, **38**, 418–419.
- 3 (a) V. L. Malinovskii, D. Wenger and R. Häner, *Chem. Soc. Rev.*, 2010, **39**, 410–422; (b) H. Kashida, T. Takatsu, K. Sekiguchi and H. Asanuma, *Chem.-Eur. J.*, 2010, **16**, 2479–2486; (c) M. Nakamura, Y. Murakami, K. Sasa, H. Hayashi and K. Yamana, *J. Am. Chem. Soc.*, 2008, **130**, 6904–6905.
- 4 (a) M. M. Rubner, C. Holzhauser, P. R. Bohlander and H. A. Wagenknecht, *Chem.-Eur. J.*, 2012, **18**, 1299–1302; (b) A. Ruiz-Carretero, P. G. A. Janssen, A. Kaeser and A. Schenning, *Chem. Commun.*, 2011, **47**, 4340–4347.
- 5 (a) D. Lindegaard, A. S. Madsen, I. V. Astakhova, A. D. Malakhov, B. R. Babu, V. A. Korshun and J. Wengel, *Bioorg. Med. Chem.*, 2008, **16**, 94–99; (b) M. E. Ostergaard and P. J. Hrdlicka, *Chem. Soc. Rev.*, 2011, **40**, 5771–5788; (c) C. Boonlua, C. Vilaivan, H. A. Wagenknecht and T. Vilaivan, *Chem.-Asian J.*, 2011, **6**, 3251–3259; (d) S. P. Sau and P. J. Hrdlicka, *J. Org. Chem.*, 2012, **77**, 5–16.
- 6 (a) T. Nguyen, A. Brewer and E. Stulz, *Angew. Chem., Int. Ed.*, 2009, **48**, 1974–1977; (b) K. Borjesson, J. G. Woller, E. Parsa, J. Martensson and B. Albinsson, *Chem. Commun.*, 2012, **48**, 1793–1795; (c) A. Mammana, G. Pescitelli, T. Asakawa, S. Jockusch, A. G. Petrovic, R. R. Monaco, R. Purrello, N. J. Turro, K. Nakanishi, G. A. Ellestad, M. Balaz and N. Berova, *Chem.-Eur. J.*, 2009, **15**, 11853–11866; (d) A. W. I. Stephenson, A. C. Partridge and V. V. Filichev, *Chem.-Eur. J.*, 2011, **17**, 6227–6238.
- 7 R. T. Ranasinghe and T. Brown, *Chem. Commun.*, 2005, 5487–5502.
- 8 R. Varghese and H. A. Wagenknecht, *Org. Biomol. Chem.*, 2010, **8**, 526–528.
- 9 R. A. Muscat, J. Bath and A. J. Turberfield, *Nano Lett.*, 2011, **11**, 982–987.
- 10 (a) A. Brewer, G. Siligardi, C. Neylon and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 777–782; (b) L. A. Fendt, I. Bouamai, S. Thöni, N. Amiot and E. Stulz, *J. Am. Chem. Soc.*, 2007, **129**, 15319–15329.
- 11 S. Preus, K. Kilså, F.-A. Miannay, B. Albinsson and L. M. Wilhelmsson, *Nucleic Acids Res.*, 2012, DOI: 10.1093/nar/gks1856.
- 12 N. Di Fiori and A. Meller, *Biophys. J.*, 2010, **98**, 2265–2272.
- 13 (a) H. Kashida, T. Hayashi, T. Fujii and H. Asanuma, *Chem.-Eur. J.*, 2011, **17**, 2614–2622; (b) D. Baumstark and H. A. Wagenknecht, *Angew. Chem., Int. Ed.*, 2008, **47**, 2612–2614.
- 14 Z. Zhang, E. M. Olsen, M. Kryger, N. V. Voigt, T. Torring, E. Gultekin, M. Nielsen, R. Mohammad Zadegan, E. S. Andersen, M. M. Nielsen, J. Kjems, V. Birkedal and K. V. Gothelf, *Angew. Chem., Int. Ed.*, 2011, **50**, 3983–3987.
- 15 G. M. Church, Y. Gao and S. Kosuri, *Science*, 2012, **337**, 1628.

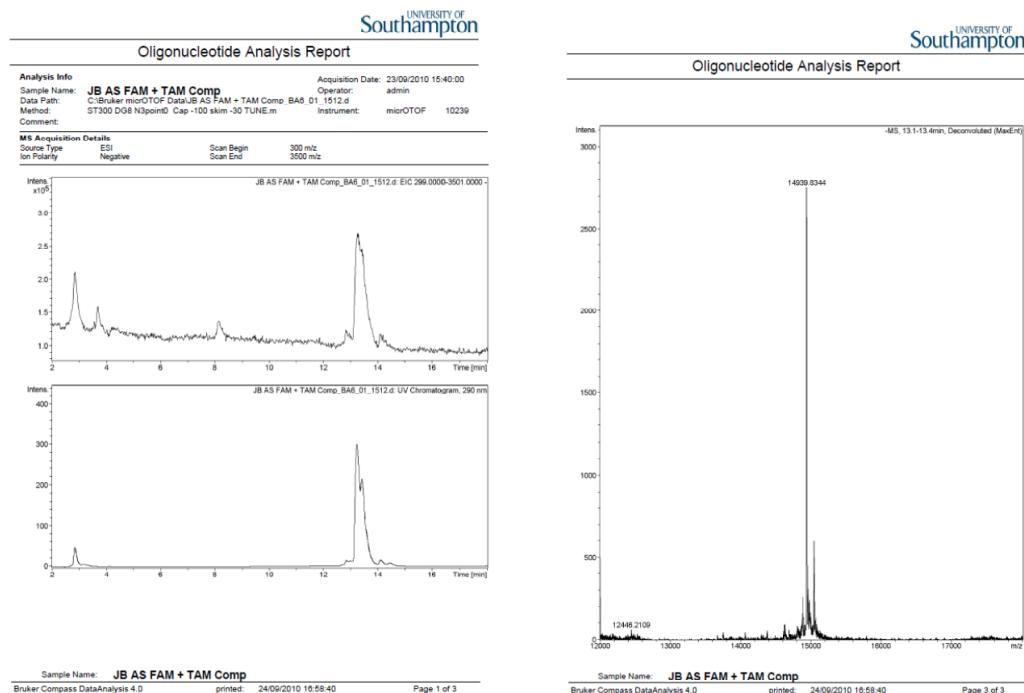
Electronic Supporting Information

## A DNA based five-state switch with programmed reversibility

Jonathan R. Burns,<sup>a</sup> Søren Preus,<sup>b</sup> Daniel G. Singleton<sup>a</sup> and Eugen Stulz\*<sup>a</sup>

### General

Solid supported DNA synthesis reagents were purchased from SAFC Proligo, DNA purification columns were purchased from Glen Research and Berry and Associates, desalting columns were purchased from GE Healthcare. DNA Melting profiles were recorded using a Varian Cary 300 Spectrophotometer and on and a Varian Eclipse fluorescence spectrometer in quartz cells with path length of 1 cm are an average of at least two denaturing-annealing cycles. The absorptions of oligonucleotide solutions were measured at 260 nm in a quartz cuvette with a path length of 1 cm. The unmodified DNA strands and the FAM/TAMRA labelled DNA was obtained from ATDBio Ltd (Southampton, UK); the FAM/TAMRA-DNA was confirmed using ESI MS (negative mode) on Bruker micrOTOF.



<sup>20</sup>  $m/z$  calcd 14940.47 Da; found 14939.83 ( $M-H^-$ ).

## Porphyrin DNA synthesis

Rigid and flexible porphyrin monomers were prepared according to published protocols.<sup>[1, 2]</sup> To generate a singularly modified zinc porphyrin on the DNA, the final acid deprotection step during solid phase DNA synthesis was omitted, after which the labelled DNA was deprotected and purified using PolyPack affinity columns by the same protocols as published previously. Porphyrin labelled DNA was purified by RP-HPLC.

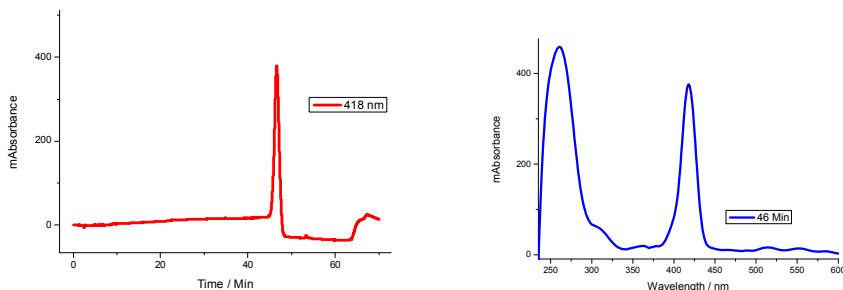


Figure S1. HPLC chromatogram of 1P, using a HFIP/MeOH gradient as described by Berova *et. al.*<sup>[3]</sup> using a Waters Xbridge OST C18 2.5  $\mu$ M 4.6 x 50 mm column (left) and UV-Vis absorption cross-section at 46 minutes<sup>10</sup> (right).

## DNA analysis

Release DNA sequences (complementary strands to **2 – 6**):

15	2a	ATTA TAAT ATTA TAAT ATTA TAAT ATTA GCA CAT TGA TTT CTC ACA CTC-5'
	3a	ATTA TAAT ATTA TAAT ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'
	4a	ATTA TAAT ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'
	5a	ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'
20	6a	CGA CAT TGA TTT CTC ACA CTC-5'

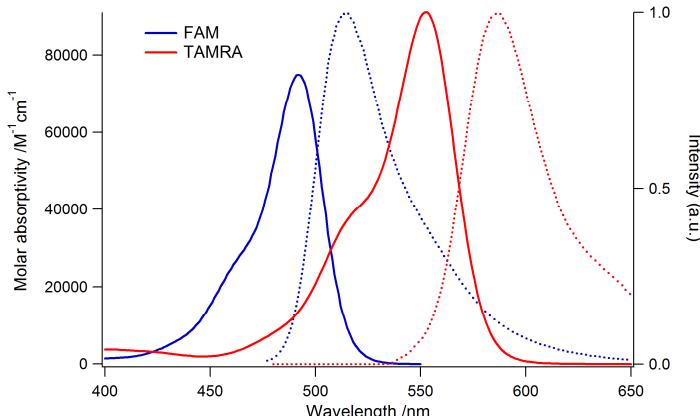
DNA was annealed using the peltier device on a Varian Cary 300 Biospectrophotometer or Varian Eclipse Spectrometer. DNA was annealed at 1 °C / min. Concentrations determined using  $\epsilon$  at 260 nm as shown in Table S1.

25 Table S1:

Strand	Sequence	$\epsilon_{260}$
1	AAAT ATTA TAAT ATTA TAAT ATTA GCT GTA ACT AAA G 3' ATTA TAAT ATTA TAAT ATTA TAAT ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'	481800 538100
2	ATTA TAAT ATTA TAAT ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'	451700
3	ATTA TAAT ATTA TAAT CGA CAT TGA TTT CTC ACA CTC-5'	365300
4	ATTA TAAT ATTA CGA CAT TGA TTT CTC ACA CTC-5'	278900
5	CGA CAT TGA TTT CTC ACA CTC-5'	191700
6	TAAT ATTA TAAT ATTA TAAT ATTA TAAT GCT GTA ACT AAA GAG TGA GAG 3'	566500
2a	TAAT ATTA TAAT ATTA TAAT ATTA GCT GTA ACT AAA GAG TGA GAG 3'	480100
3a	TAAT ATTA TAAT ATTA GCT GTA ACT AAA GAG TGA GAG 3'	393700
4a	TAAT ATTA TAAT GCT GTA ACT AAA GAG TGA GAG 3'	307300
5a	GCT GTA ACT AAA GAG TGA GAG 3'	221500
6a	AAAT ATTA TAAT ATTA TAAT ATTA GCT GTA ACT AAA G 3'	481800

### Quantum yield and overlap integral determination of the FAM and TAMRA FRET pair

The quantum yield of FAM in DNA was determined using monomeric FAM in aqueous solution as reference ( $QY = 0.95$ ) at an excitation wavelength of 495 nm. The overlap integral between the FAM emission spectrum and TAMRA absorption spectrum was calculated to be  $3.1 \times 10^{-15} \text{ M}^{-1} \text{cm}^{-1} \text{nm}^4$  yielding a critical Förster distance of  $R_0 = 57 \text{ Å}$ . The UV-Vis absorption and emission spectra are shown in Figure S2.



10 Figure S2. Molar extinction coefficient spectra of FAM and TAMRA and their normalized emission spectra.

### Quantum yield and overlap integral determination of Zn- and 2H-porphyrin DNA

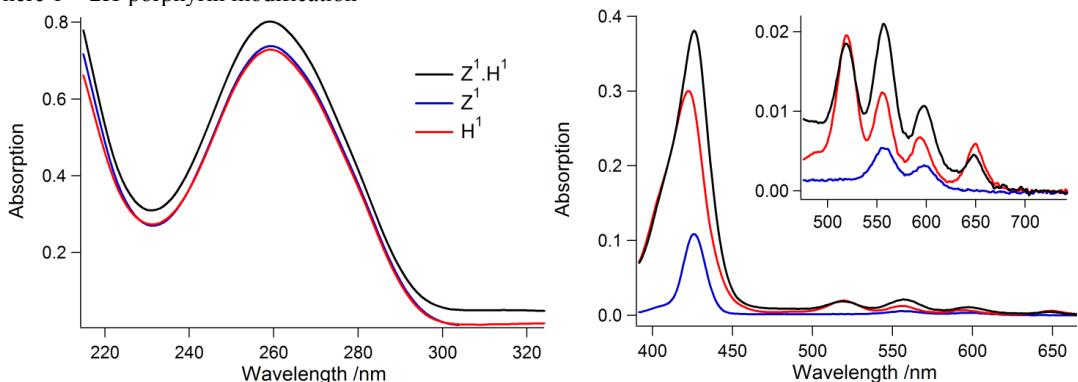
15 The quantum yield and overlap integral value for zinc- and 2H-porphyrin DNA was determined using a control 21-mer DNA sequence shown below. The UV-Vis absorption and fluorescence spectra are shown in Figure S3 and S4 respectively. The overlap between donor emission spectrum and acceptor absorption spectrum is shown in Figure S5.

20  $Z^1 = \text{Zn-porphyrin DNA} = 3' \text{ T ATT CAT ACA ACA } \underline{\text{1TG ATT TC}} 5'$

where 1 = Zn-porphyrin modification

$H^1 = \text{2H-porphyrin DNA} = 5' \text{ A TAA GTA TGT TGT AAC } \underline{\text{1AA AG}} 3'$

where 1 = 2H-porphyrin modification



25 Figure S3. UV-Vis absorption spectra of DNA region (left) and porphyrin Soret band region (right) and Q-band region expanded (inset), at 2  $\mu\text{M}$  in 0.1 M sodium phosphate buffer using a 1 cm path length cell.

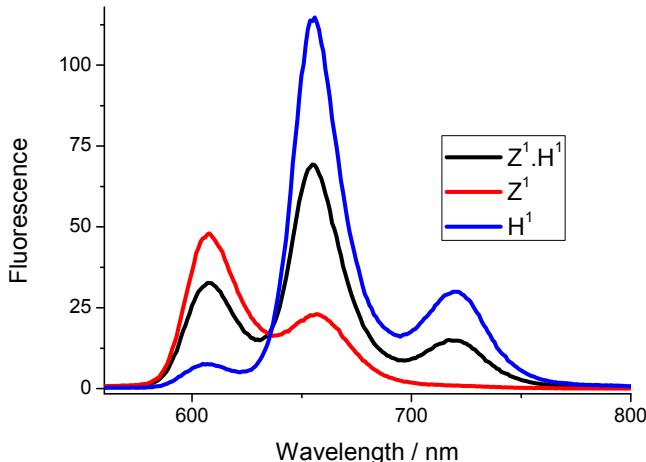


Figure S4. Fluorescence emission of Zn-porphyrin duplex DNA ( $Z^1$  red line), 2H-porphyrin duplex DNA ( $H^1$  blue line) and zinc and 2H porphyrin duplex DNA ( $Z^1.H^1$  black line), excitation at 424 nm, at 2  $\mu$ M in 0.1 M sodium phosphate buffer, 1 cm quartz cell, scanning at 600 nm per minute, 500 PMT voltage.

5

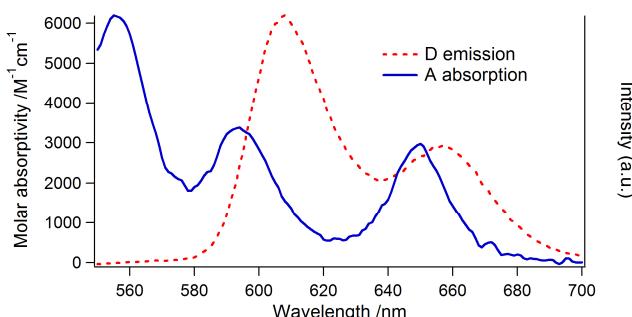


Figure S5. Overlaid Zn-porphyrin (D) emission spectrum and free base porphyrin (A) absorption spectrum.

10

### Zn/H-porphyrin DNA switch UV-Vis and Fluorescence analysis

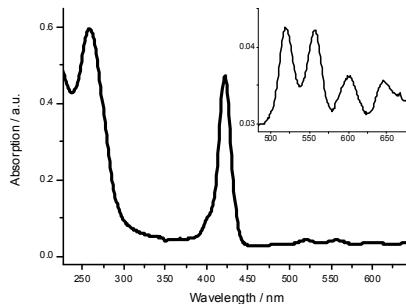


Figure S6. **1P** UV-Vis absorption spectra of DNA region and porphyrin Soret band region, and Q-band region expanded (inset), at 1  $\mu$ M in 0.1 M sodium phosphate buffer using a 1 cm path length cell.

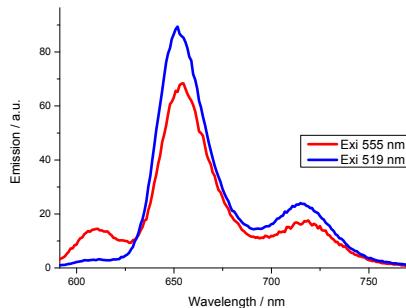


Figure S7. Fluorescence emission of **1P**, excitation at 555 nm (donor excitation, red line), and excitation at 519 nm (acceptor emission, blue line), at 2  $\mu$ M in 0.1 M sodium phosphate buffer, 1 cm quartz cell, scanning at 600 nm per minute, 600 PMT voltage.

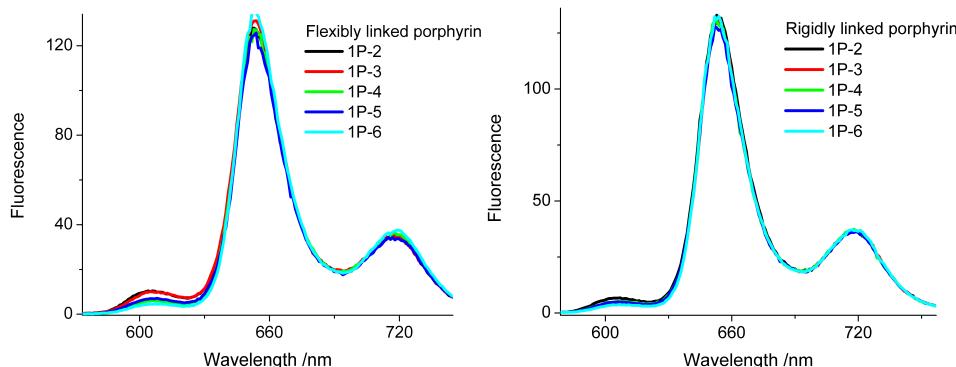


Figure S8. Emission spectra of donor+acceptor porphyrins in the DNA switch tethered via flexible (left) or rigid linkers (right) using an excitation wavelength of 426 nm. Calculated FRET efficiencies are plotted in Figure S13.

### 1F variable annealing rate UV analysis

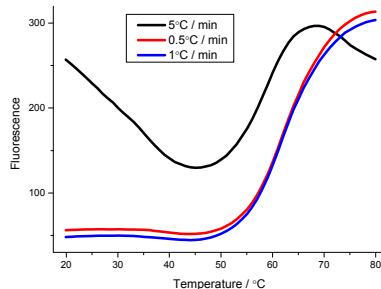


Figure S9. Fluorescence annealing profile monitoring **1F** donor emission, at 0.3  $\mu$ M in 0.1 M sodium phosphate buffer, excitation at 495 nm, emission at 510 nm, annealing at 5 °C / min (black line), 1 °C / min (blue line) and 0.5 °C / min (red line), using a 1 cm quartz cell, 500 PMT voltage.

### 1P DNA region UV melting analysis

10

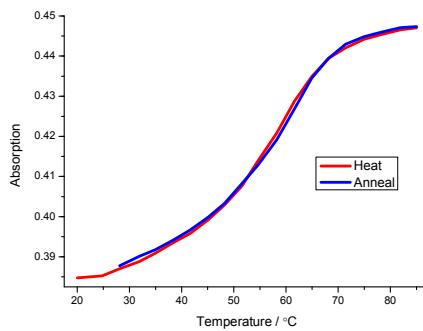


Figure S10. UV melting profile of **1P** recorded at 422 nm, at 1  $\mu$ M in 0.1 M sodium phosphate buffer, heating and annealing at 1 °C / min, 1 cm quartz cell.

## 1F Fluorescence melting analysis - acceptor emission

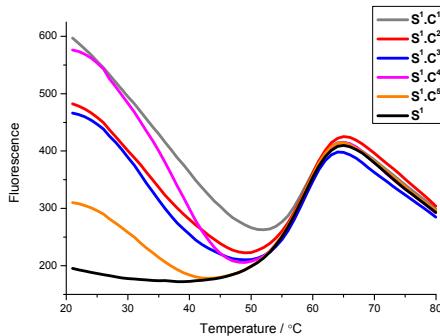


Figure S11. Fluorescence melting profile of **1F** acceptor emission combinations, at 0.3  $\mu$ M in 0.1 M sodium phosphate buffer, excitation at 495 nm, emission at 580 nm, annealing at 5  $^{\circ}$ C / min (black line), 1  $^{\circ}$ C / min (blue line) and 0.5  $^{\circ}$ C / min (red line), using a 1 cm quartz cell, 500 PMT voltage.

## Steady-state fluorescence analysis of 1F with complementary strands

10

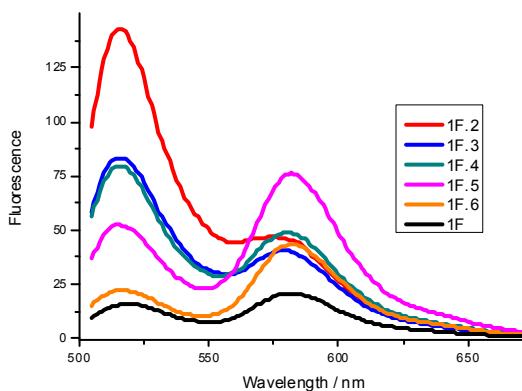


Figure S12. Steady-state fluorescence analysis of **1F** emission combinations, excitation at 495 nm, at 1  $\mu$ M in 0.1 M sodium phosphate buffer, using a 1 cm quartz cell, 500 PMT voltage. An additional reference spectrum of each sample at an excitation wavelength of 557 nm was used for the quantification of FRET efficiencies.

15

## FRET simulations

The theoretical predictions of FRET were performed using a custom build MATLAB based program described elsewhere.<sup>[4]</sup> The program simulates FRET based on an all-atom three-dimensional geometrical model of the nucleic acid constructed using individual base-pairs as building blocks and six standardized base-pair step parameters to describe each dinucleotide step in the structure (shift, slide, rise, tilt, roll and twist).<sup>[5-9]</sup> To reconstruct the structures of the adjustable strap we used the base-pair step parameters for regular B-DNA derived from the calf thymus fiber models of Arnott and coworkers<sup>[10,11]</sup> as implemented in w3DNA.<sup>[12]</sup> For the calculation of FRET we defined the base-pair coordinate frame of each of the fluorescently modified bases and implemented these in the simulation software and the geometrical model building scheme (see below). The Cartesian coordinates of both the center and direction of the transition dipole moment vectors were

equally specified within these base-pair coordinate frames allowing the calculation of the exact donor-acceptor distances of the probes positioned in the geometrical DNA models built by the program.

### Modelling the flexibly linked FAM/TAMRA probes

The definition of the local base-pair coordinate frame of the FAM/TAMRA probes attached to thymine using a hexyl linker is shown in Figure S13a and the corresponding Cartesian coordinates of transition dipole vectors are provided in Table S2. The center of the transition dipole was set in the center of the tricyclic ring and its direction pointing parallel to the long axis of the tricyclic framework. Since the hexyl linker provides diffusional mobility of the tethered probes relative to the DNA, the FAM/TAMRA dyes were positioned in their mean positions relative to the B-DNA helix as determined for the structurally similar Alexa 488 fluorophore by Sindbert *et al.* using MD simulations.<sup>[13]</sup> In this respect, the mean position is not necessarily the most probable but it is the average of all the occupiable states and was shown by Sindbert *et al.* to be a good representation of the dye position in the calculation of FRET.<sup>[13]</sup> Using this constraint for relative dipole position, the geometry of the linker and fluorophore were optimized using a force field for visual inspection only. The simulated geometries of all samples are shown in Figure S14.

It is noted that Sindbert *et al.* found that the accessible volume of the dye attached to double-stranded DNA through the hexyl linker follows the helical twist of the groove.<sup>[13]</sup> The average dye position thus has a z-displacement along the helix axis and as a result the mean pair distance for a given bp separation depends on whether the fluorophores are positioned 5'-3' or 3'-5' relative to one another. All calculated distances are shown in Table S3.

### Modelling the rigidly linked porphyrin probes

The local base-pair coordinate frame and transition dipole coordinates of the porphyrin-modified bases are shown in Figure S13b and Table S2. Since the acetylenic linker provides a rigid attachment of the probe to the base we implemented the static, AM1 optimized geometries of the porphyrin modified thymine bases. The center of the transition dipole moment was set to be in the center of the porphyrin ring.

Table S2. Cartesian coordinates of dipole centers and directions of FAM/TAMRA attached to thymine using a hexyl linker and Zn/2H-porphyrin attached to thymine through an acetylenic linker within their local base-pair coordinate frames.

	x	y	z
<i>FAM/TAMRA:</i>			
Dipole center	0.86	12.69	5.06
Dipole end	0.33	13.26	5.68
<i>Porphyrin:</i>			
Dipole center	11.94	8.37	0
Dipole end	12.87	8.75	0

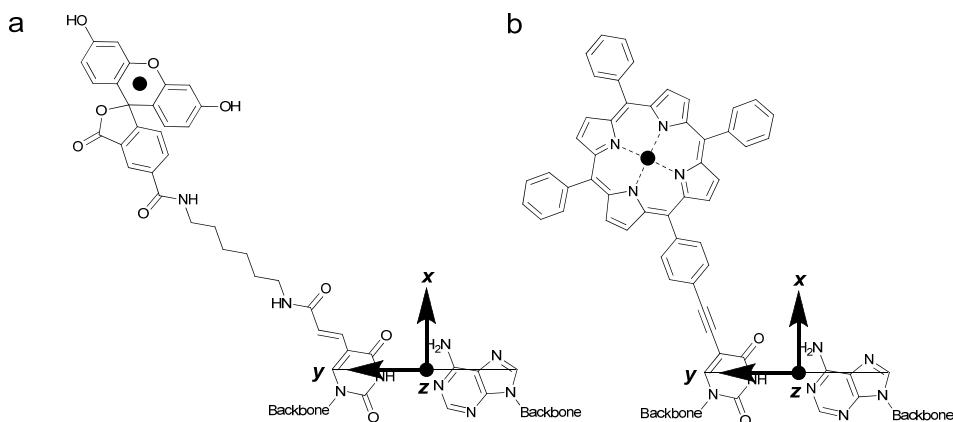


Figure S13. Definitions of local base-pair coordinate frames of the modelled dyes attached to thymine (in the Figure dyes are positioned in strand I). The origin of the coordinate frame is the mid-point of the line connecting the C8 of adenine and C6 for thymine. The *y*-axis is parallel to the C8-C6 line pointing from strand II to strand I. The *x*-axis points into the major groove while the *z*-axis completes a right-handed set.<sup>[5,9]</sup> The dipole centres are marked as black dots, and the transition dipole moments. a) FAM/TAMRA molecular framework. b) Zn/2H-porphyrin molecular framework.

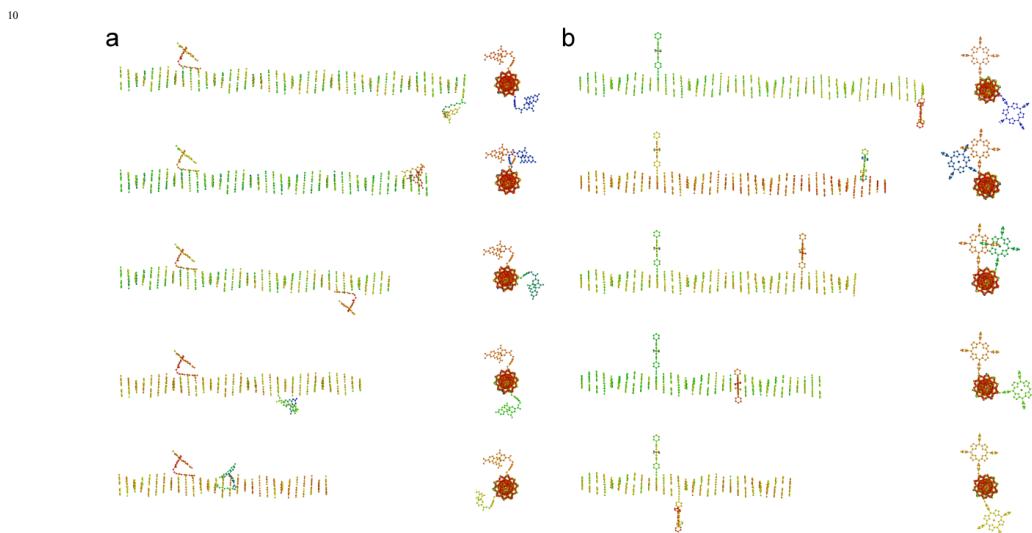


Figure S14. Side-views of simulated (ideal) three-dimensional geometries used for the theoretical predictions of FRET in the DNA switch going from sample 1.2 (top) to 1.6 (bottom). Only the spatial arrangement of the nucleobases is shown. A) FAM/TAMRA pair. B) Zn/2H-porphyrin pair. Note that Figure b shows the predicted structures of an ideally functioning switch functionalized with porphyrins, being different from what is observed experimentally (Figure S15).

Table S3. Theoretical mean-pair distances,  $R_{\text{mp}}$ ,<sup>[13]</sup> of donor-acceptor pairs in the five states of an ideal DNA switch. Samples 1F.2, 1F.3, 1F.4, 1F.5, and 1F.6 denote the FAM/TAMRA system, while 1P.2, 1P.3, 1P.4, 1P.5, and 1P.6 denote the Zn/2H porphyrin system.

Sample	$R_{\text{mp}} / \text{\AA}$	Sample	$R_{\text{mp}} / \text{\AA}$
1F.2	117	1P.2	118
1F.3	103	1P.3	92.5
1F.4	78.9	1P.4	65.5
1F.5	54.3	1P.5	44.7
1F.6	27.3	1P.6	31.9

5

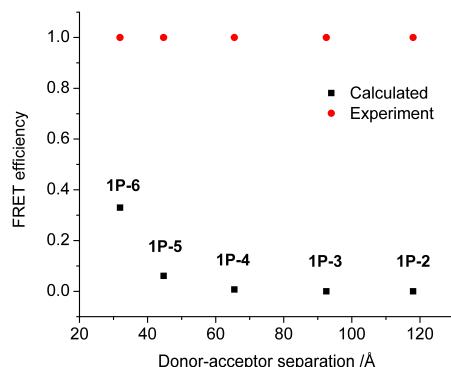


Figure S15. Measured and predicted FRET efficiencies of the DNA switch functionalized with porphyrins.  
Predicted FRET efficiencies were calculated using the distances shown in Table S3 with  $R_0 = 28.4 \text{ \AA}$ .

10

### Switching experiments with the 1F system

15

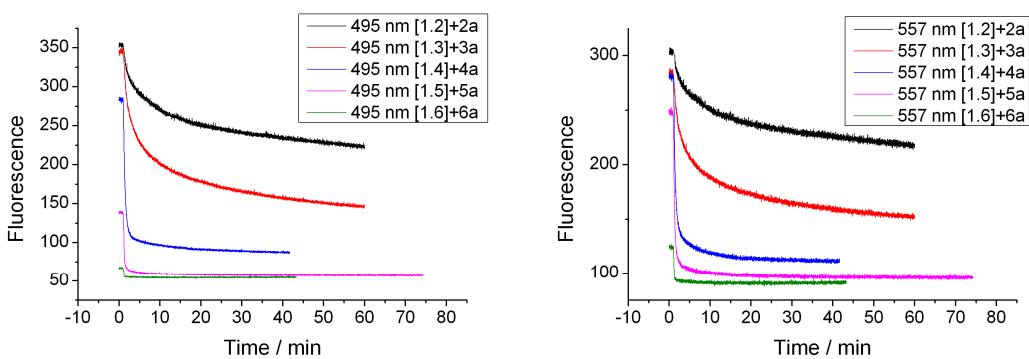


Figure S16. Release of the stabilising DNA strands (**2 – 6**) by addition of the complementary strands (**2a – 6a**; strand sequences see Table S1), monitoring the donor emission at 510 nm when exciting at 495 nm (left) and the acceptor emission at 580 nm when exciting at 557 nm (right). **1** at 0.5  $\mu\text{M}$ , stabilising strands and release strands added in 1.1 molar excess.

20

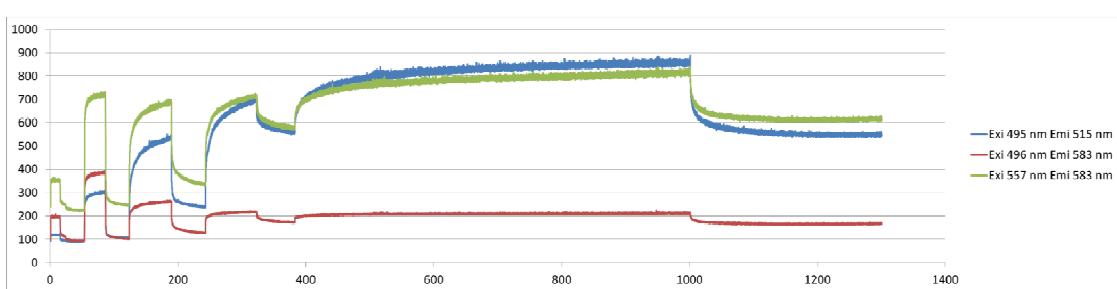


Figure S17. Switching of the **1F** system. Addition of DNA strands: see Table S4.

Table S4. Addition of stabilising and switching strands

Time (min)	Strand Added
0 to 16	1
1 to 16	6
16 to 54	6a / 2
54 to 86	2a / 6
86 to 123	6a / 3
123 to 189	3a / 6
189 to 243	6a / 4
243 to 323	4a / 6
323 to 377	6a / 5
377 to 1000	5a / 6
1000 to 1300	6a / 5

## References:

1. L. A. Fendt, I. Bouamaied, S. Thöni, N. Amiot and E. Stutz, *J. Am. Chem. Soc.*, 2007, 129, 15319-15329.
2. A. Brewer, G. Siligardi, C. Neylon and E. Stutz, *Org. Biomol. Chem.*, 2011, 9, 777-782.
3. A. Mammana, T. Asakawa, K. Bitsch-Jensen, A. Wolfe, S. Chaturantabut, Y. Otani, X. X. Li, Z. M. Li, K. Nakanishi, M. Balaz, G. A. Ellestad and N. Berova, *Bioorg. Med. Chem.*, 2008, 16, 6544-6551.
4. S. Preus, K. Kilså, F. A. Miannay, B. Albinsson and L. M. Wilhelmsson, *Nucl. Acids Res.* (2012) doi: 10.1093/nar/gks856.
5. M. A. El Hassan and C. R. Calladine, *J Mol Biol*, 1995, 251, 648-664.
6. X. J. Lu, M. A. El Hassan and C. A. Hunter, *J Mol Biol*, 1997, 273, 681-691.
7. X. J. Lu, M. A. El Hassan and C. A. Hunter, *J Mol Biol*, 1997, 273, 668-680.
8. X. J. Lu and W. K. Olson, *Nucleic Acids Res.*, 2003, 31, 5108-5121.
9. W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakkeb, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger and H. M. Berman, *J Mol Biol*, 2001, 313, 229-237.
10. S. Arnott and D. W. Hukins, *J Mol Biol*, 1973, 81, 93-105.
11. S. Arnott and D. W. L. Hukins, *Biochem Biophys Res Co*, 1972, 47, 1504-&.
12. G. Zheng, X. J. Lu and W. K. Olson, *Nucleic Acids Res.*, 2009, 37, W240-246.
13. S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Muller and C. A. Seidel, *J Am Chem Soc*, 2011, 133, 2463-2480.

## **Paper V**

*In which we gain structural and electronic insight into  
the tC probes...*



# Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family<sup>†</sup>

Søren Preus,<sup>\*a</sup> Kristine Kilså,<sup>a</sup> L. Marcus Wilhelmsson<sup>b</sup> and Bo Albinsson<sup>\*b</sup>

Received 11th January 2010, Accepted 14th May 2010

DOI: 10.1039/c000625d

Fundamental insight into the unique fluorescence and nucleobase-mimicking properties of the fluorescent nucleobase analogues of the tC family is not only vital in explaining the behaviour of these probes in nucleic acid environments, but will also be profitable in the development of new and improved fluorescent base analogues. Here, temperature-dependent fluorescence quantum yield measurements are used to successfully separate and quantify the temperature-dependent and temperature-independent non-radiative excited-state decay processes of the three nucleobase analogues tC, tC<sup>O</sup> and tC<sub>nitro</sub>; all of which are derivatives of a phenothiazine or phenoxazine tricyclic framework. These results strongly suggest that the non-radiative decay process dominating the fast deactivation of tC<sub>nitro</sub> is an internal conversion of a different origin than the decay pathways of tC and tC<sup>O</sup>. tC<sub>nitro</sub> is reported to be fluorescent only in less dipolar solvents at room temperature, which is explained by an increase in excited-state dipole moment along the main non-radiative decay pathway, a suggestion that applies in the photophysical discussion of large polycyclic nitroaromatics in general. New insight into the ground and excited-state potential energy surfaces of the isolated tC bases is obtained by means of high level DFT and TDDFT calculations. The S<sub>0</sub> potential energy surfaces of tC and tC<sub>nitro</sub> possess two global minima corresponding to geometries folded along the middle sulfur–nitrogen axis separated by an energy barrier of 0.05 eV as calculated at the B3LYP/6-311+G(2d,p) level. The ground-state potential energy surface of tC<sup>O</sup> is also predicted to be shallow along the bending coordinate but with an equilibrium geometry corresponding to the planar conformation of the tricyclic framework, which may explain some of the dissimilar properties of tC and tC<sup>O</sup> in various confined (biological) environments. The S<sub>1</sub> equilibrium geometries of all three base analogues are predicted to be planar. These results are discussed in the context of the tC bases positioned in double-stranded DNA scenarios.

## Introduction

The selection of synthetic fluorescent nucleobase analogues has grown considerably in recent years<sup>1–4</sup> with applications in areas ranging from quencher-free molecular beacons<sup>5</sup> and single-nucleotide polymorphism typing<sup>6–7</sup> to monitoring nucleic acid dynamics<sup>8–14</sup> and DNA–protein activity.<sup>15–26</sup> The ability to incorporate the fluorophore into nucleic acids as a replacement for one of the canonical bases offers a great advantage in the ability to position the reporter at a well-defined position close to or in the very site of interest. However, for by far the most fluorophores the base-stacking environment provided by double-stranded DNA introduces efficient non-radiative deactivation processes from the electronically excited states, such as charge transfer to neighbouring bases<sup>27,28</sup> and base collisions,<sup>29</sup> often

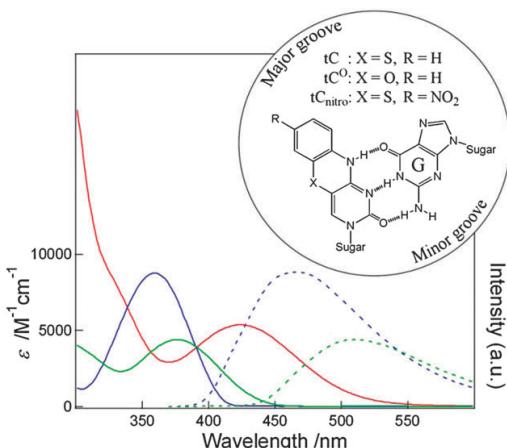
combined with a weaker oscillator strength of the lowest energy electronic transition.<sup>28,30</sup> As a result, most fluorescent base analogues are highly or partly quenched in double-stranded DNA, a property accompanied by complex intensity decays greatly varying with neighbouring bases.<sup>31–42</sup> Whereas these features are useful for reporting on the local micro-environment of DNA, the low and variable fluorescence quantum yields are unsuitable features for fluorescence anisotropy and fluorescence resonance energy transfer (FRET) studies.<sup>43</sup> Equally important, these properties limit the use of fluorescent nucleobase analogues as labels in biotechnological applications, such as real-time PCR,<sup>4</sup> and in single-molecule fluorescence studies<sup>44</sup> which could otherwise benefit from bright and photostable intrinsic reporters.

The nucleobase analogues of the tricyclic cytosine family, tC, tC<sup>O</sup> and tC<sub>nitro</sub>, constitute rare exceptions to these general characteristics (Fig. 1). UV-melting and circular dichroism measurements have shown that these analogues base-pair selectively with guanine and stabilize the B-DNA double helix compared to natural cytosine.<sup>45–47</sup> The chemical structures of tC and the nitro-substituted tC<sub>nitro</sub> are built upon a phenothiazine tricyclic framework, while tC<sup>O</sup> is derived from a phenoxazine tricyclic framework. Despite their structural similarities, the fluorescence properties of the tC bases in their monomeric

<sup>a</sup> Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen, Denmark. E-mail: spreus@nano.ku.dk

<sup>b</sup> Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden. E-mail: balb@chalmers.se; Fax: +46 317723858; Tel: +46 317723044

<sup>†</sup> Electronic supplementary information (ESI) available: Temperature dependent UV-vis absorption and emission spectra of the studied compounds. See DOI: 10.1039/c000625d



**Fig. 1** UV-vis absorption (full-drawn) and fluorescence spectra (dashed) of the monomeric forms of tC (green), tC<sup>O</sup> (blue) and tC<sub>nitro</sub> (red) in H<sub>2</sub>O. Inset: chemical structures of the tricyclic cytosine analogues in their base-pairing environment with guanine. Also shown is the direction of major and minor groove when looking down the long axis of double-stranded DNA.

forms and when incorporated into DNA vary somewhat.<sup>47–50</sup> The free nucleoside of tC has a fluorescence quantum yield of 0.13 in H<sub>2</sub>O at room temperature but increases to an average value of 0.2 in double-stranded DNA due to a slower non-radiative deactivation, relatively independent of neighbouring bases.<sup>48</sup> In contrast, the fluorescence quantum yield of tC<sup>O</sup> is 0.30 in H<sub>2</sub>O at room temperature but decreases to an average value of 0.22 in double-stranded DNA, slightly dependent on neighbouring bases, due to a lower fluorescence rate constant.<sup>47</sup> The overall high fluorescence quantum yield, combined with a molar absorptivity of the lowest energy absorption band of  $\epsilon_{\text{max}} = 9000 \text{ M}^{-1} \text{ cm}^{-1}$ , currently makes tC<sup>O</sup> on average the brightest fluorescent nucleobase analogue inside the DNA double-helix. The fluorescence decays of tC and tC<sup>O</sup> are single exponentials in double-stranded DNA which is a very convenient feature in particularly FRET measurements and strongly suggests that the tC bases have a relatively rigid and well-defined orientation inside the DNA helix.<sup>47,48</sup>

As opposed to the strong fluorescence of tC and tC<sup>O</sup>, the nitro-substituted tC<sub>nitro</sub> is virtually non-fluorescent in polar solvents at room temperature.<sup>50</sup> However, due to a low-lying intramolecular charge-transfer (CT) state the lowest energy electronic transition of tC<sub>nitro</sub> is red-shifted compared to tC and tC<sup>O</sup> (Fig. 1) which makes it useful as a FRET acceptor with tC or tC<sup>O</sup> serving as donor.<sup>45,50</sup> As demonstrated in a recent study by our groups, the well-defined position and orientation of the tC bases inside double-stranded DNA facilitates a very high control of the orientation factor in the FRET efficiency.<sup>45</sup> This attractive feature has resulted in an additional desire to expand the nucleic acid toolbox with new and improved base analogue FRET-pair combinations which again requires fundamental knowledge about the photophysical properties of these probes.

Here we provide new insight into the fluorescence properties of the isolated tC bases and their ground and excited-state structures by means of density functional theory (DFT) calculations and steady-state UV-vis absorption and fluorescence spectroscopy. The excited-state decay processes of tC, tC<sup>O</sup> and tC<sub>nitro</sub> are successfully separated into temperature-dependent and temperature-independent terms and quantified using temperature-dependent fluorescence quantum yield measurements. DFT and time-dependent DFT (TDDFT) calculations are applied to gain insight into the ground- and excited-state geometries of the tC bases which we believe are directly related to their properties in confined biological environments, such as in double-stranded nucleic acid scenarios and in protein binding pockets. Translation of the calculated potential energy surfaces into the properties of the tC bases in DNA is therefore discussed.

## Methodological section

### Chemicals

Acetonitrile, tetrahydrofuran (THF) and dioxane were of spectrophotometric grade as purchased from Sigma-Aldrich and used without further purification. 2-Methyltetrahydrofuran (MeTHF), also purchased from Sigma-Aldrich, was distilled prior use. Propylene glycol (PG) was obtained from Merck. The synthesis of the nucleosides of tC, tC<sup>O</sup> and tC<sub>nitro</sub> have previously been described<sup>45,47,51</sup> (also commercially available from Glen Research).

### UV-vis absorption and steady-state fluorescence

UV-vis absorption spectra were recorded on a Varian Cary 4000 spectrophotometer in 1 cm quartz cuvettes using pure solvent as baseline. Fluorescence spectra were recorded on a Spex Fluorolog 3 spectrophotofluorimeter (JY Horiba). The temperature-dependent fluorescence measurements were performed using an Oxford optistatDN cryostat and, at each temperature, measured after stabilization of the intensity (20 to 30 minutes). The temperature-dependent fluorescence quantum yields of the methylester<sup>49</sup> of tC in MeTHF and the nucleoside of tC<sub>nitro</sub> in PG glass, THF and 1,4-dioxane were measured at an excitation wavelength of 370 nm using quinine sulfate in H<sub>2</sub>SO<sub>4</sub> as reference ( $\Phi_f = 0.55$ ).<sup>52</sup> The fluorescence quantum yield of the nucleoside of tC<sup>O</sup> (tC<sup>O</sup>-nuc) in MeTHF was measured relative to the tC<sup>O</sup>-nuc in H<sub>2</sub>O ( $\Phi_f = 0.30$ )<sup>47</sup> at an excitation wavelength of 361 nm. Absorbances were kept below 0.05 to ensure linear response. Temperature-dependent quantum yields were calculated assuming temperature-independent oscillator strengths using the absorbance value determined at  $T = 295 \text{ K}$  for tC and tC<sup>O</sup> or at  $T = 184 \text{ K}$  for tC<sub>nitro</sub> (see spectra and discussion in ESI†). A temperature-correction was made for both the change in refractive index and molar volume of MeTHF (ESI†).

### Temperature-dependent fluorescence

It is often possible to separate the non-radiative decay rate constant of excited molecules into a temperature-dependent and a temperature-independent term. If the temperature-dependent non-radiative decay rate constant is assumed

to follow an Arrhenius-type dependency the fluorescence quantum yield is given by<sup>53</sup>

$$\Phi_f = \frac{k_f}{k_f + k_0 + A \times \exp\left(-\frac{E_a}{RT}\right)} \quad (1)$$

where  $k_f$  is the fluorescence rate constant,  $k_0$  is the temperature-independent non-radiative decay rate constant,  $A$  is the frequency factor,  $E_a$  is the activation energy of the temperature-dependent non-radiative decay process,  $R$  is the gas constant and  $T$  is the temperature. Denoting the fluorescence quantum yield as  $T$  approaches 0 K by  $\Phi_{f,0}$ , eqn (1) may be rewritten into

$$\ln\left(\frac{1}{\Phi_f} - \frac{1}{\Phi_{f,0}}\right) = -\frac{E_a}{R} T^{-1} + \ln\left(\frac{A}{k_f}\right) \quad (2)$$

which provides a means to evaluate  $E_a$  by measuring  $\Phi_f$  as a function of  $T$ .

### Quantum chemical calculations

All DFT geometry optimizations, including transition-state (TS) optimizations, were performed in the ground-state of the molecule using the B3LYP functional<sup>54–56</sup> as implemented in the Gaussian03 program package.<sup>57</sup> In the TS optimizations, the normal coordinate associated with the resulting (single) imaginary frequency was animated using GaussView. Electronic excitations were calculated using TDDFT<sup>58,59</sup> B3LYP/6-311+G(2d) as implemented in Gaussian03. Solvation effects were mimicked, where appropriate, by applying a continuum solvation shell (the CPCM model)<sup>60,61</sup> in the TDDFT calculations. The amount of HOMO → LUMO character of the lowest energy electronic transition of the investigated compounds was determined from the calculated CI-coefficients. Restricted Hartree–Fock (RHF) wavefunctions were used in all calculations.

The single-point molecular coordinates for the calculation of the potential energy surfaces were obtained from a combination of intrinsic reaction coordinate (IRC) calculations and geometry optimizations (both ground-state) starting from a TS optimized geometry.<sup>62</sup> For the IRC along the bending of tC and tC<sub>nitro</sub> the TS was the planar geometry. For the IRC following the NO<sub>2</sub> rotation of tC<sub>nitro</sub> the TS was the geometry having a NO<sub>2</sub> dihedral angle of 90° relative to the aromatic plane. For each of the isolated molecular geometries on these IRCS the singlet excited-state energies,  $E_s$ , were then determined from a TDDFT electronic excitation energy calculation (vertical excitations) as

$$E_{s_x} = E_{GS} + E_{exc} \quad (3)$$

where  $E_{GS}$  denotes the ground-state HF energy and  $E_{exc}$  is the excitation energy. The 2D PES of tC<sub>nitro</sub> was reconstructed from 68 TDDFT single point calculations which, after symmetry considerations, yielded 272 coordinates on the calculated 2D potential energy surfaces. The coordinates on the PES of tC<sup>O</sup> were extracted from a DFT B3LYP/6-311+G(2d) geometry optimization starting from the AM1 optimized, bent geometry and ending in the planar geometry of tC<sup>O</sup>.

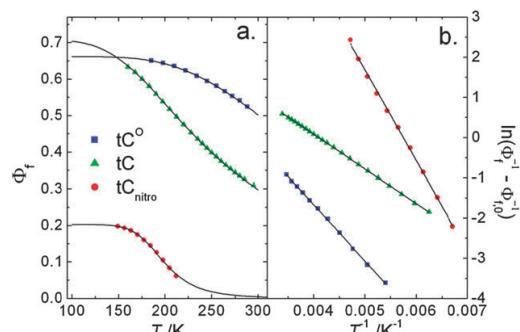
## Results

### Temperature-dependent fluorescence measurements

To quantify the deactivation pathways of the tC bases the fluorescence quantum yields were measured as a function of temperature and subsequently fitted to eqn (1) and (2) (Fig. 2a and b, respectively). Quantum yield measurements of tC and tC<sup>O</sup> were performed in liquid MeTHF (melting point  $T_G = 137$  K) from 150 K to 300 K while the quantum yield of tC<sub>nitro</sub> was monitored in PG glass ( $T_G = 214$  K) from 145 K to 210 K to exclude effects resulting from changes in solvent viscosity expected to influence the excited-state dynamics of tC<sub>nitro</sub> (*vide infra*). As the temperature is lowered the fluorescence intensity of all three fluorophores increases. The emission of tC<sub>nitro</sub> in PG glass is unstructured and centred at  $\lambda_{max} = 550$  nm (Fig. S1, ESI†). The emission of tC in MeTHF is centred at 475 nm and becomes slightly fine-structured upon lowering the temperature (Fig. S2, ESI†). The emission spectrum of tC<sup>O</sup>, in turn, displays very fine-structured vibrational modes both at RT and upon lowering the temperature (Fig. S3, ESI†).

Since the fluorescence quantum yields of tC and tC<sup>O</sup> only change slightly within the temperature interval possible to monitor in MeTHF, and in particular do not reach plateaus of constant  $\Phi_f$  at low temperatures, some considerations were needed in order to obtain the best parameters from the data. First, the values of  $\Phi_{f,0}$  were determined which resulted in proportionality between  $\ln(\Phi_f^{-1} - \Phi_{f,0}^{-1})$  and  $T^{-1}$ . The value of  $E_a$  determined from the slope of the corresponding plot was then constrained in a subsequent fit of  $\Phi_f$  to eqn (1). Fitting of  $\Phi_f$  of tC and tC<sup>O</sup> was performed by additionally constraining the values of  $k_f$  to the ones measured in H<sub>2</sub>O at room temperature.<sup>47,48</sup>

The parameters obtained in the temperature-dependent fluorescence measurements are provided in Table 1. At low temperatures the fluorescence quantum yields are found to approach maximum values of  $\Phi_{f,0} = 0.71$ , 0.66 and 0.20 for tC, tC<sup>O</sup> and tC<sub>nitro</sub>, respectively, which directly relates to the



**Fig. 2** Temperature dependence of the fluorescence quantum yield of tC, tC<sup>O</sup> and tC<sub>nitro</sub>. Measurements were performed in MeTHF for tC and tC<sup>O</sup>, and in PG glass for tC<sub>nitro</sub>. (a) Solid lines are plots of eqn (1) with the parameters shown in Table 1. (b) Solid lines are plots of the experimental data using eqn (2) with the parameters given in Table 1. Absorption and emission spectra are provided in ESI.†

**Table 1** First excited-state decay parameters of tC and tC<sup>O</sup> in MeTHF and tC<sub>nitro</sub> in PG glass

	$k_f/\text{s}^{-1}$	$k_0/\text{s}^{-1}$	$k_{\text{nr},295\text{ K}}^a/\text{s}^{-1}$	$A/\text{s}^{-1}$	$E_a/\text{eV}$	$\Phi_f^b$
tC	$4.1 \times 10^7$	$1.7 \times 10^7$	$7.6 \times 10^7$	$1.4 \times 10^9$	0.074	0.3
tC <sup>O</sup>	$8.8 \times 10^7$	$4.5 \times 10^7$	$3.9 \times 10^7$	$4.5 \times 10^9$	0.12	0.5
tC <sub>nitro</sub>	$5.4 \times 10^7$	$2.1 \times 10^8$	$1.1 \times 10^{10}$	$2.4 \times 10^{13}$	0.20	0.0

<sup>a</sup> Temperature-dependent non-radiative decay rate constant at  $T = 295\text{ K}$  calculated as  $k_{\text{nr}} = A \times \exp[-E_a/(RT)]$ . <sup>b</sup> Fluorescence quantum yield at  $T = 295\text{ K}$ .

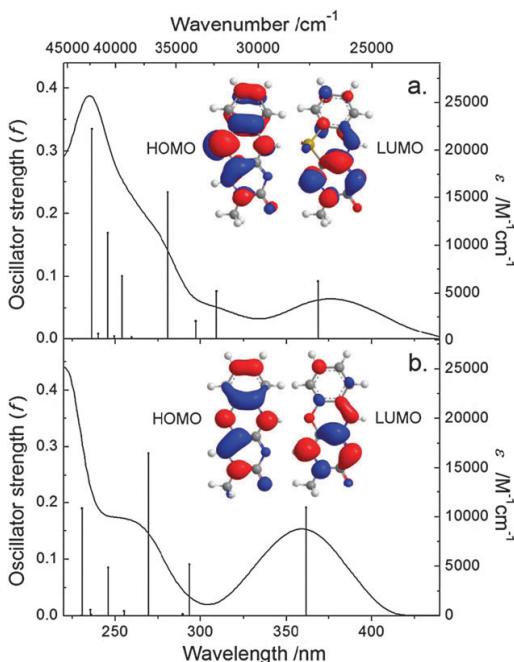
relative ratio between  $k_f$  and  $k_0$ . The fluorescence rate constants of tC, tC<sup>O</sup> and tC<sub>nitro</sub> are  $k_f = 4.1 \times 10^7\text{ s}^{-1}$ ,  $8.8 \times 10^7\text{ s}^{-1}$  and  $5.4 \times 10^7\text{ s}^{-1}$ , respectively, thus reflecting the relative magnitude of the lowest energy absorption bands<sup>45,47,49</sup> as predicted by the Strickler–Berg relation.<sup>63</sup> The temperature-independent non-radiative decay rate constants of the three investigated compounds are  $k_0 = 1.7 \times 10^7\text{ s}^{-1}$ ,  $4.5 \times 10^7\text{ s}^{-1}$  and  $2.1 \times 10^8\text{ s}^{-1}$  for tC, tC<sup>O</sup> and tC<sub>nitro</sub>, respectively. As evidenced by the combination of excellent fits to both eqn (1) and (2) in Fig. 2, the temperature-dependent non-radiative excited-state decay rate constants of all three compounds are well represented by the Arrhenius expression with activation energies of  $E_a = 0.074\text{ eV}$ ,  $0.12\text{ eV}$  and  $0.20\text{ eV}$  for tC, tC<sup>O</sup> and tC<sub>nitro</sub>, respectively. The corresponding frequency factors are  $A = 1.4 \times 10^9\text{ s}^{-1}$ ,  $4.5 \times 10^9\text{ s}^{-1}$  and  $2.4 \times 10^{13}\text{ s}^{-1}$ , which results in temperature-dependent non-radiative rate constants at room temperature ( $T = 295\text{ K}$ ) of  $k_{\text{nr},295\text{K}} = 7.6 \times 10^7\text{ s}^{-1}$ ,  $3.9 \times 10^7\text{ s}^{-1}$  and  $1.1 \times 10^{10}\text{ s}^{-1}$  for tC, tC<sup>O</sup> and tC<sub>nitro</sub>, respectively.

It is important to note that the exact values of the rate constants, activation energies and frequency factors reported here only apply for the fluorophores in the solvents in which these parameters were measured, as evidenced by the varying fluorescence quantum yields of each of the three compounds in different solvents. However, the values obtained here provide a qualitative general insight of the decay processes of the tC bases. No phosphorescence was observed from either of the investigated compounds at low temperatures in the solvents used.

### Calculated electronic spectra of tC and tC<sup>O</sup>

The electronic spectra of tC and tC<sup>O</sup> were previously calculated for the AM1 optimized geometries using the semi-empirical ZINDO/S model,<sup>47,49</sup> while TDDFT calculations were recently used to predict the electronic excitations of tC<sub>nitro</sub>.<sup>50</sup> In order to gain more quantitative insight into the electronic states of all the tC bases (*vide infra*), the level of theory is enhanced in the calculations of tC and tC<sup>O</sup> here, and the resulting spectra constitute in both cases significant improvements in the accuracy of the predicted excitation energies and oscillator strengths (Fig. 3). TDDFT with the B3LYP functional was chosen over more exact wave-function based methods due to the relatively large size of the investigated compounds.

Fig. 3a shows the results of the TDDFT B3LYP/6-311+G(2d) (+CPCM for H<sub>2</sub>O) calculated electronic spectrum of the B3LYP/6-31G(d,p) optimized geometry of tC. In general, the spectral shape of the calculated spectrum of



**Fig. 3** Ten lowest energy electronic transitions of (a) tC and (b) tC<sup>O</sup> compared to the experimental UV-vis absorption spectra in H<sub>2</sub>O (full-drawn). Frontier KS orbitals are shown as inserts. Calculations were performed on the B3LYP/6-31G(d,p) optimized geometries using TDDFT B3LYP/6-311+G(2d) with a CPCM solvation model for H<sub>2</sub>O.

tC using TDDFT agrees very well with the experimentally determined spectrum in H<sub>2</sub>O (full-drawn line), both in terms of excitation energies and intensities. The spectra calculated on each of the two bent geometries of tC (*vide infra*) were identical as expected due to mirror symmetry (data not shown). The TDDFT calculations confirm that the lowest energy absorption band of tC is due to a single electronic transition, as fluorescence anisotropy and magnetic circular dichroism measurements previously have suggested.<sup>49</sup> The predicted excitation energy of the S<sub>0</sub> → S<sub>1</sub> transition (3.36 eV) has an error of 0.05 eV compared to the peak of the absorption band in H<sub>2</sub>O (3.31 eV) and the calculated oscillator strength of 0.092 is almost identical to the experimentally determined value of 0.095. The lowest transition of tC has 87% HOMO → LUMO character with the two frontier Kohn–Sham (KS) orbitals characterized by a considerable spatial overlap (Fig. 3a insert). The KS orbitals optimized with and without the CPCM solvation model showed no significant differences (data not shown).

The TDDFT B3LYP/6-311+G(2d) (+CPCM for H<sub>2</sub>O) calculated electronic spectrum of the B3LYP/6-31G(d,p) optimized geometry of tC<sup>O</sup> is shown in Fig. 3b. The overall appearance of the calculated electronic spectrum of tC<sup>O</sup> agrees very well with the UV-vis absorption spectrum in H<sub>2</sub>O. As for tC, the calculations confirm previous experiments<sup>47</sup> showing that the lowest energy absorption band of tC<sup>O</sup> is the result of a

single electronic transition. The predicted  $S_0 \rightarrow S_1$  excitation energy of 3.43 eV is within 0.01 eV of the experimentally determined value in  $\text{H}_2\text{O}$ . The lowest energy electronic transition has 86% HOMO  $\rightarrow$  LUMO character and these frontier KS orbitals share a relatively high degree of spatial overlap (Fig. 3b inset). Again, no significant difference was observed between the KS orbitals optimized with and without the CPCM solvation model (data not shown).

## Molecular geometries and potential energy surfaces

Due to the strict conditions set by the DNA double helix on the nucleobase analogues in terms of H-bonding, base-stacking and steric hindrances, the excellent nucleobase-mimicking properties of the tC bases are related to their molecular geometries. Previous geometry optimizations of tC and tC<sup>O</sup> were performed using the semi-empirical AM1 model.<sup>47,49</sup> In order to gain more accurate and quantitative information of their molecular geometries, DFT and TDDFT calculations were performed herein.

Two local energy minima on the potential energy surface (PES) of tC were identified from a B3LYP/6-31G(d,p) conformational search while only one (the global) minimum was found for tC<sup>O</sup>. Calculations of the vibrational spectra confirmed that the optimized structures correspond to minima on the potential energy surface. As shown in Fig. 4a the two

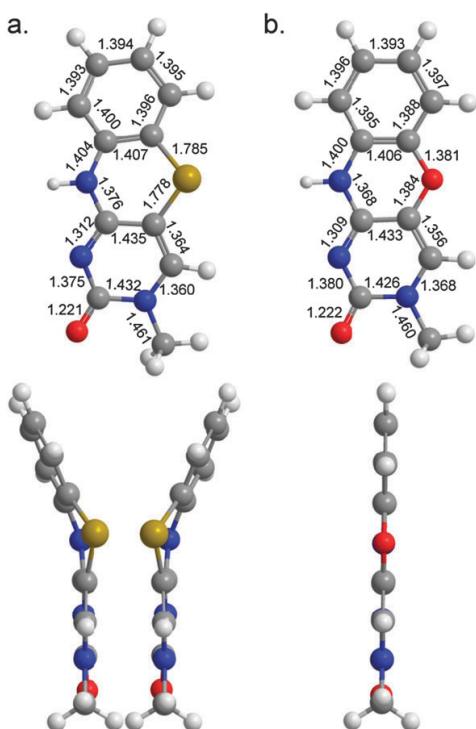
local minima found for tC are mirror images corresponding to two geometries folded  $\sim 25^\circ$  along the middle sulfur–nitrogen axis, identical to the ones found for tC<sub>nitro</sub> at the same level of theory.<sup>50</sup> The ground-state geometry of tC obtained by AM1 calculations is also the folded conformation,<sup>49</sup> and this result is supported by the X-ray structure of the parent compound phenothiazine.<sup>64,65</sup>

While AM1 optimizations of tC<sup>O</sup> show two energy minima on the PES, corresponding to the folded structures similar to those optimized for tC and tC<sub>nitro</sub>,<sup>47</sup> the only minimum on the PES of tC<sup>O</sup> identified by the B3LYP/6-31G(d,p) calculations corresponds to a planar structure of the tricyclic framework (Fig. 4b). For the planar, DFT optimized structure of tC<sup>O</sup> the IR frequency corresponding to the bending along the middle oxygen–nitrogen axis is small but real ( $\nu_{\text{bent}} = 33 \text{ cm}^{-1}$ ). At the AM1 level, on the other hand, the IR frequency of the same vibrational mode is  $34 \text{ cm}^{-1}$  and  $27i \text{ cm}^{-1}$  for the bent and planar geometries, respectively (the planar conformation of tC<sup>O</sup> was obtained by an AM1 TS optimization). The crystal structure of the tricyclic core of a tC<sup>O</sup> derivative was previously shown also to be planar.<sup>66,67</sup>

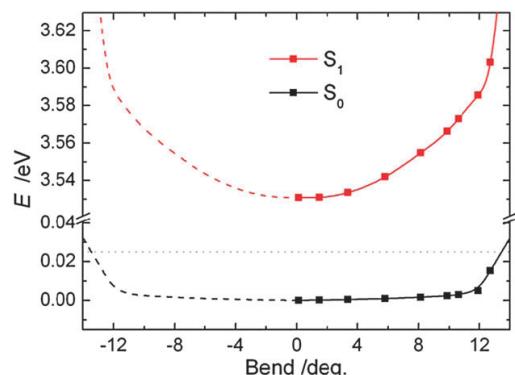
The energy barrier between the two local energy minima on the potential energy surfaces of tC and tC<sub>nitro</sub> was calculated at the DFT B3LYP/6-311G+(2d,p) level to be  $E_b = 0.056$  eV and 0.049 eV, respectively, with no improvement obtained using a larger basis set. These values were calculated as  $E_b = E_{TS} - E_{\min}$ , where the transition state energy,  $E_{TS}$ , was obtained from a TS optimization on the intrinsic reaction coordinate (IRC) between the two bent conformations and the equilibrium energy,  $E_{\min}$ , was the energy of the geometry optimized to a minimum on the PES using the same basis set. The TS was identified as the planar geometry of the tricyclic framework, the only TS on the symmetric IRC between the two local energy minima (*vide infra*). The corresponding (single) imaginary frequency of 54*i* cm<sup>-1</sup> was confirmed to be the bending along the central sulfur–nitrogen axis.

Since the ground-state potential energy surfaces of the three tC bases possess a low energy IRC following the bending along the middle sulfur/oxygen–nitrogen axis, as evidenced by the small IR frequencies reported above, the ground and excited-state potential energy surfaces of tC, tC<sup>O</sup> and tC<sub>nitro</sub> were calculated along this reaction coordinate. The PES of the ground and first excited-state of tC<sup>O</sup> calculated at the B3LYP/6-311+G(2d) level is shown in Fig. 5. The S<sub>0</sub> PES is found to be very shallow and centred around the planar equilibrium geometry, well below the average thermal energy at room temperature (dotted line) for conformations bent up to ±12°–15°. As a result of this shallow ground-state PES, the molecular framework of tC<sup>O</sup> is expected to be very flexible in terms of bending along the oxygen–nitrogen axis. The S<sub>1</sub> PES is observed to be steeper along this coordinate, but with a planar equilibrium geometry as well.

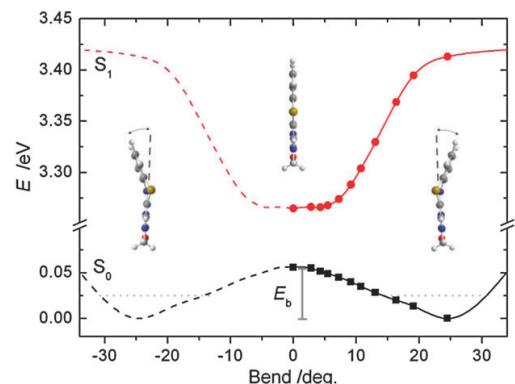
As shown in Fig. 6 the B3LYP/6-311+G(2d) calculated ground-state PES of tC possesses two minima corresponding to the conformations bent at an angle of 25° and separated by the energy barrier  $E_b$ , where the average thermal energy at room temperature is denoted as reference (0.025 eV, dotted lines). Contrary to the ground-state PES of tC, the first excited-



**Fig. 4** B3LYP/6-31G(d,p) optimized ground-state structures of (a) tC and (b) tC<sup>O</sup>. Top: front view. Bottom: side view (both local energy minima of tC are shown). Elements are labeled as follows: H white; C grey; N blue; O red; S yellow. Bond lengths are given in Ångströms.



**Fig. 5** PES of tC<sup>O</sup> following the coordinate for bending along the middle oxygen–nitrogen axis as calculated at the TDDFT B3LYP/6-311+G(2d) level. Full-drawn lines are guides for the eye, and dashed lines are mirror-images of the calculated points. The dotted grey reference line denotes the thermal energy at  $T = 295\text{ K}$ .



**Fig. 6** S<sub>0</sub> and S<sub>1</sub> potential energy surfaces of tC following the coordinate of bending along the middle sulfur–nitrogen axis. Full-drawn lines are guides for the eye and dashed lines are mirror-images of the calculated points.  $E_b = 0.05\text{ eV}$  denotes the ground-state energy barrier between the two bent geometries. The dotted grey reference line denotes the thermal energy at  $T = 295\text{ K}$ . Calculations were performed at the B3LYP/6-311+G(2d) level.

state is predicted by the TDDFT B3LYP/6-311+G(2d) calculations to possess a planar equilibrium geometry with a steep increase in energy for bending above 8° to 10°.

The fast non-radiative decay pathways of nitroaromatics are often believed to involve an internal rotation of the NO<sub>2</sub> group.<sup>68–74</sup> Besides the bending coordinate of tC<sub>nitro</sub>, the ground and excited-state energies of tC<sub>nitro</sub> were therefore additionally investigated along the NO<sub>2</sub> twist coordinate (Fig. 7). The singlet excited-state energies of tC<sub>nitro</sub> were monitored up to the third lowest in energy, the S<sub>3</sub> having 87% (n,π\*)-character when occupying the S<sub>0</sub> equilibrium geometry. The results confirm that the S<sub>0</sub> geometry is stabilized at a bent conformation of the tricyclic framework and a NO<sub>2</sub> internal rotation of 0° while the S<sub>1</sub> and S<sub>2</sub> equilibrium geometries are planar, concordant with the ground and (π,π\*) excited-state

potential energy surfaces of tC (Fig. 7a). In contrast, the S<sub>3</sub> state is predicted to go towards an energy minimum at a NO<sub>2</sub> dihedral angle of ~20° around the N–C bond at which point it is of (n,π\*) and (π,π\*) mixed character (Fig. 7b upper panel). The S<sub>3</sub> minimum energy of tC<sub>nitro</sub> is additionally found to be a slightly less bent conformation of the tricyclic framework as compared to the S<sub>0</sub> equilibrium structure (~16° of S<sub>3</sub> compared to ~25° of S<sub>0</sub>).

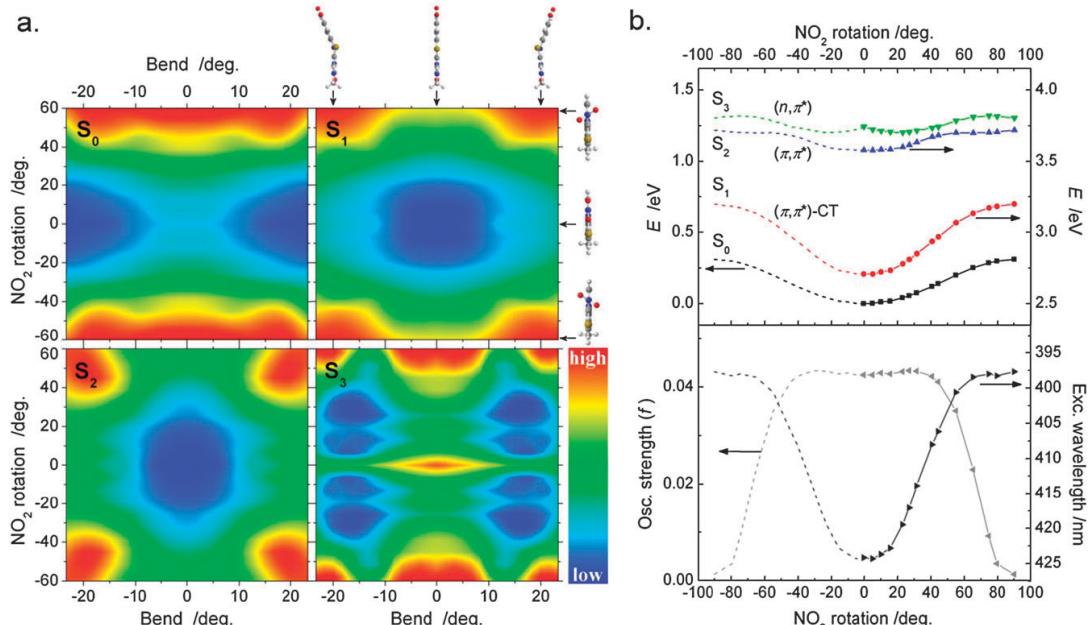
Fig. 7b shows a cross-section of the calculated potential energy surfaces of tC<sub>nitro</sub> along the NO<sub>2</sub> torsional coordinate, where excited-state classifications of the corresponding Franck–Condon excited-states are provided as inserts. The S<sub>0</sub> PES is observed to be relatively shallow along the NO<sub>2</sub> twist pathway, albeit not as shallow as that predicted for nitroperylene<sup>69</sup> and *p*-nitroaniline<sup>71</sup> where the ground-state population is believed to involve a wide distribution of NO<sub>2</sub>-twist angles. The S<sub>0</sub> → S<sub>1</sub> excitation energy of tC<sub>nitro</sub> is predicted by the TDDFT calculations to increase for conformations along the NO<sub>2</sub> torsional coordinate (Fig. 7b lower panel, black). A similar finding was previously used to explain the excited-state dynamics and fluorescence quantum yield dependency upon the excitation-wavelength of nitroperylene,<sup>69</sup> which we believe possesses similar photophysical characteristics as those of tC<sub>nitro</sub>. Whether this wavelength dependency is also a property of tC<sub>nitro</sub> remains to be investigated.

#### Solvent dipolarity dependency of the decay kinetics of tC<sub>nitro</sub>

As shown in the lower panel of Fig. 7b, the oscillator strength of the S<sub>0</sub>–S<sub>1</sub> electronic transition of tC<sub>nitro</sub> decreases drastically to zero at internal rotations of the nitro group >45°. The decrease in transition probability is the result of a change in S<sub>1</sub> character, in which the LUMO orbital becomes localized on the NO<sub>2</sub> group as the conjugation with the aromatic plane is lost. This effect is illustrated in Fig. 8a where the optimized KS LUMO of tC<sub>nitro</sub> is visualized at three different NO<sub>2</sub> twist angles. As can be concluded, the LUMO loses its contribution from the aromatic π-system and turns into a pure antibonding nitro type orbital at angles above 45° to 60°, thus decreasing the transition overlap density between the ground-state and excited-state, and the concomitant decrease in oscillator strength of the S<sub>0</sub>–S<sub>1</sub> transition (which is of 92% HOMO → LUMO character at all NO<sub>2</sub> twist angles).

The more localized LUMO in the twisted nitro configurations of tC<sub>nitro</sub> increases the excited-state dipole moment along this IRC. As a result, the S<sub>1</sub> potential energy along the nitro torsional coordinate is highly dependent on solvent dipolarity. This is illustrated in Fig. 8b where the S<sub>1</sub> energy of tC<sub>nitro</sub> along the nitro twist IRC is plotted in two different solvents of varying dipolarity as calculated at the B3LYP/6-311+G(2d) level using a polarisable continuum solvation shell in the calculations. In the non-polar cyclohexane (CHx) the potential energy curve is steep and increases to an energy at the perpendicular nitro configuration of 0.37 eV above the planar nitro configuration while this energy difference is decreased to 0.26 eV in H<sub>2</sub>O.

If the efficient non-radiative decay pathway of tC<sub>nitro</sub> involves an internal rotation of the NO<sub>2</sub> group around the



**Fig. 7** Potential energy surfaces of the ground and the three lowest electronically excited singlet states of  $\text{tC}_{\text{nitro}}$ . (a) 2D PES following the coordinates of bending along the middle sulfur–nitrogen axis (horizontal axis) and  $\text{NO}_2$  rotation (vertical axis). The shown color scale increases linearly from blue → red (low → high) as follows:  $S_0$  0.00 → 0.25 eV;  $S_1$  2.53 → 3.10 eV;  $S_2$  3.48 → 3.72 eV;  $S_3$  3.67 → 3.80 eV. (b) Cross section following the  $\text{NO}_2$  rotation coordinate at bent = 25°. Upper panel: calculated potential energy surfaces of  $S_0$ ,  $S_1$ ,  $S_2$  and  $S_3$ . Excited-state classifications are provided. Lower panel:  $S_0 \rightarrow S_1$  oscillator strength (grey) and excitation energy (black) dependency upon the rotation of the  $\text{NO}_2$  group. The full-drawn lines are guides for the eyes and the dashed lines are mirror-images of the calculated points. Calculations were performed at the TDDFT B3LYP/6-311 + G(2d) level.

C–N bond (dashed arrow in Fig. 8b), as suspected, the non-radiative deactivation process of  $\text{tC}_{\text{nitro}}$  may be expected to be driven by dipolar solvents as a result of the increase in the excited-state dipole moment along this pathway as described above. Indeed, this is supported by  $\text{tC}_{\text{nitro}}$  in less dipolar solvents such as THF and 1,4-dioxane in which  $\text{tC}_{\text{nitro}}$  is fluorescent with a quantum yield of  $\Phi_f = 0.011$  and 0.028, respectively. The emission band of  $\text{tC}_{\text{nitro}}$  in these solvents is centered around  $\lambda_{\text{max}} \approx 615$  nm which is associated with a very large Stokes shift of  $\sim 7300 \text{ cm}^{-1}$  (Fig. 9, dashed lines). In contrast, the emission of  $\text{tC}_{\text{nitro}}$  in PG glass at 200 K is centered at  $\lambda_{\text{max}} = 530$  nm corresponding to a Stokes shift of  $4800 \text{ cm}^{-1}$  as shown in Fig. 9. In all cases, the excitation spectra confirmed that the observed emission was due to  $\text{tC}_{\text{nitro}}$  itself. No emission is observed from  $\text{tC}_{\text{nitro}}$  in either of the highly dipolar solvents  $\text{H}_2\text{O}$ , ACN, DMSO or DMF.  $\text{tC}_{\text{nitro}}$  was also observed to be very weakly fluorescent in  $\text{CH}_2\text{Cl}_2$  (not shown); however, the low solubility of  $\text{tC}_{\text{nitro}}$  in this solvent hindered a quantification of the fluorescence process which we estimate to be  $\Phi_f < 0.01$ .

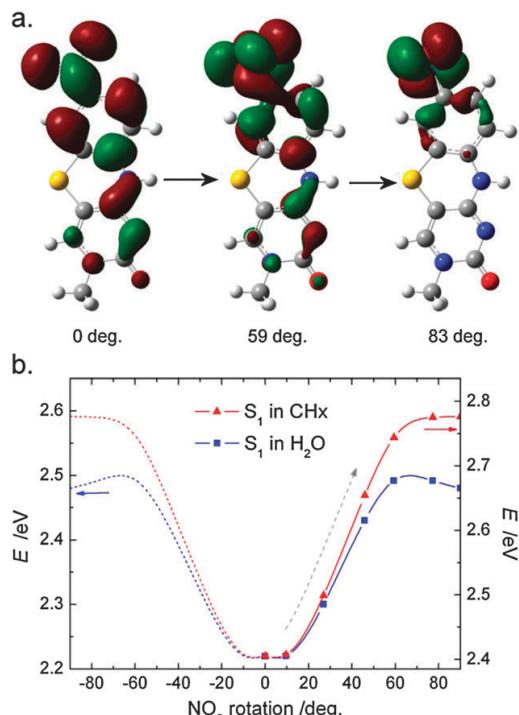
The absorption spectra of  $\text{tC}_{\text{nitro}}$  in 1,4-dioxane, THF, PG, ACN and  $\text{H}_2\text{O}$  are also shown in Fig. 9 (full-drawn lines). The absorption spectra have been normalized for the lowest energy absorption band to facilitate comparison. The absorption maximum of the neutral form of  $\text{tC}_{\text{nitro}}$  only shows minor differences in the investigated solvents. The position of the

lowest energy absorption band is  $\lambda_{\text{max}} = 420$  nm, 421 nm, 423 nm, 415 nm and 424 nm in 1,4-dioxane, THF, PG, ACN and  $\text{H}_2\text{O}$ , respectively.

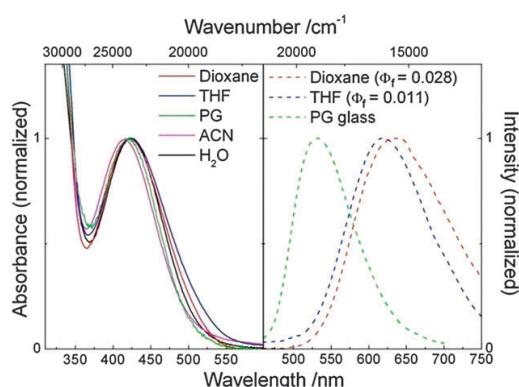
## Discussion

### Decay kinetics and potential energy surfaces of $\text{tC}$ and $\text{tC}^{\text{O}}$

Some general conclusions can be made about  $\text{tC}$  and  $\text{tC}^{\text{O}}$ . The temperature-dependent fluorescence measurements show that the total non-radiative decay rate constants of  $\text{tC}$  and  $\text{tC}^{\text{O}}$  are of the same magnitude at room temperature ( $\sim 8 \times 10^7 \text{ s}^{-1}$ ). The larger fluorescence quantum yield of  $\text{tC}^{\text{O}}$  is thus a result of a larger fluorescence rate constant, which, in turn, is due to a larger oscillator strength of the lowest energy electronic transition. However, whereas the non-radiative decay of  $\text{tC}$  is dominated by a temperature-dependent process 4 to 5 times faster than the temperature-independent non-radiative decay process, the temperature-dependent and the temperature-independent non-radiative decay processes of  $\text{tC}^{\text{O}}$  are of similar magnitude ( $\sim 4 \times 10^7 \text{ s}^{-1}$ ). The fast temperature-dependent non-radiative decay process observed for  $\text{tC}$  is due to a smaller  $E_a$  of this process compared to the  $E_a$  of the same process in  $\text{tC}^{\text{O}}$ . The frequency factor for the temperature-dependent non-radiative decay process is 3 times larger in  $\text{tC}^{\text{O}}$  relative to that in  $\text{tC}$ .



**Fig. 8** (a) Visual representation of the LUMO of tC<sub>nitro</sub> at three different degrees of NO<sub>2</sub> twist. (b) Solvent dipolarity dependency of the S<sub>1</sub> state energy of tC<sub>nitro</sub> following the NO<sub>2</sub> rotation coordinate for the flat conformation (*i.e.* bent = 0° for all calculated points). Dashed arrow denotes the proposed pathway of the main radiationless deactivation process of tC<sub>nitro</sub>. Calculations were performed using TDDFT B3LYP/6-311+G(2d). Solvation effects were simulated using the CPCM model.



**Fig. 9** Absorption and emission spectra of tC<sub>nitro</sub> in various solvents. Left: isotropic absorption spectra. Right: the emission spectra of tC<sub>nitro</sub> in propylene glycol glass at 200 K, and in THF and 1,4-dioxane at RT. All spectra have been normalized to facilitate comparison. Emission spectra were acquired using an excitation wavelength of 420 nm.

The value of  $E_a$  corresponds to an energy barrier between two molecular coordinates on the PES of the excited-state of the molecule: the energy minimum, and the point of intersection with either the electronic ground-state PES (for IC) or an excited triplet state PES (for ISC), depending on the physical nature of the temperature-dependent deactivation process. The physical equivalences of  $k_0$  and the temperature-dependent decay rate constant,  $k_{\text{nr}}(T)$ , are thus related to intersystem crossing (ISC) and IC. However, since the quantum yields of ISC are unknown, it is not possible to determine how ISC and IC relate to the fitted parameters of the tC bases. A very efficient ISC was previously reported for phenothiazine,<sup>75</sup> the parent compound of tC and tC<sub>nitro</sub>.

The low energy barrier,  $E_b$ , between the two bent conformations of tC and tC<sub>nitro</sub> reveals that these geometries frequently isomerize at room temperature. Since the reaction coordinate for the isomerization follows a vibrational mode, it is reasonable to assume a frequency factor of the order of  $A_{\text{bend}} \approx 10^{10} \text{ s}^{-1}$  which yields a rate of isomerization of  $k_{\text{iso}} \approx 10^9 \text{ s}^{-1}$ , or roughly one interconversion for every 1 ns at room temperature. In addition, the calculated excited-state potential energy curve shows that the structures of tC and tC<sub>nitro</sub> undergo a structural change from a bent to a planar conformation upon absorption of a photon. The isomerization between the two bent geometries is thus additionally activated by absorption of a photon. In an isotropic environment, however, the electronic state energies of the two isomers are identical. The radical cation of phenothiazine, important for the biological activity of its derivatives, has been predicted by DFT calculations to be planar as well, resulting in an increased aromatic resonance stability.<sup>76</sup>

It is noted that the geometrical reorganization of tC upon absorption of a photon may explain the larger Stokes shift observed for tC compared to that of tC<sup>O</sup> in H<sub>2</sub>O ( $\sim 7000 \text{ cm}^{-1}$  compared to  $\sim 6000 \text{ cm}^{-1}$ ). The potential energy surfaces predicted in Fig. 5 and 6 for tC<sup>O</sup> and tC, respectively, are additionally supported by the absorption and emission spectra of tC and tC<sup>O</sup> in MeTHF and upon lowering the temperature (ESI†). For tC no vibrational fine-structure is observed in the absorption spectrum at room temperature or by lowering the temperature, while tC<sup>O</sup> displays very distinct vibrational modes in the lowest energy absorption band at lower temperature.

#### Decay kinetics and potential energy surfaces of tC<sub>nitro</sub>

The photophysical properties of tC<sub>nitro</sub> are undoubtedly deeply associated with the nitro group. The non-fluorescent behaviour observed for tC<sub>nitro</sub> in polar solvents at room temperature is a general property of nitrosubstituted aromatic compounds. Often, the fast non-radiative excited-state deactivation is suggested to involve low frequency vibrational modes associated with large amplitude motion of the NO<sub>2</sub> group (*e.g.* see nitrobenzene,<sup>73</sup> *p*-nitroaniline,<sup>71,72</sup> and nitropropylene<sup>69</sup>). The temperature-dependent fluorescence measurements and TDDFT calculations performed here support this model in the case of tC<sub>nitro</sub>. The magnitude of the identified activation energy of the main deactivation channel of tC<sub>nitro</sub> of  $E_a = 0.2 \text{ eV}$  may very well correspond to the energy it costs to twist

the  $\text{NO}_2$  group in the first excited-state (Fig. 7 and 8). This  $E_a$  is  $\sim 0.1$  eV larger than the corresponding values of the main non-radiative deactivation processes of tC and  $\text{tC}^{\text{O}}$ ; however, in contrast to tC and  $\text{tC}^{\text{O}}$  this process completely quenches the fluorescence of  $\text{tC}_{\text{nitro}}$  due to the very rapid frequency associated with this process. Previous studies have shown that a magnitude of the frequency factor of the order of  $A \approx 10^7 \text{ s}^{-1}$  to  $10^8 \text{ s}^{-1}$  is usually seen for a spin-forbidden ISC process, while a value of  $A \approx 10^{11} \text{ s}^{-1}$  to  $10^{12} \text{ s}^{-1}$  is indicative of a spin-allowed IC process.<sup>53</sup> Comparing these values to the very large value of  $A = 2.4 \times 10^{13} \text{ s}^{-1}$  determined for the non-radiative decay process of  $\text{tC}_{\text{nitro}}$  indicates that the main deactivation channel of this chromophore is an IC process. This huge frequency factor also suggests that the efficient non-radiative decay process is activated by a vibrational or rotational mode, which is reasonably argued to involve motion of the  $\text{NO}_2$  group. Despite the fact that most nitroaromatics are found to be quenched efficiently by ISC, as discussed below, the observation of a fast  $S_1 \rightarrow S_0$  IC process as the main deactivation channel was also recently reported for nitroperylene by Mohammed and Vauthey.<sup>69</sup>

A temperature-independent deactivation process of a different origin and with a relatively large rate constant of  $k_0 = 2.1 \times 10^8 \text{ s}^{-1}$  is additionally identified in the temperature measurements of  $\text{tC}_{\text{nitro}}$  reported here. This process could be a spin forbidden ISC process which dominates the decay at lower temperatures, and possibly in non-polar solvents, due to a smaller fluorescence rate constant of  $k_f = 5.4 \times 10^7 \text{ s}^{-1}$ . In many cases, especially for small aryl moieties, the fast non-radiative decay of nitroaromatics has been assigned to an efficient ISC due to a low lying ( $n,\pi^*$ ) state. Whether ISC or IC is the main deactivation pathway of nitroaromatics was early suggested by Khalil *et al.* to relate to the singlet ( $n,\pi^*$ ) and ( $\pi,\pi^*$ ) energy splitting and thus the relative size of the aromatic system.<sup>77</sup> Indeed, many smaller nitroaromatics are phosphorescent with ISC quantum yields often approaching unity.<sup>73,74,77–82</sup> However, this may not be the case for  $\text{tC}_{\text{nitro}}$  since no phosphorescence is observed at low temperatures and the high frequency factor for the non-radiative decay of  $\text{tC}_{\text{nitro}}$  is indicative of an IC process (*vide supra*). In addition, the oscillator strength of the  $S_1 \rightarrow S_0$  transition of  $\text{tC}_{\text{nitro}}$  is much stronger than what is expected for an ( $n,\pi^*$ ) transition. The TDDFT calculations reported here predict the ( $n,\pi^*$ ) singlet-state of  $\text{tC}_{\text{nitro}}$  to lie  $\sim 1$  eV above the  $S_1$ , and associates with a considerable reorganization in this excited-state (Fig. 7). CAS-SCF calculations of nitrobenzene have previously shown a distortion of the  $\text{NO}_2$  group in the mixed ( $n,\pi^*$ ) and ( $\pi,\pi^*$ ) excited-state of this nitroaromatic compound.<sup>83</sup> A survey into the possible ISC processes of  $\text{tC}_{\text{nitro}}$ , however, is beyond the scope of this article.

In the fluorescence measurements of  $\text{tC}_{\text{nitro}}$  performed here, the very large Stokes shift of  $7300 \text{ cm}^{-1}$  observed in THF and 1,4-dioxane is strongly indicative of a CT transition, as predicted by TDDFT calculations.<sup>50</sup> The somewhat smaller Stokes shift observed for  $\text{tC}_{\text{nitro}}$  in PG glass compared to in liquid THF and 1,4-dioxane very likely comes as a result of the difference in solvent viscosity: for  $\text{tC}_{\text{nitro}}$  dissolved in PG glass, the solvent molecules are immobilized around the solute and solvent relaxation is therefore hindered during the excited-state

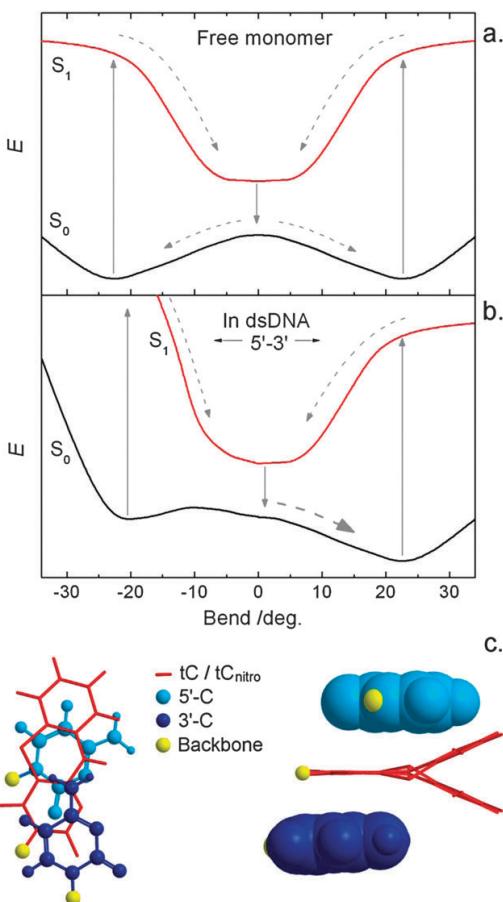
lifetime, which is not the case for  $\text{tC}_{\text{nitro}}$  in THF and 1,4-dioxane at room temperature.

The predicted increase in rotational energy of the nitro group with decreasing solvent polarity can account for the increasing fluorescence quantum yield of  $\text{tC}_{\text{nitro}}$  in less dipolar solvents. Although care should be taken when evaluating the exact quantitative energy shifts calculated in different solvation shells here, the qualitative interpretation is straightforward and indeed valid in the discussion of the solvatochromic photophysics of  $\text{tC}_{\text{nitro}}$ . As a result of the structural change accompanying the decay of the intramolecular CT state of  $\text{tC}_{\text{nitro}}$ , the excited-state dynamics of  $\text{tC}_{\text{nitro}}$  may be expected to depend not only on solvent dipolarity, but also viscosity and H-bonding properties.<sup>84,85</sup> Of particular analogy to the case of  $\text{tC}_{\text{nitro}}$ , Hicks and co-workers argued that the  $E_a$  between planar and twisted polar states of isomers of *p*-dimethylaminobenzonitrile decreases with increasing solvent polarity.<sup>84</sup> The increase in excited-state dipole moment for geometries of nitroaromatics with increasing  $\text{NO}_2$  twist angles was also predicted early by Sinha and Yates in simple theoretical models using modified Hückel theory.<sup>86</sup> Also, polar solvents have been reported to induce fast  $S_1 \rightarrow S_0$  IC in *p*-nitroaniline,<sup>87</sup> and the fluorescence quantum yield of nitroperylene is observed to increase with decreasing solvent polarity.<sup>69</sup> However, to the best of our knowledge this has not been directly linked to an increase in excited-state dipole moment as predicted here, although we believe similar considerations as those suggested here apply to other nitroaromatics as well.

## Molecular geometries of the tC bases positioned in DNA

The surrounding anisotropic nanoenvironment of the base analogues positioned in double-stranded DNA most certainly perturb their potential energy surfaces, in particular the low energy reaction coordinates associated with out-of-plane motion. The low energy IRC following the bending along the middle S–N axis of tC or  $\text{tC}_{\text{nitro}}$  and the O–N axis of  $\text{tC}^{\text{O}}$  will therefore be influenced by neighbouring bases both in terms of the type of nucleobase (pyrimidine or purine) and its position relative to the base analogue (5' or 3'). Although we have not performed careful calculations on the tC bases positioned in DNA, some features can be rationalized on the basis of the potential energy surfaces of the free monomers.

In the case of  $\text{tC}^{\text{O}}$ , the otherwise very shallow potential energy curve (Fig. 5) is probably narrowed in the rigid  $\pi$ -stacking environment of double-stranded DNA. The observed vibrationally fine-structured emission of  $\text{tC}^{\text{O}}$  only in double-stranded DNA is very likely a result of this feature.<sup>47</sup> In the case of tC and  $\text{tC}_{\text{nitro}}$ , the small energy barrier between the two bent conformations, combined with the fact that they will isomerize upon absorption of a photon, very likely facilitates an efficient selective isomerization in double-stranded DNA to the energetically preferable isomer. This is illustrated in Fig. 10 where proposed potential energy curves of tC and  $\text{tC}_{\text{nitro}}$  along the bending of the tricyclic framework are shown for the free monomeric forms (Fig. 10a) and when incorporated in between two nucleobase neighbours in a B-DNA scenario (Fig. 10b). The radiative processes between  $S_0$  and  $S_1$  are indicated as full-drawn lines while the geometrical



**Fig. 10** Hypothetical potential energy curves of ground and first excited-state of tC and tC<sub>nitro</sub> following the bending along the middle sulfur–nitrogen axis in their (a) free monomeric forms and (b) rigidly positioned in double-stranded DNA in between two representative C neighbours. Dashed and full-drawn arrows denote non-radiative and radiative processes, respectively. (c) Left: base-stacking of tC/tC<sub>nitro</sub> when looking down the long axis of double-stranded DNA. A 3'-C is shown on top of tC/tC<sub>nitro</sub> and a 5'-C is shown below tC/tC<sub>nitro</sub>. Right: same molecular coordinates as left image but projected from a side view. In the right figure the two bent conformations of tC/tC<sub>nitro</sub> are overlaid.

reorganizations occurring upon absorption and emission of a photon are shown as dashed arrows. As can be seen, the IRC along the bending is suggested to be perturbed from symmetry as a result of the highly anisotropic base-stacking environment of double-stranded DNA towards the 5' and 3' directions (Fig. 10c). In the example shown, the structure of tC (or tC<sub>nitro</sub>) is thus energetically guided into the isomer which is directed away from the 5'-C and into the major groove. Two rough single-point AM1 energy calculations of the two bent conformations of tC positioned in between C neighbours in a B-DNA geometry (*i.e.* the two configurations shown in Fig. 10c, right) supported this model, showing an energy

difference of as much as  $\sim 1$  eV between the two conformations, primarily due to an unfavorable steric interaction with the 5'-C. This example is illustrative in that other neighbouring bases will influence the potential energy surfaces of the tC bases differently.

Generally speaking, we propose that the geometries of the tC bases positioned in confined biological environments are naturally optimized to the conformation being most favorable to the overall energy, as a result of the flexible nature of the tricyclic frameworks. This insight may be an important factor in explaining the properties of the tC bases in nucleic acid contexts and their ability to adapt to different highly anisotropic and confined environments such as those found inside various nucleic acid systems<sup>46–48,88</sup> and in protein binding pockets.<sup>89–91</sup>

## Conclusions

We identified one low energy geometrical reaction coordinate on the PES of tC and tC<sup>O</sup> and two intrinsic reaction coordinates on the PES of tC<sub>nitro</sub> important for their ground and excited-state structural properties. The structure of tC<sup>O</sup> is predicted to be very flexible in terms of bending around the central oxygen–nitrogen axis but possesses a planar equilibrium geometry. The ground-state potential energy surfaces of tC and tC<sub>nitro</sub>, both having a sulfur in the central ring, possess two energy minima corresponding to geometries bent along the middle S–N axis and separated by an energy barrier of  $\sim 0.05$  eV. After excitation to the first electronically excited-state the tricyclic framework of tC and tC<sub>nitro</sub> changes to a planar equilibrium geometry from which the molecule either decays directly to the S<sub>0</sub> by emission of radiation or by a non-radiative deactivation process. The stronger fluorescence of tC<sup>O</sup> compared to tC is due to a larger oscillator strength of the lowest energy electronic transition. The total non-radiative decay rate constants of tC and tC<sup>O</sup> are of almost equal magnitude at room temperature, however, the non-radiative decay of tC is dominated by a temperature-dependent process, whereas tC<sup>O</sup> decays non-radiatively by an almost equal combination of temperature-dependent and temperature-independent processes. The lack of fluorescence from tC<sub>nitro</sub> at room temperature, on the other hand, is mainly due to an efficient temperature-dependent decay process. This is suggested to be an internal conversion process associated with rotational or vibrational modes of the NO<sub>2</sub> group as indicated by the very high frequency factor. If the NO<sub>2</sub> group twists in the excited-state along the pathway leading to IC, the excited-state dipole moment increases and may very well be the reason for the observed fluorescence from tC<sub>nitro</sub>, and related nitroaromatics, in less dipolar solvents with a fluorescence quantum yield decreasing with increasing solvent dipolarity.

As a result of low energy IRCs along the bending of the tC bases they are expected to be characterized by very flexible tricyclic frameworks. In the rigid base-stacking environment provided by double-stranded DNA the otherwise very shallow S<sub>0</sub> potential energy curve of tC<sup>O</sup> along the bending mode is expected to be steeper and governed by neighbouring bases. Due to the low energy barrier between the two bent conformations of tC and tC<sub>nitro</sub>, combined with the fact that

the excited-state equilibrium geometry is planar, we suggest that the structures of tC and tC<sub>nitro</sub> positioned in DNA are naturally optimized to the conformation being most favorable to the overall energy of the DNA helix. In general, the highly flexible tricyclic frameworks of all of the tC bases predicted here help explain their properties in nucleic acid contexts as well as their ability to adapt to different kinds of confined (biological) environments such as various nucleic acid scenarios and in protein binding pockets.

## Acknowledgements

This research is supported by the Swedish Research Council (VR), the Swedish Energy Agency, and the Danish Council for Independent Research|Natural Sciences (FNU).

## References

- J. N. Wilson and E. T. Kool, *Org. Biomol. Chem.*, 2006, **4**, 4265–4274.
- M. J. Rist and J. P. Marino, *Curr. Org. Chem.*, 2002, **6**, 775–793.
- D. W. Dodd and R. H. E. Hudson, *Mini-Rev. Org. Chem.*, 2009, **6**, 378–391.
- U. Asseline, *Curr. Org. Chem.*, 2006, **10**, 491–518.
- N. Venkatesan, Y. J. Seo and B. H. Kim, *Chem. Soc. Rev.*, 2008, **37**, 648–663.
- A. Okamoto, K. Tanaka, T. Fukuta and I. Saito, *J. Am. Chem. Soc.*, 2003, **125**, 9296–9297.
- H. A. Wagenknecht, *Ann. N. Y. Acad. Sci.*, 2008, **1130**, 122–130.
- T. Ramreddy, M. Kombrabail, G. Krishnamoorthy and B. J. Rao, *J. Phys. Chem. B*, 2009, **113**, 6840–6846.
- T. Ramreddy, B. J. Rao and G. Krishnamoorthy, *J. Phys. Chem. B*, 2007, **111**, 5757–5766.
- C. R. Guest, R. A. Hochstrasser, L. C. Sowers and D. P. Millar, *Biochemistry*, 1991, **30**, 3271–3279.
- J. M. Jean and K. B. Hall, *Biochemistry*, 2004, **43**, 10277–10284.
- M. Menger, F. Eckstein and D. Porschke, *Biochemistry*, 2000, **39**, 4505–4507.
- O. F. A. Larsen, I. H. M. van Stokkum, B. Gobets, R. van Grondelle and H. van Amerongen, *Biophys. J.*, 2001, **81**, 1115–1126.
- X. Shi, E. T. Mollova, G. Pljevaljcic, D. P. Millar and D. Herschlag, *J. Am. Chem. Soc.*, 2009, **131**, 9571–9578.
- C. H. Liu and C. T. Martin, *J. Mol. Biol.*, 2001, **308**, 465–475.
- C. Hariharan and L. J. Reha-Krantz, *Biochemistry*, 2005, **44**, 15674–15684.
- S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew. Chem., Int. Ed.*, 2008, **47**, 6661–6665.
- H. Zhang, W. Cao, E. Zakharova, W. Konigsberg and E. M. De la Cruz, *Nucleic Acids Res.*, 2007, **35**, 6052–6062.
- K. Wojtuszewski, M. E. Hawkins, J. L. Cole and I. Mukerji, *Biochemistry*, 2001, **40**, 2588–2598.
- B. W. Allan, J. M. Beechem, W. M. Lindstrom and N. O. Reich, *J. Biol. Chem.*, 1998, **273**, 2368–2373.
- B. Holz, S. Klimasauskas, S. Serva and E. Weinhold, *Nucleic Acids Res.*, 1998, **26**, 1076–1083.
- C. Hariharan, L. B. Bloom, S. A. Helquist, E. T. Kool and L. J. Reha-Krantz, *Biochemistry*, 2006, **45**, 2836–2844.
- R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, **37**, 6859–6870.
- K. D. Raney, L. C. Sowers, D. P. Millar and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 6644–6648.
- C. H. Liu and C. T. Martin, *J. Biol. Chem.*, 2002, **277**, 2725–2731.
- J. T. Stivers, K. W. Pankiewicz and K. A. Watanabe, *Biochemistry*, 1999, **38**, 952–963.
- J. M. Jean and K. B. Hall, *Biochemistry*, 2002, **41**, 13152–13161.
- J. M. Jean and K. B. Hall, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 37–41.
- E. L. Rachofsky, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 946–956.
- K. C. Thompson and N. Miyake, *J. Phys. Chem. B*, 2005, **109**, 6012–6019.
- M. Mizuta, K. Seio, A. Ohkubo and M. Sekine, *J. Phys. Chem. B*, 2009, **113**, 9562–9569.
- K. Miyata, R. Tamamushi, A. Ohkubo, H. Taguchi, K. Seio, T. Santa and M. Sekine, *Org. Lett.*, 2006, **8**, 1545–1548.
- D. C. Ward, E. Reich and L. Stryer, *J. Biol. Chem.*, 1969, **244**, 1228–1237.
- M. E. Hawkins, W. Pfleiderer, F. M. Balis, D. Porter and J. R. Knutson, *Anal. Biochem.*, 1997, **244**, 86–95.
- T. Mitsui, M. Kimoto, R. Kawai, S. Yokoyama and I. Hirao, *Tetrahedron*, 2007, **63**, 3528–3537.
- N. B. Gaiel, N. Glasser, N. Ramalanjaona, H. Beltz, P. Wolff, R. Marquet, A. Burger and Y. Mely, *Nucleic Acids Res.*, 2005, **33**, 1031–1039.
- S. G. Srivatsan, H. Weizman and Y. Tor, *Org. Biomol. Chem.*, 2008, **6**, 1334–1338.
- J. N. Wilson, Y. J. Cho, S. Tan, A. Cuppoletti and E. T. Kool, *ChemBioChem*, 2008, **9**, 279–285.
- D. A. Berry, K. Y. Jung, D. S. Wise, A. D. Sercel, W. H. Pearson, H. Mackie, J. B. Randolph and R. L. Somers, *Tetrahedron Lett.*, 2004, **45**, 2457–2461.
- A. Okamoto, Y. Saito and I. Saito, *J. Photochem. Photobiol. C*, 2005, **6**, 108–122.
- M. E. Hawkins, *Cell Biochem. Biophys.*, 2001, **34**, 257–281.
- S. Bhariw, P. Sarkar, J. D. Ballin, I. Gryczynski, G. M. Wilson and Z. Gryczynski, *Anal. Biochem.*, 2008, **377**, 141–149.
- J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, New York, 3rd edn, 2006.
- F. Ritort, *J. Phys.: Condens. Matter*, 2006, **18**, R531–R583.
- K. Börjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson and L. M. Wilhelmsson, *J. Am. Chem. Soc.*, 2009, **131**, 4288–4293.
- K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Nordén, B. Albinsson and L. M. Wilhelmsson, *Nucleic Acids Res.*, 2004, **32**, 5087–5095.
- P. Sandin, K. Börjesson, H. Li, J. Mårtensson, T. Brown, L. M. Wilhelmsson and B. Albinsson, *Nucleic Acids Res.*, 2008, **36**, 157–167.
- P. Sandin, L. M. Wilhelmsson, P. Lincoln, V. E. C. Powers, T. Brown and B. Albinsson, *Nucleic Acids Res.*, 2005, **33**, 5019–5025.
- L. M. Wilhelmsson, P. Sandin, A. Holmén, B. Albinsson, P. Lincoln and B. Nordén, *J. Phys. Chem. B*, 2003, **107**, 9094–9101.
- S. Preus, K. Börjesson, K. Kilså, B. Albinsson and L. M. Wilhelmsson, *J. Phys. Chem. B*, 2010, **114**, 1050–1056.
- P. Sandin, P. Lincoln, T. Brown and L. M. Wilhelmsson, *Nat. Protocols*, 2007, **2**, 615–623.
- J. N. Demas and G. A. Crosby, *J. Phys. Chem.*, 1971, **75**, 991–1024.
- J. B. Birks, *Organic Molecular Photophysics*, John Wiley and Sons, London, 1973.
- A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785–789.
- P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- M. J. e. a. Frisch, *GAUSSIAN 03, Revision E.01*, Gaussian, Inc., Wallington Ford, CT, 2004.
- K. Burke, J. Werschnik and E. K. U. Gross, *J. Chem. Phys.*, 2005, **123**, 62206–62209.
- M. A. L. Marques and E. K. U. Gross, *Annu. Rev. Phys. Chem.*, 2004, **55**, 427–455.
- V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, 2nd edn, 2007.
- S. J. Strickler and R. A. Berg, *J. Chem. Phys.*, 1962, **37**, 814–822.
- J. D. Bell, J. F. Blount, O. V. Briscoe and H. C. Freeman, *Chem. Commun.*, 1968, 1656–1657.
- J. J. H. McDowell, *Acta Crystallogr., Sect. B*, 1976, **32**, 5–10.
- C. J. Wilds, M. A. Maier, M. Manoharan and M. Egli, *Helv. Chim. Acta*, 2003, **86**, 966–978.
- C. J. Wilds, M. A. Maier, V. Tereshko, M. Manoharan and M. Egli, *Angew. Chem., Int. Ed.*, 2002, **41**, 115–117.

- 68 S. A. Kovalenko, R. Schanz, V. M. Farztdinov, H. Hennig and N. P. Ernsting, *Chem. Phys. Lett.*, 2000, **323**, 312–322.
- 69 O. F. Mohammed and E. Vauthey, *J. Phys. Chem. A*, 2008, **112**, 3823–3830.
- 70 J. A. Mondal, M. Sarkar, A. Samanta, H. N. Ghosh and D. K. Palit, *J. Phys. Chem. A*, 2007, **111**, 6122–6126.
- 71 V. M. Farztdinov, R. Schanz, S. A. Kovalenko and N. P. Ernsting, *J. Phys. Chem. A*, 2000, **104**, 11486–11496.
- 72 V. Kozich, W. Werncke, J. Dreyer, K. W. Brzezinka, M. Rini, A. Kumrow and T. Elsaesser, *J. Chem. Phys.*, 2002, **117**, 719–726.
- 73 M. Takezaki, N. Hirota and M. Terazima, *J. Phys. Chem. A*, 1997, **101**, 3443–3448.
- 74 R. Morales-Cueto, M. Esquivelzeta-Rabell, J. Saucedo-Zugazagoitia and J. Peon, *J. Phys. Chem. A*, 2007, **111**, 552–557.
- 75 M. Barra, G. S. Calabrese, M. T. Allen, R. W. Redmond, R. Sinta, A. A. Lamola, R. D. Small and J. C. Scaiano, *Chem. Mater.*, 1991, **3**, 610–616.
- 76 D. H. Pan and D. L. Phillips, *J. Phys. Chem. A*, 1999, **103**, 4737–4743.
- 77 O. S. Khalil, H. G. Bach and S. P. McGlynn, *J. Mol. Spectrosc.*, 1970, **35**, 455–460.
- 78 J. S. Zugazagoitia, C. X. Almora-Diaz and J. Peon, *J. Phys. Chem. A*, 2008, **112**, 358–365.
- 79 H. Ohtani, T. Kobayashi, K. Suzuki and S. Nagakura, *Bull. Chem. Soc. Jpn.*, 1980, **53**, 43–47.
- 80 B. Rusakowi and A. C. Testa, *Spectrochim. Acta, Part A*, 1971, **27**, 787–792.
- 81 R. W. Anderson, R. M. Hochstra, H. Lutz and G. W. Scott, *Chem. Phys. Lett.*, 1974, **28**, 153–157.
- 82 R. Hurley and A. C. Testa, *J. Am. Chem. Soc.*, 1968, **90**, 1949–1952.
- 83 M. Takezaki, N. Hirota, M. Terazima, H. Sato, T. Nakajima and S. Kato, *J. Phys. Chem. A*, 1997, **101**, 5190–5195.
- 84 J. M. Hicks, M. T. Vandersall, E. V. Sitzmann and K. B. Eisenthal, *Chem. Phys. Lett.*, 1987, **135**, 413–420.
- 85 Z. R. Grabowski, K. Rotkiewicz and W. Rettig, *Chem. Rev.*, 2003, **103**, 3899–4031.
- 86 H. K. Sinha and K. Yates, *J. Chem. Phys.*, 1990, **93**, 7085–7093.
- 87 C. L. Thomsen, J. Thøgersen and S. R. Keiding, *J. Phys. Chem. A*, 1998, **102**, 1062–1067.
- 88 L. M. Wilhelmsson, A. Holmén, P. Lincoln, P. E. Nielson and B. Nordén, *J. Am. Chem. Soc.*, 2001, **123**, 2434–2435.
- 89 P. Sandin, G. Stengel, T. Ljungdahl, K. Björansson, B. Macao and L. M. Wilhelmsson, *Nucleic Acids Res.*, 2009, **37**, 3924–3933.
- 90 G. Stengel, J. P. Gill, P. Sandin, L. M. Wilhelmsson, B. Albinsson, B. Nordén and D. Millar, *Biochemistry*, 2007, **46**, 12289–12297.
- 91 G. Stengel, B. W. Purse, L. M. Wilhelmsson, M. Urban and R. D. Kuchta, *Biochemistry*, 2009, **48**, 7547–7555.

**Supporting information for:**

**Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family**

Søren Preus,<sup>a</sup> Kristine Kilså,<sup>a</sup> L. Marcus Wilhelmsson,<sup>b</sup> and Bo Albinsson<sup>b</sup>

<sup>a</sup> Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen, Denmark;

<sup>b</sup> Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden

**Temperature-dependent fluorescence quantum yield measurements**

The temperature-dependent fluorescence quantum yields of tC, tC<sup>O</sup> and tC<sub>nitro</sub> were measured by acquiring absorption spectra of the compounds in an Oxford optistatDN cryostat using pure solvent in the cryostat at room temperature as baseline and acquiring emission spectra by exciting the fluorophores at 370 nm, 361 nm and 370 nm, respectively. The same setup was used for the reference fluorophore. The quantum yields at each temperature were then calculated by<sup>1</sup>

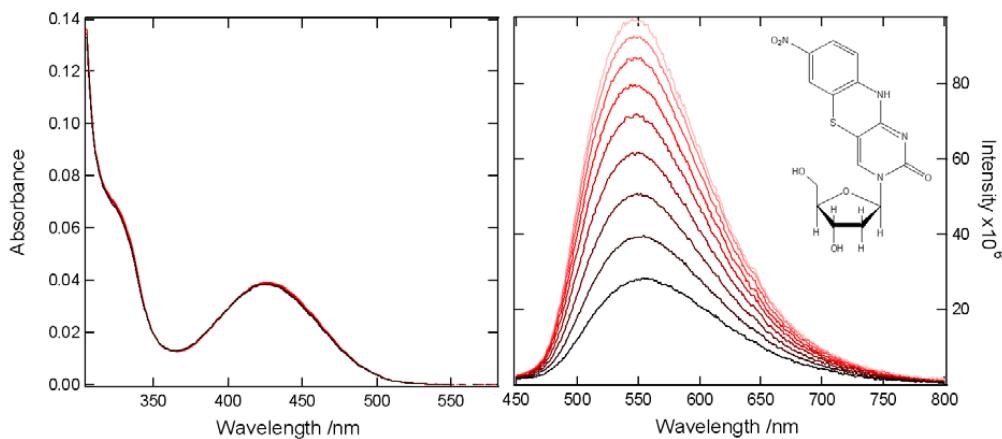
$$\Phi_f = \frac{A_{\text{ref}} \times I_f \times n_{\text{solvent}}^2}{A_f \times I_{\text{ref}} \times n_{\text{ref solvent}}^2} \times \Phi_{\text{ref}}$$

where  $A_f$  and  $A_{\text{ref}}$  are the absorbances at the excitation wavelength of the sample and the reference fluorophore, respectively,  $I$  is the integrated fluorescence intensity,  $n$  is the solvent refractive index and  $\Phi_{\text{ref}}$  is the fluorescence quantum yield of the reference compound. A temperature-correction for the refractive index of MeTHF was made by systematically adding 0.00045 to the value of  $n$  at RT per degree decrease in temperature. Since no value for the volumetric temperature expansion coefficient of MeTHF was found in the literature a simple temperature-correction for the sample volume subtraction was done using a density-temperature-relationship obtained from a linear regression of the temperature-dependent densities of MeTHF measured previously.<sup>2</sup> The temperature-corrections resulted in a change in  $\Phi_{f,0}$  from  $\Phi_{f,0} = 0.77$  to 0.71 and from  $\Phi_{f,0} = 0.70$  to 0.66 for tC and tC<sup>O</sup>, respectively.

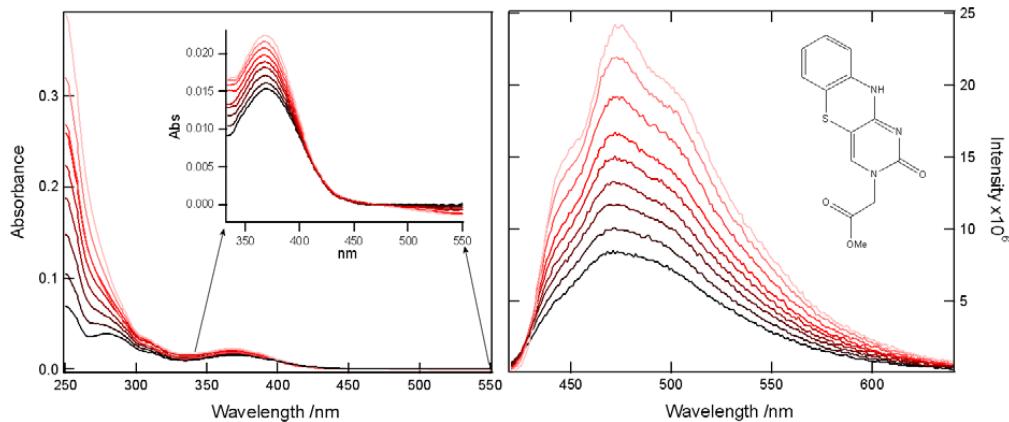
The absorption and emission spectra of each of the three fluorophores are shown in Figure S1-S3. In calculating the fluorescence quantum yields the value of the absorbance of the chromophores at the excitation wavelength was assumed to be the same at all temperatures (except for the temperature-correction for sample volume described above), and determined at  $T = 295$  K for tC and tC<sup>O</sup> and at  $T = 184$  K for tC<sub>nitro</sub>. For tC<sub>nitro</sub> the absorption spectrum (Figure S1) does not change by varying the temperature supporting the approximation in using constant absorbance. For tC and tC<sup>O</sup> the absorption spectra do change by lowering the temperature (Figure S2 and S3), however, as seen in the inserted enlargement of the spectra of tC (Figure S2, left insert) it is the light scattering of MeTHF (the solvent background) in the UV range that dominates the changes with temperature, and results in a concomitant error in the baseline used at lower temperature. The same baseline effect is seen in the absorption spectrum of the

tC<sup>0</sup> sample upon lowering the temperature (Figure S3, left). The approximation that the oscillator strength is constant with temperature might introduce slight errors in the estimated quantum yields but, nevertheless, the resulting Arrhenius fits to the varying quantum yields show excellent correlation with the model functions (Figure 2 in the article).

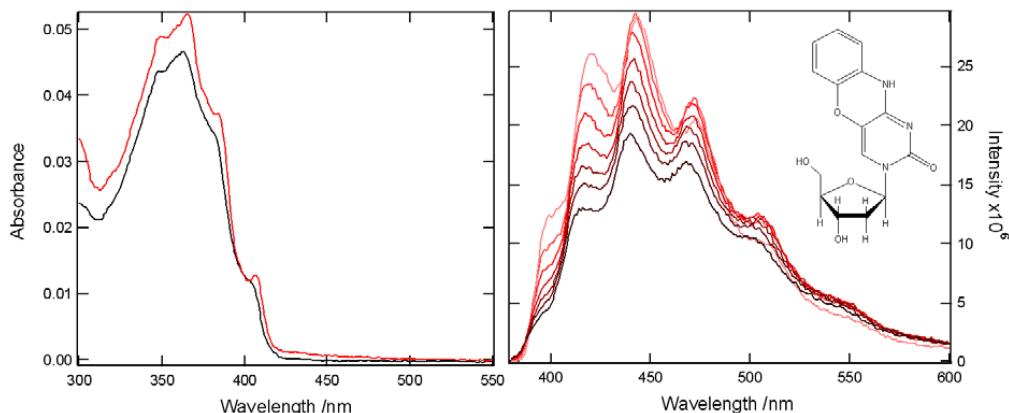
1. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd edn., Springer, New York, 2006.
2. F. Comelli, R. Francesconi, A. Bigi and K. Rubini, Journal of Chemical and Engineering Data, 2007, 52, 639-644.



**Figure 1.** Absorbance (left) and emission (right) spectra of  $t\text{C}_{\text{nitro}}$  in PG glass at various temperatures. Absorbance spectra are shown in the interval from  $T = 184$  K (black) going to  $T = 163$  K (red). Emission spectra are shown in the interval from  $T = 212$  K going to  $T = 150$  K (black→red) using an excitation wavelength of 370 nm and same slit bandwidths at all T.



**Figure 2.** Spectral evolution of the  $t\text{C}$  in MeTHF as the temperature is lowered from  $T = 295$  K to  $T = 153$  K (black→red). Left: Absorption spectra. Insert shows a magnified area of the spectrum. Right: Fluorescence spectra acquired using an excitation wavelength of 370 nm and same slit bandwidths at all temperatures.



**Figure 3.** Left: Absorption spectra acquired of the nucleoside of  $t\text{C}^0$  at 295 K (black) and at 218 K (red). Right: Spectral evolution of  $t\text{C}^0$  emission in MeTHF as the temperature is lowered from 295 K to 175 K (black→red). Spectra were acquired using an excitation wavelength of 361 nm and same slit bandwidths at all temperatures.



## **Paper VI**

*In which a new fluorescent adenine analogue is  
presented...*



# Quadracyclic Adenine: A Non-Perturbing Fluorescent Adenine Analogue

Anke Dierckx,<sup>[a]</sup> Francois-Alexandre Miannay,<sup>[a]</sup> Nouha Ben Gaied,<sup>[b]</sup> Søren Preus,<sup>[c]</sup> Markus Björck,<sup>[a]</sup> Tom Brown,<sup>[b]</sup> and L. Marcus Wilhelmsson<sup>\*[a]</sup>

**Abstract:** Fluorescent-base analogues (FBAs) comprise a group of increasingly important molecules for the investigation of nucleic acid structure and dynamics as well as of interactions between nucleic acids and other molecules. Here, we report on the synthesis, detailed spectroscopic characterisation and base-pairing properties of a new environment-sensitive fluorescent adenine analogue, quadracyclic adenine (qA). After developing an efficient route of synthesis for the phosphoramide of qA it was incorporated into DNA in high yield by using standard solid-phase synthesis procedures. In DNA qA serves as an adenine analogue that preserves the B-form and, in contrast to most currently available FBAs, maintains or even increases the

stability of the duplex. We demonstrate that, unlike fluorescent adenine analogues, such as the most commonly used one, 2-aminopurine, and the recently developed triazole adenine, qA shows highly specific base-pairing with thymine. Moreover, qA has an absorption band outside the absorption of the natural nucleobases ( $>300$  nm) and can thus be selectively excited. Upon excitation the qA monomer displays a fluorescence quantum yield of 6.8% with an emission maximum at 456 nm. More importantly, upon incorporation

into DNA the fluorescence of qA is significantly less quenched than most FBAs. This results in quantum yields that in some sequences reach values that are up to fourfold higher than maximum values reported for 2-aminopurine. To facilitate future utilisation of qA in biochemical and biophysical studies we investigated its fluorescence properties in greater detail and resolved its absorption band outside the DNA absorption region into distinct transition dipole moments. In conclusion, the unique combination of properties of qA make it a promising alternative to current fluorescent adenine analogues for future detailed studies of nucleic acid-containing systems.

**Keywords:** DNA • DNA solid-phase synthesis • duplex stability • fluorescence • fluorescent adenine analogue

## Introduction

Fluorescence is an invaluable tool for researchers studying the structure and dynamics of nucleic acids and their interactions with other biomolecules. The main advantage of fluorescence over other standard techniques, such as X-ray crystallography, NMR spectroscopy, and electrophoresis, is its high sensitivity, versatility as well as the ability to moni-

tor the dynamics of DNA/RNA. However, the use of fluorescence to study nucleic acids requires some form of labelling with a fluorescent marker since the natural nucleobases are virtually non-fluorescent ( $\Phi_f \sim 10^{-4}$ ).<sup>[1]</sup> For this reason, there is an ongoing effort to develop different types of fluorescent probes with properties suited for various purposes. Most available dyes are tethered covalently to the nucleic acids by a linker, for example, fluorescein, rhodamine, the Alexa-dyes and the Cy-dyes. Since these fluorophores reside outside of the base stack, these are referred to as external modifications. A second class, namely internal modifications, comprise fluorescent nucleic acid base analogues (FBAs).<sup>[2]</sup> These molecules resemble one of the natural nucleobases, show at least some base pairing to one of the natural bases through hydrogen bonds, do not seriously perturb the natural DNA or RNA structure, and most importantly, are significantly fluorescent. Having the fluorescent probe inside the base stack has prominent advantages compared to externally attached bulky adducts, which can significantly perturb the DNA duplex structure and also interfere with the interaction with other molecules. Furthermore, FBAs in contrast to external probes, can be firmly stacked and hence report on the local structure and dynamics of the DNA duplex rather than the dynamics of the probe moiety itself.

- [a] A. Dierckx,<sup>\*</sup> Dr. F.-A. Miannay,<sup>\*</sup> M. Björck,  
Prof. Dr. L. M. Wilhelmsson  
Department of Chemical and Biological Engineering  
Physical Chemistry, Chalmers University of Technology  
41296 Gothenburg (Sweden)  
Fax: (+46) 31-772-3858  
E-mail: marcus.wilhelmsson@chalmers.se
- [b] Dr. N. Ben Gaied, Prof. Dr. T. Brown  
School of Chemistry, University of Southampton  
SO17 1BJ Southampton (UK)
- [c] S. Preus  
Department of Chemistry, University of Copenhagen  
Universitetsparken 5, 2100 Copenhagen (Denmark)
- [†] These authors contributed equally to this work.
- Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201103419>.

Over recent decades several research groups have focused on the development of FBAs with the above-mentioned properties. However, synthesising the optimal FBA has proved challenging. Instead, each class of FBA holds specific advantages depending on the kind of experiment one wants to perform. Further development and characterisation of novel FBAs will hopefully lead to a more complete understanding of the relationship between their structural and fluorescence properties.<sup>[3]</sup> The base-discriminating analogues designed by Saito and co-workers,<sup>[4]</sup> the pyrimidine analogues designed by Tor et al.<sup>[5]</sup> and the cytosine analogue pyrrolo dC<sup>[6]</sup> as well as its derivatives<sup>[7]</sup> are just a few examples of thoroughly studied FBAs. Other classes, such as the pteridines designed by Hawkins et al. comprise the guanine analogues 3-MI<sup>[8]</sup> and 6-MI and the adenine analogues 6-MAP and DMAP<sup>[9]</sup> as their most promising members.<sup>[10]</sup> The size-expanded DNA alphabet designed by Kool et al. seems practical for evaluating steric effects in DNA duplex structures among other things.<sup>[11]</sup> Another unique group of FBAs is the tricyclic cytosine analogues tC and tC<sup>O</sup>, the emission of which is virtually insensitive to their microenvironment.<sup>[12]</sup> In addition tC<sup>O</sup>, together with the non-emissive cytosine analogue tC<sub>nitro</sub> have been shown to function as the first nucleobase FRET pair, and yield detailed distance and orientation information about the DNA duplex.<sup>[13]</sup> In contrast to this latter class, all other FBAs show sensitivity to their immediate surroundings.

One of the most popular and widely applied FBAs is the adenine analogue 2-aminopurine (2-AP). The high quantum yield as a monomer in water (0.68) is decreased dramatically upon incorporation into nucleic acids.<sup>[14]</sup> This sensitivity of the emission to the microenvironment has been successfully used in many applications, of which some examples are given below. Even though it has found use in numerous studies, 2-AP causes a moderate destabilisation of the DNA duplex.<sup>[15]</sup> Furthermore, it shows base-pairing to both thymine and cytosine and hence is not an ideal adenine analogue.<sup>[16]</sup> It also shows a considerable amount of dynamics inside DNA.<sup>[17]</sup> These properties can be a substantial drawback in studies requiring a high level of detail and a low level of perturbation to the DNA duplex.

Of the numerous studies involving FBAs over the past decades, only a few will be highlighted here to illustrate the importance of these fluorophores (for more examples see recent reviews by Sinkeldam et al.<sup>[2b]</sup> and Wilhelmsson et al.<sup>[2a]</sup>). Many applications involve the use of the change in emission upon interaction of another molecule with DNA or RNA. This can yield information on conformational changes both in the protein and the DNA/RNA. An illustration is the elegant study of the kinetics of promoter binding and open complex formation upon interaction of bacteriophage T7 RNA polymerase with DNA containing 2-AP modified promoters.<sup>[18]</sup> Another example, which generated information on the dissociation of the multimeric form of the HIV-1 integrase complex upon binding to DNA, involves the guanine analogue 3-MI.<sup>[19]</sup> Other applications focus on the structure of the DNA itself instead of interactions with

other molecules. For instance the fluorescent adenosine analogue 6MAP has been used to analyse the pre-melting transitions of DNA A tracts.<sup>[20]</sup> FBAs are now also starting to appear in the field of nanotechnology. A first report of this involves the cytosine analogue tC<sup>O</sup> sensing the local stability of self-assembling DNA hexagons.<sup>[21]</sup> Finally, FBAs are also finding applications in the RNA field. This is nicely illustrated by the use of 2-AP as a probe for research on small RNAs and RNA elements that control gene expression.<sup>[22]</sup>

Here we present the detailed and optimised experimental procedure for the synthesis of the new fluorescent adenine analogue, quadracyclic adenine (qA, Figure 1). The synthesis

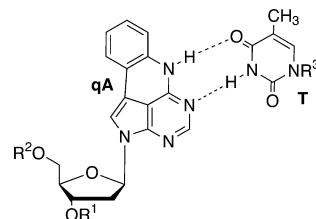


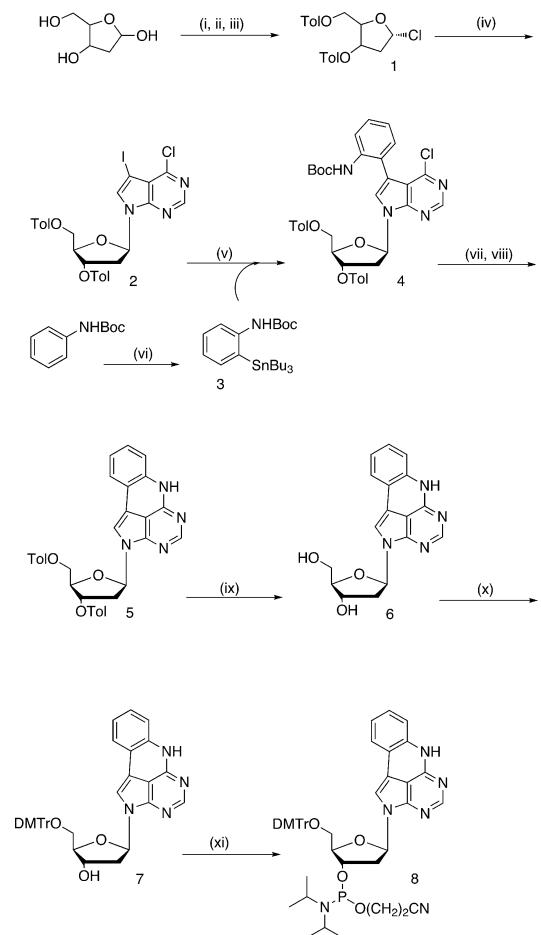
Figure 1. Structure of the qA monomer in the putative qA-T base pair. Measurements on the qA monomer were performed with  $R^1=R^2=H$ . For qA-T incorporated into DNA,  $R^1$ ,  $R^2$  and  $R^3$  represent the rest of the DNA structure.

of qA (the 2'-deoxyriboside of 2,6-dihydro-2,3,5,6-tetraazaa-ceanthrylene;  $\beta$  anomer) has been reported by Matteucci et al. but no experimental procedure or analytical data were described.<sup>[23]</sup> Furthermore, we have found this nucleobase analogue to be emissive and therefore we now report on its fluorescent properties and present a detailed photophysical characterisation of qA both as a monomer and as a constituent of single- and double-stranded DNA. As a monomer, qA shows a lowest energy absorption band well-resolved from the natural DNA bases and we have located the transition dipole moments over this absorption band. Surprising results were obtained when investigating the quantum yields upon exciting over these different absorption transitions. Furthermore, we find that qA is moderately fluorescent as a monomer in water ( $\Phi_f=6.8\%$ ,  $\lambda_{ex}=337\text{ nm}$ ) compared to other adenine analogues, such as 2-AP.<sup>[14]</sup> However, upon incorporation into DNA, quantum yields reach maximum values, which are higher than those corresponding to 2-AP.<sup>[14]</sup> We also report on the non-perturbing properties of qA in DNA duplex systems. On average qA slightly stabilises DNA duplexes and CD measurements suggest that the DNA duplex adopts the B-form. Furthermore, qA is a selective adenine analogue since it shows specific base pairing to thymine only, in contrast to other fluorescent adenine analogues, such as 2-AP and triazole adenine.<sup>[16,24]</sup> To the best of our knowledge qA is the only fluorescent adenine analogue with such favourable properties.

## Results and Discussion

**Synthesis of the quadracyclic 2'-deoxyadenosine analogue, qA:** The synthesis of the quadracyclic analogue of 2'-deoxyadenosine (qA; 2'-deoxyriboside of 2,6-dihydro-2,3,5,6-tetraazaacanthrylene;  $\beta$  anomer) has previously been reported by Matteucci et al. but no experimental procedures were described.<sup>[23]</sup> Here we report the synthesis of the corresponding phosphoramidite monomer (Scheme 1) and oligonucleotides containing qA (see the Supporting Information).

The synthesis of the qA phosphoramidite (**8**) was initiated with the conversion of the 2'-deoxyribose into 3,5-ditoluoyl-1-chloro-deoxyribose (predominantly  $\alpha$  anomer; **1**)<sup>[25]</sup> in



Scheme 1. Synthesis of the fluorescent quadracyclic phosphoramidite monomer (**8**). i) acetyl chloride, MeOH; ii) *p*-toluenesulfonyl chloride, pyridine, 0°C; iii) acetyl chloride/acetic acid/H<sub>2</sub>O, acetic acid (36% for 3 steps); iv) 7-chloro-6-iodo-purine, NaH, CH<sub>3</sub>CN, 50°C (73%); v) Pd(PPh<sub>3</sub>)<sub>4</sub>, CuI, CsF, DMF, 100°C (60%); vi) SnBu<sub>3</sub>Cl, *tert*-butyllithium, THF, -78°C to -20°C (65%); vii) DBU, DABCO, DMF, 75°C; viii) 25% TFA/DCM (46% over 2 steps); ix) NaOMe/MeOH, (61%); x) DMTCl, pyridine (68%); xi) PCl[(OCH<sub>2</sub>)<sub>2</sub>CN]N(iPr)<sub>2</sub>, DIPEA, CH<sub>2</sub>Cl<sub>2</sub> (79%).

three steps with an overall yield of 36% (Scheme 1). Condensation of 6-chloro-7-deaza-7-iodopurine with the chloro sugar by using sodium hydride<sup>[26]</sup> gave the desired compound (**2**) in 73% yield. The major isomer was consistent with the  $\beta$  anomer by using proton NMR spectroscopy. Moreover, the final product that was later incorporated into oligonucleotides gave stable duplexes, further confirming the correct isomer. The sugar–base condensation step was followed by a Stille coupling. In the original work Matteucci et al.<sup>[23]</sup> used trimethylstannyl-*N*-(Boc)aniline obtained by *ortho*-functionalisation of the *N*-(Boc)aniline.<sup>[27]</sup> In our hands this step was low yielding and we attempted to overcome this problem using tributyltin chloride. Our first attempts, following the original procedure, gave very low yields and changing the catalyst, the solvent and the temperature did not improve the situation.

A procedure was then followed that employed a combination of cesium fluoride and cuprous salts to enhance the reaction rate.<sup>[28]</sup> The use of Cu<sup>I</sup> and highly polar solvents (NMP or DMF) in the Stille reaction is well-documented. The rate accelerating effect of Cu<sup>I</sup> is attributed to a preliminary transmetalation reaction with the organostannane that generates a more reactive organo–copper intermediate. In addition, the Bu<sub>3</sub>SnCl by-product can be efficiently transformed into insoluble Bu<sub>3</sub>SnF, which is easily removed from the reaction mixture by filtration. In our case, a high temperature was needed to consume all the starting material to give a respectable yield of **4** (60%). Formation of compound **5** was achieved by using a mixture of DABCO and DBU. Under these conditions partial deprotection of the *tert*-butyloxycarbonyl protecting group occurred. Therefore, the Boc-intermediate was not isolated but was fully deprotected by using a 25% solution of TFA in DCM. Deprotection of the toluoyl group was then performed by using a solution of NaOMe (0.5 M) in MeOH, which gave the free qA nucleoside (**6**) in 61% yield. It was crucial to purify this nucleoside by column chromatography in order to achieve successful tritylation in the next step. Purified compound **6** was tritylated by using dimethoxytrityl chloride in pyridine to give (**7**) in 68% yield. This was finally converted to the qA phosphoramidite monomer (**8**) in 79% yield.

### Spectroscopic characterisation of the qA monomer

**Steady state absorption and fluorescence spectra:** The steady state absorption and emission spectra of the quadracyclic adenine (qA) monomer excited at 337 nm in water and methanol are shown in Figure 2. The absorption spectrum is characterised by a sharp peak at 335 nm with an extinction coefficient of 5000 M<sup>-1</sup> cm<sup>-1</sup> (in water) and a shoulder at 300 nm; this indicates the presence of several different transitions ( $S_0 \rightarrow S_2$  and  $S_0 \rightarrow S_3$ , respectively, as discussed below) and a lower energy spectral feature above 370 nm. The absorption of qA at 335 nm is red-shifted compared to the natural nucleobases, which absorb up to 300 nm, and therefore allows for easy selective excitation. This is an advantage compared to, for example, the fluorescent triazole adenine,

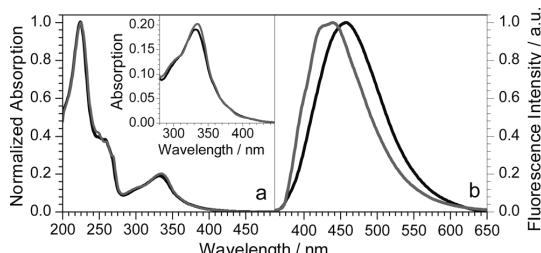


Figure 2. a) Normalised absorption, and b) emission spectra of the qA monomer in water (black) and methanol (grey) when excited at 337 nm at 25°C. Inset: lowest energy absorption band of qA outside the region of DNA absorption.

$A^T$ , ( $\lambda_{abs,max} = 282$  nm)<sup>[24]</sup> or the series of 7-substituted adenines designed by Seela et al., of which the most fluorescent members have their absorption centred around 280 nm.<sup>[29]</sup>

As can be seen in Figure 2 the two emission spectra of qA in water and methanol show a maximum emission peak at 456 and 440 nm, respectively. Furthermore, the spectral shape and position of the emission are independent of the excitation energy; this suggests that emission always occurs from the same energy level ( $S_1$ ; see the Supporting Information). qA shows a lower fluorescence in water ( $\Phi_f = 6.8\%$ ,  $\lambda_{ex} = 337$  nm) than other fluorescent adenine analogues, such as 2-AP (68%),<sup>[14]</sup> 6-MAP (39%), DMAP (48%),<sup>[9]</sup> xA (44% in methanol),<sup>[11a]</sup> hxy<sup>7</sup>c<sup>7</sup>A<sub>d</sub> (27%)<sup>[30]</sup> or triazole adenine (61%).<sup>[24]</sup> However, as will be discussed below, qA is a promising FBA when located inside DNA.

**Transition dipole moments:** To obtain detailed information about the transition dipole moments of qA, we recorded the steady-state excitation spectrum and fundamental anisotropy of the qA monomer in an H<sub>2</sub>O/ethylene glycol matrix (1:2 mixture) at -100°C (Figure 3). As in the absorption spectrum recorded for the free monomer in solution (Figure 2), this excitation spectrum (Figure 3, dashed line) shows a sharp peak around 330 nm, a weaker one at 300 nm, and a longer wavelength absorption transition centred at 390 nm with fine structure due to the vibrational levels. The fundamental anisotropy,  $r_0$  (Figure 3), shows three different main plateaus between 280 and 450 nm. The lowest energy plateau is located between 390 and 450 nm ( $r_0 = 0.36$ ), the second occurs around the absorption maximum at 330 nm ( $r_0 = 0.25$ ) and the third around 300 nm ( $r_0 = 0.05$ ). The low energy plateau shows an  $r_0$  value close to the theoretical maximum of 0.4. This suggests that, in this particular spectral region, the absorption and emission transition dipole moments are parallel.

The reason why this experimental value is slightly lower could be due to vibrational torsions and slight geometrical changes in the molecule between the moment of excitation and emission. Furthermore, as was mentioned in the previous section, all emission spectra of qA are virtually independent of the excitation wavelength; this suggests that the

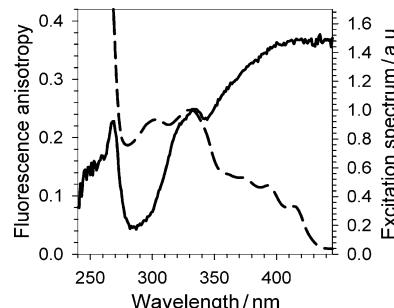


Figure 3. Excitation anisotropy spectrum (black line) and excitation spectrum (dashed line) of qA in an H<sub>2</sub>O/ethylene glycol matrix (1:2 mixture) at -100°C. The table shows the results of the TDDFT calculations on the three lowest energy electronic transitions of qA.

emission transition dipole moment is invariable over this spectral region (see the Supporting Information). As a result of this, the anisotropy recorded will give a direct measure of the angles between a certain absorption transition moment and the emission transition moment. Consequently, the three different plateaus mentioned above, suggest that there are three distinct absorption transition dipole moments: one around 400 nm that is parallel to the emission transition moment ( $S_0 \rightarrow S_1$ ), and two others ( $S_0 \rightarrow S_2$  and  $S_0 \rightarrow S_3$ , respectively) with different angles relative the constant emission dipole moment. We also noticed a bulge located around the second plateau at 340 nm. This observation might be attributed to vibrational features situated between the electronic states  $S_2$  and  $S_1$ .

Theoretical calculations of the lowest singlet energy excitations of qA were performed on the B3LYP/6-31G(d,p) optimised geometry by using TDDFT B3LYP/6-311+G(2d) including a CPCM solvation shell for H<sub>2</sub>O as implemented in Gaussian 03. Above 280 nm these calculations show three significant transitions centred at 283.15 nm ( $f=0.1049$ ;  $S_0 \rightarrow S_3$ ), 329.84 nm ( $f=0.1968$ ;  $S_0 \rightarrow S_2$ ) and 349.51 nm ( $f=0.0594$ ;  $S_0 \rightarrow S_1$ ). Comparing these values to the transitions suggested by the absorption and excitation spectra as well as the fundamental anisotropy, we observe that they match very well, except that they are somewhat blue-shifted. This kind of spectral shift between theoretical calculations and experiments is common and has also been observed for other FBAs.<sup>[12a,31]</sup> Importantly, the relative positions of the transitions and the trend in the calculated oscillator strengths correspond very well to the intensity of the spectral bands of both the absorption and excitation spectra of the qA mono-

mer. This is further evidence for the existence of three different transitions in the low energy band of qA.

**Fluorescence properties of qA:** The fluorescence quantum yields of qA obtained for different excitation wavelengths between 295 and 377 nm are shown in Figure 4. Surprisingly

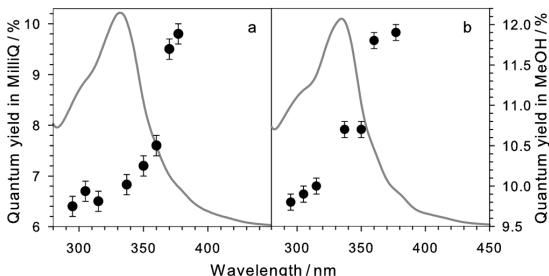


Figure 4. Steady-state absorption spectrum and fluorescence quantum yields of qA for different excitation wavelengths between 295 and 377 nm in: a) MilliQ water, and b) methanol. All measurements were performed at  $25(\pm 0.5)^\circ\text{C}$ .

the fluorescence quantum yields in water decrease gradually from 9.8 to 6.4 % with excitation at shorter wavelengths (see the Supporting Information). When exciting at wavelengths further into the red region even higher quantum yields are detected. However, these values are difficult to accurately determine since the excitation energy severely overlaps the emission spectrum at these wavelengths. Nevertheless, irrespective of the excitation wavelength the spectral envelope of the emission and its position stays virtually the same for the monomer both in water and in methanol (see the Supporting Information). Furthermore, the time-resolved fingerprint of qA in water is defined by a single time component of 3.2 ns ( $\chi^2 = 1.684$ ), which is insensitive to excitation (295, 330 and 377 nm) and emission wavelengths (456 and 520 nm; see the Supporting Information). This suggests that regardless of which state is being populated upon excitation, the fluorescence always originates from deactivation of the same electronic state ( $S_1$ ). Thus, internal conversion from the high-energy electronic states ( $S_2$ ,  $S_3$ ) occurs followed by a deactivation of the excited  $S_1$  population through fluorescence or non-radiative processes.

The decrease in quantum yield from 377 to 300 nm is stepwise in methanol (Figure 4b) and more continuous in water (Figure 4a); this suggests a gradually changing overlap between the electronic transitions  $S_0 \rightarrow S_3$ ,  $S_0 \rightarrow S_2$  and  $S_0 \rightarrow S_1$ . This indicates a more efficient non-radiative deactivation to the ground state when populating the  $S_2$  or  $S_3$  states rather than the  $S_1$  state. The highest quantum yield value is obtained when mostly populating the  $S_1$  state ( $\lambda_{\text{exc}} = 377 \text{ nm}$ ); when the population ratio is increased between  $S_2$  and  $S_1$  (i.e., exciting at lower

wavelengths) the quantum yield decreases. Upon excitation further into the blue to  $S_3$  ( $\lambda_{\text{exc}} = 295 \text{ nm}$ ) the fluorescence quantum yield in water does not change significantly (6.8 and 6.4 %, respectively). To explain these observations we suggest a hypothesis in which part of the highly excited population could relax and then be trapped in a non-radiative parasite state. Such an energy trap state could be localised between the  $S_2$  and  $S_1$  states. Indeed, the decrease in quantum yield of at least 30 % between the  $S_1$  and  $S_2$  states clearly indicates that a decrease of deactivation from  $S_1$  through the fluorescence path occurs. In conclusion, we suggest a deactivation pathway of the qA monomer in which at least 30 % of the molecules that are excited to higher electronic states ( $S_2$ ,  $S_3$ ) get trapped in a “parasite state” and the rest relax to the lowest vibrational level of  $S_1$  and then further to  $S_0$  by emission or internal conversion and vibrational relaxation. It has been observed that the quantum yield of the virtually non-fluorescent natural mononucleotide, dGMP, also varies with the excitation energy. In that case, populating higher energy electronic states was suggested to open non-radiative deactivation pathways.<sup>[32]</sup>

### Structure and stability of DNA duplexes containing qA

**UV duplex melting experiments:** In order to investigate the spectroscopic and structural properties of qA inside DNA, eleven 10-mer oligonucleotides were synthesised, each with one qA base incorporated opposite T (Table 1). To evaluate the effect of both purines and pyrimidines neighbouring qA at the 3' and 5' ends, these eleven sequences have variations in the bases directly flanking qA. Furthermore, we designed a twelfth 10-mer DNA duplex containing two qA molecules in order to investigate the structural influence and the photophysical properties of qA in this case (sequence CT<sub>2</sub>TA).

The melting temperatures of the different duplexes containing qA ( $T_m^{\text{qA}}$ ) as well as those of the corresponding nat-

Table 1. Melting temperatures of the qA modified DNA duplexes ( $T_m^{\text{qA}}$ ). Melting temperatures of the corresponding natural duplexes are also listed ( $T_m^{\text{A}}$ ) as well as the corresponding difference in melting temperature with the modified duplexes ( $\Delta T_m$ ).

DNA sequence <sup>[a,b]</sup>	Neighbouring bases <sup>[c]</sup>	$T_m^{\text{qA}}$ [°C] <sup>[d]</sup>	$T_m^{\text{A}}$ [°C] <sup>[d]</sup>	$\Delta T_m$ [°C]
5'-d(CGCAA(qA)ATCG)-3'	<b>AA</b>	46.3	46.6	-0.3
5'-d(CGCAG(qA)ATCG)-3'	<b>GA</b>	48.3	48.3	0.0
5'-d(CGCAG(qA)GTCG)-3'	<b>GG</b>	51.9	51.5	0.4
5'-d(CGCAG(qA)CTCG)-3'	<b>GC</b>	52.3	50.8	1.5
5'-d(CGCAA(qA)CTCG)-3'	<b>AC</b>	52.7	50.8	1.9
5'-d(CGCAT(qA)ATCG)-3'	<b>TA</b>	48.5	45.3	3.2
5'-d(CGCAC(qA)ATCG)-3'	<b>CA</b>	53.2	49.9	3.3
5'-d(CGCAT(qA)GTCG)-3'	<b>TG</b>	48.7	45.3	3.4
5'-d(CGCAC(qA)CTCG)-3'	<b>CC</b>	58.2	52.9	5.3
5'-d(CGCAC(qA)TTCG)-3'	<b>CT</b>	56.1	49.6	6.5
5'-d(CGATT(qA)TGCG)-3'	<b>TT</b>	50.6	43.7	6.9
5'-d(CGC(qA)T(qA)ATCG)-3'	<b>CT, TA</b>	55.6	46.4	9.2

[a] The natural sequences contain an adenine instead of (qA). [b] Duplexes were obtained by annealing the sequences listed to their natural complementary oligonucleotides. [c] Sequences are named by the bases neighbouring qA, in which the purines are shown in bold and the pyrimidines in italic. [d] Samples were prepared in phosphate buffer (pH 7.5, 200 mM Na<sup>+</sup>, 17 mM phosphate).

ural duplexes ( $T_m^A$ ) are listed in Table 1. In general qA has a stabilising effect on the duplex structure, which translates to an average rise in melting temperature of approximately 3°C. Moreover, the extent of this stabilisation is dependent on the bases flanking qA, which allows for selecting the desired stability of a qA-containing duplex. When qA is flanked by two purines (**AA**, **GA** and **GG**), the stability of the duplex structure on average is equal to that of the natural duplexes. A pyrimidine flanking qA on the 3' side (**GC** and **AC**) causes a slight stabilisation of the duplexes (1.7°C on average), whereas a pyrimidine neighbouring on the 5' side (**TA**, **CA** and **TG**) causes a moderate stabilisation (3.3°C on average). DNA duplexes containing qA flanked by two pyrimidines have the highest stability (**CC**, **CT** and **TT**), which is shown by an average rise in melting temperature of 6.2°C compared to the natural duplexes. In general we can conclude that the stability of the duplexes is increased most when qA is flanked by a pyrimidine on the 5' side.

When investigating the structural overlap of qA and its neighbours in sequences **CA** and **CC**, an overlap between the extended ring system of qA and the cytosine flanking it on the 5' side is observed. This overlap is not present for natural adenine in the same sequence (data not shown). The increased overlap arising when qA instead of natural A is flanked by a pyrimidine on the 5' side might explain the higher stability of such duplexes. The overall stabilising effect of qA can thus be explained by better  $\pi$ - $\pi$  stacking with its neighbours due to the extended ring system compared to natural adenine. Furthermore, the additional rings extended on C6 and N7 of adenine are well accommodated in the major groove in contrast to extensions on the C8-position, which are known to destabilise the duplex structure.<sup>[24,33]</sup> It should also be noted that sequence **CT/TA** containing two qAs shows a rise in melting temperature (9.2°C) that is virtually equal to the linear combination of the stabilisations caused by qA in the sequences **CT** and **TA** (9.7°C).

The UV melting results determined for qA are very similar to earlier observations for the tricyclic cytosine analogues tC and tC<sup>O</sup>, containing a similar two-ring extension towards the major groove.<sup>[12a,34]</sup> Both tC and tC<sup>O</sup> increase the melting temperature of 10-mer duplexes by about 3°C under similar conditions (phosphate buffer, pH 7.5, [Na<sup>+</sup>] = 50 mM). Furthermore, as for qA, the most stable duplexes are formed when they are flanked by a pyrimidine on the 5' side. Also in the case of tC and tC<sup>O</sup>, a better overlap between the extended ring systems and the adjacent 5' pyrimidine has been suggested as the reason for the increased duplex stability.<sup>[12a,34]</sup>

Importantly, qA is a good adenine analogue since it causes no destabilisation of the DNA structure. On the contrary, it slightly stabilises the duplexes depending on the neighbouring bases. This property distinguishes qA from most other fluorescent base analogues. For example, the widely used adenine analogue 2-aminopurine (2-AP) has been shown to cause a destabilisation of 10°C in 10-mer duplexes (100 mM KCl, 0.1 mM EDTA, 21  $\mu$ M DNA duplexes,

decamer).<sup>[15b]</sup> Also, another study involving 14-mer duplexes confirmed the destabilising effect of 2-AP by a drop in melting temperature of 6°C (15 mM sodium citrate, pH 7.25, 50  $\mu$ M DNA duplexes).<sup>[15a]</sup> However, other studies have reported a more limited destabilisation of duplexes due to incorporation of 2-AP.<sup>[17,35]</sup> Besides 2-AP, other adenine analogues, such as the size-expanded adenine analogue xA, cause a destabilisation of 4.2–4.9°C when centrally substituted into 12-mer duplexes (100 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM Na-PIPES, pH 7.0, 5.0  $\mu$ M DNA).<sup>[11c]</sup> Also, the pteridine adenine analogues, 6MAP and DMAP, on average have shown a slight (2.4°C) to moderate destabilisation (4.6°C) in 21-mer duplexes (10 mM NaCl, 10 mM Tris, pH 7.5), respectively.<sup>[9]</sup> Furthermore, another purine analogue, the pteridine guanine analogue 3-MI, shows an average drop in melting temperature of 8.3°C, which is comparable to a single base mismatch under similar conditions.<sup>[10b]</sup> As for qA, the modified 7-deaza-adenine analogues developed by Seela et al. have also been shown to stabilise DNA duplexes.<sup>[30,36]</sup> However, to the best of our knowledge, no reports have been published regarding the fluorescence properties of these analogues inside DNA.

**Circular dichroism:** CD spectra of the samples listed in Table 1 were used to evaluate any structural perturbations to the B-DNA helix caused by the incorporation of qA (Figure 5). All duplexes show a typical B-DNA signature, which in the region between 180 and 300 nm, is characterised by a positive band at 275 nm, a negative band at 240 nm, a band that is less negative or positive at 220 nm and a narrow negative band in the region between 220 and 190 nm, preceded by a large positive peak at 180–190 nm. Surprisingly there is no specific signal corresponding to the lowest energy absorption band of qA centred around 337 nm. For other common base analogues, such as tC or 2-

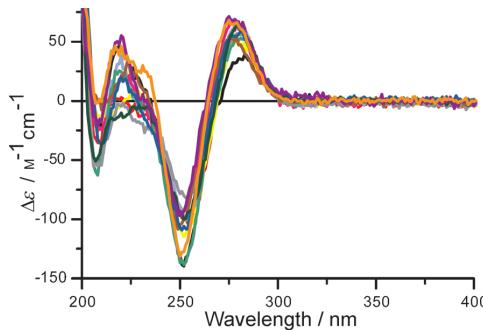


Figure 5. CD spectra of the twelve qA modified duplexes. Duplexes are named as the bases neighbouring qA (purines are shown in bold and pyrimidines in italic): **AA** (black), **GG** (red), **GA** (yellow), **CT** (light blue), **TG** (pink), **AC** (light green), **CC** (grey), **TA** (brown), **CA** (dark blue), **GC** (dark green), **CT/TA** (violet) and **TT** (orange). Duplexes were prepared at a concentration of 2.5  $\mu$ M in phosphate buffer (pH 7.5, 200 mM Na<sup>+</sup>, 17 mM phosphate) and measured at 25°C;  $\Delta\epsilon$  on the  $\Delta\epsilon$  axis refers to the qA/duplex concentration (2.5  $\mu$ M).

aminopurine, a CD band appears for the low energy transition.<sup>[34–35]</sup> However, similar observations to those made here for qA have been reported for the structurally comparable cytosine analogue tCO<sup>[12a]</sup>. We have not yet found a satisfactory explanation for this phenomenon.

The CD signatures of all the modified duplexes containing qA were compared to those of the corresponding non-modified duplexes, and were shown to be virtually identical (see the Supporting Information). Overall it can be concluded that the incorporation of one or several qA molecules preserves the overall B-DNA structure.

**Base pairing specificity of qA:** To investigate the base-pairing specificity of qA, we synthesised oligonucleotides to combine with three of the modified sequences (**GA**, **CT** and **TA**) to produce duplexes containing an adenine, guanine or cytosine instead of thymine opposite to qA. The sequences were chosen to evaluate the influence of having purines, pyrimidines or a combination of both directly neighbouring mismatched qA. The melting temperatures of the modified as well as the unmodified mismatched duplexes are listed in Table 2. For comparison, the melting temperatures of the matched duplexes with qA and A opposite to thymine are also shown. When comparing the difference in melting temperature between the modified matched and mismatched duplexes ( $T_m^{qA-T}$  and  $T_m^{qA-X}$ ) all three sequences are significantly destabilised when qA is placed opposite to G, C or A instead of thymine. This indicates that qA exhibits selective base-pairing with thymine, and hence, is a specific adenine analogue. In contrast, a less specific base-pairing pattern has been reported for other fluorescent base analogues. The adenine analogues 2-AP and triazole adenine, A<sup>T</sup>, for example, are able to form stable base pairs with cytosine and adenine, respectively, besides the conventional base pairing with thymine.<sup>[16,24]</sup> Furthermore, the base discriminating fluorescent analogue BPP can form stable base pairs both with adenine and guanine.<sup>[4]</sup>

Interestingly, qA and natural adenine show the same order of preference for base-pairing partners. When comparing the destabilisations in Table 2, both for qA and A, a more stable mismatch is formed with guanine compared to the mismatches with cytosine and adenine. The same order of thermal stability has been reported before in a study by Gaffney and Jones for mismatches in a 9-mer oligonucleotide.<sup>[37]</sup> The fact that qA shows the same base pairing preference as natural adenine is an important indication that the base-pairing pattern upon modification with qA is indeed unperturbed.

The quantum yields of the duplexes containing a qA mismatch as well as the corresponding matched duplexes are listed in Table 3. On average the quantum yield increases threefold for the mismatched sequences compared to the matched qA-T sequences. We have observed that the average quantum yield of qA in single strands is higher or virtu-

Table 2. Melting temperatures of sequences containing quadricyclic adenine ( $T_m^{qA-X}$ ) and natural adenine ( $T_m^{A-X}$ ) opposite to guanine, adenine or cytosine (X). Melting temperatures of matched sequences with qA ( $T_m^{qA-T}$ ) and A ( $T_m^{A-T}$ ) are compared with the mismatched cases.

Seq. <sup>[a]</sup>	X <sup>[b]</sup>	$T_m^{qA-T}$ [°C] <sup>[c]</sup>	$T_m^{qA-X}$ [°C] <sup>[c]</sup>	$\Delta T_m^{qA}$ [°C]	$T_m^{A-T}$ [°C] <sup>[c]</sup>	$T_m^{A-X}$ [°C] <sup>[c]</sup>	$\Delta T_m$ [°C]
<b>GA</b>	G		38.8	9.5		35.6	12.4
	A	48.3	34.2	14.1	48.3	32.7	15.4
	C		34.1	14.2		28.7	19.3
<b>CT</b>	G		47.6	8.5		38.4	11.1
	A	56.1	40.5	15.6	49.6	33.1	16.4
	C		39.3	16.8		31.5	18.1
<b>TA</b>	G		37.0	11.5		31.2	12.4
	A	48.5	33.5	15.0	45.3	27.3	16.3
	C		32.8	15.8		24.9	18.7

[a] Sequences are named by the bases neighbouring qA and A. Full sequences are listed in Table 1. [b] Base opposite qA or A in the mismatched sequences. [c] Samples containing duplexes (2.5 μM) were prepared in phosphate buffer (pH 7.5, 200 mM Na<sup>+</sup>, 17 mM phosphate).

Table 3. Quantum yields of the mismatched ( $\Phi^{qA-X}$ ) and matched ( $\Phi^{qA-T}$ ) modified duplexes.

Seq. <sup>[a]</sup>	X <sup>[b]</sup>	$\Phi^{qA-T}$ [%] <sup>[c,d]</sup>	$\Phi^{qA-X}$ [%] <sup>[c,e]</sup>
<b>GA</b>	G		3.0
	A	0.5	1.6
	C		1.6
<b>CT</b>	G		0.5
	A	0.2	0.5
	C		0.5
<b>TA</b>	G		0.4
	A	0.2	0.5
	C		0.4

[a] Sequences are named by the bases neighbouring qA and A. Full sequences are listed in Table 1. [b] Base opposite of qA or A in the mismatched sequences. [c] Samples containing duplexes (2.5 μM) were prepared in phosphate buffer (pH 7.5, 200 mM Na<sup>+</sup>, 17 mM phosphate). [d] Fluorescence quantum yields were determined relative to the reference quinine sulfate in 0.5 M H<sub>2</sub>SO<sub>4</sub> ( $\Phi=0.55$ ) at 25°C, and are averages of at least two independent measurements. [e] Fluorescence quantum yields determined at 13°C were very similar (maximum 11% difference) to values presented here.

ally equal to their corresponding mismatched duplexes. Thus, we suggest that the increase in quantum yield might be due to the weaker base-pairing between qA and the mismatching bases. This will lead to less rigid or less extensive base-stacking and an increased opening rate for the mismatched qA base pairs relative to qA-T. In spite of these perturbations, the CD spectra of the modified mismatched and matched duplexes are virtually identical (data not shown); this indicates that the overall B-DNA structure remains unchanged in these mismatched 10-mers.

**Spectroscopic and photophysical characterisation of qA in DNA:** The fluorescence quantum yields and lifetimes of qA incorporated into the twelve single-stranded sequences, as

Table 4. Fluorescence quantum yields and average fluorescence lifetimes of the twelve single- and double-stranded oligonucleotides containing qA, referred to as ss and ds, respectively.

Neighbouring bases <sub>ss</sub> <sup>[a]</sup>	$\Phi_{ss}$ [%] <sup>[b,c]</sup>	$\langle \tau \rangle_{ss}$ [ns] <sup>[b,d]</sup>	Absorption max <sub>ss</sub> [nm] <sup>[b]</sup>	Emission max <sub>ss</sub> [nm] <sup>[b]</sup>	Neighbouring bases <sub>ds</sub> <sup>[a]</sup>	$\Phi_{ds}$ [%] <sup>[b,c]</sup>	$\langle \tau \rangle_{ds}$ [ns] <sup>[b,d]</sup>	Absorption max <sub>ds</sub> [nm] <sup>[b]</sup>	Emission max <sub>ds</sub> [nm] <sup>[b]</sup>
<b>AA</b>	5.8	1.3	337	432	<b>AA</b>	0.6	0.7	334	434
<b>GG</b>	5.5	2.0	337	433	<b>GG</b>	0.6	0.9	336	435
<b>GA</b>	4.4	2.0	337	432	<b>GC</b>	0.6	1.4	339	453
<b>AC</b>	2.0	0.6	338	434	<b>GA</b>	0.5	0.5	336	438
<b>GC</b>	1.6	1.3	337	438	<b>CA</b>	0.3	0.5	337	444
<b>CA</b>	1.0	0.8	338	433	<b>CT;TA</b>	0.3	1.3	337	457
<b>TA</b>	0.6	0.7	338	434	<b>AC</b>	0.2	0.4	337	441
<b>CT;TA</b>	0.4	0.6	337	439	<b>TA</b>	0.2	0.6	337	451
<b>TG</b>	0.4	0.6	337	437	<b>TG</b>	0.2	0.6	334	453
<b>CC</b>	0.4	0.4	338	440	<b>CT</b>	0.2	1.3	331	459
<b>CT</b>	0.3	0.5	337	443	<b>CC</b>	0.2	0.8	337	452
<b>TT</b>	0.2	0.7	338	434	<b>TT</b>	0.2	0.7	336	452

[a] Full sequences are listed in Table 1. [b] Samples were prepared in phosphate buffer (pH 7.5, 200 mM Na<sup>+</sup>, 17 mM phosphate). [c] Fluorescence quantum yields were determined relative to the reference quinine sulfate in 0.5 M H<sub>2</sub>SO<sub>4</sub> ( $\Phi_f=0.55$ ) at an excitation wavelength of 337 nm at 25°C, and are averages of three independent measurements. [d] Fluorescence decays were recorded by using an excitation wavelength of 330 nm and monitoring emission at 456 nm. For each sample, two independent measurements were obtained to confirm the reproducibility of the results. Decays were convoluted by using the IRF and fitted to a bi- or tri-exponential function. Detailed fitting parameters are presented in the Supporting Information.

well as the corresponding duplexes are listed in Table 4. Also, the absorption and emission maxima of each sequence are presented. The emission maxima of qA in both single- and double-stranded sequences (except duplexes **CT** and **CT;TA**) show a blue-shift compared to the free qA monomer in water ( $\lambda_{em,max}=456$  nm). This is according to expectations upon introducing qA to a less polar environment since TDDFT calculations suggest an increased dipole moment in the excited state.<sup>[38]</sup> Furthermore, slight variations for single-stranded samples (432–443 nm) and large shifts (434–459 nm) for duplexes can be observed between various sequences. In contrast, the absorption maxima are virtually the same for all single-stranded sequences (337–338 nm) and only slightly red-shifted compared to the free qA monomer ( $\lambda_{abs,max}=335$  nm). For the modified duplexes these numbers show more variation, and range from 331 to 339 nm, with slight red-shifts (except sequences **CT**, **AA** and **TG**) compared to the qA monomer. Similar observations have been made for the FBAs 2-AP<sup>[14]</sup> and 3-MI.<sup>[8]</sup> Moreover, redshifts are common for intercalating dyes, such as YO, YOYO<sup>[39]</sup> and ethidium bromide.<sup>[40]</sup>

A general decrease in the quantum yields and average fluorescence lifetimes are observed upon incorporation of qA into oligonucleotides. The extent of the decrease is highly dependent on the immediately surrounding bases (vide infra). Lifetimes of qA inside DNA were fitted to two or three exponentials, for which detailed fitting parameters are shown in the Supporting Information. The average lifetimes of qA vary between 0.4–2.0 ns in single strands and between 0.4–1.4 ns in duplexes (Table 4). The average lifetimes for all single- and double-stranded molecules (1.0 and 0.8 ns, respectively) are approximately three- and four-times shorter, respectively, than the lifetime of the qA monomer in water ( $\tau=3.2$  ns;  $\chi^2=1.684$ ). The observed general decreases in quantum yield and average lifetime upon incorporation into DNA are common features of most fluorescent base analogues.<sup>[2a,6,10a,14,24]</sup> A range of theories have been

suggested in the literature to explain this phenomenon. It has previously been shown that, for example, in single-stranded systems, the base-stacking effect could produce significant charge transfer by the molecular orbital overlap with the surrounding natural bases.<sup>[41]</sup>

In the double-stranded structure there is also the effect of the opposite strand to consider. Particularly in the case of 2-AP, the results of Ward et al. suggest that this quenching could be due to base-pairing.<sup>[14]</sup> Another explanation for the decrease in fluorescence intensity and the shortening of fluorescence lifetimes in a DNA-context could be the improved accessibility of a conical intersection (CI) between the excited state and the ground state<sup>[42]</sup> induced by the immediate surroundings in the DNA. It has been proven that such CIs occur in all natural nucleobases and in various fluorescent base analogues,<sup>[43]</sup> such as 2-AP.<sup>[44]</sup> Even though this hypothesis needs to be verified by theoretical calculations on qA, it is possible that in the monomeric form, the CI between S<sub>1</sub> and S<sub>0</sub> is not accessible to the excited population because of energy barriers that occur on the potential energy surface. These energy barriers could disappear when qA is incorporated into DNA systems, and thus a larger part of the excited population could relax non-radiatively through the CI. This kind of mechanism could explain the short time-components found here in the fitting parameters (see the Supporting Information).

An average decrease in fluorescence quantum yield of elevenfold was observed upon incorporation of the qA monomer ( $\Phi_{water,\lambda ex337\text{ nm}}=6.8\%$ ) into single-stranded DNA (Table 4). Moreover, qA shows a further sensitivity to its microenvironment as the quantum yields, average lifetimes and emission maxima of the sequences differ substantially depending on the bases neighbouring qA. As can be observed in Table 4, qA shows the highest quantum yield in oligonucleotides when it is surrounded by two purines (**AA**, **GG**, **GA**), followed by the group of sequences with only one purine neighbouring qA at the 5' side (**AC**, **GC**) and by the

group with one purine on the 3' side (**CA**, **TA**, **TG**). The lowest quantum yield values were recorded for the group of sequences (**CC**, **CT** and **TT**) with two pyrimidines flanking qA.

For the corresponding duplexes, a further fourfold decrease in average quantum yields can be observed compared to the single-stranded sequences (Table 4). Also, shorter average lifetimes were observed for half of the sequences (**AA**, **GG**, **GA**, **AC**, **CA** and **TA**). As for single strands, the highest quantum yield values were observed among the group of sequences with two purines flanking qA (**AA**, **GG**, **GA**). The **GC** sequence can also be found among the higher values, but all other double-stranded sequences show lower and similar quantum yields (0.2–0.3%). It should be noted that duplexes (**AA**, **GG** and **GA**) with two purines flanking qA, having the same stability as natural duplexes, are the sequences that on average show the highest quantum yield both in single- and double-stranded DNA. In contrast, sequences **CC**, **TT** and **CT** are the sequences that have the highest duplex stability and the lowest quantum yields. This can indicate a more firm stacking of qA with the neighbouring bases in the latter sequences, and hence explain an increased quenching efficiency.

The fact that **GG** is one of the sequences with the highest quantum yield is in stark contrast to observations made for the majority of fluorescent base analogues. Guanine is the natural base with the lowest oxidation potential and, thus, the highest susceptibility to oxidation.<sup>[45]</sup> This means that, depending on the oxidation potential of the neighbouring base analogue, G can transfer an electron into it, thereby quenching the excited state by photo-induced electron transfer (PET). For several common base analogues, such as the adenine analogues 2-aminopurine,<sup>[41c]</sup> triazole adenine,<sup>[24]</sup> the cytosine analogue 1C<sup>O[12a]</sup> and the base-discriminating analogue BPP,<sup>[4]</sup> a decreased quantum yield has been observed due to neighbouring guanines. It has also been suggested that the pteridine analogues 6MAP, DMAP and 3MI, act as excited state acceptors, which are preferentially quenched by purines.<sup>[46]</sup> We have not measured the redox potentials of qA and therefore cannot draw any conclusions regarding possible PET. One should remember though that even if the redox potential of qA had been determined, structural parameters, such as stacking with neighbouring bases and relative orientation, are also expected to play an important role for PET. Thus, knowledge of the redox potential of the qA monomer would give a clue to possible PET but could not tell the full story. However, as for many other FBAs we believe the quenching of qA inside DNA to be due to a combination of various mechanisms, all influenced by the surrounding sequence and base stacking. Therefore, we are currently investigating the details of the reduced lifetimes using a range of time-resolved techniques to obtain a more complete picture about the complex quenching mechanisms.

Importantly, qA shows significant fluorescence inside DNA compared to other adenine analogues. The maximum quantum yield value reached in single strands (5.8%) is approximately four-times higher than corresponding values re-

corded for nucleic acids containing 2-AP,<sup>[14]</sup> and also slightly higher compared to 6MAP.<sup>[9]</sup> In double-stranded DNA the maximum value recorded for qA (0.6%) is comparable to that of 2-AP.<sup>[14]</sup> The fluorescence decays of single- and double-stranded systems containing qA, just as for 6-MAP and DMAP, show more complex fitting parameters, and the dominant time component of those decays is shorter than that of the free monomer itself.<sup>[9]</sup> Furthermore, we recorded fluorescence melting curves for duplex **AA**; this yielded a virtually identical melting temperature to that recorded with UV/Vis melting ( $T_{m,UV}=46.3^{\circ}\text{C}$ ;  $T_{m,F}=46.8^{\circ}\text{C}$ ). This indicates no pre-melting at the qA-modified site and illustrates the applicability of qA as a fluorescent thymine-specific and non-perturbing adenine analogue.

## Conclusion

In conclusion, quadraacyclic adenine, qA, is an environment-sensitive non-perturbing fluorescent adenine analogue. The analogue is easy to selectively excite because its lowest energy excitation maximum (335 nm) is well resolved from the natural DNA absorption. Surprisingly, the quantum yield of qA varies with excitation energy, despite the fact that its emission always occurs from the  $S_1$  state. The free monomer of qA is moderately fluorescent ( $\Phi_{\text{water},\lambda_{\text{ex}}=337\text{ nm}}=6.8\%$ ) in comparison with other adenine analogues, but upon incorporation into oligonucleotides it reaches quantum yield values of up to four-times higher than corresponding values recorded for 2-AP. Furthermore, inside double-stranded DNAs the quantum yield of qA is of the same order as 2-AP in duplex systems.<sup>[14]</sup> Most importantly, the DNA duplex stability is either enhanced or preserved depending on the bases flanking qA. This is in contrast to the destabilisation caused by other fluorescent adenine analogues, such as 2-AP, 6MAP, DMAP, xA or triazole adenine, and is therefore a promising feature.<sup>[9,11c,14,24]</sup> To the best of our knowledge these properties of qA are unprecedented for fluorescent adenine analogues. Additionally, the highly selective base pairing between qA and thymine is an advantageous feature for many purposes. For the most common adenine analogue 2-AP and also for the recently developed triazole adenine, a lack of specificity has been reported.<sup>[14,24]</sup> In combination, these favourable properties of qA make it an interesting alternative to 2-AP in fluorescent experiments that require detailed measurements and a need to maintain the integrity of the DNA double helix. As a result of the promising properties found here, several derivatives of qA are currently being synthesised, and are aimed at even higher quantum yields but maintain the important ability to form stable duplexes and specific pairing with thymine.

## Experimental Section

**General:** The synthesis route to the quadraacyclic adenine monomer is presented in the results section (Scheme 1). Detailed synthesis proce-

dures, compound characterisation, incorporation of qA into oligonucleotides and purification of oligonucleotides are described in Scheme S1 in the Supporting Information.

**Sample preparation:** DNA samples were prepared in sodium phosphate buffer ( $\text{pH } 7.5$ , 200 mM  $\text{Na}^+$ , 17 mM phosphate) and their concentrations were determined by recording the absorption at 260 nm. The extinction coefficients of the various sequences were calculated by summation of the extinction coefficients of the individual natural bases and qA. This sum was multiplied by 0.9 to correct for base stacking interactions. The extinction coefficient of qA was determined to be  $10000 \text{ M}^{-1} \text{ cm}^{-1}$  at 260 nm in water. This was obtained by measuring the absorption of a solution with a known concentration (86.4  $\mu\text{M}$ ) of qA. The extinction coefficients that were used for the different bases at 260 nm were  $\epsilon_T = 9300$ ,  $\epsilon_C = 7400$ ,  $\epsilon_G = 11800$ ,  $\epsilon_A = 15300$  and  $\epsilon_{qA} = 10000 \text{ M}^{-1} \text{ cm}^{-1}$ . From these extinction coefficients, those for the different modified sequences were calculated to be:  $\epsilon_{GA} = 96800$ ,  $\epsilon_{CT} = 87400$ ,  $\epsilon_{GC} = 89600$ ,  $\epsilon_{CA} = 92800$ ,  $\epsilon_{GG} = 93600$ ,  $\epsilon_{CC} = 85700$ ,  $\epsilon_{TA} = 94500$ ,  $\epsilon_{AA} = 99900$ ,  $\epsilon_{AC} = 92800$ ,  $\epsilon_{TG} = 91400$ ,  $\epsilon_{TT} = 93100$  and  $\epsilon_{CTTA} = 103500 \text{ M}^{-1} \text{ cm}^{-1}$ . Extinction coefficients of the unmodified oligonucleotides were calculated in the same way.

**UV and fluorescence melting experiments:** Samples containing modified duplex DNA (2.5  $\mu\text{M}$ ) were prepared by mixing buffered solutions containing the single-stranded complementary strands, of which the non-modified strand was in 15 % excess. The excess was to assure hybridisation of all the modified strands for further fluorescence measurements. For natural DNA duplexes, the same procedure was applied by using equimolar amounts of single-stranded DNA. For UV melting experiments, duplex samples were heated to  $90^\circ\text{C}$ , and annealed by being cooled to  $5^\circ\text{C}$  at the rate of  $1^\circ\text{C min}^{-1}$ . Next, melting curves were recorded during heating of the samples to  $90^\circ\text{C}$ , followed by cooling to  $5^\circ\text{C}$  at a rate of  $0.5^\circ\text{C min}^{-1}$ . The temperature was kept at  $90^\circ\text{C}$  for 5 min between heating and cooling. The melting curves were measured on a Varian Cary 4000 spectrophotometer with a programmable multi-cell temperature block by using an absorption wavelength of 260 nm. Melting temperatures were calculated by using the half maximum and the maximum of the first derivative of the melting curves. The values presented in this paper are averages of at least two independent measurements. The fluorescence melting curves for duplex **AA** were recorded on a Varian Cary Eclipse instrument. Samples were annealed by being heated to  $90^\circ\text{C}$ , followed by cooling to  $15^\circ\text{C}$  at a rate of  $0.5^\circ\text{C min}^{-1}$ . Next, melting curves were recorded while the samples were heated to  $90^\circ\text{C}$ , followed by cooling to  $15^\circ\text{C}$  also at a rate of  $0.5^\circ\text{C min}^{-1}$ . The melting temperature was calculated as the maximum of the first derivative of the melting curve.

**Circular dichroism:** CD spectra of 2.5  $\mu\text{M}$  duplex samples, which were prepared as described above, were recorded on a Chirascan CD spectrophotometer. The spectra were recorded at  $25^\circ\text{C}$  for a wavelength range of 200 to 420 nm. Corrections for background contributions were made and final spectra are averages of at least five measurements at a scan rate of  $2 \text{ nm s}^{-1}$ .

#### Fluorescence measurements

**Steady state fluorescence measurements:** For all steady-state fluorescence measurements, a SPEX fluorolog 3 spectrophotofluorimeter (JY Horiba) was employed. Buffered solutions of qA-modified DNA duplexes were prepared as described above ( $\text{pH } 7.5$ , 200 mM  $\text{Na}^+$ , 17 mM phosphate). Samples containing the monomer in Milli-Q water or in methanol and buffered solutions of single-stranded modified oligonucleotides were set to an absorption of approximately 0.03 at the excitation wavelength. Quantum yields were determined relative to the quantum yield of quinine sulfate ( $\Phi_I = 0.55$ ) in  $\text{H}_2\text{SO}_4$  (0.5 M) at  $25^\circ\text{C}$ .<sup>[47]</sup> For the duplexes and single-stranded oligonucleotides an excitation wavelength of 337 nm was used and emission spectra were recorded between 340 and 700 nm. The quantum yield of the qA monomer in Milli-Q water and methanol was recorded as a function of seven different excitation wavelengths between 295 and 377 nm. The emission of the quinine sulfate reference was recorded between 305 and 700 nm.

Steady-state excitation spectra were recorded by exciting between 240 and 450 nm and monitoring the emission fixed at 456 nm. Glan polarisers were used to obtain the polarised excitation spectra necessary to calcu-

late the anisotropy,  $r$ , from Equation (1), in which the instrument correction factor is given by Equation (2) and  $I_{XY}$  is the excitation spectrum for which the subscripts  $X$  and  $Y$  denote polarisation directions of the excitation and emission light, respectively, and H and V refer to horizontal and vertical, respectively.

$$r = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}} \quad (1)$$

$$G = \frac{I_{HV}}{I_{HH}} \quad (2)$$

The anisotropy of the immobile qA monomer in  $\text{H}_2\text{O}$ /ethylene glycol (1:2 mixture) glass at  $-100^\circ\text{C}$ , also known as the fundamental anisotropy ( $r_0$ ) for a certain transition ( $i$ ) is related to the angle between the absorbing and the emitting transition moments ( $\alpha_i$ ) as given by Equation (3).

$$r_{oi} = \frac{1}{5}(3\cos^2 \alpha_i - 1) \quad (3)$$

**Time-resolved fluorescence measurements:** Time-resolved fluorescence measurements were performed by using time-correlated single photon counting (TCSPC). The excitation source was the third harmonic (330 or 295 nm) of a Tsunami mode-locked Ti-sapphire laser running at 80 MHz. The UV pulses (1–2 ps) were generated in a Tsunami frequency tripler by using two 0.5 mm type I BBO crystals. The fluorescence of the samples was spectrally filtered at 456 nm by a monochromator and detected by a microchannel plate photomultiplier Hamamatsu R3809U-50. The counts were fed into a multichannel analyser with 4096 channels (Lifespec, Edinburgh Analytical Instruments) where a minimum of 10000 counts were recorded in the top channel. All fluorescence decays were recorded in a time window of 10 ns and the spectral resolution of the monochromator was set from 5 to 11 nm. The intensity data were convoluted with the instrument response function (IRF), the full width half maximum (FWHM) of which is given at 50 ps and fitted to two- or three-exponential expressions with the program Fluofit Pro v.4 (PicoQuant, GmbH) for the free monomer in solution and the modified single strands. For the modified duplexes, a fitting program created by Thomas Gustavsson from the Francis Perrin laboratory at CEA-Saclay was used. This uses a simple normalised analytical expression:  $F(t) = \text{const} + a \times \exp(-t/t_1) + (1-a) \times b \times \exp(-t/t_2) + (1-a) \times (1-b) \times c \times \exp(-t/t_3) + (1-a) \times (1-b) \times (1-c) \times \exp(-t/t_4)$ . To describe the fluorescence decays in a quantitative way, we performed a nonlinear fitting/deconvolution procedure using bi- and triexponential functions, convoluted with the IRF. The average decay time with  $j=1, 2$  or 3 is given by Equation (4).

$$\langle \tau \rangle = \frac{\sum_{i=1}^j a_i \tau_i}{\sum_{i=1}^j a_i} \quad (4)$$

#### Acknowledgements

This work was supported by the Swedish Research Council and Olle Engkvist Byggmästare Foundation. Furthermore, we would like to thank Thomas Gustavsson from the Francis Perrin Laboratory CEA-Saclay in France for the use of his software to fit the fluorescence decays of the modified duplexes.

- [1] D. Onidas, D. Markovitsi, S. Marguet, A. Sharonov, T. Gustavsson, *J. Phys. Chem. B* **2002**, *106*, 11367–11374.
- [2] a) L. M. Wilhelmsson, *Q. Rev. Biophys.* **2010**, *43*, 159–183; b) R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579–2619.
- [3] S. Preus, K. Kilså, L. M. Wilhelmsson, B. Albinsson, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8881–8892.

- [4] A. Okamoto, Y. Saito, I. Saito, *J. Photochem. Photobiol. C* **2005**, *6*, 108–122.
- [5] a) N. J. Greco, Y. Tor, *J. Am. Chem. Soc.* **2005**, *127*, 10784–10785; b) S. G. Srivatsan, Y. Tor, *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053; c) S. G. Srivatsan, N. J. Greco, Y. Tor, *Angew. Chem.* **2008**, *120*, 6763–6767; *Angew. Chem. Int. Ed.* **2008**, *47*, 6661–6665; d) S. G. Srivatsan, H. Weizman, Y. Tor, *Org. Biomol. Chem.* **2008**, *6*, 1334–1338.
- [6] D. A. Berry, K. Y. Jung, D. S. Wise, A. D. Sercel, W. H. Pearson, H. Mackie, J. B. Randolph, R. L. Somers, *Tetrahedron Lett.* **2004**, *45*, 2457–2461.
- [7] a) F. Wojciechowski, R. H. E. Hudson, *J. Am. Chem. Soc.* **2008**, *130*, 12574–12575; b) R. H. E. Hudson, R. D. Viirre, Y. H. Liu, F. Wojciechowski, A. K. Dambenieks, *Pure Appl. Chem.* **2004**, *76*, 1591–1598.
- [8] S. L. Driscoll, M. E. Hawkins, F. M. Balis, W. Pfleiderer, W. R. Laws, *Biophys. J.* **1997**, *73*, 3277–3286.
- [9] M. E. Hawkins, W. Pfleiderer, O. Jungmann, F. M. Balis, *Anal. Biochem.* **2001**, *298*, 231–240.
- [10] a) M. E. Hawkins, *Cell Biochem. Biophys.* **2001**, *34*, 257–281; b) M. E. Hawkins, W. Pfleiderer, F. M. Balis, D. Porter, J. R. Knutson, *Anal. Biochem.* **1997**, *244*, 86–95.
- [11] a) A. T. Krueger, E. T. Kool, *J. Am. Chem. Soc.* **2008**, *130*, 3989–3999; b) H. B. Liu, J. M. Gao, E. T. Kool, *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402; c) J. M. Gao, H. B. Liu, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 11826–11831.
- [12] a) P. Sandin, K. Börjesson, H. Li, J. Mårtensson, T. Brown, L. M. Wilhelmsson, B. Albinsson, *Nucleic Acids Res.* **2008**, *36*, 157–167; b) P. Sandin, L. M. Wilhelmsson, P. Lincoln, V. E. C. Powers, T. Brown, B. Albinsson, *Nucleic Acids Res.* **2005**, *33*, 5019–5025.
- [13] K. Börjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson, L. M. Wilhelmsson, *J. Am. Chem. Soc.* **2009**, *131*, 4288–4293.
- [14] D. C. Ward, E. Reich, L. Stryer, *J. Biol. Chem.* **1969**, *244*, 1228–1237.
- [15] a) O. V. Petruskene, S. Schmidt, A. S. Karyagina, I. I. Nikolskaya, E. S. Gromova, D. Cech, *Nucleic Acids Res.* **1995**, *23*, 2192–2197; b) D. G. Xu, K. O. Evans, T. M. Nordlund, *Biochemistry* **1994**, *33*, 9592–9599.
- [16] a) S. M. Watanabe, M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6429–6433; b) L. C. Sowers, Y. Boulard, G. V. Fazakerley, *Biochemistry* **2000**, *39*, 7613–7620; c) S. M. Law, R. Eritja, M. F. Goodman, K. J. Breslauer, *Biochemistry* **1996**, *35*, 12329–12337.
- [17] A. Dallmann, L. Dehmel, T. Peters, C. Mugge, C. Griesinger, J. Tuma, N. P. Ernsting, *Angew. Chem.* **2010**, *122*, 6126–6129; *Angew. Chem. Int. Ed.* **2010**, *49*, 5989–5992.
- [18] R. P. Bandwar, S. S. Patel, *J. Biol. Chem.* **2001**, *276*, 14075–14082.
- [19] E. Deprez, P. Tauc, H. Leh, J. F. Mouscadet, C. Auclair, M. E. Hawkins, J. C. Brochon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10090–10095.
- [20] K. E. Augustyn, K. Wojtuszewski, M. E. Hawkins, J. R. Knutson, I. Mukerji, *Biochemistry* **2006**, *45*, 5039–5047.
- [21] P. Sandin, J. Tumpane, K. Börjesson, L. M. Wilhelmsson, T. Brown, B. Nordén, B. Albinsson, P. Lincoln, *J. Phys. Chem. C* **2009**, *113*, 5941–5946.
- [22] a) S. D. Gilbert, C. D. Stoddard, S. J. Wise, R. T. Batey, *J. Mol. Biol.* **2006**, *359*, 754–768; b) J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. J. Lilley, D. A. Lafontaine, *Chem. Biol.* **2006**, *13*, 857–868.
- [23] C. A. Buhr, M. D. Matteucci, B. C. Froehler, *Tetrahedron Lett.* **1999**, *40*, 8969–8970.
- [24] A. Dierckx, P. Dinér, A. H. El-Sagheer, J. D. Kumar, T. Brown, M. Grotli, L. M. Wilhelmsson, *Nucleic Acids Res.* **2011**, *39*, 4513–4524.
- [25] M. Hoffer, *Chem. Ber.* **1960**, *93*, 2777–2781.
- [26] Z. Kazimierczuk, H. B. Cottam, G. R. Revankar, R. K. Robins, *J. Am. Chem. Soc.* **1984**, *106*, 6379–6382.
- [27] J. M. Muchowski, M. C. Venuti, *J. Org. Chem.* **1980**, *45*, 4798–4801.
- [28] S. P. H. Mee, V. Lee, J. E. Baldwin, *Angew. Chem.* **2004**, *116*, 1152–1156; *Angew. Chem. Int. Ed.* **2004**, *43*, 1132–1136.
- [29] F. Seela, M. Zulauf, M. Sauer, M. Deimel, *Helv. Chim. Acta* **2000**, *83*, 910–927.
- [30] F. Seela, M. Zulauf, *Chem. Eur. J.* **1998**, *4*, 1781–1790.
- [31] S. Preus, K. Börjesson, K. Kilså, B. Albinsson, L. M. Wilhelmsson, *J. Phys. Chem. B* **2010**, *114*, 1050–1056.
- [32] F. A. Miannay, T. Gustavsson, A. Banyasz, D. Markovitsi, *J. Phys. Chem. A* **2010**, *114*, 3256–3263.
- [33] R. G. Eason, D. M. Burkhardt, S. J. Phillips, D. P. Smith, S. S. David, *Nucleic Acids Res.* **1996**, *24*, 890–897.
- [34] K. C. Engman, P. Sandin, S. Osborne, T. Brown, M. Billeter, P. Lincoln, B. Nordén, B. Albinsson, L. M. Wilhelmsson, *Nucleic Acids Res.* **2004**, *32*, 5087–5095.
- [35] N. P. Johnson, W. A. Baase, P. H. von Hippel, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3426–3431.
- [36] F. Seela, H. Thomas, *Helv. Chim. Acta* **1995**, *78*, 94–108.
- [37] B. L. Gaffney, R. A. Jones, *Biochemistry* **1989**, *28*, 5881–5889.
- [38] J. R. Lakowicz, in *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York, **2006**, pp. 205–235.
- [39] A. Larsson, C. Carlsson, M. Jonsson, B. Albinsson, *J. Am. Chem. Soc.* **1994**, *116*, 8459–8465.
- [40] M. J. Waring, *J. Mol. Biol.* **1965**, *13*, 269–282.
- [41] a) J. N. Wilson, Y. J. Cho, S. Tan, A. Cuppoletti, E. T. Kool, *Chem.-BioChem.* **2008**, *9*, 279–285; b) J. M. Jean, K. B. Hall, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 37–41; c) O. J. G. Somsen, v. A. Hock, v. H. Amerongen, *Chem. Phys. Lett.* **2005**, *402*, 61–65.
- [42] *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy* (Eds.: W. Domcke, D. R. Yarkony, H. Köppel), World Scientific Publishing, **2004**.
- [43] “Quantum Mechanical Studies of the Photophysics of DNA and RNA Bases”: K. A. M. Kistler, S. Matsika, in *Challenges and Advances in Computational Chemistry and Physics*, Vol. 7 (Ed.: T.-S. Y. Lee, D. M. York), Springer, Heidelberg (Germany), **2009**, pp. 285–339.
- [44] a) S. Perun, A. L. Sobolewski, W. Domcke, *Mol. Phys.* **2006**, *104*, 1113–1121; b) J. X. Liang, S. Matsika, *J. Am. Chem. Soc.* **2011**, *133*, 6799–6808.
- [45] S. Fukuzumi, H. Miyao, K. Ohkubo, T. Suenobu, *J. Phys. Chem. A* **2005**, *109*, 3285–3294.
- [46] M. Narayanan, G. Kodali, Y. J. Xing, M. E. Hawkins, R. J. Stanley, *J. Phys. Chem. B* **2010**, *114*, 5953–5963.
- [47] W. H. Melhuish, *J. Phys. Chem.* **1961**, *65*, 229–235.

Received: October 31, 2011

Published online: March 21, 2012