FLUORESCENCE QUENCHING IN 2-AMINOPURINE-LABELED MODEL DNA SYSTEMS

by

Jacob Michael Remington

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Chemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2019

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Kohler for helping me succeed at research during both my undergraduate and graduate studies. He has also encouraged me to continue pursuing mathematics as an undergraduate and molecular simulations as a graduate student. I would also like to my advisor Professor Callis for always being there to listen to my newest idea and providing insightful lessons about fluorescence quenching along the way. The rest of my graduate committee, professors Rebane, Grumstrup, and Stadie have provided excellent scientific dialogue and coursework. Professor Martin McCullagh has also offered his expertise to teach me how to do molecular dynamics simulation and I am very thankful for that help. My wife Emily has given me support at every stage and has been instrumental for me completing my thesis. Finally, my parents Sherry and Jim provided me with an upbringing that encouraged free thought, cooperation, and hard work which has allowed me to get to where I am today. This work was funded in parts by Grant No. CHE-1465277 from the NSF and by NSF grant CHE-1817500. I also appreciate the funding provided by MSU's Graduate School and the Office of the Vice President for Research and Economic Development

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

LIST OF ABBREVIATIONS

2Ap: 2-Aminopurine
A: Adenine
C: Cytosine
CD: Conformational Dynamics
CG: Conjugate Gradient
COM: Center Of Mass
CPG: Controlled Pore Glass
CT: Charge Transfer
DNMP: Dinucleoside Monophosphates
ET: Electron Transfer
FES: Free Energy Surface
FRET: Förster Resonance Energy Transfer
FWHM: Full Width at Half Maximum
G: Guanine
HMSM: Hidden Markov State Model
HP: Helical Parameters
IRF: Instrument Response Function
MD: Molecular Dynamics
MSM: Markov State Model
ODE: Ordinary Differential Equation
PES: Potential Energy Surface
QM: Quantum Mechanics
SASA: Solvent Accessible Surface Area
SD: Steepest Decent
SH: Structural Heterogeneity
ssDNA: Single-Stranded DNA
T: Thymine
TA: Transient Absorption
TCSPC: Time-Correlated Single-Photon Counting
TICA: Time-Lagged Independent Coordinate Analysis
TRF: Time-Resolved Fluorescence
UV: Ultraviolet

ABSTRACT

For the last 50 years changes to the fluorescence properties of 2-aminopurine have been used to probe the structure and dynamics of DNA. 2-Aminopurine's utility has arisen from the quenching of its emission when π-stacked with neighboring nucleobases. In the time-domain, the emission decay profile of 2-aminopurine requires multiple exponential decay components to model. Despite its extensive usage, the microscopic origin of the decay heterogeneity is not clear. In this thesis, steady-state absorption, fluorescence, and time-resolved fluorescence results are compared to multiple microsecond molecular dynamics simulations of 2-aminopurine-labeled adenine containing single-stranded DNA oligomers of varying length and position of the 2-aminopurine probe. First, previous reports of ultrafast electron transfer in π-stacked adenine oligomers are used to build a new model for quenching of 2-aminopurine that is π-stacked with adenine. For dinucleotides, a static distribution of unstacked structures combined with a distance dependent electron transfer mechanism is posited to explain the disperse emission decay timescales. Investigating the dinucleotides with molecular dynamics simulations analyzed with Markov state models quantify the structural heterogeneity of the dinucleotides. At least seven structures are sampled that could alter the quenching of 2-aminopurines's fluorescence. The Markov state models also demonstrate the timescales for transitions between these structures range from 1.6 to 25 ns, suggesting 2-aminopurine, with its monomer-like lifetime of 10 ns, is sensitive to the conformational dynamics of the dinucleotides as well. This dual fluorescence quenching and molecular dynamics simulation approach is extended to 2-aminopurine labeled trinucleotides and 15 base oligomers to interrogate the position dependent structural heterogeneity and conformational dynamics in these systems. Both shifts in the experimental absorption spectra, and molecular dynamics simulations agree that the interior base is more likely to be stacked than the exterior bases. Time-resolved emission experiments reveal emission from 2-aminopurine is quenched faster on the 5' end relative to the 3' end, in agreement with the faster stacking kinetics observed for bases on the 5' end relative to the 3' end obtained from molecular dynamics simulation. These results suggest that the time-resolved emission from 2-aminopurine may serve as an experimental observable for calibration of the dynamical properties predicted by molecular dynamics simulation.

CHAPTER ONE

INTRODUCTION

Electromagnetic radiation from the sun is absorbed by the earth and its inhabitants providing essential energy for life to exist. While absorption of the UV portion of this light causes most organic molecules to undergo photochemical reactions, the four canonical nucleobases that compose DNA, adenine (A), thymine (T), cytosine (C), and guanine (G), are highly photostable due to efficient ultrafast (< ps) internal conversion pathways that transform the UV photon energy into heat.[1-2] Yet the DNA double helix is not just the sum of its parts, and based on the structure of the DNA at the instant of photoexcitation, new channels for deactivation are possible.[3] Understanding the different structures DNA can adopt, i.e. its structural heterogeneity (SH), is therefore an important aspect to recognize how DNA achieves its primary function to protect life's genetic information. Also crucial to DNAs function is the ability for the stored information to be read. In eukaryotes this is achieved by a structural transition from its coiled, fully base paired, and double stranded form, to two locally unwound and unpaired single strands of DNA (ssDNA) with the aid of RNA polymerases.[4] The transitions between distinct structures, or conformational dynamics (CD), is ubiquitous in biological molecules.[5-9] Here, the problem of quantifying the structural and dynamical properties of a particular form of DNA, ssDNA oligomers, will be addressed. It will be shown that both picosecond time-resolved fluorescence (TRF) measurements of oligomers labeled with fluorescence probe 2-aminopurine (2Ap) and molecular dynamics (MD) simulations can quantify the SH and CD present in these systems.

2Ap is a widely used fluorescence probe because it is a constitutional isomer of A, has a lowest energy absorption band red-shifted ~50 nm from the canonical nucleobases, and its emission quantum yield is strongly affected by interactions with nearby bases.[10] It is well understood that base stacking of 2Ap with any of A, T, C, or G, causes emission quenching[11] Specifically, the quenching mechanism is likely due to electron transfer to form dark excited states which can deactivate non-radiatively.[12-14] Interaction with the neighbor nucleobases also causes minor shifts to 2Ap's absorption and emission bands.[15-16] In TRF measurements the emission decay of a free 2Ap nucleobase is monoexponential with an excited state lifetime of 10 ns in aqueous solvent.[17] However, when 2Ap is incorporated into a DNA oligomer, its emission decay is modeled with three to five discrete exponential components.[10, 12, 14, 18-2224] Generally the leading models in the literature assume the decay heterogeneity arises from a combination of the SH and CD of the DNA.[10] However, the multi-exponential emission decay profile is not consistent with the fact that the SH of these systems is commonly considered using only two-state models in which the bases are either stacked or unstacked.[23-25] In Chapter 3 it will be shown that these two viewpoints can be reconciled by leveraging knowledge from the excited state deactivation of the natural base containing oligomers.

A leading model for the excited state deactivation of the natural base containing dinucleotides assumes that when the nucleobases are π-stacked, their interaction with UV light initially forms a Frenkel exciton that rapidly relaxes to charge transfer (CT) states between only two bases within 200 fs.[26] These CT states then relax via back electron transfer on a 10-100s of ps timescale based on their identity.[27] Finally, only when the

DNA is denatured so that the bases are not stacked, does the deactivation mechanism resemble the native bases.[26] These pathways are described in Figure 1.1 along with sample π-stacked and extended structures. In longer ssDNA oligomers the deactivation is dominated by the same pathways and excitation is presumed to only extend to two bases if they are π-stacked.[28] Finally, the structure of double stranded DNA can also affect UV-deactivation when two bases (of particular identities) in adjacent strands are base paired, allowing the formation of proton coupled electron transfer states[29].

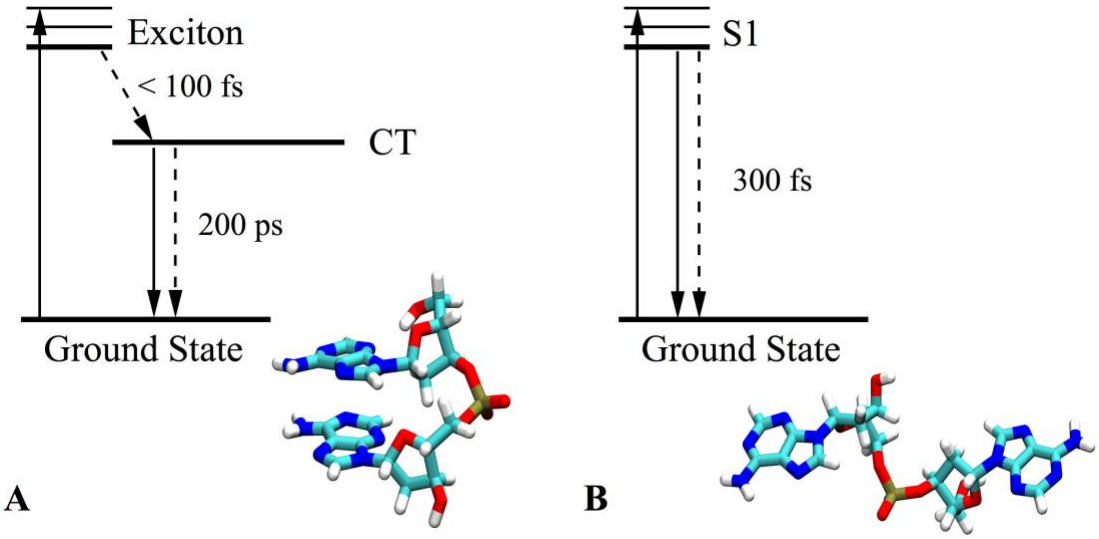

Figure 1.1 (A) Sample stacked and (B) unstacked structures of dA_2 along with the corresponding dominate excited state decay pathways following UV excitation. Radiative transitions are solid arrows while non-radiative transitions are dashed. Timescales shown are the lifetimes of the given states.

The observation that CT state formation is so rapid between two π-stacked A suggests that a 2Ap π-stacked with A ought to form CT states on a timescale too rapid for detection in picosecond TRF measurements. Thus in Chapter 3 it is demonstrated that the emission decay from the 2Ap and A dinucleotides is consistent with a model where

fluorescence is solely from unstacked bases that are quenched with a rate that depends on the distance between the two residues.[14] Despite the agreement of the new model proposed by us with TRF measurements at room temperature[14] it is a purely static model that neglects the possibility for transitions between structures that modulate the decay rate of 2Ap on timescales less that 10 ns. This assumption is not supported by temperature dependent measurements which show the quenching is generally faster at higher temperatures,[14, 30] even though the fewer stacked bases are expected at higher temperatures.[25, 31]

To independently check whether the 2Ap-labeled dinucleotides sample only two structures which interconvert on timescales greater than 10 ns timescales, MD simulations are performed in Chapter 4 with widely accepted nucleic acid forcefields.[32-34] These MD forcefields approximate the Born-Oppenheimer potential energy surface (PES) with pair-wise potentials that are parameterized to reproduce chosen experimental observables.[35-36] A total of 10 μs of simulation time for each 2Ap-labeled dinucleotide demonstrate the two-state model for dinucleotides is insufficient and at least seven distinct structures are sampled.[37] Furthermore, structures intermediate between stacked and unstacked ones are observed that could affect the quenching rate of 2Ap. These results are consistent with, and improve upon, other simulations which predict multiple structures for similar model RNA dinucleotides.[38]

Not only do the MD simulations reveal the SH present in dinucleotides, they also provide insight into the CD. Previous studies have shown that the longest timescale processes in biomolecular CD are well described by a Markovian description of the dynamics, e.g. one where the transition rate between any two structures is time

independent.[39] These memoryless models are called Markov state models (MSMs) and have become quite popular in recent years.[40-44] By constructing MSMs from the MD simulations, the assumption that the transition rates between the seven structures are much slower than 10 ns is tested. The simulations in Chapter 4 demonstrate that this approach predicts multiple CD timescales ranging from 1.6 to 25 ns, well within the timescale of 2Ap emission.[37] To combine the SH and CD from the MSM with a quenching model for 2Ap, the master equation approach is used to predict 2Ap's excited state population, and is shown to predict multiple sub 10 ns decay lifetimes for the 2Ap-labeled dinucleotides. Not only does this study demonstrate the importance of SH and CD in 2Ap quenching, it also suggests that TRF measurements of 2Ap could be used as an experimental observable to test the dynamical properties predicted by MD simulation.

To further test the utility of the dual TRF and MD simulation approach, studies are described in Chapter 5 for 2Ap-labeled trinucleotides and 15-base oligomers otherwise containing only A. By placing 2Ap at different positions along the ssDNA oligomers, the position dependent SH and CD is explored. Steady-state fluorescence measurements of similar 2Ap-labeled pentamers have been reported previously and demonstrate the 2Ap at the interior positions are more likely to be stacked than 2Ap at the ends.[45] This is in agreement with transient absorption (TA) studies of the stacking heterogeneity in ssDNA A containing oligomers which found that A at the interior positions is more likely to be stacked.[28] However, experimental TRF measurements of similar 30-base oligomers have shown 2Ap at these interior positions are actually quenched slower than 2Ap at the ends.[46] Steady-state and TRF measurements shown in

Chapter 5 reproduce these findings in the 15-base oligomers, but 2Ap at the interior position of the trinucleotide is actually quenched faster than 2Ap on the ends. After describing a novel nearest neighbor stacking model for the ssDNA that includes the seven structures predicted for the dinucleotides, 15.5 microseconds of MD simulation were run for the trinucleotide system. These simulations support the conclusions that bases are stacked more often at the interior of the strand. Building a MSM from the trinucleotide simulations demonstrates multi-scale stacking kinetics are consistent with the position dependent trends in the experimental TRF measurements. Finally, the ability of current standard MD simulations to completely sample the PES for the 15 base oligomers is commented on.

1.1 References

1. Pecourt, J. M. L.; Peon, J.; Kohler, B., Ultrafast Internal Conversion of Electronically Excited RNA and DNA Nucleosides in Water. *J. Am. Chem. Soc.* **2000,** *122*, 9348-9349.

2. Pecourt, J. M. L.; Peon, J.; Kohler, B., DNA Excited-State Dynamics: Ultrafast Internal Conversion and Vibrational Cooling in a Series of Nucleosides. *J. Am. Chem. Soc.* **2001,** *123*, 10370-10378.

3. Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernàndez, C. E.; Kohler, B., DNA Excited-State Dynamics: From Single Bases to the Double Helix. In *Annu. Rev. Phys. Chem.*, Annual Reviews: Palo Alto, 2009; Vol. 60, pp 217-239.

4. Hahn, S., Structure and Nechanism of the RNA Polymerase II Transcription Machinery. *Nat. Struct. Mol. Biol.* **2004,** *11*, 394-403.

5. Dror, R. O.; Mildorf, T. J.; Hilger, D.; Manglik, A.; Borhani, D. W.; Arlow, D. H.; Philippsen, A.; Villanueva, N.; Yang, Z. Y.; Lerch, M. T.; Hubbell, W. L.; Kobilka, B. K.; Sunahara, R. K.; Shaw, D. E., Structural Basis for Nucleotide Exchange in Heterotrimeric G Proteins. *Science* **2015,** *348*, 1361-1365.

6. Zeng, X. J.; Zhang, L. Y.; Xiao, X. C.; Jiang, Y. Y.; Guo, Y. Z.; Yu, X. Y.; Pu, X. M.; Li, M. L., Unfolding Mechanism of Thrombin-Binding Aptamer Revealed by Molecular Dynamics Simulation and Markov State Model. *Sci Rep* **2016,** *6*, 14.

7. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct Observation of Hierarchical Protein Dynamics. *Science* **2015,** *348*, 578-581.

8. Uversky, V. N., Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **2016,** *291*, 6681-6688.

9. Liao, C.; May, V.; Li, J., PAC1 Receptors: Shapeshifters in Motion. *Journal of molecular neuroscience : MN* **2018**.

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

11. Jean, J. M.; Hall, K. B., 2-Aminopurine Fluorescence Quenching and Lifetimes: Role of Base Stacking. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 37-41.

12. Wan, C.; Xia, T.; Becker, H. C.; Zewail, A. H., Ultrafast Unequilibrated Charge Transfer: A New Channel in the Quenching of Fluorescent Biological Probes. *Chem. Phys. Lett.* **2005,** *412*, 158-163.

13. Madhavan Narayanan, G. K., Yangjun Xing, and Robert J. Stanley, Photoinduced Electron Transfer Occurs between 2-Aminopurine and the DNA Nucleic Acid Monophosphates: Results from Cyclic Voltammetry and Fluorescence Quenching. *J. Phys. Chem. B.* **2010,** *114*, 10573-10580.

14. Remington, J. M.; Philip, A. M.; Hariharan, M.; Kohler, B., On the Origin of Multiexponential Fluorescence Decays from 2-Aminopurine-Labeled Dinucleotides. *J. Chem. Phys.* **2016,** *145*, 155101.

15. Evans, K.; Xu, D.; Kim, Y.; Nordlund, T. M., 2-Aminopurine optical spectra: Solvent, pentose ring, and DNA helix melting dependence. *J. Fluoresc.* **1992,** *2* (4), 209-216.

16. Kervin Evans, D. X., Younsik Kim, Thomas M. Nordlund, 2-Aminopurine Optical Spectra: Solvent, Pentose Ring, and DNA Helix Melting Dependence. *J. Fluoresc.* **1992,** *2*, 209-216.

17. D. C. Ward, E. R. a. L. S., Fluorescence Studies of Nucleotides and Polynucleotides: I. Formycin, 2-Aminopurine Riboside, 2,6-Diaminopurine Riboside, and Their Derivatives. *J. Biol. Chem.* **1969,** *244*, 1228-1237.

18. Greiner, V. J.; Kovalenko, L.; Humbert, N.; Richert, L.; Birck, C.; Ruff, M.; Zaporozhets, O. A.; Dhe-Paganon, S.; Bronner, C.; Mély, Y., Site-Selective Monitoring

of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry* **2015,** *54*, 6012-6020.

19. Voltz, K.; Léonard, J.; Touceda, P. T.; Conyard, J.; Chaker, Z.; Dejaegere, A.; Godet, J.; Mély, Y.; Haacke, S.; Stote, R. H., Quantitative Sampling of Conformational Heterogeneity of a DNA Hairpin Using Molecular Dynamics Simulations and Ultrafast Fluorescence Spectroscopy. *Nucleic Acids Res.* **2016,** *44*, 3408-3419.

20. Tolbert, M.; Morgan, C. E.; Pollum, M.; Crespo-Hernández, C. E.; Li, M. L.; Brewer, G.; Tolbert, B. S., HnRNP Al Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation. *J. Mol. Biol.* **2017,** *429*, 2841-2858.

21. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

22. Jean, J. M.; Krueger, B. P., Structural Fluctuations and Excitation Transfer between Adenine and 2-Aminopurine in
Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B.* **2006,** *110*, 2899-2909.

23. Pörschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

24. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 2315-2328.

25. Pörschke, D., Dynamics of Nucleic-Acid Single-Strand Conformation Changes - Oligoriboadenylic and Polyriboadenylic Acids. *Eur. J. Biochem.* **1973,** *39*, 117-126.

26. Chen, J.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

27. Takaya, T.; Su, C.; de La Harpe, K.; Crespo-Hernàndez, C. E.; Kohler, B., UV Excitation of Single DNA and RNA Strands Produces High Yields of Exciplex States Between Two Stacked Bases. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 10285-10290.

28. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B.* **2012,** *116*, 10266-10274.

29. Zhang, Y. Y.; Li, X. B.; Fleming, A. M.; Dood, J.; Beckstead, A. A.; Orendt, A. M.; Burrows, C. J.; Kohler, B., UV-Induced Proton-Coupled Electron Transfer in Cyclic DNA Miniduplexes. *J. Am. Chem. Soc.* **2016,** *138*, 7395-7401.

30. Somsen, O. J.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Structural Heterogeneity in DNA: Temperature Dependence of 2-Aminopurine Fluorescence in Dinucleotides. *Chemphyschem* **2005,** *6*, 1622-1627.

31. Kang, H.; Chou, P. J.; Johnson, W. C.; Weller, D.; Huang, S. B.; Summerton, J. E., Stacking Interactions of ApA Analogs with Modified Backbones. *Biopolymers* **1992,** *32*, 1351-1363.

32. Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007,** *92*, 3817-3829.

33. Chen, A. A.; García, A. E., High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *P. Natl. Acad. Sci. USA* **2013,** *110*, 16820-16825.

34. Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8*, 2506-2520.

35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

36. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, et al *AMBER 2018*, University of California, San Francisco, 2018.

37. Remington, J. M.; McCullagh, M.; Kohler, B., Molecular Dynamics Simulations of 2-Aminopurine-Labeled Dinucleoside Monophosphates Reveal Mutliscale Stacking Kinetics. *J. Phys. Chem. B* **2019,** *123*, 2291-2304.

38. Hayatshahi, H. S.; Henriksen, N. M.; Cheatham, T. E., Consensus Conformations of Dinucleoside Monophosphates Described with Well-Converged Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018,** *14*, 1456-1470.

39. Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R., Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and Beta-hairpin Peptide. *J. Phys. Chem. B.* **2004,** *108* (21), 6582-6594.

40. Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S., Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009,** *131*, 124101.

41. Pande, V. S.; Beauchamp, K.; Bowman, G. R., Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010,** *52*, 99-105.

42. Chodera, J. D.; Noé, F., Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

43. Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011,** *134*, 174105.

44. Husic, B. E.; Pande, V. S., Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018,** *140*, 2386-2396.

45. Steven P. Davis, M. M., Andreia Williams, and Thomas M. Nordlund, Position Dependence of 2-Aminopurine Spectra in Adenosine Pentadeoxynucleotides. *J. Fluoresc.* **2003,** *13*, 249-259.

46. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B.* **2007,** *111*, 5757-5766.

CHAPTER TWO

GENERAL METHODS

Preface

The contents of this chapter describe the details of both the experimental (Sections 2.1-3) and theoretical (Sections 2.4 - 5) procedures and the various data analysis techniques used to understand the SH and CD of the ssDNA oligomers. The spectroscopic techniques used are described in Section 2.1 while the methods for sample preparation are in Section 2.2. Section 2.3 is devoted to the analysis of experimental data, focusing on the mathematical relationships describing excited state models. Section 2.4 describes details for the MD simulations used, while Section 2.5 outlines the analysis methods used to condense the vast amounts of data generated via MD simulation.

2.1 Spectroscopy

2.1.1 UV/Visible Absorption Measurements

All UV/Visible absorption spectra were obtained with a Lambda 25 spectrometer (PerkinElmer, Inc.) using a 1 cm pathlength Spectrosil® quartz cuvette. Care was taken to avoid damage and smudging of the windows to ensure consistent reflectivity. To background a scan, the absorbance was set to zero at all wavelengths when only air was in the reference and sample holders, and then a scan was run with the solution identical to the DNA solution but not containing the DNA placed in the sample holder. These scans were then subtracted from scans which included the DNA to obtain the DNA spectrum.

2.1.2 UV/Visible Fluorescence Measurements

Steady-state emission spectra were recorded with a fluorometer (Fluorolog 3, Horiba Jobin Yvon). The emission was collected at 90° with respect to the excitation light. The obtained raw emission spectra were background subtracted by recording an emission spectra without any solute present but the same solvent. Then the emission spectra were corrected by multiplication by a wavelength-dependent correction factor accounting for the sensitivity of the detector. This correction factor was determined by comparison of measured emission spectra to reference spectra and is described elsewhere.[47] The absorbance was always kept below 0.01 O.D.

2.1.3 Time Resolved Fluorescence Measurements

Fluorescence decay curves were obtained using an emission lifetime spectrometer (FluoTime200, PicoQuant Photonics North America) that uses the time-correlated single-photon counting (TCSPC) technique. Excitation pulses were provided by a widely tunable femtosecond laser oscillator (Chameleon Ultra II, Coherent) with an 80 MHz pulse repetition rate. Third harmonic pulses at 310 nm were produced using an auto-tracking harmonic generator. The laser beam was passed through an electro-optic modulator (Model 350-80, Conoptics) connected to a bias/drive controller (Model 25D, Conoptics). A synchronous countdown timer (Model 305, Conoptics) selected every 30th pulse. The polarization of the excitation pulse was set to vertical before the sample. The focused excitation pulse in these experiments had a spot size of 115 ± 10 μm (FWHM) and a power of 330 ± 10 μW, resulting in a maximum peak intensity of about 5.4 MW/cm^2. The light emitted at ~90° was sent through a 350 nm (model XLU0350, Asahi

Spectra, Torrance, CA) long pass filter, then through a polarizer set at the magic angle from vertical, and then through a monochromator to pick out the desired emission wavelength. Finally, the emitted light was detected with a microchannel plate photomultiplier tube (model R3809U-50, Hamamatsu Corp.) positioned past the exit slit of a monochromator. PicoHarp software (PicoQuant) was used to record the data with an 8 ps bin width. A Ludox® scattering solution was used to record the instrument response function (IRF), which was measured to be 65 ± 1 ps (FWHM).

2.2 Sample Preparation

2.2.1 Samples

The deoxyribose form of 2Ap (2-amino-9-(β-D-2-deoxyribofuranosyl)purine, ≥97% purity, was purchased from Berry & Associates (Dexter, MI). The ssDNA oligomers labeled with 2AP used in these studies were provided by professor Mahesh Hariharan (Indian Institute of Science Education and Research Thiruvananthapuram). These samples were synthesized on a K&A Laborgeraete DNA synthesizer employing the automated solid-phase phosphoramidite chemistry. The 2Ap and 2′-deoxyadenosine phosphoramidite were purchased from Glen research and used as received. The fluorescent label 2Ap was incorporated into the dinucleotides on a 1 μmol scale with appropriate controlled pore glass (CPG) beads used as 3′ solid support as reported earlier.[48] Oligonucleotides were de-protected and isolated from the solid-phase support using concentrated ammonium hydroxide with stirring for 24 h at room temperature. The 2Ap-labeled dinucleotides were then purified by reverse phase high performance liquid chromatography (Prominence HPLC System, Shimadzu) with a Phenomenex column

(Luna 5u C18(2) 100A, 250 × 10 mm), employing a gradient of 20 mM ammonium acetate buffer and acetonitrile (flow rate of 1 ml/min) as mobile phase.

2.2.2 Solvents

When aqueous buffers were used as a solvent, a 100 mm NaCl, 100 mM tris-HCl buffer (pH 7.45) made with water from an ultrapurifier (Synergy, EMD Millipore) was used. For the trinucleotides and 15 base oligomers the buffer had a pH of 6.98. This was done by dissolving tris in the water and then adjusting the pH with HCl. The pH was determined with a MI-410 Micro Combination pH electrode (Microelectrodes Inc., Bedford NH) with an OrionStar A121 pH meter (Thermo Scientific, Waltham, MA). When HPLC grade methanol (EMD Chemicals, Philadelphia, PA) was used as a solvent no additional salt or buffer was added. The 2Ap concentration was adjusted by dilution to provide an absorbance of about 0.01 at its peak absorption wavelength 310 nm.

2.3 Experimental Data Analysis

2.3.1 Fluorescence Quantum Yields

To calculate the emission quantum yield of 2Ap in a oligomer Φ_s, its integrated emission spectra, F_s, was compared to the integrated emission of a 2Ap reference sample, F_r, which has a known reference quantum yield, Φ_r, of 0.68 in water.[17] These quantities are related through the following equation:

$$\Phi_s = \Phi_r \left(\frac{F_s}{F_r}\right) \tag{2.1}$$

In Equation 2.1 it is assumed that the index of refraction of the aqueous buffer is the same as the buffer used in ref 17 and that the two samples have the same absorbance at the excitation wavelength.

2.3.2 Excited State Modeling

Here various excited state models are defined that will be used to fit experimental TRF curves. Generally, these models focus on the 2Ap excited state manifold which we assume is dominated by the excited state population of the S1 state that produces emission observable in the TRF experiments. When a bulk sample of 2Ap is excited by an instantaneous pulse of light, its time dependent excited state population, $P(t)$, is subject to parallel deactivation via the radiative, k_r, and non-radiative, k_{nr}, rates. The time dependent rate of light emitted by this population is therefore just $P(t)$ multiplied by k_r. Here we describe models for the $P(t)$ acknowledging that the signals measured are proportional to the aforementioned product of $P(t)$ and k_r, with proportionality constant depending on the precise experimental setup.

When 2Ap is isolated in aqueous solvent its excited state population decays via the solution to a first-order, linear, one-dimensional, ordinary differential equation (ODE), i.e. a single exponential decay:

$$P(t) = P(0)e^{-t/\tau} \tag{2.2}$$

In Equation 2.2 there are two parameters, the initial population (proportional to the amplitude, A, used in fitting) and the lifetime, τ, that is the inverse sum of k_r and k_{nr}. However, when 2Ap is attached to a ssDNA molecule, the environment surrounding the probe can be different when the ssDNA is in different conformations providing different k_{nr} for the distinct populations. In this case there are n different populations of excited 2Ap and in the limit that the structures are static the excited state population is given by Equation 2.3.

$$P(t) = \sum_{i=1}^{n} P_i(0)e^{-t/\tau_i} \tag{2.3}$$

Here the *n* pre-exponential factors are the initial excited state populations of 2Ap in each conformation (again proportional to the A_i used to fit experimental data) whose relative values are equal to the ground state equilibrium distribution according to the Franck-Condon principle. Similarly the lifetimes, τ_i, are different for each conformation because of the differences in k_{nr}. It is noted that Equation 2.3 is the solution to an uncoupled first-order, linear, *n*-dimensional, ODE.

However, Equation 2.3 and the interpretation that fitted A_i are proportional to the excited state population of 2Ap in a particular ssDNA structure is not correct if the structures of the ssDNA are not static and instead dynamically interconvert. To model this case, a coupled, first-order, linear, *n*-dimensional, ODE must be solved.[49] The steps to do so are shown in the following for completeness and its importance to this dissertation. The ODE to solve is the following:

$$\frac{d}{dt}\begin{bmatrix} P_1(t) \\ \vdots \\ P_n(t) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\tau_1} - \sum_{j\neq 1} k_{1j} & \cdots & k_{n1} \\ \vdots & \ddots & \vdots \\ k_{1n} & \cdots & -\frac{1}{\tau_n} - \sum_{j\neq n} k_{nj} \end{bmatrix} \begin{bmatrix} P_1(t) \\ \vdots \\ P_n(t) \end{bmatrix} \tag{2.4}$$

The additional parameters here are the n^2-n off-diagonal k_{ij} representing the rate constant for transitions from structure *i* to structure *j*. The subtraction of the sum of off-diagonal rows in Equation 2.4 ensures that the system satisfies detailed balance. The matrix in Equation 2.4 will be called the rate matrix, $\boldsymbol{K}$, and the population vector will be denoted with an arrow. First, Equation 2.4 is rewritten in matrix form:

$$\frac{d}{dt}\vec{P} = \boldsymbol{K}\vec{P} \tag{2.5}$$

An eigenvalue decomposition for the matrix $\boldsymbol{K}$ is computed to produce the diagonal matrix $\boldsymbol{\Lambda}$ containing the eigenvalues, the matrix $\boldsymbol{V}$ containing the eigenvectors, and its inverse $\boldsymbol{V^{-1}}$:

$$\frac{d}{dt}\vec{P} = \boldsymbol{V\Lambda V^{-1}}\vec{P} \tag{2.6}$$

Multiplying each side of Equation 2.6 by $\boldsymbol{V^{-1}}$ and introducing the coordinate transformed population vector $\vec{Q} = \boldsymbol{V^{-1}}\vec{P}$ decouples the ODE:

$$\frac{d}{dt}\vec{Q} = \boldsymbol{\Lambda}\vec{Q} \tag{2.7}$$

Now Equation 2.7 can be readily solved to provide time dependent solutions to Q_i(t):

$$Q_i(t) = Q_i(0)e^{-t/\lambda_i} \tag{2.8}$$

Transforming back to the original population vectors provides the final solution for the excited state population where V_i is the i^{th} left-eigenvector of $\boldsymbol{K}$:

$$P(t) = \sum_{i=1}^{n}\left(\vec{P}(0)\cdot V_i\right)e^{-t/\lambda_i} \tag{2.9}$$

Comparison of Equations 2.3 and 2.9 demonstrates two key differences that are crucial for proper modeling of dynamical fluorescence quenching. First, the lifetimes in Equation 2.3 are not the eigenvalues in Equation 2.9, which means the lifetimes obtained through fitting of Equation 2.3 to experimental data don't necessarily extract the conformational specific decay rates for the 2Ap probe. Secondly, the pre-exponential factors in Equation 2.9 are the dot products of the initial population vector onto the eigenspace of the rate matrix, which again implies that fits of Equation 2.3 to extract the initial population of 2Ap in different structures are incorrect when the structures interconvert. However, while it would be ideal to fit Equation 2.4 to experimental TRF data, it has too many free

parameters, and instead the possibility of independently estimating the interconversion rates from MD simulations will be explored in Chapters 4 and 5 for specific systems.

A feature of the proceeding analysis is the presumption that there are a finite number of structures the 2Ap-labeled ssDNA can adopt. However, as has been argued by others[50] and again by us in Chapter 3, there could in fact be a whole continuum of structures providing slightly different k_{nr} values for 2Ap. The natural extension of the discrete model in Equation 2.3, again assuming no interconversion, is:

$$P(t) = \int_0^{\tau_{max}} f(\tau) e^{-t/\tau} d\tau \tag{2.10}$$

Equation 2.10 is closely related to the Laplace transformation of the lifetime distribution, $f(\tau)$, but with slightly modified integration bounds that we proposed to account for the finite lifetime of a free 2Ap probe. A physically insightful way to think about the $f(\tau)$ is to consider a functional relationship between a structural parameter, x, and k_{nr} , so that an experimental estimate of $f(\tau)$ can be used to estimate $f(x)$, thereby providing structural insight. For an example see Chapter 3 where x is the center of mass distance between 2Ap and A and k_{nr} decays exponentially with distance according to a simple electron transfer rate model.

2.3.3 Modeling Time Resolved Fluorescence Traces

Possibly the most important aspect that should be considered when trying to model TRF data of 2Ap-labeled oligomers obtained with the TCSPC technique described in Section 2.1.3 is the convolution of the excitation pulse with excited state model:[51]

$$F(t) = \int_{-\infty}^{t} (R(t' - \gamma) - F_{IRF}) M(t - t') dt' + F_{\text{bkg}} \quad (2.11)$$

$M(t)$ is proportional to the chosen excited state population models, $P(t)$, described in Section 2.3.2., F_{bkg} is the background emission that represents a constant offset, and the excitation pulse, or IRF descried in 2.1.3, is $R(t)$. Also γ is the time-shift parameter between excitation and emission traces, while F_{IRF} is a constant offset for the IRF (usually <10 counts).

In Equation 2.11, $F(t)$ is the model to be compared to the experimental signal, $S(t)$, by minimizing the reduced chi-squared parameter, χ^2, through adjustment of the p parameters in the model:

$$\chi^2 = \frac{1}{m-p} \sum_{i=1}^{m} \frac{1}{\sqrt{S(t_i)}} \left(S(t_i) - F(t_i)\right)^2 \quad (2.12)$$

It was found through contact with an engineer at PicoQuant that the software provided to fit experimental decays evaluates Equation 2.11 by directly applying Gaussian quadrature. For the case when $M(t)$ decays faster than the typical length of the excitation pulse, this produces incorrect results for $F(t)$. Furthermore, approximating the integral in Equation 2.10 with quadrature also can give improper results if $f(\tau)$ has a tail that slowly decays at large τ. Instead of using this software we chose to write Matlab® scripts that apply a simple approximation to the integral in Equation 2.11. First we take into account that $R(t)$is a measured histogram with m bins:

$$R(t) = \begin{cases} R_1 & t_0 \le t < t_1 \\ R_2 & t_1 \le t < t_2 \\ \vdots & \vdots \\ R_m & t_{m-1} \le t < t_m \\ 0 & \text{Otherwise} \end{cases} \tag{2.13}$$

Here it is noted that Equation 2.13 is a 0th order approximation for $R(t)$, and higher order approximations obtained via linear or polynomial interpolation could be used in the following. Inserting Equation 2.13 into the integral in Equation 2.11 allows one to pull out the constants, R_1, and rewrite the result as a matrix equation:

$$\begin{bmatrix} F(t_1) \\ F(t_2) \\ \vdots \\ F(t_m) \end{bmatrix} = \begin{bmatrix} \int_{t_0}^{t_1} M(t_1 - t')dt' & 0 & \dots & 0 \\ \int_{t_0}^{t_1} M(t_2 - t')dt' & \int_{t_1}^{t_2} M(t_2 - t')dt' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \int_{t_0}^{t_1} M(t_m - t')dt' & \int_{t_1}^{t_2} M(t_m - t')dt' & \dots & \int_{t_{m-1}}^{t_m} M(t_m - t')dt' \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{bmatrix} \tag{2.14}$$

In Equation 2.14 and the following, the background and time-shift parameters are neglected for the sake of clarity. For $M(t)$ given by Equations 2.2 or 2.3, the integrals in the above matrix can be easily solved using simple variable substitution. To compare the ability of this method to an exact result, we consider the $R(t)$ given by a Gaussian shaped pulse of width Δ centered at the origin and $M(t)$ given by Equation 2.2 with unity preexponential factor, which has an exact solution to the integral in Equation 2.11 given by:

$$\frac{1}{2} e^{\frac{t^2}{4\ln(2)\Delta}} erfc\left(\Delta\frac{\sqrt{\ln(2)}}{\tau} - \frac{t}{2\Delta\sqrt{\ln(2)}}\right) \tag{2.15}$$

The results of this comparison are shown in Figure 2.1 for decreasing τ, demonstrating the numerical approach is sufficiently accurate for relevant parameters to the TCSPC data.

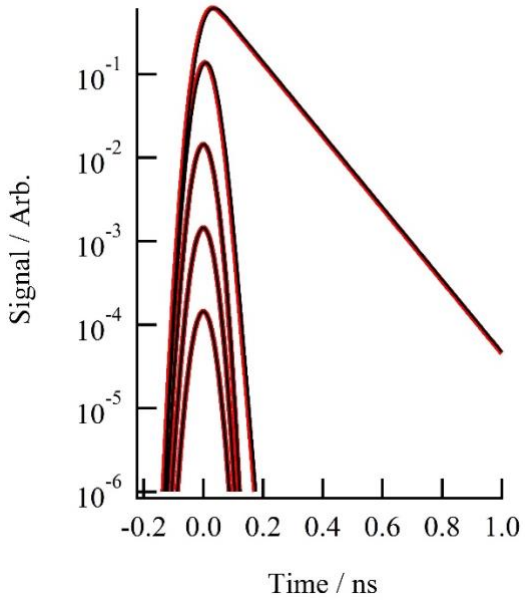

Figure 2.1. The exact result of the convolution of a gaussian pulse and single exponential decay provided by Equation 2.15 (red) is compared to the numerical solution provided by Equation 2.14 (black). The parameters for these curves are Δ = 65 ps, and τ = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} ns. The width of each bin was set to 8 ps for this calculation. Note the logarithmic y-axis.

The evaluation of the integrals in Equation 2.14 for lifetime distributions (Equation 2.10) is not necessarily so straight forward and requires additional care. Plugging Equation 2.10 into an arbitrary integral in Equation 2.14 gives:

$$\int_{t_{j-1}}^{t_j}\int_0^{\tau_{max}} f(\tau)e^{-\frac{t_i-t'}{\tau}}\,d\tau\,dt' \tag{2.16}$$

By changing the order of integration and integrating with respect to t' this can be simplified further:

$$\int_0^{\tau_{max}} \tau f(\tau)\left(e^{-\frac{t_j-t_i}{\tau}} - e^{-\frac{t_{j-1}-t_i}{\tau}}\right)d\tau \tag{2.17}$$

To evaluate Equation 2.17 for any $f(\tau)$ the trapezoidal rule with I linearly spaced grid points was used to give:

$$\sum_{k=1}^{I-1} \frac{\tau_{l+1}-\tau_l}{2}\left[\tau_{k+1}f(\tau_{k+1})\left(e^{-\frac{t_j-t_i}{\tau_{k+1}}}-e^{-\frac{t_{j-1}-t_i}{\tau_{k+1}}}\right)+\tau_k f(\tau_k)\left(e^{-\frac{t_j-t_i}{\tau_k}}-e^{-\frac{t_{j-1}-t_i}{\tau_k}}\right)\right] \quad (2.18)$$

Figure 2.2 shows the use of Equation 2.18 for increasing I applied to the bi-lorentzian $f(\tau)$ that will be described in Chapter 3. Figure 2.2 clearly shows that by I=50 the numerical solution converges. To be extra careful the number of grid points was increased to 200 for all fits shown in Chapter 3.

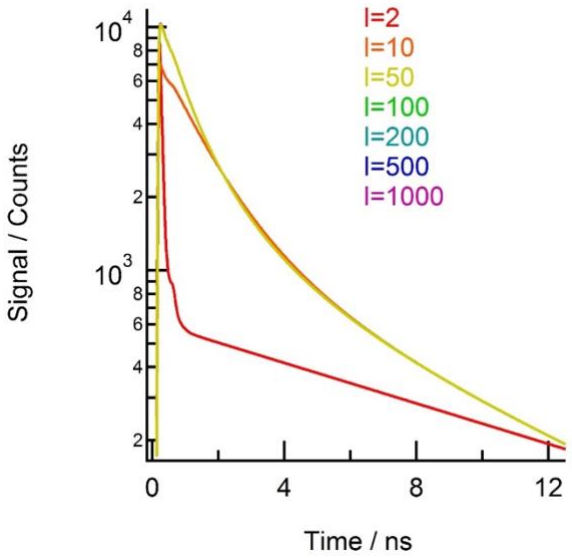

Figure 2.2. The accuracy Equations 2.14 and 2.18 to approximate Equation 2.11 is shown for increasing number of trapezoidal grid points, I.

2.4 Molecular Dynamics Simulations

2.4.1 Force Field, Structure Generation, and Solvent Model

To approximate the PES for the ssDNA oligomers, the widely used Amber ff99bsc0 forcefield[32] was used along with refined base stacking parameters[33], χ dihedral parameters,[34] and ε and ξ backbone torsion parameters.[52] While the A residue used in the simulations has preset charges provided in this Amber forcefield, the 2Ap residue does

not. The restrained electrostatic potential (RESP) fitting procedure was used to derive point charges for the 2Ap base.[53] A methyl group was used as a cap for the nucleobase during the charge fitting procedure. The nonstandard nucleobase was optimized at the B3LYP/6-31G* level of theory and, subsequently, the electrostatic potential was calculated using the HF/6-31G* level of theory. Both geometry optimization and electrostatic potential calculations were performed in the GAMESS program.[54-55] Atomic charges were fit to the electrostatic potential using the RESP module in Amber and are shown in Table 2.1 and Figure 2.3. Also included in Table 2.1 are the atom types. DNA residues of the 2Ap base were developed by combining the parameterized sugar backbone with the parameters for the nonstandard nucleobase.

Table 2.1 Atomic Charges and Assigned Atom Types for 2Ap Used for the Simulations.

Atom Name	Atom Type	Charge / e
N9	N*	-0.047936
C8	C2	0.150546
H8	H5	0.165748
N7	NB	-0.598370
C5	CB	0.156676
C6	CA	0.157719
H6	H4	0.150049
N1	NC	-0.674915
C2	CQ	0.879792
N2	N2	-0.930086
H21	H	0.394869
H22	H	0.394869
N3	NC	-0.682870
C4	CB	0.378610
Methyl Cap	----	0.105299

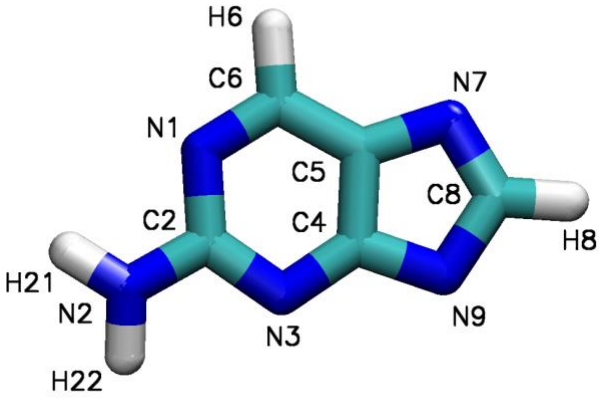

Figure 2.3 Atom names for 2Ap residue used in the simulations.

Except for the simulations of the trinucleotide and 15-base oligomer, which are described in Section 5.2, the initial structures were generated using LEaP. The initial structures for all oligomers were neutralized with the number Na^+ ions[56] equal to one minus the length of the oligomer. For the majority of simulations, the rigid TIP3P water model[57] was used. All oligomers were solvated using enough water molecules to fill a rectangular box such that no solute atom was less than 12 Å from the edge of the box using the solvatebox command. For the dinucleotide simulations this was roughly 1600 waters and the trinucleotide simulations required roughly 2100 waters. Finally, based on the box volume, enough Na^+ and Cl^- ions[56] were added to simulate a salt concentration of slightly more than 100 mM. After equilibration (see section 2.4.2) the box volume decreases and the concentration was recalculated to ensure the resultant concentration was as close to 100 mM as integer ion counts allow.

2.4.2 Simulation Protocol and Thermalization

Once the initial structures and solvent were generated with either the methods described in Section 2.4.2 or Section 5.2, the systems were energy minimized in eight stages that alternated using 1000 steps of the steepest decent (SD) and conjugate gradient (CG) methods. Every two stages positional restraints on the solute were reduced to ensure unreasonable structures did not result. The exact parameters for these stages are shown in Table 2.2.

Table 2.2 Parameters for Minimization of DNA Structures

Stage	Type	Restraints / kcal mol^{-1} $Å^{-2}$
1	SD	500
2	CG	500
3	SD	50
4	CG	50
5	SD	5
6	CG	5
7	SD	0
8	CG	0

Once minimized, the structures are essentially at 0 K as they have no velocity. To prepare initial conditions for the long MD simulations four thermalization stages were performed at a temperature of 300 K and pressure of 1 atm where again the restraints on the solute were slowly decreased. The temperature was controlled using the weak-coupling algorithm[59] with a 10 ps time constant. The pressure was controlled using the isotropic Monte Carlo barostat with a relaxation time of 1 ps. The parameters for these thermalization stages are shown in Table 2.3.

Table 2.3 Parameters for Thermalization of DNA Structures

Stage	*Length / ns*	*Restraints / kcal mol^{-1} Å^{-2}*
1	0.2	500
2	0.2	50
3	1	5
4	1	0

Finally, long Molecular dynamics simulations of all oligomers were run using the GPU-accelerated Amber package for explicit solvent particle mesh Ewald summation based simulations.[36] A 10 Å cut-off was used for non-bonding interactions. The SHAKE algorithm was used to constrain bonds involving hydrogen,[35] allowing a 2 fs time step to be utilized for the simulations. For the dinucleotide simulations all coordinates were saved every 2 ps, while the longer oligomers were only saved every 10 ps. These simulations were conducted using NVIDIA Pascal P100 GPU nodes on the Ohio State Supercomputing Center Owens Cluster.[60]

2.5 Theoretical Data Analysis

2.5.1 Calculation of Collective Variables

Structures were first stripped of water and ions using the CPPTRAJ program.[61] The coordinates for each were analyzed with in-house Python scripts utilizing the MDAnalysis package,[62-63] and scipy.[64] Still, the large amount of data produced by the simulations necessitated its distillation into a manageable amount of information. A classic way to achieve this is to perform a mapping of the solute molecules high dimensional cartesian coordinate vector onto a small number of collective variables. A simple example of this is the mapping from twelve dimensions to one dimension when a

dihedral angle is calculated from three atomic positions. Here, eight dihedral angles per base along backbone of the ssDNA that connect the two bases were calculated using a previously published algorithm.[65] A global coordinate used to a measure the surface area of the ssDNA was the solvent accessible surface area (SASA). The SASA were calculated with CPPTRAJ using the molsurf command with a probe of 1.4 Å.[66]

Another mapping that was found to be helpful at disentangling the SH of ssDNA was the calculation of the helical parameters (HP).[67] The six HP for ssDNA are the coordinates needed for the rigid body transformations that bring successive 5' nucleobases into overlap with the next base along the strand. In this way a *N* base oligomer requires 6(*N*-1) HPs. The first step for calculating the HP is to find the center of mass (COM) position for each base in the strand as if it were a free base (i.e., by ignoring the ribose group). Next, three orthogonal unit vectors, $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, were found for each purine nucleobase using the position vectors ($\vec{x}$) for the N1, N7, and N9 atoms. The first axis, $\vec{v}_1$, was calculated by normalizing ($\vec{x}_{\mathrm{N1}} - \vec{x}_{\mathrm{N9}}$), then $\vec{v}_3$ as ($\vec{x}_{\mathrm{N7}} - \vec{x}_{\mathrm{N9}}$)/|($\vec{x}_{\mathrm{N7}} - \vec{x}_{\mathrm{N9}}$)| $\times$ $\vec{v}_1$, and finally $\vec{v}_2$ as $\vec{v}_1 \times \vec{v}_3$. For a nucleobase in a double-stranded B-form DNA structure, $\vec{v}_1$ and $\vec{v}_2$ point toward the complementary strand toward and the major groove, respectively, while $\vec{v}_3$ is directed in the 3' to 5' direction.[67-70] This is demonstrated in Figure 2.4 which shows the three vectors for the two As at interior of a B-form DNA crystal structure.[70] The three distance parameters (shift, slide, rise) were calculated by projecting the vector defined by the difference between the two center of mass vectors onto the three 5' base axes. However, instead of the usual three angular parameters (roll, tilt, and twist), the projection of the 3' base axes onto the respective 5' base axes were used so that the projection of the two base normal vectors represents the cosine of the

angle between the two nucleobase planes, cos(θ). By constructing the parameters this way, it is easy to see the distance parameters translate the COM positions to the same point while the angular parameters align the three base axis.

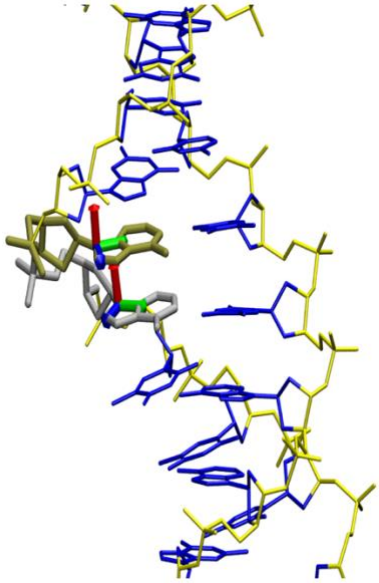

Figure 2.4 The three base axis are shown for two As on the interior of the B-DNA dodecamer[70] as green, blue, and red respectively. The 5' A is shown in gold while the 3' A is shown as gray. For the other bases the backbone is yellow while the bases are blue.

The use of HPs and backbone dihedral angles for a three base ssDNA is still a 36 dimensional dataset, and is far too large to systematically characterize the different structures and transitions that characterize its SH and CD. A method to further reduce the dimensionality of the dataset while still preserving the kinetically relevant structures is time-lagged independent coordinate analysis (TICA).[71-72] TICA determines a linear subspace which maximizes the time-lagged covariance of a time-structured dataset. The key parameter that makes TICA advantageous is the lag-time τ_{TICA} which accounts for the fact that if the ssDNA is initially in structure A, after τ_{TICA} time later it may have transitioned into structure B. In this way the direction that points between the two

structures gains extra covariance and will likely be captured in one of the TICA components. When TICA was used here it was used within the Pyemma package.[73]

2.5.2 Markov State Models

MSM construction was done primarily within the Pyemma package using python scripts.[73] The first stage in MSM construction is the clustering of chosen data set into microstates. On the order of 3000 microstates were used for each system studied. A back of the envelope calculation where each dimension in the data set distinguishes between 2 distinct structures, suggests that 3000 microstates is sufficient to cluster an 11 dimension dataset (2^{11}~3000). Clustering was done using either the regular space or *k*-means clustering algorithm. The purpose of clustering the data is that the cluster centers can be used to define a Voronoi tessellation of the coordinate space which partitions the data into the different microstates. This allows the continuous trajectory provided by the coordinates to be turned into a discrete one. In this way transitions between different states can be readily counted. A custom method for discretizing the trajectories was also developed outside of Pyemma, which starts by manually defining regions of a 2D coordinate space. This was achieved by simply clicking on a plotted 2D FES and saving the coordinates that were clicked on using python's event object. These coordinates define a polygon, which the shapely[74] package can use to determine if coordinates at each frame of the trajectory are inside of or not. This custom method was used to define discrete trajectories for the SASA transition path averages described in Section 4.2.2

The next step to construct a MSM was to evaluate the implied timescales[75] for the discrete trajectory. This involves calculating the inverse eigenvalues of the transition rate

matrix generated for various lag-times τ_{lag} to determine a τ_{lag} where the inverse eigenvalues are independent of τ_{lag}. Nominally, the transition rate matrices estimated are dependent on τ_{lag} because they contain the rates for transitions that occur between each state over a period of τ_{lag}. However, when the memoryless assumption about the kinetics holds, the transition rates won't depend strongly on τ_{lag}.[75] Unfortunately, the converse is not true so another more robust test, the Chapman-Kolmogorow test was also used.[43] The Chapman-Kolmogorow test calculates the conditional probabilities that an initial state *i* is found in state *j* at a later time predicted by a MSM with a fixed lag time versus MSMs estimated with longer lag times. If these two predictions are similar than the memory is likely lost on a time shorter than τ_{lag}. Using the above considerations, a lag time for MSM construction was chosen for each system that to our best ability ensured memory effects were neglected.

As the MSMs constructed usually involved 3000 states, additional simplifications in the form of hidden MSMs (HMSMs) were used to gain more understanding of the ssDNAs CD. HMSMs are an approximation to the MSM which coarse grains the microstates into a much smaller number of macrostates. HMSMs were calculated using the algorithm provided within Pyemma for a number of states equal to one plus the number of dominate timescales observed in the implied timescales plot. Additionally, inspection of the relevant FES to determine how many local minima were present was used to determine the number of macrostates as well.

2.6 References

17. D. C. Ward, E. R. a. L. S., Fluorescence Studies of Nucleotides and Polynucleotides: I. Formycin, 2-Aminopurine Riboside, 2,6-Diaminopurine Riboside, and Their Derivatives. *J. Biol. Chem.* **1969,** *244*, 1228-1237.

32. Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007,** *92*, 3817-3829.

33. Chen, A. A.; García, A. E., High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. USA* **2013,** *110*, 16820-16825.

34. Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8*, 2506-2520.

35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

36. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, et al *AMBER 2018*, University of California, San Francisco, 2018.

43. Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011,** *134*, 174105.

47. Stark, C. Excited State Processes in Ruthenium(II) Polypyridyl Complexes and Cerium Oxide Nanoparticles. Montana State University, Bozeman, Montana, 2016.

48. Letsinger, R. L.; Wu, T., Use of a Stilbenedicarboxamide Bridge in Stabilizing, Monitoring, and Photochemically Altering Folded Conformations of Oligonucleotides. *J. Am. Chem. Soc.* **1995,** *117*, 7323-7328.

49. Steinfeld, J. I.; Francisco, J. S.; Hase, W. L., *Chemical Kinetics and Dynamics*. Prentice Hall: Upper Saddle River, N.J., 1999.

50. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of Lifetime Distributions from Fluorescence Decays with Application to DNA-Base Analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

51. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. Kluwer Academic/Plenum: New York, 1999.

52. Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E.; Otyepka, M.; Jurečka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013,** *9*, 2339-2354.

53. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

54. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomicand Molecular Electronic-Structure System. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

55. Gordon, M. S.; Schmidt, M. W., *Advances in Electronic Structure Theory: GAMESS a Decade Later*. Elsevier Science Bv: Amsterdam, 2005; p 1167-1189.

56. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B.* **2008,** *112*, 9020-9041.

57. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

58. Wu, Y. J.; Tepper, H. L.; Voth, G. A., Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006,** *124*, 12.

59. Andrea, T. A.; Swope, W. C.; Andersen, H. C., The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water. *J. Chem. Phys.* **1983,** *79*, 4576-4584.

60. Center, O. S., Ohio Supercomputer Center. 1987.

61. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

62. R. J. Gowers, M. L., J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, et al *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulation*, SciPy: Austin, TX, 2016.

63. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.

64. Jones, E.; Oliphant, E.; Peterson, P.; others SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ (accessed 2017-).

65. Blondel, A.; Karplus, M., New Formulation for Derivatives of Torsion Angles and Improper Torsion Angles in Molecular Mechanics: Elimination of Singularities. *J. Comput. Chem.* **1996,** *17*, 1132-1141.

66. Connolly, M. L., Analytical Molecular-Surface Calculation. *J. Appl. Crystallogr.* **1983,** *16*, 548-558.

67. Lu, X. J.; Olson, W. K., 3DNA: a Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic. Acids. Res.* **2003,** *31*, 5108-5121.

68. Diekmann, S., Definitions and Nomenclature of Nucleic-Acid Structure Parameters. *J. Mol. Biol.* **1989,** *205*, 787-791.

69. Lu, X. J.; ElHassan, M. A.; Hunter, C. A., Structure and Conformation of Helical Nucleic Acids: Analysis Program (SCHNAaP). *J. Mol. Biol.* **1997,** *273*, 668-680.

70. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA Dodecamer - Conformation and Dynamics. *P. Natl. Acad. Schi-Biol.* **1981,** *78*, 2179-2183.

71. Schwantes, C. R.; Pande, V. S., Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013,** *9*, 2000-2009.

72. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013,** *139*, 015102.

73. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernàndez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015,** *11*, 5525-5542.

74. S., G.; others *Shapely: Manipulation and Analysis of Geometric Objects*, toblerity.org: 2007—.

75. Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R., Describing Protein Folding Kinetics by Molecular Dynamics Dimulations. 2. Example Spplications to Alanine Dipeptide and Beta-Hairpin Peptide. *J. Phys. Chem. B.* **2004,** *108*, 6582-6594.

CHAPTER THREE

ON THE ORIGIN OF MULTIEXPONENTIAL FLUORESCENCE DECAYS FROM 2-AMINOPURINE LABELED DINUCLEOTIDES

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Jacob M. Remington

Contributions: Collected and analyzed all steady-state and time-resolved data. Constructed all figures, besides Figure 1, and the primary draft of the document.

Co-Author: Abbey M. Philip

Contributions: Synthesized the samples studied in the manuscript along with Mahesh Hariharan. Provided a description of the synthesis included in the methods section. Constructed Figure 1.

Co-Author: Mahesh Hariharan

Contributions: Synthesized the dinucleotide samples studied in the manuscript along with Abbey M. Philip. Provided insight about the distance dependent ET model.

Co-Author: Bern Kohler

Contributions: Provided experimental equipment and guidance, along with key insights for the modeling of the data. Wrote and aided in the preparation of the manuscript.

Manuscript Information

Jacob M. Remington, Abbey M. Philip, Mahesh Hariharan, and Bern Kohler

The Journal of Chemical Physics

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
_x__ Published in a peer-reviewed journal

AIP Publishing
In Volume 145, 155101 (2016)
DOI: 10.1063/1.4964718

ON THE ORIGIN OF MULTIEXPONENTIAL FLUORESCENCE DECAYS FROM 2-AMINOPURINE LABELED DINUCLEOTIDES

Jacob M. Remington,[1] Abbey M. Philip,[2] Mahesh Hariharan,[2] and Bern Kohler[1]

[1]Department of Chemistry, Montana State University, Bozeman, Montana 59715, USA
[2]School of Chemistry, Indian Institute of Science Education and Research Thiruvananthapuram, CET Campus, Sreekaryam, Thiruvananthapuram, Kerala 695016, India

Abstract

The fluorescent probe 2-aminopurine (2Ap) has been used for decades to study local conformational fluctuations in DNA. Steady-state and time-resolved measurements of 2Ap fluorescence have been used to predict specific conformational states through suitable modeling of the quenching of the fluorescence of a 2Ap residue incorporated site-specifically into a DNA strand. The success of this approach has been limited by a lack of understanding of the precise factors responsible for the complex, multiexponential decays observed experimentally. In this study, dinucleotides composed of 2Ap and adenine were studied by the time-correlated single-photon counting technique to investigate the causes of heterogeneous emission kinetics. Contrary to previous reports, we argue that emission from 2Ap that is stacked with a neighboring base contributes negligibly to the emission signals recorded more than 50 ps after excitation, which are instead dominated by emission from unstacked 2Ap. We find that the decay kinetics can be modeled using a continuous lifetime distribution, which arises from the inherent distance dependence of electron transfer rates without the need to postulate a small

number of discrete states with decay times derived from multiexponential fits. These results offer a new perspective on the quenching of 2Ap fluorescence and expand the information that can be obtained from experiments.

3.1 Introduction

Fluorescent probes are invaluable for understanding the structural dynamics of biological macromolecules, including DNA.[76] In these experiments, changes in the intensity and lifetime of emission from a well-characterized chromophore are used to infer structural changes that take place on time scales commensurate with excited-state deactivation. The success of this approach relies on understanding the deactivation pathways of the isolated chromophore and developing realistic models that explain how the monomeric pathways are modified or how new pathways are created when the probe is incorporated into specific DNA sequences. The first step is often achieved by comparing high-level quantum chemical calculations with experimental measurements of excited state energies and lifetimes. The second step is the most challenging because labeled DNA molecules are too large for high-level calculations. In this case, more experimental work is needed to explore the possible decay pathways.

2-Aminopurine (2Ap), a constitutional isomer of adenine (A), is the most widely used fluorescent probe molecule for studying DNA structure, dynamics, and DNA/protein interactions. Jones and Neely[10] recently reviewed its role as a fluorescent probe of DNA/protein interactions. 2Ap minimally disturbs native DNA structure and has a high emission quantum yield that is strongly attenuated by interactions with nearby bases.[77] Excitation of the lowest energy absorption band of 2Ap in an oligonucleotide

results in multi-exponential florescence decay kinetics with lifetimes ranging from < 100 ps to 10 ns.[78] A continuous distribution of decay rates has also been used to fit the emission signals.[50, 79]

Here, the emission dynamics from two simple 2Ap-labeled dinucleotides (Figure 3.1), 5′-(2Ap)A-3′ and 5′-A(2Ap)-3′, are investigated by the time-correlated single-photon counting (TCSPC) technique. We report complex, multiexponential emission decays similar to ones observed in other 2Aplabeled DNA compounds, but we provide an interpretation that differs from the conventional picture, which assigns subnanosecond emission lifetimes to different stacked, or partially stacked, conformations. The conventional explanation is difficult to reconcile with the observation that excited-state lifetimes of dinucleotides containing natural nucleobases are independent of the precise stacking conformation.[26] Drawing on this result and on the emerging understanding of excited state dynamics of native bases, we propose that emission on the TCSPC time scale is dominated by emission from unstacked conformations.

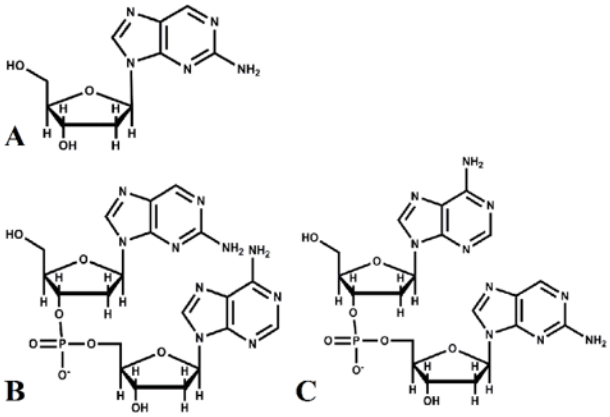

Figure 3.1Structures of the samples studied. (a) The 2′-deoxynucleoside of 2Ap used as a reference compound, (b) the 5′-labeled dinucleotide, 5′-(2Ap)A-3′, and (c) the 3′-labeled dinucleotide, 5′-A(2Ap)-3′.

Emission quenching when 2Ap is stacked with another residue is nearly universally believed to occur by electron transfer (ET),[10, 13, 21, 80] but the factors that govern emission by a 2Ap residue that is unstacked or poorly stacked with neighboring bases at the instant of excitation are less clear. Proposed quenching mechanisms in this case include solvation-dependent intersystem crossing to a dark $^3n\pi^*$ state,[81] or stacking within the excited state lifetime.[21, 82-83] Our preliminary analysis points to a major role for distance dependent ET between 2Ap and a nearby base, suggesting that emission decays can be used in principle to derive information about interbase distances in nucleic acids. Correctly interpreting 2Ap fluorescence decays could provide new opportunities to use this versatile fluorophore as a shortrange molecular ruler capable of measuring distances between 6 and 14 Å.

3.2 Methods

3.2.1 Samples

The deoxyribose form of 2-aminopurine (2-amino-9-(β -D-2-deoxyribofuranosyl)purine, ≥ 97% purity, hereafter, 2Ap) was purchased from Berry & Associates (Dexter, MI). Oligonucleotides were synthesized on a K&A Laborgeraete DNA synthesizer employing the automated solid-phase phosphoramidite chemistry. The 2-aminopurine and 2′ -deoxyadenosine phosphoramidite were purchased from Glen research and used as received. The fluorescent label 2Ap was incorporated into the dinucleotides on a 1 μ mol scale with appropriate controlled pore glass (CPG) beads used as 3′ solid support as reported earlier.[48] Oligonucleotides were de-protected and isolated from the solid-phase support using concentrated ammonium hydroxide with stirring for 24 h at room temperature. The 2Ap-labeled dinucleotides were then purified by reverse phase high performance liquid chromatography (Prominence HPLC System, Shimadzu) with a Phenomenex column (Luna 5u C18(2) 100A, 250 x 10 mm), employing a gradient of 20 mM ammonium acetate buff er and acetonitrile (flow rate of 1 ml/ min) as mobile phase. A 100 mM tris-HCl, 100 mm NaCl buffer (pH 7. 45 ± 0. 05) made with water from an ultrapurifier (Synergy, EMD Millipore) was used to prepare aqueous solutions for fluorescence measurements. The 2Ap concentration was 0.25 μM, resulting in an absorbance of 0.013 at 310 nm.

3.2.2 Instrumentation

A UV/ Vis spectrophotometer (Lambda 25, Perkin Elmer) was used to record absorption spectra. Steady-state fluorescence spectra were recorded with a fluorometer (Fluorolog 3, Horiba Jobin Yvon). Fluorescence decay curves were obtained using an emission lifetime spectrometer (FluoTime200, PicoQuant Photonics North America) that uses the time-correlated single-photon counting (TCSPC) technique. Excitation pulses were provided by a widely tunable femtosecond laser oscillator (Chameleon Ultra II, Coherent) with an 80 MHz pulse repetition rate. Third harmonic pulses tunable between 230 and 360 nm were produced using an auto-tracking harmonic generator. For this study, excitation pulses with a center wavelength of 310 nm were used. The laser beam was passed through an electrooptic modulator (Model 350-80, Conoptics) connected to a bias/ drive controller (Model 25D, Conoptics). A synchronous countdown timer (Model 305, Conoptics) selected every 30^{th} pulse. The polarization of the excitation pulse was set to vertical before the sample. The focused excitation pulse had a spot size of 115 μ m (FWHM) and a power of 330 μ W, resulting in a maximum peak intensity of 5.4 MW/ cm.[10] The light emitted at ~ 90◦ was sent through a 350 nm longpass filter (XUL0350 – Longpass, Asahi Spectra USA), and a polarizer set at the magic angle from vertical before detection with a microchannel plate photomultiplier tube (model R3809U-50, Hamamatsu Corp.) positioned past the exit slit of a monochromator. PicoHarp software (PicoQuant) was used to record the data with an 8 ps bin width. A Ludox® scattering solution was used to record the instrument response function (IRF), which was measured to be 64 ps (FWHM).

3.2.3 Modeling

Emission decay curves were fit by nonlinear least squares optimization using either the Fluofit program (version 4.5, PicoQuant) to fit sums of exponentials or an in-house algorithm for lifetime distribution fitting written for the Matlab program. Both programs compare the experimental decays to the decay function, D(t) , given by the convolution of the measured instrument response function (IRF) and a kinetic model function, M(t) ,

$$D(t) = B_{dec} + \int_{-\infty}^{t} [IRF(t-\gamma) - B_{irf}] M(t-\tau) d\tau \quad . \qquad (3.1)$$

In Eq. (1), the parameters B_{irf} and B_{dec}, account for dark counts in the measured IRF and measured decay, respectively. These parameters were set to the counts recorded before the first excitation pulse and were not optimized during fitting. The time-shift parameter, γ, shifts the IRF signal by a constant amount in order to give the IRF and the fluorescence decay curves a common time origin. The time-shift parameter in these fits always assumed a reasonable value between -10 ps and 10 ps.

The first model used was a sum of four (n = 4) exponentials,

$$M(t) = \sum_{i=1}^{n} A_i e^{-\frac{t}{\tau_i}} , \qquad (3.2)$$

where A_i are amplitudes and τ_i are rate constants. Fits obtained with Eqs. (3.1) and (3.2) were compared with fits using a continuous lifetime distribution,[79, 84]

$$M(t) = \int_{0}^{t_f} f(\tau) e^{-\frac{t}{\tau}} d\tau \quad . \qquad (3.3)$$

In Eq. (3.3), the upper limit of integration, τ_f, was set to the lifetime of the 2'-deoxynucleoside of 2Ap measured at the same temperature (Table A.1.5). This choice

restricts the possible lifetimes to be less than that of the free monomer consistent with the expectation that interactions with adenine in the dinucleotide result solely in lifetime shortening. The space of possible *f(τ)* that can satisfactorily fit a given experimental emission signal is vast, and finding *f(τ)* is a numerically ill-posed problem.[85] To circumvent this difficulty, we approximate the shape of $f(\tau)$, by a sum of two Lorentzian distributions,

$$f(\tau) = \sum_{j=1}^{2} \frac{A_j}{2\pi} \cdot \frac{\Delta_j}{(\tau - \mu_j)^2 + \left(\frac{\Delta_j}{2}\right)^2}, \tag{3.4}$$

where A_j is the amplitude of the j^{th} Lorentzian component, Δ_j is the full width at half maximum, and μ_j is the most probable lifetime (mode). This lifetime distribution has been used previously to fit emission decays of 2Ap-labeled oliognucleotides.[79] However, even this choice does not insure uniqueness. In order to avoid the trivial case in which the Δ_j approach very small values, causing Eq. (3.3) to become a sum of discrete exponential terms, a range of initial parameter values were tried until the *f(τ)* determined by Marquardt-Levenberg optimization was unimodal. Including the time-shift parameter in Eq. (3.1), the lifetime distribution model in Eqs. (3.3) and (3.4) has 7 adjustable parameters. To further reduce the number of fitting parameters, Δ_1 and μ_1 were linked in simultaneous fits to decays from the 5'- and 3'-labeled oligonucleotides at each temperature, yielding 11 adjustable parameters for every pair of decay curves. For both models, the amplitude parameters were restricted to nonnegative values. To compare the overall dynamics of the two samples, amplitude weighted mean lifetimes, $\langle\tau\rangle$, were calculated with Eq. 3.5.

$$\langle \tau \rangle = \frac{\sum_{i=1}^{n} A_i \tau_i}{\sum_{i=1}^{n} A_i} \tag{3.5}$$

Fits were performed using the first 4500 points of each measured decay trace (0 – 36 ns). The fit quality was assessed from the reduced χ^2 statistic defined in Eq. (3.6) and from visual inspection of the residuals weighted according to Poisson statistics as shown in Eq. (7),

$$\chi^2 = \left(\frac{1}{N-p} \right) \sum_{i=1}^{N} (\mathrm{Res}_i)^2 , \tag{3.6}$$

$$\mathrm{Res}_i = \frac{D(t_i) - Y_i}{\sqrt{Y_i}}, \tag{3.7}$$

where Y_i, are the observed counts in each of N channels, p is the number of adjustable parameters, and $D(t_i)$ is the model decay function from Eq. (3.1) evaluated at discrete time points. Asymptotic standard errors were calculated with the Fluofit program for the sum of exponential fits, and 2σ values are reported throughout the text. Errors reported for the lifetime distributions were calculated using support plane analysis.[86]

3.2.4 Experimental Protocols

Steady-state emission quantum yields were estimated with Eq. (3.8) using the 2Ap monomer ($\Phi_{2Ap} = 0.68$)[87] as a fluorescence reference,

$$\Phi = \Phi_{2Ap} \left(\frac{F}{F_{2Ap}} \right) \tag{3.8}$$

In Eq. (3.8), F and F_{2Ap} are the fluorescence spectra integrated between 330 and 500 nm of equal absorbance solutions of the dinucleotide and the 2Ap monomer, respectively.

Before integration, each spectrum was corrected for the wavelength dependence of the detector and by subtracting emission from a buffer-only solution.

Fluorescence decays of the dinucleotides were recorded as a function of temperature at an emission wavelength of 370 nm with excitation at 310 nm. Data acquisition was stopped for each trace upon reaching a pre-set maximum number of counts (10,000, unless otherwise stated). In addition, time-resolved emission spectra (TRES) were assembled from decays recorded by scanning the emission wavelength in 5 nm steps. For each decay trace, the integration time was held constant at 1 minute. These traces were then rescaled by a correction factor to account for the wavelength dependence of the detector. The constancy of the excitation laser intensity was confirmed by recording the IRF with a dilute scattering solution before and after each sample run. The area under the IRF was found to vary by no more than 1% from run to run.

3.3 Results

Upon excitation at 310 nm, the emission traces at 370 nm for 5'-(2Ap)A-3', and 5'-A(2Ap)-3' at 20 °C decay non-exponentially as seen previously for 2Ap/A-containing di-[88] and oligonucleotides[89] (Figure 3.2). Overall, the signals from the two dinucleotides are very similar, but some subtle differences can be seen, especially between 3 and 10 ns. The temperature dependence of the fluorescence decays of 2Ap, 5'-(2Ap)A-3', and 5'-A(2Ap)-3' was investigated at λ_{ex} = 310 nm and λ_{em} = 370 nm every 5 °C between 10 and 60 °C. Results at the lowest and highest temperatures are shown in Figure 3.2. Increasing the temperature increases the rate of decay of the emission from both the monomer and the labeled dinucleotides, but emission by the latter depends somewhat

more strongly on temperature, especially near 5 ns. For all samples, the slope of the decay curves at long times approaches that of the 2Ap monomer.

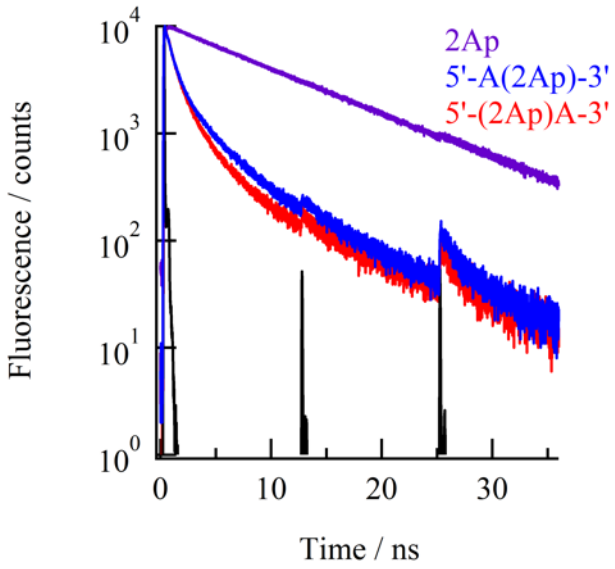

Figure. 3.2. Normalized TCSPC traces of 2Ap, 5'-A(2Ap)-3', and 5'-(2Ap)A-3' (from top to bottom at 10 ns) at 20 °C. Emission from both 2Ap-labeled dinucleotides decays more rapidly than from the monomer and in a non-single-exponential manner. The IRF (black curve) shows the satellite pulses responsible for the weak re-excitation evident at 12.5 and 25 ns. For all fluorescence decays, λ_{ex} = 310 nm and λ_{em} = 370 nm.

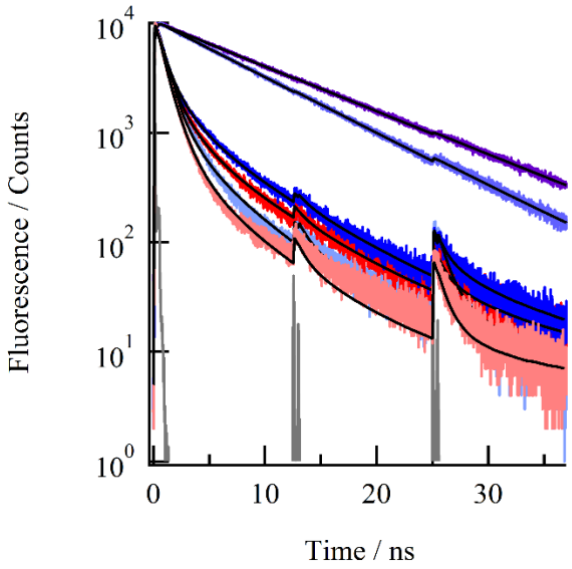

FIG. 3. Emission decay curves (λ_{ex} = 310 nm and λ_{em} = 370 nm) in order of the relative signal intensity at 10 ns: 2Ap at 10 ℃ and 60 ℃ (dark and light purple, respectively); 5'-A(2Ap)-3' (blue) and 5'-(2Ap)A-3' (red) at 10 ℃; 5'-A(2Ap)-3' (light blue) and 5'-(2Ap)A-3' (light red) at 60 ℃. Best fit curves are shown by solid black lines. A single exponential model was used to fit the 2Ap traces, while the bi-Lorentzian lifetime distribution model in Eqs. (3.3) and (3.4) was used to fit all dinucleotides traces. The IRF is shown in gray.

TRES of 5'-(2Ap)A-3', 5'-A(2Ap)-3', and 2Ap were measured at 20 ℃ by stepping λ_{em} between 330 and 500 nm and recording the emission decays at each wavelength while maintaining constant integration time and intensity of the excitation laser (Figure 3.4). The three samples had the same absorbance at the excitation wavelength of 310 nm to allow comparison of the signal amplitudes (Figure A.1.1). The maximum strength of the emission near time zero is attenuated in the dinucleotides by a factor of 1.4. Comparing the two dinucleotides, the emission from the 3' label is attenuated by the same factor as the 5' label (Figure 3.4). The peak wavelength of the time-zero emission spectrum is the same for the dinucleotides and the 2Ap monomer to within ± 5 nm. In fact, the TRES of the three samples are identical in shape at all times as demonstrated by the agreement of the spectra recorded at 0 ns and the scaled 9 ns

spectrum (Figure 3.4). Steady-state fluorescence of the samples was measured with excitation at 310 nm, see supplementary material, to quantify the steady-state emission quantum yields (Figure A.1.3). The steady-state fluorescence spectra have the same shape as the 2Ap reference spectrum. The fluorescence quantum yields determined for 5'-(2Ap)A-3' and 5'-A(2Ap)-3' are 0.072 ± 0.009 and 0.061 ± 0.012, respectively.

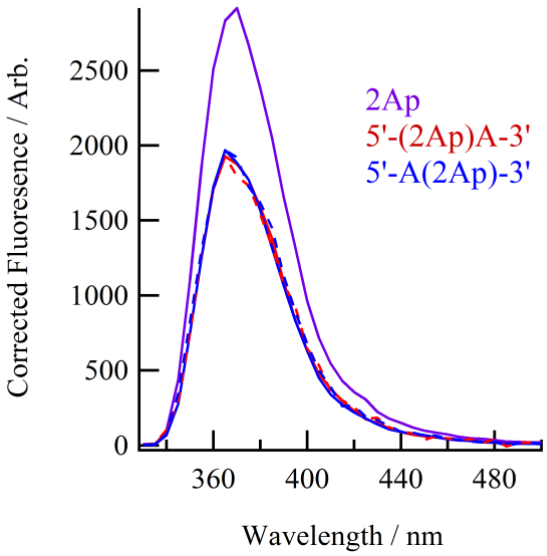

Figure 3.4. TRES of equal absorbance solutions of 2Ap (uppermost curve: 0 ns), 5'-(2Ap)A-3' (solid red: 0 ns, dashed red: 9 ns), and 5'-A(2Ap)-3' (solid blue: 0 ns, dashed blue: 9 ns). The spectra at 9 ns have been scaled for comparison with the 0 ns spectra, showing that both dinucleotides have nearly identical normalized emission spectra.

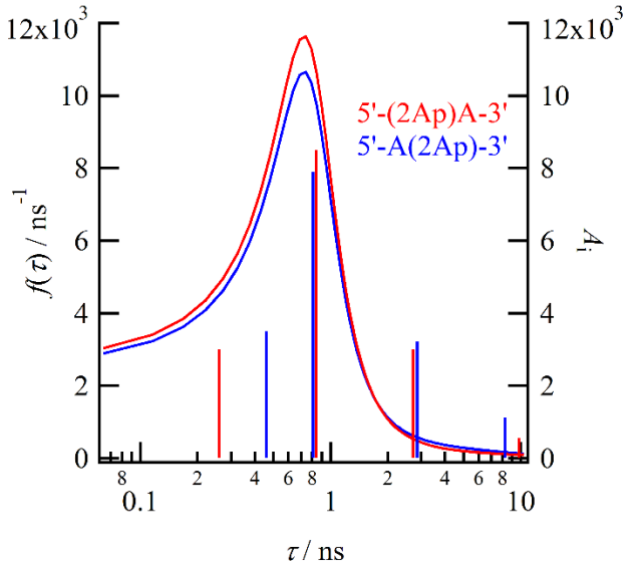

Figure 3.5. Comparison of a four exponential fit (vertical lines represent the amplitudes of the various decay components, right vertical axis) and a bi-Lorentzian lifetime distribution fit (continuous lines, left vertical axis) to emission from 5'-(2Ap)A-3' (red, uppermost curve) and 5'-A(2Ap)-3' (blue) at 10 °C.

The best-fit parameters from fits to the 10 °C emission traces of each dinucleotide are shown in Figure 3.5 for the sum-of-exponentials model, Eq. (3.2), and the continuous lifetime distribution model, Eq. (3.3). All fitting parameters are listed in Tables A.1.1 – A.1.4 in Supplementary Material. The amplitude-weighted mean lifetimes calculated from the sum of exponentials model using eq. 3.5 are 1.43 ± 0.08 ns and 1.68 ± 0.08 ns for the 5' and 3' labels, respectively. This difference is due mainly to the increased amplitude of τ_4 in the 3' sample, but the fit parameters are otherwise rather similar for both dinucleotides.

The lifetime distributions for both dinucleotides are unimodal. Because of the similar appearance of the signals from the 5'- and 3'-labeled dinucleotides the Δ_j and k_j parameters of the dominant Lorentzian component were linked during simultaneous (global) fitting to the decays from both samples. Consequently, the second Lorentzian

accounts chiefly for the subtle differences between the emission decays from the two samples. Figure 3.5 shows that the four exponential model qualitatively approximates the distribution of discrete lifetimes. Both models also yield comparable residuals at temperatures lower than 20 °C (Figure A.1.4).

The lifetime distribution model was also used to fit the temperature-dependent decay traces of both dinucleotides (Figure 3.6). At all temperatures, unimodal distributions were obtained. For both samples, the peak in the lifetime distribution narrows somewhat and increases in amplitude as the temperature is increased. The bi-Lorentzian fits are also slightly worse at higher temperatures as seen from their greater χ^2 values (Tables A.1.1, A.1.2).

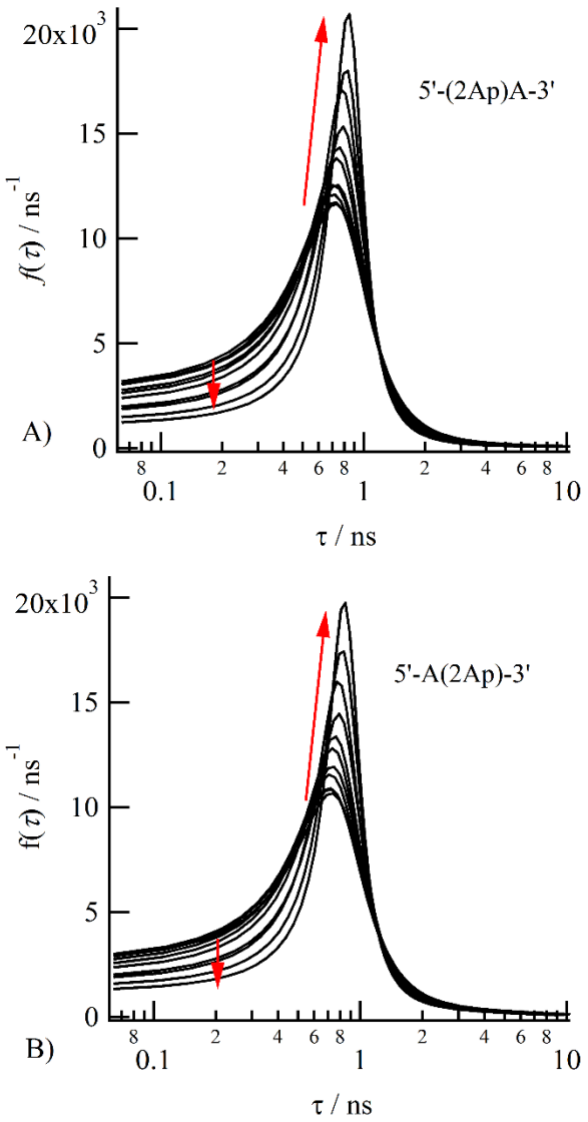

Figure 3.6. Best-fit lifetime distributions for A) 5'-(2Ap)A-3' and B) 5'-A(2Ap)-3' solutions as a function of temperature. The arrows indicate the changes that occur as the temperature is increased from 10 to 60 °C in steps of 5 °C.

3.4 Discussion

3.4.1 Comparison with Previous Time-Resolved Emission Measurements

The 5'-labeled dinucleotide was studied previously by time-resolved emission spectroscopy.[88, 90-91] Our traces at times greater than 1 ns are similar to ones in Ref. [88]. Our kinetics at intermediate times depend strongly on temperature (Figure 3.3) in agreement with Ref. [88]. However, in previous TCSPC studies of the 5'-labeled dinucleotide, a short 25-31 ps component was reported at room temperature.[88, 91] An 80 ps time constant was measured in fluorescence up-conversion experiments, although exact buffer conditions were not reported, and may differ from our conditions.[90] In contrast, our fit with the sum-of-exponentials model yields a shortest lifetime of 320 ps at 20 °C. This lifetime is more similar to lifetimes near 500 ps seen in longer $(A)_n$ oligomers that contain a 2Ap probe.[89]

To explore these discrepancies, we re-fit our 20 °C data for the 5'-(2Ap)A-3' sample, constraining the two time constants in the biexponential fit function to have the values reported in Ref. [88] (31 and 590 ps). The resulting residuals (Figure A.1.4, bottom panel) resemble ones reported in Ref. [88], but have more non-random character than the residuals from our best fit with a 320 ps lifetime (Figure A.1.4, top panel). This leads us to question the need for a sub-100 ps component, but we acknowledge the difficulties of resolving lifetimes shorter than the IRF.[92]

Although we do not observe a tens of ps decay component, there is a prompt loss of emission that can be seen by comparing emission from absorbance-matched samples of 2Ap and the dinucleotides (Figure 3.4). The loss in the maximum signal near time zero

indicates the existence of an ultrafast decay channel, which is consistent with previous reports of rapidly forming dark-states in 2Ap-labeled oligomers.[93] Quenching of excited states that occurs too quickly to resolve using a given measurement technique has been termed quasi-static quenching,[94] and the origins of this channel will be discussed next.

3.4.2 Influence of Stacked 2Ap on TCSPC Decay Traces

Understanding of excited electronic states of DNA strands made of native bases has advanced greatly during the past decade.[95] We consider next how these recent insights can be used to construct a new model for the fluorescence quenching of stacked 2Ap. Because 2Ap is a simple isomer of adenine it is logical that fluorescence quenching pathways for 2Ap emission in DNA could closely resemble those of native bases. A key observation from studies of native DNA is the speed of forming charge transfer (CT) states. When a dinucleotide composed of two A residues is in a π-stacked conformation, photoexcitation creates a Frenkel exciton that rapidly decays to a CT state in less than ~200 fs.[96] The formation of CT states in DNA strands made of natural bases is strongly gated by base stacking as these states are only observed when two or more bases are stacked at the instant of photon absorption.[97] Disrupting base stacking by adding methanol to the $(dA)_2$ dinucleotide completely eliminates the long-lived CT state.[96]

The redox properties of 2Ap are similar to those of the canonical bases, and cyclic voltammetry measurements predict that photoexcited 2Ap can oxidize or reduce A.[98] Based on the favorable energetics, photoexcitation of a π-stacked 2Ap/A pair is expected to result in ultrafast ET, rapidly quenching the bright $^1\pi\pi^*$ state of 2Ap.[98-99] This concept

has been used previously to explain the 50% loss in initial fluorescence from the 5'-(2Ap)A-3' dinucleotide.[90]

Ultrafast ET quenching of the 2Ap $^1\pi\pi^*$ state in stacked bases has an important consequence for TCSPC measurements of 2Ap-labeled oligomers: Emission from the excited state of 2Ap, hereafter denoted 2Ap*, should be virtually undetectable whenever 2Ap is π-stacked with A (or C, T, or G). The maximum signal from two emitters with the same radiative rate will be proportional to their lifetimes when these are much shorter than the IRF. The FWHM of the IRF in a fast TCSPC lifetime instrument such as ours is on the order of 100 ps or 1000 times longer than the lifetime of approximately 100 fs that we believe describes 2Ap* in a base stack. This implies that the maximum signal amplitude from stacked 2Ap* will be 1000 times smaller than the maximum signal from the 2Ap monomer. This assumes equal excitation rates and a radiative rate of 2Ap* that is the same for the monomer and for 2Ap in a base stack. Thus, our experimental conditions that result in 10^4 maximum counts from monomeric 2Ap suggest a maximum signal of 10 counts from stacked 2Ap, a value that falls below the noise floor of approximately 100 counts, and is thus too weak to detect.

The CT state could still contribute to the TCSPC signal *after* interbase ET as analogous CT states in native base containing DNA have been shown to emit over hundreds of ps, albeit with low radiative rates.[100-101] According to a computational study, the oscillator strength of the (2Ap)A CT state is hundreds to thousands of times weaker than that of the 2Ap $^1\pi\pi^*$ state.[102] Because emission signals are proportional to the product of the number of excited molecules times their radiative decay rate, it is also necessary to know the amounts of stacked and unstacked conformations present. The

equilibrium constant describing stacking is less than unity (see below), indicating that the signal from CT state emission will be hundreds to thousands of times weaker than that of unstacked 2Ap residues. These considerations predict that emission from the bright $^1\pi\pi^*$ state of 2Ap will completely mask any weak emission from CT states.

Another important property of CT states in DNA oligomers made of native bases is their strongly red-shifted emission spectra relative to the emission by individual nucleotides.[103] This agrees with computational predictions that CT states lie lower in energy than the monomer $^1\pi\pi^*$ states.[104-106] Accordingly, emission by a 2Ap/A CT state should be strongly red-shifted compared with emission from monomeric 2Ap. A state bearing some of these properties has been observed in 2Ap oligomers although it appears to have a much longer decay lifetime than even monomer 2Ap, and is only observed when excited at 360 nm.[107-108] Here, the TRES from 2Ap-labeled dinucleotides recorded by the TCSPC technique 0 and 9 ns after photoexcitation strongly resemble the spectrum of the 2Ap monomer (Figure 3.4). In other studies, steady-state emission from 2Ap-labeled oligomers is red-shifted by a few nm from the monomer spectrum.[15] Such shifts were not observed in our time-resolved measurements, possibly due to monomer-like hydration of the bases in the dinucleotides studied here. Given that the TRES coincide with the 2Ap monomer spectra (Figure 3.4), it is unlikely that the CT state rapidly populated after photoexcitation of stacked 2Ap contributes appreciably to the observed emission decay.

A further possibility is that the dark CT state decay could repopulate the bright monomer excitonic state.[109] This is more likely when the CT state and the bright excited state of 2Ap have similar energies, but computational studies[102] and studies on the natural

bases suggest that the CT state is at least 1 eV lower in energy. In aqueous solution, especially for a short dinucleotide, the nearby dipoles are expected to stabilize the CT state on time scales similar to those needed to make and break hydrogen bonds. We propose that none of the observed four exponential decay components, or any part of the lifetime distribution (Figure 5), can be attributed to stacked conformations of the oligomer.

3.4.3 Influence of Base Stacking Equilibria on 2Ap Decay Traces

Our model, which assumes that stacked 2Ap does not contribute to the emission measured by the TCSPC technique, allows estimation of the equilibrium constant, K_s, that describes the unstacked to stacked transition of the dinucleotide,

$$\text{unstacked dinucleotide} \rightleftharpoons \text{stacked dinucleotide} \qquad K_s = \frac{[\text{stacked}]}{[\text{unstacked}]}. \qquad (9)$$

If only unstacked 2Ap residues contribute to the measured emission, then the signal strength at any time is proportional to the number of unstacked conformers that are still excited. Because all unstacked conformers are still excited at zero time, the strength of the signal at time zero from the dinucleotide divided by the strength of the signal at time zero from an equal absorbance solution of 2Ap monomer gives the fraction of unstacked conformers when both samples initially have the same number of excited states and identical radiative rates. In this case, K_s is given by,

$$K_s = \frac{M_{2\text{Ap}}(0)}{M_{\text{dinuc}}(0)} - 1. \qquad (3.10)$$

In Eq. 3.10, the kinetic model function, *M*, from Eq. 3.1 describes the initial, deconvoluted signal from the 2Ap and dinucleotide samples. Emission from 2Ap and the

2Ap-labeled dinucleotides was recorded from equal-absorbance solutions under otherwise identical experimental conditions in order to produce the same number of excited states in each sample. The values of $M(0)$ were determined using Eq. (3.3) and from the results of fitting the bi-Lorentzian model to the various emission decays are shown in the supplementary materials (Figure A.1.5). K_s values of 0.28 ± 0.03 and 0.27 ± 0.03 were determined for the 5'- and 3'-labeled samples, respectively, using the signals at 370 nm, the wavelength of maximum emission. Very similar values were obtained from fits with Eq. (3.2), indicating that the method is insensitive to the precise model used for the unstacked population, as long as it fits the data well.

To the best of our knowledge this is the first estimate of the base stacking equilibrium for 2Ap in single-stranded DNA. Stacking appears to be less favored than in the adenine-adenine dinucleotide, d(ApA). The value of K_s for d(ApA) has been reported to be 1.8[110] and 11.5[111] from circular dichroism (CD) experiments, 4.5[112] and 5.6[113], according to molecular dynamics (MD) simulations, and a value of 5.7[111] was determined from UV-vis absorption experiments. These higher values suggest that the altered position of the amino group in 2Ap compared to A may interfere with base stacking. Consistent with this notion some experiments on dsDNA have suggested that 2Ap slightly disrupts the DNA helix.[114]

Results from a combined molecular dynamics simulation / fluorescence lifetime study of 2Ap-labeled hairpins suggest that 2Ap is stacked with neighboring bases in approximately 50% of conformers.[115] This value, which corresponds to an equilibrium constant of approximately unity, was observed in MD simulations in which 2Ap was positioned within the hairpin stem. However, stacking was observed less often when 2Ap

was located in the more flexible loop region in line with the lower stacking equilibrium constants reported here.

If quasi-static processes were the only source of quenching in the 5'-labeled dinucleotide, then its fluorescence quantum yield should be equal to the quantum yield of the 2Ap monomer (0.68[87]) multiplied by the fraction of unstacked bases. The latter quantity is determined to be 0.79 from the calculated K_s value of 0.27, resulting in a predicted quantum yield of 0.54. This value is nearly an order of magnitude greater than the observed steady-state fluorescence quantum yield of 0.072, indicating that unstacked 2Ap residues are subject to additional quenching pathways. These pathways give rise to the multi-exponential decay kinetics in Figures 3.2 and 3.3, and the remainder of this paper is focused on proposing a new mechanism to describe this quenching.

3.4.4 A Model for Fluorescence Quenching of an Unstacked 2Ap*/A Pair

In general, fitting a sum-of-exponentials model to the fluorescence decay of a 2Ap-labeled DNA strand assumes that a Markov state model[116] accurately captures the dynamics on the conformational landscape of the oligomer. In such a description, the experimentally determined decay rates are the eigenvalues of the underlying kinetic scheme created by subdividing different thermodynamically stable regions of the conformational landscape into states and defining interconversion rates between them. These states can be distinguished through time-domain fluorescence if they have different nonradiative decay rates for the 2Ap probe perhaps due to different rates of base stack formation. Contrary to the coarse graining inherent in the Markov state model approach is the supposition that the decay kinetics arise from a *continuous* distribution of

heterogeneous micro-environments,[117] each with a different nonradiative decay rate for the 2Ap probe (Figure 5). Distinguishing between these two cases using time-resolved emission traces alone is very difficult in the presence of experimental noise[118]. However, MD simulations have found that the conformational landscape of a dinucleotide is well described by a two-state model that includes a stacked state and an unstacked state with a continuous range of interbase separations (>6 Å).[113] This suggests that a Markov state model with 3 to 5 distinct unstacked conformations is inappropriate for unstacked 2Ap. We believe that a continuous approach is more realistic for describing unstacked 2Ap. We also suggest that the inherent distance dependence of ET rates naturally suggests how the nonradiative decay rates of 2Ap can vary for different microenvironments.

A very crude model for long-range ET predicts that ET rates decay exponentially with donor-acceptor distance, r,[119]

$$\frac{k_{ET}(r)}{k_{ET,0}} = e^{-\beta(r-r_0)} \tag{3.11}$$

where r_o is the distance at which the ET rate has the maximum value $k_{ET,0}$, and β is a solvent-dependent parameter describing the attenuation of the electronic coupling between donor and acceptor. β values of 0.2 to 0.9 Å^{-1} have been reported for π-stacked DNA systems in which donor and acceptor are separated by two or more intervening stacked bases.[120-122] Larger values of 1.0 to 1.4 Å^{-1} characterize ET in the interior of proteins,[123] while values as high as 1.6 to 1.7 Å^{-1} have been reported for free molecules in aqueous solution.[124-126] Because the electronic coupling between *unstacked and solvent-separated* 2Ap and A is of interest here, a β value of 1.7 Å^{-1}, which is appropriate for a water-like environment, is used. Similarly, r_o is taken to be 3.5 Å, the van der Waals

contact distance between stacked bases, and $k_{ET,0}$ is set to 1/100 fs^{-1}, a rough limit for ET at minimal donor-acceptor separation.[119] It is important to note that Eq. 3.11 is a rough approximation that does not account for the potential distance dependence of the reorganization energy, a quantity that has been estimated for well-stacked, double-stranded, B-form DNA[127], but is uncertain for unstacked dinucleotides. The implicit assumption in Eq. 3.11 that the driving force for ET is independent of distance is justified by the large dielectric constant of water.

In order to explore a particularly simple limit, we assume that transitions between the different microstates can be neglected during the lifetime of 2Ap*. In this case, the decay rate of 2Ap* is determined entirely by the interbase separation, and the lifetime distribution $f(\tau)$ (Figures 3.5, 3.6) contains information about the 2Ap*/A interbase distance distribution, $P(r)$. The latter can be extracted from the former using the inverse of Eq. (3.11) as shown in Eq. (3.12) with the appropriate variable substitution from Eq. (3.13).

$$r = r_0 - \frac{1}{\beta} ln \left[\frac{\frac{1}{\tau} - \frac{1}{\tau_{2Ap}}}{k_{ET,0}} \right] \tag{3.12}$$

$$\frac{dr}{d\tau} = \frac{1}{\beta \tau^2 \left(\frac{1}{\tau} - \frac{1}{\tau_{2Ap}} \right)} \tag{3.13}$$

where τ_{2Ap} is the lifetime of the 2Ap monomer (Table A.1.5). The distance-dependent ET model in Eqs. (3.11)–(3.13) was applied to the experimentally determined lifetime distribution shown in Figure 3.5 to determine the interbase separation distribution for both samples (Figure 3.7). The results reveal that the heterogeneous decay kinetics of unstacked 2Ap∗ in a dinucleotide are consistent with a distribution of

distances between the bases with the most probable separation located between 8 and 10 Å. The distance *r* in eq. 11 is usually understood to be the distance between molecular centers of mass. We note that the distance between glycosidic nitrogen atoms in RNA dinucleotides was observed to fluctuate between approximately 6 and 10 Å in molecular dynamics simulations carried out in organic solvents where little base stacking is expected.[128] We do not claim that the crude model behind Eqs. 11 – 13 is able to accurately predict distances, but the results in Fig. 7 do show that varying the electron transfer parameters over a range of possible values robustly predicts that bases in unstacked conformations are separated by distances that are physically reasonable for a dinucleotide. This supports our conjecture that distance-dependent quenching of the emission from unstacked 2Ap* is responsible for the observed heterogeneous emission decay signals. From trial fits of the emission decays using other lifetime distributions, we found that the location of the most likely interbase distance is essentially the same provided that a good fit is obtained as judged by the residuals. Earlier, Fogarty et al.[84] showed that the location of the maximum of the lifetime distribution is nearly the same for two different rate distributions.

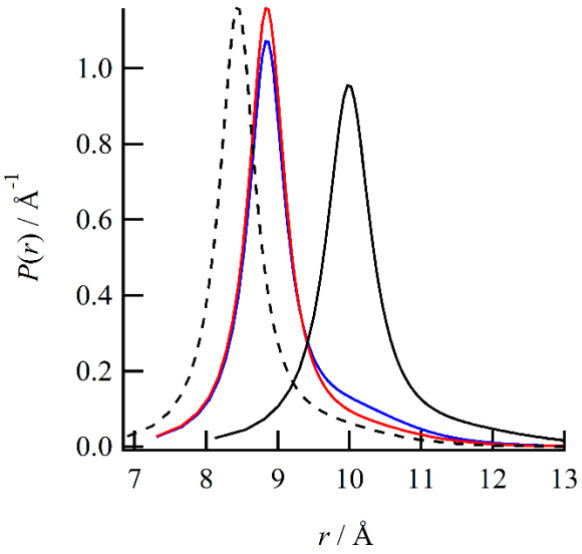

Figure 3.7. The interbase-separation distribution of unstacked 5'-(2Ap)A-3' (top, red curve) and 5'-A(2Ap)-3' (bottom, blue curve) at 10 °C calculated from the rate distribution in Figure 4 using Eqs. (12) and (13). For the red and blue curves, β = 1.7 Å^{-1}, and $k_{ET,0}$ = (100 fs)$^{-1}$. The solid black and dashed curves, respectively, demonstrate the effect of changing β to 1.4 Å^{-1} and $k_{ET,0}$ to (200 fs)$^{-1}$ for the 5'-(2Ap)A-3' sample.

The 2Ap and A residues are similar in size and mode of attachment to the DNA backbone, suggesting that exchanging their positions should have little to no effect on their mutual separation. The very similar signals observed for the 5'- and 3'-labeled dinucleotides are consistent with this supposition and our model in which emission arises from unstacked conformations that undergo quenching by ET at a distance. If stacked states of the dinucleotide were important, then greater differences might be expected due to differences in how a 5'- vs. 3'-A stacks with 2Ap.

Our model differs sharply from ones in which motions following excitation are responsible for quenching. Thus, Barton and co-workers concluded that dynamic stacking of an initially unstacked 2Ap is a dominant decay pathway for photoexcited 2Ap.[129-131] However, the adoption of a fully stacked geometry from an initially unstacked conformation is difficult to complete within the 10 ns lifetime of 2Ap*, according to

results from temperature-jump experiments that indicate that 100 ns is needed for stack formation in $(dA)_2$ at room temperature.[132] These stacking times are somewhat longer than the 20 ns unstacking lifetime for self-associating dimethyladenosine dimers at room temperature.[133] In contrast, MD simulations of trinucleotides found much shorter stacking times of 10 to 100 ps during 10 ns tragectories.[134] Importantly, our model incorporates the notion that the bases do not need to be fully stacked for quenching to occur. Instead, modest movement toward reduced interbase separation can greatly increase the probability of quenching through the exponential dependence on distance in Eq. 11.

Even though base stacking and unstacking may occur at modest rates, our model can be refined by permitting diffusion during the lifetime of 2Ap*. The temperature-dependent narrowing of the lifetime distribution (Figure 5) is similar to what has been observed previously in the distance dependence of linked naphthalene-dansyl FRET pairs (Figure 8 in Ref. [135]). In the naphthalene-dansyl system, diffusion-assisted energy transfer was invoked to explain the narrowing of the radial distribution with respect to temperature, and we suggest that a similar process operates in 2Ap-labeled DNA. Future work is needed to understand the precise effects of diffusion in the context of our distance-dependent quenching model.

3.5 Conclusions

This paper has presented a new model for understanding 2Ap fluorescence quenching that draws heavily on understanding of native base photophysics obtained over the past decade. Due to ultrafast ET between stacked bases, 2Ap in a labeled DNA strand is expected to contribute significantly to the emission detected in TCSPC experiments

only when it is not stacked with another nucleobase. Consequently, the multi-exponential decay of emission from the 2Ap probe observed with the TCSPC technique arises from bases that are initially unstacked at the time of photoexcitation.

This model, which challenges the commonly held viewpoint, predicts that there is a nearly instantaneous (on the TCSPC time scale) loss of excited-state population that can be used to quantify the degree of base stacking in the immediate vicinity of the 2Ap probe. The lifetime distribution extracted from emission decays recorded at a fixed temperature can then be interpreted using a simple distance-dependent ET model, according to which unstacked 2Ap undergoes ET with the closest base at a rate that depends exponentially on their mutual separation. Using reasonable parameters to model distance-dependent ET rates, and assuming that changes in the interbase separation occur slowly relative to the lifetime of 2Ap* (negligible diffusion), the emission dynamics are consistent with an interbase distance in most unstacked conformations of between 6 and 10 Å. However, the temperature dependence of the lifetime distribution indicates that some diffusion is occurring. The framework described here provides a starting point for developing models that better capture the interplay between conformational dynamics and distance-dependent emission quenching.

ASSOCIATED CONTENT

See supplementary material (Appendix A) for steady-state spectra, background TCSPC signals, fit residuals, and all fit parameters.

AUTHOR INFORMTION

E-mail: mahesh@iisertvm.ac.in. Phone: +91 (0)471 2778101 (M.H.).
E-mail: kohler.40@osu.edu. Phone: +1 614-688-2635 (B.K.).

ACKNOWLEDGMENTS

This work was funded by Grant No. CHE-1465277 from the NSF. M.H. thanks the Nanobiotechnology Task Force, DBT, Government of India, for financial support (Grant No. BT/PR5761/NNT/106/599/2012). J.R. received financial support from MSU's Graduate School and the Office of the Vice President for Research and Economic Development

REFERENCES

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

13. Madhavan Narayanan, G. K., Yangjun Xing, and Robert J. Stanley, Photoinduced Electron Transfer Occurs between 2-Aminopurine and the DNA Nucleic Acid Monophosphates: Results from Cyclic Voltammetry and Fluorescence Quenching. *J. Phys. Chem. B.* **2010,** *114*, 10573-10580.

15. Evans, K.; Xu, D.; Kim, Y.; Nordlund, T. M., 2-Aminopurine Optical Spectra: Solvent, Pentose Ring, and DNA Helix Melting Dependence. *J. Fluoresc.* **1992,** *2* (4), 209-216.

21. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

26. Chen, J.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

48. Letsinger, R. L.; Wu, T., Use of a Stilbenedicarboxamide Bridge in Stabilizing, Monitoring, and Photochemically Altering Folded Conformations of Oligonucleotides. *J. Am. Chem. Soc.* **1995,** *117*, 7323-7328.

50. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of Lifetime Distributions from Fluorescence Decays with Application to DNA-Base Analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

76. Gohlke, L. B. a. J. r., Fluorescence Probes for Structure. *Annu. Rev. Biochem.* **1972,** *41*, 841-868.

77. T. M. Nordlund, S. A., L. Nilsson, and R. Rigler, A. Graslund, L. W. McLaughlin, Structure and Dynamics of a Fluorescent DNA Oligomer Containing the EcoRI Recognition Sequence: Fluorescence, Molecular Dynamics, and NMR Studies. *Biochemistry* **1989,** *28*, 9095-9103.

78. Jean, J. M.; Hall, K. B., Stacking-Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry* **2004,** *43*, 10277-10284.

79. Bharill, S.; Sarkar, P.; Ballin, J. D.; Gryczynski, I.; Wilson, G. M.; Gryczynski, Z., Fluorescence intensity decays of 2-aminopurine solutions: Lifetime distribution approach. *Anal. Biochem.* **2008,** *377*, 141-149.

80. Torsten Fiebig, C. W., and Ahmed H. Zewail, Femtosecond Charge Transfer Dynamics of a Modified DNA Base: 2-Aminopurine in Complexes with Nucleotides. *Chemphyschem* **2002,** *3*, 781-788.

81. Lobsiger, S.; Blaser, S.; Sinha, R. K.; Frey, H. M.; Leutwyler, S., Switching on the Fluorescence of 2-Aminopurine by Site-Selective Microhydration. *Nat. Chem.* **2014,** *6*, 989-993.

82. Melanie A. O'Neill, J. K. B., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

83. Melanie A. O'Neill, J. K. B., DNA-Mediated Charge Transport Requires Conformational Motion of the DNA Bases: Elimination of Charge Transport in Rigid Glasses at 77 K. *J. Am. Chem. Soc.* **2004,** *126*, 13234-13235.

84. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of lifetime distributions from fluorescence decays with application to DNA-base analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

85. Siemiarczuk, A.; Wagner, B. D.; Ware, W. R., Comparison of the Maximum-Entropy and Exponential Series Methods for the Recovery of Distributions of Lifetimes From Fluorescence Lifetime Data. *J. Phys. Chem.* **1990,** *94*, 1661-1666.

86. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3 ed.; Springer: Berlin Heidelverg, 2006.

87. Ward, D. C.; Reich, E.; Stryer, L., Fluorescence Studies of Nucleotides and Polynucleotides. I. Formycin, 2-Aminopurineriboside, 2,6-Diaminopurineriboside, and their Derivatives. *J. Biol. Chem.* **1969,** *244*, 1228-1237.

88. Somsen, O. J. G.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Structural Heterogeneity in DNA: Temperature Dependence of 2-Aminopurine Fluorescence in Dinucleotides. *ChemPhysChem* **2005,** *6*, 1622-1627.

89. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B* **2007,** *111*, 5757-5766.

90. Wan, C. Z.; Xia, T. B.; Becker, H. C.; Zewail, A. H., Ultrafast Unequilibrated Charge Transfer: A New Channel in the Quenching of Fluorescent Biological Probes. *Chem. Phys. Lett.* **2005,** *412*, 158-163.

91. Somsen, O. J. G.; Trinkunas, G.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Local Diffusive Dynamics in DNA - A time-Resolved Fluorescence and Molecular-Dynamics Study of Dinucleotides with 2-Aminopurine. *J. Lumin.* **2006,** *119*, 100-104.

92. Becker, W., *Advanced Time-Correlated Single Photon Counting Techniques*. Springer: Berlin Heidelberg, 2005; p 415.

93. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

94. Callis, P. R., Binding Phenomena and Fluorescence Quenching. I: Descriptive Quantum Principles of Fluorescence Quenching Using a Supermolecule Approach. *J. Mol. Struct.* **2014,** *1077*, 14-21.

95. Chen, J. Q.; Zhang, Y. Y.; Kohler, B., Excited States in DNA Strands Investigated by Ultrafast Laser Spectroscopy. In *Photoinduced Phenomena in Nucleic Acids Ii: DNA Fragments and Phenomenological Aspects*, Barbatti, M.; Borin, A. C.; Ullrich, S., Eds. Springer-Verlag Berlin: Berlin, 2015; Vol. 356, pp 39-87.

96. Chen, J. Q.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

97. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B* **2012,** *116*, 10266-10274.

98. Narayanan, M.; Kodali, G.; Xing, Y. J.; Stanley, R. J., Photoinduced Electron Transfer Occurs between 2-Aminopurine and the DNA Nucleic Acid Monophosphates: Results from Cyclic Voltammetry and Fluorescence Quenching. *J. Phys. Chem. B* **2010,** *114*, 10573-10580.

99. Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H., Femtosecond Direct Observation of Charge Transfer Between Bases in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97*, 14052-14055.

100. Stuhldreier, M. C.; Temps, F., Ultrafast Photo-Initiated Molecular Quantum Dynamics in the DNA Dinucleotide d(ApG) Revealed by Broadband Transient Absorption Spectroscopy. *Faraday Discuss.* **2013,** *163*, 173-188.

101. Skowron, D. J.; Zhang, Y.; Beckstead, A. A.; Remington, J. M.; Strawn, M.; Kohler, B., Subnanosecond Emission Dynamics of AT DNA Oligonucleotides. *ChemPhysChem* **2016,** *in press*.

102. Liang, J. X.; Nguyen, Q. L.; Matsika, S., Exciplexes and Conical Intersections Lead to Fluorescence Quenching in Pi-Stacked Dimers of 2-Aminopurine With Natural Purine Nucleobases. *Photochem. Photobiol. Sci.* **2013,** *12*, 1387-1400.

103. Kwok, W. M.; Ma, C. S.; Phillips, D. L., Femtosecond Time- and Wavelength-Resolved Fluorescence and Absorption Spectroscopic Study of the Excited States of Adenosine and an Adenine Oligomer. *J. Am. Chem. Soc.* **2006,** *128*, 11894-11905.

104. Stuhldreier, M. C.; Schüler, C.; Kleber, J.; Temps, F., Femtosecond Fluorescence Measurements of the Adenine Dinucleotide: Direct Observation of the Excimer State. In *Ultrafast Phenomena XVII Proceedings of the 17th International Conference,The Silvertree Hotel and Snowmass Conference Center, Snowmass, Colorado, United States, July 18-23, 2010* Chergui, M.; Jonas, D.; Riedle, E.; Schoenlein, R.; Taylor, A., Eds. Oxford University Press: New York, 2011; pp 553-555.

105. Liang, J. X.; Matsika, S., Pathways for Fluorescence Quenching in 2-Aminopurine pi-Stacked with Pyrimidine Nucleobases. *J. Am. Chem. Soc.* **2011,** *133*, 6799-6808.

106. Banyasz, A.; Gustavsson, T.; Onidas, D.; Changenet-Barret, P.; Markovitsi, D.; Improta, R., Multi-Pathway Excited State Relaxation of Adenine Oligomers in Aqueous Solution: A Joint Theoretical and Experimental Study. *Chem. Eur. J.* **2013,** *19*, 3762-3774.

107. Bonnist, E. Y. M.; Jones, A. C., Long-Wavelength Fluorescence from 2-Aminopurine-Nucleobase Dimers in DNA. *ChemPhysChem* **2008,** *9*, 1121-1129.

108. Rist, M.; Wagenknecht, H.-A.; Fiebig, T., Exciton and excimer formation in DNA at room temperature. *ChemPhysChem* **2002,** *3*, 704-707.

109. Kuzmin, M. G.; Soboleva, I. V.; Dolotova, E. V., Transient Exciplex Formation Electron Transfer Mechanism. *Adv. Phys. Chem.* **2011,** *2011*, 813987.

110. Kang, H.; Chou, P. J.; Johnson, W. C.; Weller, D.; Huang, S. B.; Summerton, J. E., Stacking Interactions of ApA Analogs With Modified Backbones. *Biopolymers* **1992,** *32*, 1351-1363.

111. Fornasiero, D.; Kurucsev, T., Vibronic Exciton Interactions - Resolution and Interpretation of the Temperature-Dependent Circular-Dichroism and Absorption-Spectra of ApA and of dApdA. *Eur. J. of Biochem.* **1984,** *143*, 1-7.

112. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Thoery Comput.* **2015,** *11*, 2315-2328.

113. Jafilan, S.; Klein, L.; Hyun, C.; Florian, J., Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B* **2012,** *116*, 3613-3618.

114. Dallmann, A.; Dehmel, L.; Peters, T.; Mugge, C.; Griesinger, C.; Tuma, J.; Ernsting, N. P., 2-Aminopurine Incorporation Perturbs the Dynamics and Structure of DNA. *Angew. Chem. Int. Edit.* **2010,** *49*, 5989-5992.

115. Voltz, K.; Leonard, J.; Touceda, P. T.; Conyard, J.; Chaker, Z.; Dejaegere, A.; Godet, J.; Mely, Y.; Haacke, S.; Stote, R. H., Quantitative Sampling of Conformational Heterogeneity of a DNA Hairpin Using Molecular Dynamics Simulations and Ultrafast Fluorescence Spectroscopy. *Nucleic Acids Res.* **2016,** *44*, 3408-3419.

116. Chodera, J. D.; Noe, F., Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struc. Biol.* **2014,** *25*, 135-144.

117. Albery, W. J.; Bartlett, P. N.; Wilde, C. P.; Darwent, J. R., A General-Model for Dispersed Kinetics in Heterogeneous Systems. *J. Am. Chem. Soc.* **1985,** *107*, 1854-1858.

118. Vix, A.; Lami, H., Protein Fluorescence Decay - Discrete Components or Distribution of Lifetimes - Really No Way Out of the Dilemma. *Biophys. J.* **1995,** *68*, 1145-1151.

119. Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L., Nature OF Biological Electron-Transfer. *Nature* **1992,** *355*, 796-802.

120. Arkin, M. R.; Stemp, E. D. A.; Holmlin, R. E.; Barton, J. K.; Hormann, A.; Olson, E. J. C.; Barbara, P. F., Rates of DNA-mediated Electron Transfer Between Metallointercalators. *Science* **1996,** *273*, 475-480.

121. Murphy, C. J.; Arkin, M. R.; Jenkins, Y.; Ghatlia, N. D.; Bossmann, S. H.; Turro, N. J.; Barton, J. K., Long-Range Photoinduced Electron-Transfer Through a DNA Helix. *Science* **1993,** *262*, 1025-1029.

122. Lewis, F. D.; Wu, T. F.; Zhang, Y. F.; Letsinger, R. L.; Greenfield, S. R.; Wasielewski, M. R., Distance-Dependent Electron Transfer in DNA Hairpins. *Science* **1997,** *277*, 673-676.

123. Tezcan, F. A.; Crane, B. R.; Winkler, J. R.; Gray, H. B., Electron Tunneling in Protein Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 5002-5006.

124. Langen, R.; Chang, I. J.; Germanas, J. P.; Richards, J. H.; Winkler, J. R.; Gray, H. B., Electron-Tunneling in Proteins - Coupling Through a Beta-Strand. *Science* **1995,** *268*, 1733-1735.

125. Lin, J. P.; Balabin, I. A.; Beratan, D. N., The Nature of Aqueous Tunneling Pathways Between Electron-Transfer Proteins. *Science* **2005,** *310*, 1311-1313.

126. Ponce, A.; Gray, H. B.; Winkler, J. R., Electron Tunneling Through Water: Oxidative Quenching of Electronically Excited Ru(tpy)(2)(2+) (tpy=2,2 ': 6,2 "-terpyridine) by Ferric Ions in Aqueous Glasses at 77 K. *J. Am. Chem. Soc.* **2000,** *122*, 8187-8191.

127. Tavernier, H. L.; Fayer, M. D., Distance Dependence of Electron Transfer in DNA: The Role Of The Reorganization Energy and Free Energy. *J. Phys. Chem. B* **2000,** *104*, 11541-11550.

128. Norberg, J.; Nilsson, L., Solvent Influence on Base Stacking. *Biophys. J.* **1998,** *74*, 394-402.

129. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

130. O'Neill, M. A.; Barton, J. K., DNA-Mediated Charge Transport Requires Conformational Motion of the DNA Bases: Elimination of Charge Transport in Rigid Glasses at 77 K. *J. Am. Chem. Soc.* **2004,** *126*, 13234-13235.

131. O'Neill, M. A.; Barton, J. K., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

132. Porschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

133. Heyn, M. P.; Nicola, C. U.; Schwarz, G., Kinetics of Base-Stacking Reaction of N6,N6-Dimethyladenosine - Ultrasonic-Absorption and Dispersion Study. *J. Phys. Chem.* **1977,** *81*, 1611-1617.

134. Jean, J. M.; Krueger, B. P., Structural Fluctuations and Excitation Transfer Between Adenine and 2-Aminopurine in Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B* **2006,** *110,* 2899-2909.

135. Lakowicz, J. R.; Wiczk, W.; Gryczynski, I.; Szmacinski, H.; Johnson, M. L., Influence of End-To-End Diffusion pn Intramolecular Energy-Transfer as Observed by Frequency-Domain Fluorometry. *Biophys. Chem.* **1990,** *38,* 99-109.

CHAPTER FOUR

MOLECULAR DYNAMICS SIMULATIONS OF 2-AMINOPURINE-LABELED DINUCLEOSIDE MONOPHOSPHATES REVEAL MULTISCALE STACKING KINETICS

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Jacob M. Remington

Contributions: Ran all MD simulations and analyzed all the data. Constructed draft document and kept it up to date.

Co-Author: Martin McCullagh

Contributions: Provided parameters and guidance for MD simulations. Also helped prepare the final text.

Co-Author: Bern Kohler

Contributions: Provided key insights for the simulation and modeling of the data. Wrote and aided in the preparation of the manuscript.

Manuscript Information

Jacob M. Remington, Martin McCullagh, and Bern Kohler

Journal of Physical Chemistry B

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
_x__ Published in a peer-reviewed journal

American Chemical Society
In Volume 123, 2291-2304 (2019)
DOI: 10.1021/acs.jpcb.8b12172

MOLECULAR DYNAMICS SIMULATIONS OF 2-AMINOPURINE-LABELED DINUCLEOSIDE MONOPHOSPHATES REVEAL MULTISCALE STACKING KINETICS

Jacob M. Remington,[1] Martin McCullagh,[2] and Bern Kohler[3]

[1]Department of Chemistry, Montana State University, Bozeman, Montana 59715, USA
[2]Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, USA
[3]Department of Chemistry and Biochemistry, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210, USA

Abstract

Molecular dynamics (MD) simulations of 2-aminopurine (2Ap)-labeled DNA dinucleoside monophosphates (DNMPs) were performed to investigate the hypothesis that base stacking dynamics occur on timescales sufficiently rapid to influence the emission signals measured in time-resolved fluorescence experiments. Analysis of multiple microsecond-length trajectories shows that the DNMPs sample all four coplanar stacking motifs. In addition, three metastable unstacked conformations are detected. A hidden Markov-state model (HMSM) was applied to the simulations to estimate transition rates between the stacked and unstacked states. Transitions between different stacked states generally occur at higher rates when the number of nucleobase faces requiring desolvation is minimized. Time constants for structural relaxation range between 1.6 and 25 ns, suggesting that emission from photoexcited 2Ap, which has an excited-state lifetime of 10 ns, is sensitive to base stacking kinetics. A master equation model for the excited-state population of 2Ap predicts multiexponential emission decays that reproduce the sub-10 ns emission decay lifetimes and amplitudes seen in

experiments. Combining MD simulations with HMSM analysis is a powerful way to understand the dynamics that influence 2Ap excited-state relaxation and represents an important step toward using observed emission signals to validate MD simulations.

4.1 Introduction

The adenine isomer 2-aminopurine (2Ap) is a highly fluorescent-modified nucleobase that has been successfully used as a probe of nucleic acid structure and dynamics (Figure 4.1).[10, 18-20] As a monomer in aqueous solution, 2Ap is highly emissive, and a monoexponential decay with a lifetime of 10 ns is observed in time-resolved fluorescence (TRF) measurements. However, when 2Ap is incorporated into a DNA strand, its emission is strongly quenched, and the measured emission signals exhibit multiple exponential decay components.[11, 22, 30, 46, 50, 78, 136-139] This complexity is already observed in dinucleoside monophosphates (DNMPs), the shortest possible DNA strands with just two interacting nucleosides joined by a phosphate group. For example, studies of 2Ap-labeled single-stranded DNA dinucleotides show that three to five exponentials are needed to fit the measured emission signals between 100 ps and 50 ns.[14, 30] Even though the 2Ap fluorophore interacts with only a single nucleobase to which it is joined via the usual phosphate linkage found in nucleic acids, detailed understanding of the mechanisms responsible for fluorescence quenching is still lacking.

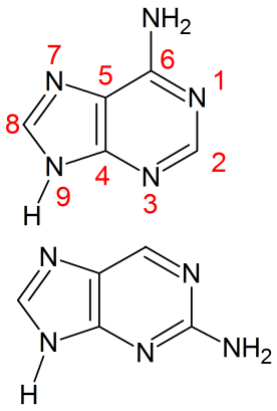

Figure 4.1. Chemical structures of adenine (top, with ring numbering) and its fluorescent isomer 2Ap (bottom). The α face (explained in the results section) is oriented toward the viewer in each.

It is widely believed that base stacking quenches fluorescence by a 2Ap label in an oligonucleotide.[12, 14, 21-22] The observed complexity of the emission signals likely reflects an underlying diversity of structures, which transition between stacked and unstacked structures over multiple timescales. However, the structures of dinucleotides and DNMPs are commonly considered using two-state models in which the bases are either stacked or unstacked.[23-25] These models have been used to estimate base stacking and unstacking rates from ultrasound and temperature-jump measurements,[23, 25, 140-141] but it is not clear that the resulting rates accurately capture the underlying dynamics. Recent molecular dynamics (MD) simulations of RNA DNMPs indicate that base stacking plays out on a complex structural landscape that involves six or more distinct regions of conformational space.[38, 142] Structural heterogeneity was also found by ab initio calculations of the stacking potential energy surface, which identified 10− 18 minima for two stacked bases without any backbone present.[143]

There is general agreement that the complex TRF signals seen in 2Ap-labeled oligonucleotides arise from multiple DNA structures (structural heterogeneity) and transitions between these structures (conformational dynamics).[10] The role of conformational dynamics is most strongly supported by temperature-dependent measurements, which show that increasing the temperature actually decreases the mean emission decay lifetime of the 2Ap label.[14, 30] This is not predicted by a static model as increasing the temperature decreases the number of stacked bases. Although both static quenching and dynamic quenching are known to be important, a model that correctly predicts the balance between them for a particular 2Ap-labeled oligomer is lacking.

Markov-state models (MSMs) can be used to estimate the kinetics of interconversion among the many structures sampled by biomolecules.[41-43, 144] MSMs work by assuming that the memory of previous structures is lost on a much shorter timescale than a typical structural residency time.[145] By using a lag time to neglect short-time memory effects, the long-time probability of transitioning between any two structures is a time-independent rate constant that can be estimated by counting transitions and dividing by the residency time of the initial state.[146] While discretizing the conformational space into disjoint regions provides a way to count transitions, more advanced methodologies have been developed based on network models.[147-148] Unfortunately, the full conformational space for a DNMP has too many degrees of freedom to make discretization practical, so projections onto reduced dimensional free energy surfaces (FESs) are required.[149] The MSMs constructed on a reduced dimensional FES are then termed projected Markov models.[150] This has the practical advantage of reducing model complexity so that the important conformations of the biomolecule can

be identified. Noé et al. have shown that projected Markov models can be approximated by hidden Markov-state models (HMSMs), which assume that conformations are not necessarily disjoint on the FES, and it is this approach that will be used here to quantify the structural heterogeneity and conformational dynamics of the DNMPs 5′-(2Ap)A-3′ and 5′-A(2Ap)-3′ .[145, 150]

Pinamonti et al.[142] recently used MSMs to analyze MD simulations of several RNA oligonucleotides and showed that structural conversions occur over multiple timescales ranging from a few nanoseconds to 80 ns. Despite the dynamical complexity revealed in this work, the authors compared the stacking kinetics inferred from the simulations with temperature-jump experiments that predict first-order folding kinetics. This begs the question of whether other experimental observables might provide a better basis for evaluating the simulated dynamics. Here, we propose that the multiexponential TRF signals from 2Ap-labeled oligonucleotides can provide a sensitive test of the multiscale stacking kinetics seen in simulations like those in ref [142]. To explore this possibility, MD simulations of 5′-(2Ap)A-3′ and 5′-A(2Ap)-3′ DNA DNMPs were performed and analyzed with HMSMs. The twin aims of our study are to (1) quantify the structural heterogeneity and conformational dynamics predicted by the current Amber force fields using HMSMs and (2) assess the role of structural dynamics in TRF quenching experiments.

4.2 Methods

4.2.1 MD Simulations

MD simulations of all DNMPs were run using the GPU-accelerated Amber package for explicit solvent particle mesh Ewald summation-based simulations.[36] Amber force field parameters ff99bsc0[32] were used along with refined base stacking parameters,[33] χ dihedral parameters,[34] and ε and ζ backbone torsion parameters.[52] The restrained electrostatic potential (RESP) fitting procedure was used to derive point charges for the 2Ap base.[53] A methyl group was used as a cap for the nucleobase during the charge fitting procedure. The nonstandard nucleobase was optimized at the B3LYP/6-31G* level of theory, and subsequently, the electrostatic potential was calculated using the HF/6-31G* level of theory. Both geometry optimization and electrostatic potential calculations were performed in the GAMESS program.[54-55] Atomic charges were fit to the electrostatic potential using the RESP module in Amber (Table A.2.1 and Figure A.2.1). Also included in Table A.2.1 are the atom types. DNA residues of the 2Ap base were developed by combining the parameterized sugar backbone with the parameters for the nonstandard nucleobase. Initial structures were generated using LEaP, and a single Na^+ counterion[56] was then added to each DNMP. The DNMP was solvated using enough TIP3P water molecules (1594 and 1590 waters for the 5′ - and 3′ -labeled samples) to fill a rectangular box such that no solute atom was less than 12 Å from the edge of the box. Na^+ and Na^- ions[56] were then added to simulate a salt concentration of 100 mM.

These initial structures were put through a multistage minimization and equilibration procedure. During each of four minimization and four equilibration stages,

restraints on the solute atomic positions were reduced from very strong to nonexistent. This resulted in a density near 1.000 ± 0.005 g mL^{-1} and a final rectangular box volume of 35 X 35 X 40 $Å^3$. The fourth equilibration stage was run for 1 ns with the same settings used in production runs. Specifically, a constant temperature of 300 K was achieved using the weak-coupling algorithm[59] with a 10 ps time constant, and the isotropic Monte Carlo barostat with a relaxation time of 1 ps maintained constant pressure. The SHAKE algorithm was used to constrain bonds involving hydrogen,[35] allowing a 2 fs time step to be utilized for the simulations. A 10 Å cutoff was used for nonbonding interactions. Finally, five trajectories were run for 2 μ s for each DNMP with all coordinates saved every 2 ps. Simulations were conducted using NVIDIA Pascal P100 GPU nodes on the Ohio State Supercomputing Center Owens Cluster.[60]

4.2.2 Postprocessing

Structures were first stripped of water and ions using the CPPTRAJ program.[61] The coordinates for each were analyzed with in-house Python scripts utilizing the MDAnalysis package.[62-63] The center of mass (COM) position was calculated for each base in the DNMP as if it were a free base (i.e., by ignoring the ribose group). The COM positions for the two bases in this study, adenine and 2-aminopurine, are illustrated in Figure 2. Next, the six coordinates needed for the rigid body transformations that bring the 5' nucleobase into overlap with the 3' base were calculated. To do this, three orthogonal unit vectors, $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, were found for each purine nucleobase using the position vectors ($\vec{x}$) for the N1, N7, and N9 atoms (Figure 4.2). The first axis, $\vec{v}_1$, was calculated by normalizing $(\vec{x}_{\mathrm{N1}} - \vec{x}_{\mathrm{N9}})$, then $\vec{v}_3$ as $(\vec{x}_{\mathrm{N7}} - \vec{x}_{\mathrm{N9}})/|(\vec{x}_{\mathrm{N7}} - \vec{x}_{\mathrm{N9}})| \times \vec{v}_1$, and

finally $\vec{v}_2$ as $\vec{v}_1 \times \vec{v}_3$. For a nucleobase in a double-stranded B-form DNA structure, $\vec{v}_1$ and $\vec{v}_2$ point toward the major groove and toward the complementary strand, respectively, while $\vec{v}_3$ is directed in the 3' to 5' direction.[67-70] The three distance parameters (shift, slide, rise) were calculated by projecting the vector defined by the difference between the two center of mass vectors onto the three 5' base axes. However, instead of the usual three angular parameters (roll, tilt, and twist), the projection of the 3' base axes onto the respective 5' base axes were used so that the projection of the two base normal vectors represents the cosine of the angle between the two nucleobase planes, cos(θ).

It was found that the important conformational states can be easily visualized by projecting the six-dimensional space onto the two-dimensional space formed by cos(θ) and the COM distance between the two nucleobases multiplied by the sign of the rise helical parameter (a one-dimensional displacement vector). FESs were calculated by the negative logarithm of the normalized 2d histogram formed by binning the trajectories along these two coordinates. For each DNMP the zero for free energy was set to the bin with the largest number of counts. Additional FESs were calculated for the $\chi_{5'}$ and $\chi_{3'}$ dihedrals defined in Figure 2. Following convention,[151] the χ glycosidic bond torsion angles are classified as *syn* or *anti* depending on their values: *syn*: (–90°, 90°), *anti*: (–180°, –90°) or (90°, 180°).

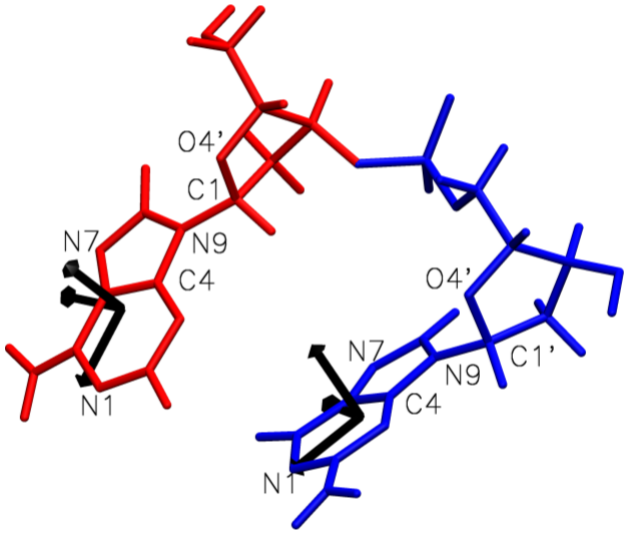

Figure 4.2. The atomic coordinates used to calculate the helical parameters and glycosidic dihedral angles illustrated for the 5'-A(2Ap)-3' DNMP. The 5' base is shown in red, and the 3' base is shown in blue. The three orthogonal axes for each base defined using N9, N7, and N1 are shown in black with the origin positioned at the COM. For the dihedral angles about the glycosidic bonds, $\chi_{5'}$ and $\chi_{3'}$, the O4', C1', N9, and C4 atomic coordinates were used.[65] In the extended structure shown, the rise is negative, while $\cos(\theta)$ is positive.

Both standard MSMs and HMSMs were constructed using the Pyemma Package.[73] First, the six rigid body transformation coordinates and the $\chi_{5'}$ and $\chi_{3'}$ dihedrals from each independent 2 μs run were normalized by their range and shifted to have zero mean. While directly calculating the mean value of a periodic variable such as an angle can yield incorrect values of the mean, it sufficed here for preprocessing the data for clustering. A dataset including all five of the runs was then clustered using the regular space clustering algorithm with a minimum distance of 0.18 and a stride of 200 ps, resulting in around 2000 cluster centers.[43] Implied timescales were calculated and a lag time of 500 ps was chosen for reversible MSM estimation for both DNMPs (Figure A.2.2A and B). At this lag time, at least six timescales were apparent for both samples, although the ratio between the 6th and 7th timescale was only 1.2. Additional HMSMs

were also constructed using other lag times (Figure A.2.3A-C for lag times of 0.50, 0.75, and 1.25 ns), and assuming a different number of states (Figure A.2.4A-D for 7, 8, 9, or 10 states). A Chapman-Kolmogorov test was conducted assuming seven metastable states (Figure A.2.5).[43] The MSMs were then coarse grained to seven states and the resulting memberships investigated to ensure good separation of states for both DNMPs, 5'-A(2Ap)-3' (Figure A.2.6A) and 5'-(2Ap)A-3' (Figure A.2.6B).[152] The seven states include the four possible types of stacks (i.e., $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, and $\beta\beta$, see the results section for further details), two states denoted as sheared conformations, and a large number of conformations in which the bases are generally quite distant (Ex, for extended conformation).

The static occupation probabilities were taken from the equilibrium eigenvector of the reversible HMSM constructed using all five datasets. The transition probability matrix encoding the probability of transitioning between the seven states over a 500 ps window was converted to a rate matrix by dividing by the 500 ps lag time. Furthermore, the diagonal elements were adjusted to yield a zero row sum in order to conserve population (see Tables 2 and 3). To estimate error bars for the interconversion rates, the procedure was repeated for the individual 2 μs runs. The errors listed in the text and in the tables are the sample standard deviations calculated from the set of 2 μs trajectories and are hereafter referred to as 1σ errors. Investigating the FESs (Figure 4.3) calculated for individual runs showed that the αβ stack (see Results section for a definition) was visited for all of the 5′ -A(2Ap)-3′ runs but only three of the five 5′-(2Ap)A-3′ runs. For each of the eight runs where all four stacks were visited, transition probability matrices

were calculated and used to obtain 1σ errors for each DNMP using in-house Python scripts that utilize the Scipy and Numpy packages.[64]

To investigate changes in the solvent accessible surface area (SASA) during structural transitions, the SASA of the entire DNMP was first calculated with CPPTRAJ using the molsurf command with a probe of 1.4 Å.[66] Then the trajectories were discretized using Shapely,[74] and the regions are indicated in Figure A.2.7A to get sequences of conformations visited. A search was conducted to find transitions between any pair of conformations that occurred within 500 ps using a sliding window. Transitions were only saved when the initial/final states were visited for at least 75% of a 25 ps padding time inserted before and after the 500 ps transition time. The SASA during these 550 ps transition paths was averaged over all transitions of a particular type to produce the curves shown in Figure 4.6 in the Results section. This procedure was repeated using different regions (Figure A.2.7B) and different parameters to accept transitions (padding time of 30− 70 ps and visitation limit of 70− 80%) to verify that similar results were obtained (Figure A.2.8).

The creation and diagonalization of master equations to predict the population of electronically excited 2Ap were also done using in-house Python scripts that utilize the Scipy and Numpy packages.[64] To calculate the amplitudes associated with each predicted decay rate, $A_i = (\vec{N}(0)\cdot v_i)(\vec{O}\cdot v_i)$ is used,[153] where $\vec{N}(0)$ is the initial excited-state population, assumed to be just the ground-state equilibrium obtained from the HMSMs, and $\vec{O}$ is a vector containing the relative radiative rates in each conformation. For the purposes here, $\vec{O}$ was set to a vector of ones, corresponding to the assumption that the radiative rate in each conformation is identical.

4.3 Results

A robust algorithm for identifying base-stacked conformations of the DNMPs is a prerequisite for analyzing both structural heterogeneity and base stacking kinetics. This algorithm must also identify the different stacking motifs that arise because of the two distinct faces of each nucleobase. Following Rose,[154] the two faces of purine bases are denoted by α and β , according to whether the five-membered imidazole ring is at 9 or 3 o' clock, respectively, when the vector pointing from C4 to C5 is oriented toward 12 o' clock.[26] A DNMP can therefore adopt any of four possible cofacial stacking motifs, depending on whether the α or β face of the 5′ base is stacked against the α or β face of the 3′ base. Henceforth, these will be referred to as $\beta\alpha$, $\beta\beta$, $\alpha\alpha$, and $\alpha\beta$, where $\beta\alpha$ indicates that the β face of the 5′ base is stacked against the α face of the 3′ base. For example, the DNMP shown in Figure 4.2 has $\alpha\beta$ stacking, and the β faces of the nucleobases face the viewer.

The six helical parameters[67] describing the rigid body transformations bringing the 5′ base axes into alignment with 3′ base axes (Figure 4.2) encode the face−face stacking motifs in the sign of the rise and the angle between the two nucleobase planes, θ (see Figure 4.2). The sign of the rise indicates whether the 5′ base is showing its α (rise < 0) or β face (rise > 0) to the 3′ base. Similarly, the sign of $\cos(\theta)$ determines if the 3′ base is showing its α ($\cos(\theta) > 0$) or β face ($\cos(\theta) < 0$) to the 5′ base. Here, we use the COM distance multiplied by the sign of the rise (henceforth, this product will be referred to as displacement) and the cosine of θ to define a 2D FES that can easily distinguish the four stacks (Figure 4.3).

Both Figure 4.3 A,B reveal four minima where the bases are close (|displacement| < 5.5 Å) and nearly coplanar (0.75 < | cos(θ)|), corresponding to the four cofacial stacking motifs of the two sequence isomers. A great diversity of unstacked conformations contributes to the FESs in Figure 4.3 , but even here, some structure is apparent. When the rise is negative, relatively low-energy structures are found that have displacement values between −7.5 and −5 Å and 0.3 < | cos(θ)| < 1. These are designated as sheared conformations because of their resemblance to structures seen in simulations of RNA DNMPs by Hayatshahi et al.[38] These structures have intermediate proximity/coplanarity and are denoted with an asterisk, $\alpha\alpha$* and $\alpha\beta$*. The remaining regions of the 2D coordinate space represent the additional extended conformations of the DNMP, which will be denoted by Ex.

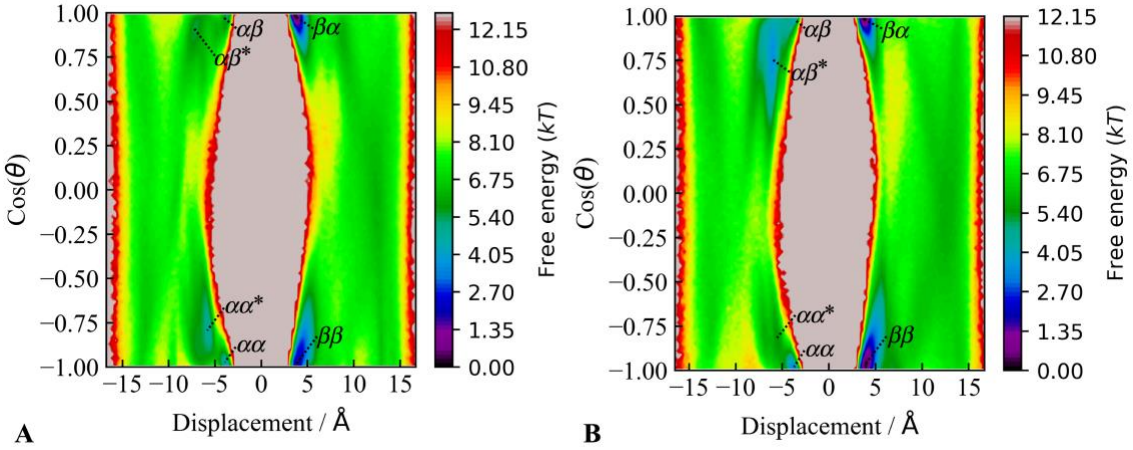

Figure 4.3. FESs calculated from five 2 μs MD simulations of (A) 5'-A(2Ap)-3' and (B) 5'-(2Ap)A-3'. The displacement plotted on the x-axis is the COM distance between nucleobases multiplied by the sign of the rise helical parameter, and θ is the angle between the nucleobase planes. The four cofacial stacking motifs and the two sheared conformations are labeled according to the text with dashed lines positioned near their respective minima.

Although the 2D surfaces in Figure 4.3 are informative, they lack the atomic detail that MD simulations provide. Sample structures for the seven states taken from near the minima in Figure 4.3 are shown in Figure 4.4 . The extended structure shown was selected arbitrarily from the broad distribution of structures that characterize that state. Examining these structures reveals key differences between the two most likely stacks; in the *βα* stack, the twist is modest (36.3°), whereas the *ββ* stack has a large twist value (155°). Mean twist, $\chi^{3'}$, and $\chi^{5'}$ values are tabulated in Table A.2.2 . This twist arises because the dihedral angle $\chi^{3'}$ is in the *syn* orientation for *βα* and *anti* for *ββ* .[155] On the other hand, $\chi^{5'}$ is *anti* for αα and *syn* for *αβ* stacks.

In the sheared conformations, $\chi^{5'}$ adopts values similar to the ones observed in the highly populated ββ and βα stacks. Partitioning the 2D FESs along $\chi^{3'}$ and $\chi^{5'}$ for both DNMPs in Figure A.2.9 demonstrates how these coordinates partition the

displacement/angle space shown in Figure 4.3 (and in Figure A.2.10). The images in Figure 4.4 show that sheared conformations involve noncovalent interaction of the 3′ nucleobase face with the O4′ atom on the deoxyribose sugar backbone. The solvent around these sheared conformations does not appear to penetrate between the 3′ nucleobase face and the backbone, while both faces of the 5′ base are solvated.

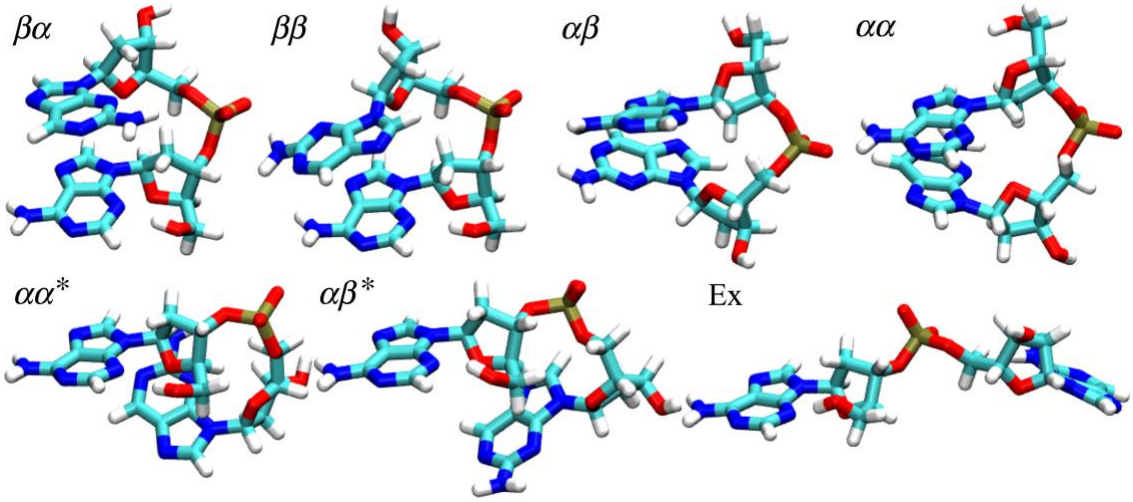

Figure 4.4 Representative structures of the seven metastable conformational states of the 5'-A(2Ap)-3' DNMP taken near the minima in Figure 4.3A. The labels for each structure are defined in the text. In each structure, the 5' base (adenine) is always shown in the same orientation with its amino group pointing to the left.

Now that the structural heterogeneity in these DNMPs has been classified, the conformational dynamics can be estimated using HMSMs.[150] First, the use of Markovian kinetics is justified qualitatively by investigating the stochastic behavior of the COM distance and angle time traces (Figure 4.5). Figure 4.5 reveals the separation of timescales between infrequent transitions among the different metastable states (e.g., the $\beta\beta \rightarrow$ Ex transition seen near 50 ns) and the rapid fluctuations about the metastable minima that precede these transitions (e.g., at times between 30 and 50 ns). First, ~2000

state MSMs were constructed using the eight-dimensional dataset defined by the six rigid body transformations plus $\chi^{3'}$ and $\chi^{5'}$. A variety of other coordinates were tried including ones from a time-lagged independent coordinate analysis[72] of all atomic positions in the reference frame defined by the 5′ base axes, all backbone dihedral angles, and a coordinate set that included only the displacement and cos(θ), but all produced similar MSMs. The MSMs were then coarse-grained into seven-state HMSMs. Further validation and construction of the MSMs/ HMSMs are described in the Appendix 2 (see Figures A.2.2− A.2.6); however, it is noted here that the seven aforementioned structures coincide with six timescales at a 500 ps lag time and fall out naturally from the metastable memberships in the HMSMs (Figure A.2.6).

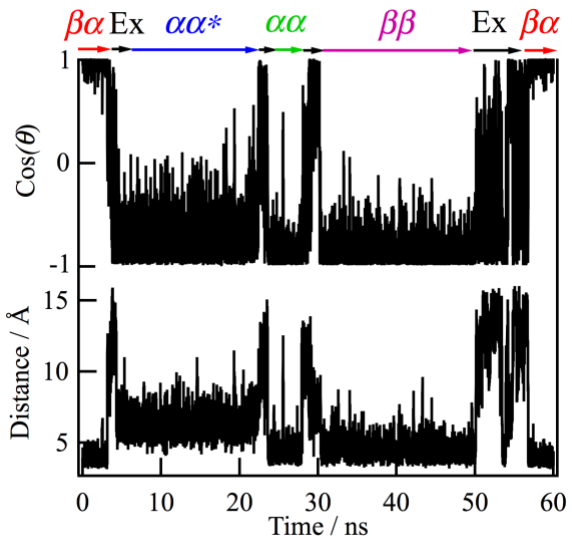

Figure 4.5. A 60 ns portion of the center of mass distance time trace (bottom) and the cosine of the angle between the two nucleobase normal vectors (top) for the 5'-A(2Ap)-3' DNMP. The structures visited during the 60 ns are shown above.

The equilibrium eigenvectors from the resultant HMSMs, which represent the probability of being in a particular conformation, are shown in Table 4.1. In agreement with Figure 4.3 , the *ββ* and *βα* stacks are the most populated structures for both DNMPs. In 5′-A(2Ap)-3′, *βα* is the most probable stack, whereas *ββ* has more population in 5′-(2Ap)A-3′ . On the other hand, the *αβ* and *αα* stacks are both rare with populations near 1%. The former is the stacking conformation found in crystal structures of double-stranded B-DNA. The unstacked conformations (Un) are roughly equally distributed between the *αβ** , *αα** , and Ex conformations. Calculating the stacking equilibrium constant, K_s = [St]/[Un], from the HMSM equilibrium populations gives values of 2.6 ± 0.8 for 5′-A(2Ap)-3′ and 2.7 ± 0.4 for 5′-(2Ap)A-3′. The general location for each of the seven structures is shown in Figure 4.3, and the values of $\chi^{3'}$ and $\chi^{5'}$ are also shown in Table 4.1.

Table 4.1. Relative probabilities and characteristic structural parameters for the conformations in the HMSMs.[a]

Sample	**State**	**Prob. / %**	**Sign(rise)**	**Sign(cos(*θ*))**	**$\chi_{5'}$**	**$\chi_{3'}$**
5’-A(2Ap)-3’	*βα*	43.7 (7.6)	+	+	*syn*	*syn*
	ββ	25.7 (8.5)	+	–	*syn*	*anti*
	αα	1.5 (1.1)	–	–	*anti*	*syn*
	*αα**	5.8 (3.2)	–	–	*syn*	*syn*
	αβ	1.1 (0.3)	–	+	*anti*	*anti*
	*αβ**	8.9 (1.4)	–	+	*syn*	*anti*
	Ex	13.3 (6.0)	+/–	+/–	*syn*/*anti*	*syn*/*anti*
5’-(2Ap)A-3’	*βα*	23.1 (2.4)	+	+	*syn*	*syn*
	ββ	47.1 (3.3)	+	–	*syn*	*anti*
	αα	1.4 (1.3)	–	–	*anti*	*syn*
	*αα**	7.5(1.5)	–	–	*syn*	*syn*
	αβ	1.3 (3.2)	–	+	*anti*	*anti*
	*αβ**	12.2 (3.2)	–	+	*syn*	*anti*
	Ex	7.4 (1.8)	+/–	+/–	*syn*/*anti*	*syn*/*anti*

[a]Values in parentheses are 1σ errors estimated from independent trajectories.

For both DNMPs, the rate matrices containing the individual transition rates between the different structures were estimated from a single dataset formed by joining the five independent 2 μs trajectories, while the 1σ errors were estimated from HMSM analyses of the five 5′-A(2Ap)-3′ trajectories and three 5′-(2Ap)A-3′ trajectories because two of the latter trajectories did not visit the *αβ* state. The rates are shown in units of μs^{-1} in Tables 4.2 and 4.3. Inspection of Tables 4.2 and 4.3 shows that rates from extended and sheared conformations to stacked states (i.e., stacking transitions from any of the Un states to any of the stacked states *αβ*, *αα*, *βα*, or *ββ*) are generally greater than unstacking rates. The stacking rates have a broad range from $<1\ \mu s^{-1}$ (*αα** → *αβ*) to $\sim 200\ \mu s^{-1}$ (*αβ** → *ββ*). All four stacking/unstacking transitions associated with *αα** ↔ *αβ* and *αβ** ↔ *αα* have very small rates. Interconversions between the different unstacked conformations occur at moderate rates that exceed $10\ \mu s^{-1}$.

For both DNMPs, transitions for which both the rise and cos(*θ*) change sign generally have larger rates than ones when only the rise changes sign. In Tables 4.2 and 4.3, the former are highlighted in blue, whereas the latter are shown in red text. The favored transitions also include the aforementioned *αβ** → *ββ* stacking transition, which has the largest rate for 5′-A(2Ap)-3′. Finally, with a few exceptions, transitions when only cos(*θ*) changes sign (purple in Tables 4.2 and 3) have larger rates than when only the rise changes sign (red in Table 4.2).

Table 4.2. Transition rates (in μs^{-1}) for 5'-A(2Ap)-3'.[a,b,c]

Initial \ Final	$\beta\alpha$	$\beta\beta$	$\alpha\alpha$	$\alpha\beta$	$\alpha\alpha$*	$\alpha\beta$*	Ex
$\beta\alpha$	-51 (7)	2.3 (0.8)	0.7 (0.5)	0 (0)	7 (19)	2.6 (1.4)	38 (21)
$\beta\beta$	3.8 (2.7)	-96 (7)	0.1 (0.7)	1.7 (1.1)	1.7 (2.9)	83 (9)	5.6 (2.6)
$\alpha\alpha$	20 (22)	2 (5)	-192 (86)	32 (37)	18 (89)	5 (8)	116 (70)
$\alpha\beta$	0 (0)	39 (11)	44 (31)	-185 (80)	0 (0)	94 (48)	8 (84)
$\alpha\alpha$*	51 (77)	7 (7)	5 (5)	0 (0)	-262 (89)	15 (25)	183 (82)
$\alpha\beta$*	13 (5)	240 (48)	0.9 (1.6)	12 (4)	10 (34)	-372 (41)	97 (36)
Ex	127 (61)	11 (6)	14 (8)	1 (54)	81 (35)	66 (56)	-300 (51)

[a]Values in parentheses are 1σ errors estimated from independent trajectories.

[b]With the exception of transitions involving the Ex state, transition rates are color-coded, depending on whether rise (red), cos(θ) (purple), or both rise and cos(θ) change sign (blue).

[c]For clarity, values less than 0.5 μs^{-1} are shown as zero.

Table 4.3 Transition rates (in μs^{-1}) for 5'-(2Ap)A-3'.[a,b,c]

Initial \ Final	$\beta\alpha$	$\beta\beta$	$\alpha\alpha$	$\alpha\beta$	$\alpha\alpha$*	$\alpha\beta$*	Ex
$\beta\alpha$	-60 (7)	7.6 (1.8)	1.1 (0.7)	0 (0)	44 (17)	1.6 (1.0)	6 (23)
$\beta\beta$	3.7 (1.4)	-889 (6)	0 (0)	1 (16)	3.0 (1.5)	49.1 (6.2)	32 (14)
$\alpha\alpha$	17 (26)	7 (20)	-100 (38)	34 (24)	41.7 (0.3)	0 (0)	0 (0)
$\alpha\beta$	0 (0)	41 (87)	37 (11)	-200 (220)	4 (11)	68 (62)	56 (88)
$\alpha\alpha$*	137 (56)	19 (14)	7.8 (0.4)	0 (6)	-315 (14)	33 (12)	117 (66)
$\alpha\beta$*	3.1 (1.2)	190 (31)	0 (0)	7 (48)	21 (3)	-317 (25)	96 (35)
Ex	18 (50)	200 (120)	0 (0)	10 (73)	119 (24)	157 (87)	-500 (11)

[a]Values in parentheses are 1σ errors estimated from independent trajectories.
[b]With the exception of transitions involving the Ex state, transition rates are color-coded, depending on whether rise (red), cos(θ) (purple), or both rise and cos(θ) change sign (blue).
[c]For clarity, values less than 0.5 μs^{-1} are shown as zero.

Tables 4.2 and 3 were obtained by enforcing detailed balance with respect to the observed distribution of conformations, which ensures that an eigenvalue of zero is associated with the equilibrium eigenvector.[40, 43] The reciprocals of the remaining six nonzero eigenvalues are the relaxation times. These values (with 1σ errors) are calculated to be 2.09 (0.26), 2.63 (0.41), 4.32 (0.92), 5.4 (1.1), 7.0 (2.5), and 23.0 (1.4) ns for 5′-A(2Ap)-3′ and 1.66 (0.22), 2.71 (0.35), 3.72 (0.52), 4.7 (1.5), 11.1 (6.6), and 20.6 (5.3) ns for 5′-(2Ap)A-3′. The eigenvectors associated with each of the relaxation times[43] identify the structural transitions occurring on that timescale. This is demonstrated graphically in Figure A.2.11 , where the interconverting structures result in oppositely signed amplitudes in the eigenvector. The eigenvectors show that the slowest modes are associated with transitions between different stacking motifs and the intermediate modes

resemble various unstacked to stacked transitions, while the fastest modes represent interconversion between the extended and sheared conformations. The general trends described above were also observed when lag times of 0.75 and 1.25 ns were used to construct the HMSMs (Figure A.2.3).

4.4 Discussion

4.4.1 Stacking Heterogeneity of the DNMPs

Previous MD studies of RNA and DNA DNMPs have compared two-state stacked/unstacked equilibrium constants predicted by customized force fields to experimental measurements.[24, 38, 156] However, as Pinamonti et al. pointed out,[142] the experiments suggesting two-state folding may capture only the slowest component of the dynamics. As discussed below, our MD simulations support the conclusion that base stacking is an inherently multiscale phenomenon. We propose that the famously multiexponential fluorescence decay signals observed in the studies of 2-Ap-containing DNA strands[10] are evidence for this and can thus be used to rigorously test the dynamical predictions made by MD simulations.

The predicted K_s values of 2.6 ± 0.8 for 5′-A(2Ap)-3′ and 2.7 ± 0.4 for 5′-(2Ap)A-3′ fall within the wide range of experimental values reported for $(dA)_2$ using circular dichroism spectroscopy (1.8−11.5).[31, 157] The K_s values determined here are lower than the value of 4.49 ± 0.84 for $(dA)_2$ determined by Brown et al.[24] from MD simulations of DNA DNMPs using Amber force fields that lacked the stacking corrections used here. These corrections are needed to avoid overstabilizing base stacking. Our K_s values are larger than the value of 1.0 ± 0.03 found by Hayatshahi et

al.[38] for the RNA DNMP 5′-AA-3′ in simulations that used stacking corrections. Experiments have shown greater stacking for the DNA 5′-AA-3′ DNMP than for the RNA form.[24, 158]

Hayatshahi et al.[38] also sought to quantify the structural heterogeneity of RNA DNMPs using temperature replica exchange MD simulations. Using *K*-means clustering of the heavy atom coordinate space, they found that 5′-AA-3′ adopts five different structures, which they designated A-form, ladder, inverted, sheared, and extended. These structures correspond, respectively, to the *αβ*, *ββ*, *βα*, *αβ**, and Ex structures of this study. It is unclear whether their analysis missed the *αα** sheared conformation and the rare *αα* stack, or if these structures are only observed in DNA DNMPs like the ones studied here.

The most probable stacks observed here, *ββ* and *βα*, were also the most probable in the simulations in ref [38]. Our 2D $\chi^{5'}$ and $\chi^{3'}$ FESs (Figure A.2.9) are consistent with their 1D histograms for $\chi^{5'}$ and $\chi^{3'}$ (ref [38], Figure S7), which are both bimodal and favor positive (*syn*) values of $\chi^{5'}$ and $\chi^{3'}$. Figure A.2.10 shows how these variables partition the displacement and cos(θ) space. Furthermore, Hayatshahi et al.[38] found that reversing the base sequence in purine-containing DNMPs (e.g., AG vs GA) does not introduce new structures but perturbs the relative population of each structure. This is consistent with the finding here that 5′-A(2Ap)-3′ and 5′-(2Ap)A-3′ sample similar structures but with different probabilities (Table 4.1).

The key structural difference between *βα* and *ββ* stacks is the *syn* versus anti value of the glycosidic bond dihedral angle, $\chi^{3'}$, a change that is accompanied by a reversal in the sign of cos(θ) (Table 4.1). Table 4.1 reveals a further pattern in the $\chi^{5'}$ and $\chi^{3'}$ values: the sheared *αα** adopts similar $\chi^{5'}$, $\chi^{3'}$ values as the *βα* stack, and similar

glycosidic bond dihedral angles are observed for the sheared conformation $\alpha\beta$* and the $\beta\beta$ stack. This suggests that these sheared conformations are misfolded states where $\chi^{5'}$ and $\chi^{3'}$ are ready to form a more stable stack ($\beta\alpha/\beta\beta$), but the 3′ base happens to have positive, instead of negative, rise and interacts with the backbone instead of with the 5′ base. Additionally, the rare stacks ($\alpha\alpha$, $\alpha\beta$) and their corresponding sheared conformations ($\alpha\alpha$*, $\alpha\beta$*) have different $\chi^{5'}$ values but otherwise exhibit similar values for the remaining conformational degrees of freedom explored here.

Insight into the factors influencing base stacking can be gained from the above observations. For instance, if the influence of the backbone can be neglected, and if favorable electrostatics[159] and dispersion[160] are the most important factors stabilizing base stacking in a particular stacking motif, then the observation that $\beta\alpha$ is the most populated stack in 5′-A(2Ap)-3′ suggests that $\alpha\beta$ should be the most populated stack in 5′-(2Ap)A-3′ simply because both structures have the β-face of adenine stacking with the α-face of 2Ap. However, this was not observed, and the $\alpha\beta$ stack of 5′-(2Ap)A-3′ was visited a mere 1% of the simulation time (see Table 4.1). In fact, inspection of Tables 4.2 and 3 shows that permuting the A and 2Ap residues changes the relative population of each of the various stacking types only modestly. Thus, it may be incorrect to ignore the backbone. Instead, it is more important whether the backbone conformation exposes the α or β face of the 5′ base to the α or β face of the 3′ base than whether the 5′ base is A or 2Ap. More specifically, Figures S9 and S10 show that each type of stack has characteristic *anti*/*syn* values of $\chi^{5'}$ and $\chi^{3'}$. We propose that the entropy of the backbone is the decisive factor explaining the increased propensity to form $\beta\alpha$ and $\beta\beta$ stacks over $\alpha\beta$ and $\alpha\alpha$ stacks and that the electrostatics and dispersion forces are minor contributors.

This is consistent with the idea that the backbone conformational entropy has a large influence on stacking energetics.[161]

4.4.2 Conformational Dynamics of the DNMPs.

Experimentally, the kinetics of base stacking and unstacking were investigated by temperature-jump UV−visible, acoustic attenuation, and NMR spectroscopies in the 1970s.[25, 140-141, 162-164] The interpretation of these experiments emphasized transitions between just two states: random coil (or, as we refer to it here, "extended") and stacked. On the other hand, MD simulations reveal a larger number of possibly kinetically trapped states, so it is difficult to make comparisons with the experiment.[142] We examine this issue carefully below when we address DNMP emission decays, but we first discuss the characteristic timescales for conformational transitions known from various experiments.

Monomeric purine nucleosides switch between *syn* and *anti* conformations about χ with time constants of a few nanoseconds (2.7 ns for adenosine).[165-166] The estimated rates for transitions in the DNMPs (Tables 4.2 and 4.3) where only one of $\chi^{5'}$ or $\chi^{3'}$ changes are much slower than 370 μs^{-1} = (1/2.7) ns^{-1}, a difference that may be due to the steric hindrance introduced by the second nucleobase when the bases are stacked or by the ribose when they are sheared. Building in complexity, the aggregation of two nucleobase monomers in solution to yield a dimer was found to occur with a second order rate constant for stacking in the range of 10^8–10^9 M^{-1} s^{-1} and a first-order unstacking rate on the order of 10^7 s^{-1}.[140, 167] According to Tables 4.2 and 4.3, transitions between the most common stack of each DNMP and the extended state occur at a rate of approximately 3 x 10^7 s^{-1} in reasonable agreement with experiment.

To our knowledge, direct experimental observations of base stacking kinetics have not been reported for DNA DNMPs, but RNA DNMPs were investigated via temperature-jump measurements 40 years ago revealing a single relaxation timescale of 29 ± 6 ns for AA at 300 K.[23] Ref [23] also reported a longer timescale of 100− 300 ns for dA_2 and poly(A), but later measurements on poly(A) did not report a 100s of ns time constant. As pointed out by Pinamonti et al.,[142] it is possible that the high electric field used in the older cable discharge experiments led to artifacts. The single experimental timescale for AA is comparable to the longest timescale of 23.0 ± 1.4 ns for 5′-A(2Ap)-3′ predicted by HMSMs, which is due mostly to transitions between ββ and βα stacks, although the sheared conformation $\alpha\beta^*$ also contributes (Figure A.2.11A). The 5′-(2Ap)A-3′ sample also includes a contribution from the sheared conformation $\alpha\alpha^*$ (Figure A.2.11B). The contribution of $\alpha\beta^*$ or $\alpha\alpha^*$ to the slowest relaxation process may be significant because these sheared conformations ought to display significantly different hypochromicities than the stacks because of the disruption of π−π interactions.

An idealized relaxation experiment was simulated using the full MSMs and the dynamical fingerprint methodology.[153] To mimic a temperature-jump measurement with UV−visible absorption detection, stacked conformations were classified using a COM distance cutoff of 5 Å (distance < 5 Å) and a θ cutoff of 45° ($|\cos(\theta)| > 0.707$). This distance cutoff distinguishes between the stacks and sheared conformations found in the negative displacement region of Figure 4.3, whereas the angular cutoff captures much of the stacked regions in the FES in Figure 3. Then, the initial nonequilibrium population resulting from a 2.5 K temperature jump[168] was determined using the Boltzmann distribution for 302.5 K inferred from the relative probabilities in the MSM. The

observation vector, which maps the average values for an experimental observable onto each conformation, was set to 0.9 for stacks and 1 for unstacked conformations. Physically, this corresponds to the stacked conformations having a 10% lower UV−visible absorption cross section. The value of 10% was selected based on the percentage hypochromicity of 9.4% reported for the RNA DNMP ApA at its peak absorption wavelength.[169]

The simulated signals for both samples are shown in Figure A.2.12, and the timescales and amplitudes are listed in Table A.2.3. The majority of the relaxation for both samples is due to the shorter <10 ns timescales, while the slowest timescales make up 10% of the signal for 5′ -A(2Ap)-3′ , but only 2% for 5′-(2Ap)A-3′. We note that the amplitudes of the slowest timescales are sensitive to the exact values of hypochromicity chosen for the $\beta\beta$ and $\beta\alpha$ stacks, which could in fact be slightly different. There are no experimental temperature-jump data to compare to, but the simulated signals predict that higher time resolution measurements will detect multiscale dynamics with a difference between shortest and longest time constants of a factor of less than 20. This suggests that the earlier temperature-jump experiments, which had a time resolution of 10− 50 ns,[25] may have missed the faster stacking time constants predicted here by the MSM. Indeed, reasonable agreement is seen between the longest timescale from the MSM of these 2Ap-labeled DNMPs, 17− 21 ns (Table A.2.3), and the experimental value for AA, 29 ± 6 ns.[23] We propose below that TRF measurements of 2Ap-labeled oligomers with their higher time resolution offer an experimental observable sensitive to the faster relaxation timescales.

Next, we discuss the structure of the transition rate matrices in Tables 4.2 and 4.3 and attempt to explain the observed trends with microscopic models. Several factors, the $\chi^{5'}$ and $\chi^{3'}$ angles and SASA, partially explain why transitions for which both the cos(θ) and rise change sign are more probable than when only one of these variables changes sign (Tables 4.2 and 4.3). First, while there are many degrees of freedom in the nucleic acid backbone, Tables 4.1−3 show that consideration of just the $\chi^{5'}$ and $\chi^{3'}$ angles is very helpful for disentangling the structural heterogeneity and kinetics. Transitions between $\alpha\alpha^* \leftrightarrow \beta\alpha$ (the double arrow indicates transitions in either direction) and those between $\alpha\beta^* \leftrightarrow \beta\beta$ do not require significant changes in $\chi^{5'}$ or $\chi^{3'}$. Accordingly, the kinetic rates for the $\alpha\alpha^* \leftrightarrow \beta\alpha$ and $\alpha\beta^* \leftrightarrow \beta\beta$ transitions in Tables 4.2 and 4.3 are large relative to other transitions out of $\alpha\alpha^*/\beta\alpha$ or $\alpha\beta^*/\beta\beta$. On the other hand, transitions such as $\alpha\alpha^* \leftrightarrow \beta\beta$ and $\alpha\beta^* \leftrightarrow \beta\alpha$, which require a large change in the dihedral angles, are relatively slow. Furthermore, transitions that require both $\chi^{5'}$ and $\chi^{3'}$ to switch between *syn* and *anti*, such as $\alpha\alpha \leftrightarrow \beta\beta$ and $\alpha\beta \leftrightarrow \beta\alpha$, occur with almost zero frequency. The large asymmetry observed in some cases between the forward and backward transition rates (e.g., a factor of 5 for $\alpha\alpha^* \leftrightarrow \beta\alpha$) can be explained by the fact that the higher energy conformation is closer to the transition state.

The barriers associated with *syn*/*anti* transitions in $\chi^{5'}$ and $\chi^{3'}$ do not fully explain the differences in rates when only one of $\chi^{5'}$ or $\chi^{3'}$ changes. For example, both $\alpha\beta^* \rightarrow \alpha\alpha^*$ and $\alpha\beta^* \rightarrow \beta\alpha$ only change $\chi^{3'}$ from *anti* to s*yn*, yet $\alpha\beta^* \rightarrow \alpha\alpha^*$ occurs 10 times more rapidly (Tables 4.2 and 3). The same holds for the $\alpha\alpha^* \rightarrow \alpha\beta^*$ and $\alpha\alpha^* \rightarrow \beta\beta$ transitions, although the difference is more modest. Instead, the trends can be explained using the concept of SASA[170] and transition-state solvation shell considerations.[171] When the

DNMP transitions between conformations requiring a change in sign of both the rise and cos(θ) (blue entries in Tables 4.2 and 3), the bases can roll around each other with minimal increase in solvent exposure. This is in contrast to transitions in which all four nucleobase faces are exposed to the solvent as when only the sign of the rise changes (Tables 4.2 and 3, red entries). The same arguments suggest that when only cos(θ) changes sign (Tables 4.2 and 3, purple entries), T-shaped structures would be observed, which are characterized by an intermediate degree of solvent exposure. Rare T-shaped structures were observed in the simulations but had populations of <1% and were therefore not classified as a separate state during HMSM construction (Figure A.2.6).

Other factors being equal, we propose that transitions between states that involve the smallest increase in SASA along the reaction coordinate occur at the fastest rates. This is supported by an estimation of ensemble-averaged SASA values for the structural transitions shown in Figure 4.6, which are observed to occur over a period of 500 ps in the 5′-A(2Ap)-3′ DNMP. The 500 ps period was chosen to mimic the lag time used to estimate the kinetic rates in Table 4.2. This graph shows that the sheared conformations have a SASA value roughly 20 Å^2 larger than the stacked structures (compare the SASA value of the initial state near 0 ps with the SASA value of the final state near 500 ps). Comparing the SASA traces at 250 ps provides an estimate of the SASA of the transition state. These traces reveal that on average transitions for which the rise changes sign (red curves in Figure 4.6) increase their SASA values by 20 Å^2 before folding into the final structure. This is in contrast to transitions where both the rise and cos(θ) change sign (blue curves in Figure 4.6). For these transitions, SASA increases by less than 10 Å^2 en route to the final state. The nucleobases thus appear to minimize the exposed

hydrophobic surface area along the transition path. The fastest transitions from one type of stack to another occur when one base is able to “roll around” the other, reducing the number of nucleobase faces that are fully exposed to the solvent and thereby reducing desolvation when refolding takes place.

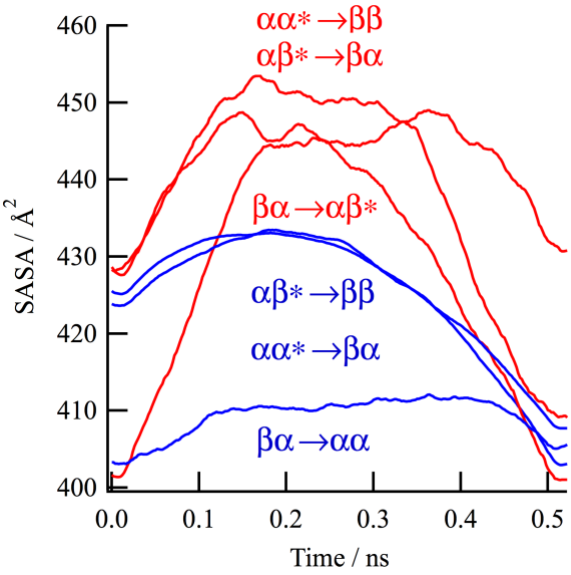

Figure 6. SASA values averaged over all 500 ps transition paths for the transitions indicated are shown for 5'-A(2Ap)-3'. The labels are ordered from bottom to top by the SASA values at 250 ps. The color coding matches Table 4.2 to stress that transitions where both cos(θ) and rise change sign (blue) tend to expose less surface area than when only the rise changes sign (red).

4.4.3 Implications of Structural Heterogeneity and Conformational Dynamics on the TRF of 2Ap

The multiexponential TRF decays from 2Ap-labeled oligomers are commonly attributed to structural heterogeneity and conformational dynamics. While this is likely true, it is still unclear how to extract information about either from measured emission signals. In the past, structural heterogeneity in 2Ap-labeled oligomers has rarely been quantified via an atomistic description like the one provided by MD simulation,[19, 22] and, to our knowledge, the conformational dynamics of 2Ap-labeled oligomers have never been directly quantified. Voltz et al.[19] found that fluorescence quenching in 2Ap-labeled DNA hairpins on timescales <90 ps correlates with stacking populations derived from 400 ns MD simulations.[19] Furthermore, they inferred that the dynamic quenching observed on longer timescales could be due to gating motions that occur with a rate of

circa 25 μs^{-1}.[19] Indeed, this rate is consistent with the structural isomerization rates observed here (Tables 4.2 and 3). However, these authors did not address dynamic aspects of quenching such as the stacking rate(s) predicted by their MD simulations.

Somsen et al. fit a stretched exponential plus a single exponential to the TRF data of 2Ap-labeled dinucleotides, including 5′-(2Ap)A-3′.[172] They qualitatively compared the diffusive dynamics implied by the stretched exponential[173] to fluctuations in the COM distance and in θ obtained from short ($\sim$ 30 ns) MD simulations.[172]. Finally, Jean and Krueger estimated energy-transfer efficiencies from 10 ns MD simulations of 2Ap-labeled trinucleotides.[22] They demonstrated that a briefly visited unstacked conformation drastically reduces the Coulombic coupling between 2Ap and A that is necessary for energy transfer from A to 2Ap relative to stacked conformations. These previous MD simulations of 2Ap-labeled oligonucleotides were unable to adequately sample the full structural heterogeneity (and dynamics), but this is now a more realistic goal in light of current computing power and a motivation behind our study.

The seven distinct structures shown in Figure 4.4 provide the basis for interpreting the TRF experiments. This is accomplished using a master equation (eq A.2.1) that has been slightly modified from the dynamical fingerprint model for MSMs[153] by including terms in the rate matrix to account for radiative and nonradiative excited-state deactivation. Modification of the dynamical fingerprint model[153] is necessary because the experimental observable in a TRF measurement is proportional to the time-dependent excited-state population, represented hereafter as 2Ap*. Because the proportionality constant relating population to the TRF signal includes the radiative decay rate of 2Ap* , the observation vector[153] must include the relative radiative decay

rate of 2Ap* in each particular conformation. To reduce the number of free parameters, the radiative decay rate in each conformation was set equal to the observed decay rate for the 2Ap monomer, (1/9.89) ns^{-1}.[14] Following the Franck−Condon principle, the initial population distribution for 2Ap* was set equal to the ground-state equilibrium distribution (Table 4.1).

Underlying our master equation approach are a few assumptions. First, the electronic excitation is assumed to be localized on 2Ap, allowing the 2Ap* population to be partitioned into the seven DNMP conformations. The localization of excitation is justified by the quite different vertical excitation energies of 2Ap (4.0 eV) and A (4.8 eV), which inhibit delocalized states. Another assumption is that the conformational dynamics simulated in the ground-state fully describe the dynamics of electronically excited conformers. In this case, the conformational transition rates determined from the MD simulations (Tables 4.2 and 3) can be used as the off-diagonal elements in the master equation for 2Ap*. This second assumption may not be true on short <100 ps timescales as nuclear motion brings the system from the Franck−Condon region to the excited-state minima, but is likely true on the slower timescales associated with stacking/ unstacking transitions.

Given a model for emission quenching of 2Ap (i.e., a choice of nonradiative decay rate, k_{nr}, for each conformation), the time-dependent population of 2Ap* can be predicted and compared to experiment. Even assuming that all 2Ap* conformations have identical radiative decay rates, there are still a large number of free parameters because of the conformation-specific rates of nonradiative decay. Additional simplifying assumptions about the remaining parameters were therefore made, leading to the three

models, which make the predications summarized in Figure 4.7, and which are now introduced.

The first model (model 1) assumes that excited states of 2Ap deactivate faster than conformational transitions can take place (Figure 4.7A). For this purely static model, the values of $k_{i,j}$ in eq A.2.1 are set equal to zero to completely suppress conformational dynamics. The only information from the MD simulations used in this model is the equilibrium population of the seven conformations. Model 1 thus tests the assumption frequently made in the literature that the relative amplitudes of the multiple emission lifetimes measure the populations of 2Ap* in various conformations.[10] Model 1 predicts up to three distinct lifetimes, depending on the nonradiative decay rates chosen for the stacked, sheared, and extended conformations.

Calculations of electron-transfer parameters[174-175] could be used to guide the choice of k_{nr} values, but we chose values heuristically based on the following considerations. The four stacks are expected to have the highest *k*nr values because of enhanced electron transfer mediated by π−π stacking[14, 22, 78] or because of reduced solvent exposure of 2Ap.[81] On the other hand, excited states of extended conformations ought to be quenched at significantly lower rates. The sheared conformations are postulated to have intermediate quenching rates, reflecting their intermediate solvent exposure, base−base distance, and orientation (cos(θ)). Guided by these considerations, it was found that k_{nr} values of 2.0, 0.33, and 0 ns^{-1} for the stacked, sheared, and extended conformations, respectively, yielded multiexponential decays similar to experiment (compare green and black curves in Figure 4.7D).

Model 2 (Figure 4.7 B) assumes that stacked conformations contribute negligibly to the emission signals as discussed in ref [14]. In this limit, emission arises solely from the three unstacked conformations ($\alpha\alpha$*, $\alpha\beta$*, and Ex), and the 2Ap* population at any instant is the sum of these three populations. In ref [14], nonradiative decay was assumed to occur via electron transfer, quenching excited states in unstacked conformations at a rate that depends exponentially on the interbase separation. In contrast to the static model in ref [14], we assume here that electron-transfer rates are zero for each of the unstacked conformations. In this case, quenching is purely dynamic, and k_{nr} is determined solely by the sum of rates with which each unstacked conformation forms stacks because 2Ap* is assumed to instantly quench upon stacking. The three-state master equations for this model are obtained from data in Tables 4.2 and 3. The corresponding rate matrices are shown in Table A.2.4A,B for 5′-A(2Ap)-3′ and 5′-(2Ap)A-3′, respectively.

Model 2 predicts (Figure 4.7 B) that emission decays from both samples are dominated by a ~ 4 ns time constant, but this compares poorly with the multiexponential decays seen experimentally (also compare the red and black curves in Figure 4.7 D). The 2 and 4 ns lifetimes likely reflect the time needed for base stacking as they are comparable to the net stacking times of 1.7 and 2.1 ns calculated for the two DNMPs as the reciprocal of the sum of all of the stacking rates in Tables 4.2 and 3, respectively. Model 2 also does not reproduce the ~ 10 ns lifetime component seen in the experiments (Figure 4.7 D). This is because the three unstacked conformations interconvert more rapidly (Tables 4.2 and 3), preventing 2Ap* population from remaining in a particular unstacked conformation. Better agreement between the dominant ~ 4 ns lifetime in

model 2 and the intermediate 1−3 ns experimental lifetimes could be achieved by introducing a nonzero k_{nr} for the sheared conformations.

The final model, model 3, combines static and dynamic quenching characteristics from models 1 and 2. In this hybrid model, we assume that transitions to stacked states do not instantly quench 2Ap*. Instead, the conformation-specific nonradiative decay rates, k_{nr}, are set to the same values used in model 1: 2.0 ns^{-1} for stacks, 0.33 ns^{-1} for sheared conformations, and 0 ns^{-1} for the extended conformation. In model 3, 2Ap* can emit in all seven conformations and all 42 structural interconversion rates predicted by the HMSMs are included. The rate matrices for the 2Ap* population are shown in Table A.2.5A,B for 5′-A(2Ap)-3′ and 5′-(2Ap)A-3′, respectively. In principle, this model could allow 2Ap* that is in a stack to remain electronically excited upon unstacking, although this is found to be unimportant because the unstacking rates (see Tables 4.2 and 3) are all much lower than the 2.0 ns^{-1} nonradiative decay rate used for 2Ap* in stacks. Model 3 predicts seven emission decay time constants (see Table A.2.6), but these were consolidated into three values (Figure 4.7 C) using an amplitude-weighted average over similar time constants.

Model 3 correctly captures the multiexponential character of the emission decays at times less than 5 ns seen experimentally, but, like model 2, it fails to capture the weak monomer-like tail (compare blue and black curves in Figure 4.7 D). The amplitudes in model 3 (and model 1) could be altered by changing the radiative decay rates for stacks and partially stacked conformations. For example, the overestimate of the amplitude of the fastest timescales in Figure 4.7 C could be corrected by reducing the radiative decay rate for stacked conformations.

Decay components predicted by model 3 do not arise from specific conformational states as in model 1. This is seen by comparing the eigenvectors from models 1 and 3 (Figure A.2.13). Because the master equation for model 1 is a diagonal matrix (not shown), the eigenvectors are the standard basis vectors. Inspection of the eigenvectors from model 3 (Figure A.2.13) shows that they are generally not standard basis vectors. Of the three groups of timescales in Figure 4.7 C, standard basis vectorlike eigenvectors are only seen within the group corresponding to emission from the stacked conformations (decay of 2Ap* in the $\beta\alpha$ and $\beta\beta$ structures). This is expected because the ratio of the sum of transition rates out of these structures to the excited-state deactivation rates is smallest. The slowest timescale predicted by model 3 appears to be 2Ap* in a quasi-equilibrium distribution of conformations, all of which decay on the 2−3 ns timescale. Besides the slowest timescale, the other eigenvectors all include transitions between structures, indicating that the amplitudes of these timescales are not due solely to the population of 2Ap* in a particular conformation. The conformational dynamics that are obtained through MD simulation are thus critical for predicting emission decays by model 3.

In conclusion, some final comments about the suitability of the three models for predicting experimental TRF traces from MD simulations processed using MSM/HMSMs are warranted. First, only model 1, which neglects the structural transition rates estimated by the HMSMs, captures the weak ~ 5% amplitude monomer-like tail seen experimentally. While this could mean that the MD simulations are overestimating the transition rates, another possibility is that a small population of the 2Ap monomer

exists in the experimental samples because of cleaved dinucleotides as suggested by Somsen et al.[30]

On the other hand, the purely dynamical description provided by model 2 fails to capture the 300−500 ps lifetime component seen experimentally (Figure 4.7 D). On face value, this suggests that the MD simulations are actually underestimating base stacking rates. However, the use of a 500 ps lag time to construct the MSM/HMSMs makes it impossible to identify structural dynamics on shorter timescales. The predicted quenching occurring with time constants of between 2 and 4 ns is due to stacking of initially unstacked 2Ap*, but capturing faster quenching events will require a more refined analysis of the MD simulations or a description of the dynamics that goes beyond the Markovian assumption.[176-178] Indeed, the FESs in Figure 4.3 show that barriers out of the Ex conformation are roughly equal to $k_B T$, suggesting that a subset of trajectories leaving Ex may have memory.

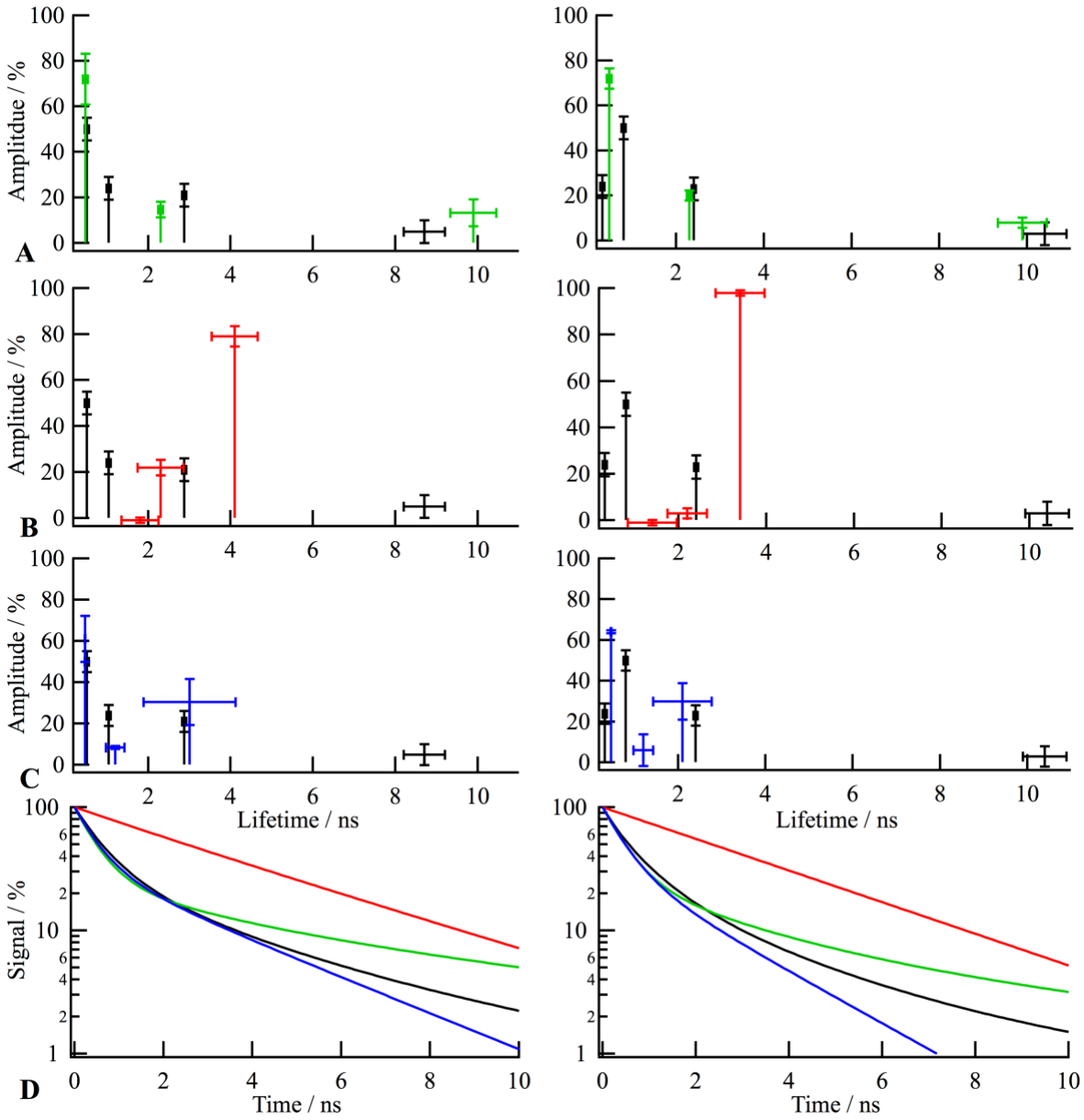

Figure 4.7. Experimental (black) and predicted (colored) lifetimes and amplitudes (A, B, and C) and experimental TRF signals (D) for 5'-A(2Ap)-3' (left) and 5'-(2Ap)A-3' (right). Shown are three models described in the text: model 1 (green, A), model 2 (red, B), and model 3 (blue, C).

While the multiexponential emission decays predicted by models 1 and 3 give better agreement with experiment, each relies to a degree on the introduction of arbitrarily selected values for the nonradiative decay rates of the sheared and stacked

conformations. Clearly, independent estimates of these quantities would be helpful; however, we point out that there is a strong argument for favoring model 3 over model 1. In particular, the MD simulation results clearly do not support the assumption made in model 1 that deactivation rates of excited 2Ap are much shorter than conformational transition rates. The experimental TRF signals establish that significant numbers of excited states decay between 1 and 50 ns, but conformational transitions occur over similar timescales according to the HMSM analysis made in this study. Excited-state deactivation and conformational transitions thus occur on strongly overlapped timescales and emission quenching cannot be modeled without a thorough understanding of the multiscale dynamics of base stacking and unstacking.

4.5 Conclusions

The possibility that the observed decay heterogeneity in TRF traces of 2Ap-labeled DNMPs is due to both structural heterogeneity and conformational dynamics was explored via HMSM analysis of all-atom MD simulations in explicit solvent. At least seven different metastable structures were observed, including structures that have not been identified previously in MD simulations of DNMPs. The conformational dynamics of transitions between the seven metastable states reveal multiple timescales, which are roughly commensurate with ones seen in fluorescence decays from 2Ap-labeled DNMPs. Kinetic models of base stacking on the nanosecond timescale require more than just two states as there are distinct metastable unstacked structures (Ex, $\alpha\beta$*, $\alpha\alpha$*) and stacked structures ($\alpha\beta$, $\alpha\alpha$, $\beta\alpha$, $\beta\beta$). Although there are 42 different interconversion rates between these states, it was shown that there is an underlying structure to the interconversion rate

matrix; transitions where both the rise and cos(θ) change sign occur far more rapidly than when only the rise or cos(θ) change sign, and this was rationalized by changes in the solvent exposed surface area upon stacking. Future work is needed to investigate whether these findings can be extended to other DNMPs and longer oligonucleotides.

Three models for 2Ap emission quenching were used to predict emission decays from the MD data and to explore static versus dynamic effects on emission quenching. Only the purely static model was able to predict the weak ($\sim$ 5% amplitude) monomer-like component because of the relatively rapid stacking rates of the three unstacked conformations. However, the purely static model can be discarded because the MD results show that the conformational transitions that are expected to quench excited states of 2Ap* take place more rapidly than expected from past experiments. Indeed, the timescale for transitions between the seven metastable states strongly overlaps with the characteristic timescales of the TRF signals. A model in which nonradiative decay occurs solely due to unstacked-to-stacked transitions predicts emission decays with 2−4 ns decay times for both 2Ap-containing DNMPs, although limitations in either the MSM/HMSM methodology or the current implementation of it prohibit the identification of faster stacking events.

The best agreement with the experiment is currently obtained with a model that combines structural heterogeneity in the form of conformer-specific rates of nonradiative decay with dynamics that describe transitions among conformational states prior to excited-state relaxation. However, because the conformer-specific nonradiative decay rates are unknown, more refined quantum mechanical-based estimates are needed in the future to reach stronger conclusions about the quenching mechanism. Correctly

interpreted, the experimental TRF signals can be expected to provide a window on the rapid folding and unfolding events in single-stranded RNA and DNA oligonucleotides that are a target of atomistic MD simulations.

ASSOCIATED CONTENT

See supplementary material (Appendix B) for atom names for 2Ap used in the MD simulations; validation of the MSMs and HMSMs, including implied timescales; eigenvectors for HMSMs obtained at increasing lag times (0.5, 0.75, and 1.25 ns); memberships for increasing number of states (7, 8, 9, and 10); Chapman−Kolmogorov test; fuzzy memberships for the final HMSMs; two regions used to construct discrete trajectories for the SASA calculation; demonstration of how the parameters in this calculation influence the results; FESs for the $\chi^{5'}$ and $\chi^{3'}$ dihedrals in both DNMPs; scatterplots to demonstrate how they partition the displacement and cos(θ) space; eigenvectors for the two HMSMs used in the main text; signals from the simulated temperature-jump experiment; eigenvectors for model 3; description of the master equation used for prediction of the 2Ap and 2Ap* population shown along with a generic form for the master equation and a short proof that it satisfies detailed balance; atomic charges and assigned atom types for 2Ap; average values for $\chi^{5'}$, $\chi^{3'}$, and the twist helical parameter for the seven conformations in each DNMP; six slowest relaxation times and amplitudes for the simulated temperature-jump experiment; rate equations for model 2 and model 3; and all seven timescales predicted by model 3.

AUTHOR INFORMATION

E-mail: Martin.McCullagh@colostate.edu. Phone: +1 (970) 491-3572 (M.M.).
E-mail: kohler.40@osu.edu. Phone: +1 614-688-2635 (B.K.).

ACKNOWLEDGMENTS

We thank Russell Davidson for the parameterization of the 2Ap residue. This work was supported in part by NSF grant CHE-1817500 (to B.K.).

REFERENCES

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

11. Jean, J. M.; Hall, K. B., 2-Aminopurine Fluorescence Quenching and Lifetimes: Role of Base Stacking. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 37-41.

12. Wan, C.; Xia, T.; Becker, H. C.; Zewail, A. H., Ultrafast Unequilibrated Charge Transfer: A New Channel in the Quenching of Fluorescent Biological Probes. *Chem. Phys. Lett.* **2005,** *412*, 158-163.

14. Remington, J. M.; Philip, A. M.; Hariharan, M.; Kohler, B., On the Origin of Multiexponential Fluorescence Decays from 2-Aminopurine-Labeled Dinucleotides. *J. Chem. Phys.* **2016,** *145*, 155101.

18. Greiner, V. J.; Kovalenko, L.; Humbert, N.; Richert, L.; Birck, C.; Ruff, M.; Zaporozhets, O. A.; Dhe-Paganon, S.; Bronner, C.; Mély, Y., Site-Selective Monitoring of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry* **2015,** *54*, 6012-6020.

19. Voltz, K.; Léonard, J.; Touceda, P. T.; Conyard, J.; Chaker, Z.; Dejaegere, A.; Godet, J.; Mély, Y.; Haacke, S.; Stote, R. H., Quantitative Sampling of Conformational Heterogeneity of a DNA Hairpin Using Molecular Dynamics Simulations and Ultrafast Fluorescence Spectroscopy. *Nucleic Acids Res.* **2016,** *44*, 3408-3419.

20. Tolbert, M.; Morgan, C. E.; Pollum, M.; Crespo-Hernández, C. E.; Li, M. L.; Brewer, G.; Tolbert, B. S., HnRNP Al Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation. *J. Mol. Biol.* **2017,** *429*, 2841-2858.

21. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

22. Jean, J. M.; Krueger, B. P., Structural Fluctuations and Excitation Transfer between Adenine and 2-Aminopurine in
Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B.* **2006,** *110*, 2899-2909.

23. Pörschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

24. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 2315-2328.

25. Pörschke, D., Dynamics of Nucleic-Acid Single-Strand Conformation Changes - Oligoriboadenylic and Polyriboadenylic Acids. *Eur. J. Biochem.* **1973,** *39*, 117-126.

26. Chen, J.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

30. Somsen, O. J.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Structural Heterogeneity in DNA: Temperature Dependence of 2-Aminopurine Fluorescence in Dinucleotides. *Chemphyschem* **2005,** *6*, 1622-1627.

31. Kang, H.; Chou, P. J.; Johnson, W. C.; Weller, D.; Huang, S. B.; Summerton, J. E., Stacking Interactions of ApA Analogs with Modified Backbones. *Biopolymers* **1992,** *32*, 1351-1363.

32. Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007,** *92*, 3817-3829.

33. Chen, A. A.; García, A. E., High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *P. Natl. Acad. Sci. USA* **2013,** *110*, 16820-16825.

34. Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8*, 2506-2520.

35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

36. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, et al *AMBER 2018*, University of California, San Francisco, 2018.

38. Hayatshahi, H. S.; Henriksen, N. M.; Cheatham, T. E., Consensus Conformations of Dinucleoside Monophosphates Described with Well-Converged Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018,** *14*, 1456-1470.

40. Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S., Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009,** *131*, 124101.

41. Pande, V. S.; Beauchamp, K.; Bowman, G. R., Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010,** *52*, 99-105.

42. Chodera, J. D.; Noé, F., Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

43. Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011,** *134*, 174105.

46. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B.* **2007,** *111*, 5757-5766.

50. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of Lifetime Distributions from Fluorescence Decays with Application to DNA-Base Analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

52. Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E.; Otyepka, M.; Jurečka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013,** *9*, 2339-2354.

53. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

54. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomicand Molecular Electronic-Structure System. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

55. Gordon, M. S.; Schmidt, M. W., *Advances in Electronic Structure Theory: GAMESS a Decade Later*. Elsevier Science Bv: Amsterdam, 2005; p 1167-1189.

56. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B.* **2008,** *112*, 9020-9041.

59. Andrea, T. A.; Swope, W. C.; Andersen, H. C., The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water. *J. Chem. Phys.* **1983,** *79*, 4576-4584.

60. Center, O. S., Ohio Supercomputer Center. 1987.

61. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

62. R. J. Gowers, M. L., J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, et al *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulation*, SciPy: Austin, TX, 2016.

63. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.

64. Jones, E.; Oliphant, E.; Peterson, P.; others SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ (accessed 2017-).

65. Blondel, A.; Karplus, M., New Formulation for Derivatives of Torsion Angles and Improper Torsion Angles in Molecular Mechanics: Elimination of Singularities. *J. Comput. Chem.* **1996,** *17*, 1132-1141.

66. Connolly, M. L., Analytical Molecular-Surface Calculation. *J. Appl. Crystallogr.* **1983,** *16*, 548-558.

67. Lu, X. J.; Olson, W. K., 3DNA: a Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic. Acids. Res.* **2003,** *31*, 5108-5121.

68. Diekmann, S., Definitions and Nomenclature of Nucleic-Acid Structure Parameters. *J. Mol. Biol.* **1989,** *205*, 787-791.

69. Lu, X. J.; ElHassan, M. A.; Hunter, C. A., Structure and Conformation of Helical Nucleic Acids: Analysis Program (SCHNAaP). *J. Mol. Biol.* **1997,** *273*, 668-680.

70. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA Dodecamer - Conformation and Dynamics. *P. Natl. Acad. Schi-Biol.* **1981,** *78*, 2179-2183.

72. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013,** *139*, 015102.

73. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernàndez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015,** *11*, 5525-5542.

74. S., G.; others *Shapely: Manipulation and Analysis of Geometric Objects*, toblerity.org: 2007—.

78. Jean, J. M.; Hall, K. B., Stacking-Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry* **2004,** *43*, 10277-10284.

81. Lobsiger, S.; Blaser, S.; Sinha, R. K.; Frey, H. M.; Leutwyler, S., Switching on the Fluorescence of 2-Aminopurine by Site-Selective Microhydration. *Nat. Chem.* **2014,** *6*, 989-993.

136. Rachofsky, E. L.; Osman, R.; Ross, A. J. B., Probing Structure and Dynamics of DNA with 2-Aminopurine: Effects of Local
Environment on Fluorescence. *Biochemistry* **2001,** *40*, 946-956.

137. Larsen, O. F. A.; van Stokkum, I. H. M.; Gobets, B.; van Grondelle, R.; van Amerongen, H., Probing the Structure and Dynamics of a DNA Hairpin by Ultrafast Quenching and Fluorescence Depolarization. *Biophys. J.* **2001,** *81*, 1115-1126.

138. Rai, P.; Cole, T. D.; Thompson, E.; Millar, D. P.; Linn, S., Steady-State and Time-Resolved Fluorescence Studies Indicate an Unusual Conformation of 2-Aminopurine Within ATAT and TATA Duplex DNA Sequences. *Nucleic Acids Res.* **2003,** *31*, 2323-2332.

139. Somsen, O. J.; van Hoek, A.; van Amerongen, H., Fluorescence Quenching of 2-Aminopurine in Dinucleotides. *Chem. Phys. Lett.* **2005,** *402*, 61-65.

140. Pörschke, D.; Eggers, F., Thermodynamics and Kinetics of Base-Stacking Interactions. *Eur. J. Biochem.* **1972,** *26*, 490-498.

141. Heyn, M. P.; Nicola, C. U.; Schwarz, G., Kinetics of Base-Stacking Reaction of N6,N6-Dimethyladenosine - Ultrasonic-Absorption and Dispersion Study. *J. Phys. Chem.* **1977,** *81*, 1611-1617.

142. Pinamonti, G.; Zhao, J. B.; Condon, D. E.; Paul, F.; Noé, F.; Turner, D. H.; Bussi, G., Predicting the Kinetics of RNA Oligonucleotides Using Markov State Models. *J. Chem. Theory Comput.* **2017,** *13*, 926-934.

143. van Mourik, T.; Hogan, S. W. L., DNA Base Stacking Involving Adenine and 2-Aminopurine. *Struct. Chem.* **2016,** *27*, 145-158.

144. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C., Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007,** *126*, 155101.

145. Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. In *Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G. R.; Pande, V. S.; Noé, F., Eds. Springer-Verlag Berlin, Heidelberger Platz 3, D-14197 Berlin, Germany: 2014; Vol. 797.

146. Wales, D. J., *Energy Landscapes: With Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press: The Edinburgh Building, Cambridge CB2 2RU, UK, 2003; p 671.

147. Rao, F.; Caflisch, A., The Protein Folding Network. *J. Mol. Biol.* **2004,** *342*, 299-306.

148. Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C., Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.* **2007,** *126*, 155102.

149. Wales, D. J.; Bogdan, T. V., Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B.* **2006,** *110*, 20765-20776.

150. Noé, F.; Wu, H.; Prinz, J. H.; Plattner, N., Projected and Hidden Markov Models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* **2013,** *139*, 184114.

151. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. Recommendations 1982. *Eur. J. Biochem.* **1983,** *131*, 9-15.

152. Kube, S.; Weber, M., A Coarse Graining Method for the Identification of Transition Rates Between Molecular Conformations. *J. Chem. Phys.* **2007,** *126*, 024103.

153. Keller, B. G.; Prinz, J. H.; Noé, F., Markov Models and Dynamical Fingerprints: Unraveling the Complexity of Molecular Kinetics. *Chem. Phys.* **2012,** *396*, 92-107.

154. Rose, I. A.; Hanson, K. R.; Wilkinson, K. D.; Wimmer, M. J., A Suggestion for Naming Faces of Ring Compounds. *P. Natl. Acad. Schi-Biol.* **1980,** *77*, 2439-2441.

155. Sundaralingam, M.; Westhof, E., *The Nature of the Mobility of the Sugar and its Effects on the Dynamics and Functions of RNA and DNA*. 1981; p 301-326.

156. Jafilan, S.; Klein, L.; Hyun, C.; Florián, J., Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B.* **2012,** *116*, 3613-3618.

157. Fornasiero, D.; Kurucsev, T., Vibronic Exciton Interactions - Resolution and Interpretation of the Temperature-Dependent Circular-Dichroism and Absorption-Spectra of ApA and of dApdA. *Eur. J. Biochem.* **1984,** *143*, 1-7.

158. Neidle, S., *Principles of Nucleic Acid Structure*. Elsevier Academic Press Inc: San Diego, 2008; p 1-289.

159. Newcomb, L. F.; Gellman, S. H., Aromatic Stacking Interactions in Aqueous-Solution - Evidence that Neither Classical Hydrophobic Effects nor Dispersion Forces are Important. *J. Am. Chem. Soc.* **1994,** *116*, 4993-4994.

160. Šponer, J.; Riley, K. E.; Hobza, P., Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases. *Physical Chemistry Chemical Physics* **2008,** *10*, 2595-2610.

161. Mak, C. H., Unraveling Base Stacking Driving Forces in DNA. *J. Phys. Chem. B.* **2016,** *120*, 6010-6020.

162. Brennan, M.; Kustin, K., Ultrasonic Attenuation in Basic Purine Solutions. *J. Phys. Chem.* **1972,** *76*, 2838-2841.

163. Engel, J. D.; Hippel, P. H. V., Effects of Methylation on Stability of Nucleic-Acid Conformations - Studies at Monomer Level. *Biochemistry* **1974,** *13*, 4143-4158.

164. Garland, F.; Christian, S. D., Thermodynamic and Kinetic-Model of Sequential Nucleoside Base Aggregation in Aqueous-Solution. *J. Phys. Chem.* **1975,** *79*, 1247-1252.

165. Hemmes, P. R.; Oppenheimer, L.; Jordan, F., Ultrasonic Relaxation Evaluation of Thermodynamics of Syn-Anti Glycosidic Isomerization in Adenosine. *J. Am. Chem. Soc.* **1974,** *96*, 6023-6026.

166. Rhodes, L. M.; Schimmel, P. R., Nanosecond Relaxation Processes in Aqueous Mononucleoside Solutions. *Biochemistry* **1971,** *10*, 4423-4433.

167. Polacek, R.; Buckin, V. A.; Eggers, F.; Kaatze, U., Kinetics of Base-Stacking Interactions and Proton Exchange of 6-Methylpurine Aqueous Solutions. *J. Phys. Chem. A.* **2004,** *108*, 1867-1872.

168. Hoffman, G. W., Nanosecond Temperature-Jump Apparatus. *Rev. Sci. Instrum.* **1971,** *42*, 1643-1647.

169. Warshaw, M. M.; Tinoco, I., Optical Properties of 16 Dinucleoside Phosphates. *J. Mol. Biol.* **1966,** *20*, 29-38.

170. Lee, B.; Richards, F. M., Interpretation of Protein Structures - Estimation of Static Accessibility. *J. Mol. Biol.* **1971,** *55*, 379-400.

171. Wallqvist, A.; Berne, B. J., Molecular-Dynamics Study of the Dependence of Water Solvation Free-Energy on Solute Curvature and Surface-Area. *J. Phys. Chem.* **1995,** *99*, 2885-2892.

172. Somsen, O. J. G.; Trinkunas, G.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Local Diffusive Dynamics in DNA - A Time-Resolved Fluorescence and Molecular-Dynamics Study of Dinucleotides with 2-Aminopurine. *J. Lumin.* **2006,** *119*, 100-104.

173. Edholm, O.; Blomberg, C., Stretched Exponentials and Barrier Distributions. *Chem. Phys.* **2000,** *252*, 221-225.

174. Callis, P. R., Predicting Fluorescence Lifetimes and Spectra of Biopolymers. In *Methods in Enzymology*, Johnson, M. L.; Brand, L., Eds. Elsevier Academic Press Inc: San Diego, 2011; Vol. 487, pp 1-38.

175. Blumberger, J., Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **2015,** *115*, 11191-11238.

176. Burshtein, A. I., Non-Markovian Theories of Transfer Reactions in Luminescence and Chemiluminescence and Photo- and Electrochemistry. *Adv. Chem. Phys.* **2004,** *129*, 105-418.

177. Min, W.; Luo, G. B.; Cherayil, B. J.; Kou, S. C.; Xie, X. S., Observation of a Power-Law Memory Kernel for Fluctuations Within a Single Protein Molecule. *Phys. Rev. Lett.* **2005,** *94*, 198302.

178. Suárez, E.; Adelman, J. L.; Zuckerman, D. M., Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. *J. Chem. Theory Comput.* **2016,** *12*, 3473-3481.

CHAPTER FIVE

TIME-RESOLVED FLUORESSENCE AND MOLECULAR DNYNAMICS SIMULATIONS OF 2-AMINOPURINE-LABELED SINGLE-STRANDED DNA

5.1 Introduction

The primary focus so far in this dissertation has been on 2Ap-labeled ssDNA oligomers containing only two bases. While the simplicity of these systems has allowed for a very detailed characterization of the emission quenching mechanism, 2Ap is often used to help determine the local stacking dynamics of DNA oligomers that contain more than two bases.[179-180] As a means to extend the models built so far to these more biologically relevant systems, the steady-state absorption/emission and TRF signals along with MD simulations of 2Ap-labeled ssDNA dinucleotides, trinucleotides, and 15 base oligomers are described in this chapter. Specifically, oligomers of varying length (n = 2-15 bases) and varying position of the 2Ap label (interior, 5', or 3' end) were studied. This corresponds to a 2Ap nucleoside, termed d2Ap; two dinucleotides, 5'-(2Ap)A-3' and 5'-A(2Ap)-3'; three trinucleotides, 5'-(2Ap)AA-3', 5'-A(2Ap)A-3', and 5'-AA(2Ap)-3'; and three 15 base oligomers, 5'-(2Ap)A_{14}-3', 5'-A_{14}(2Ap)-3', and 5'-A_7(2Ap)A_7-3'. Technically these samples are poly(nucleoside phosphates) because they contain n-1 phosphate groups (e.g. 5'-(2Ap)A-3' is a dinucleoside monophosphate), but they will be referred to here as polynucleotides. Besides extending the current models, the results obtained herein also increase the understanding of the SH and CD of ssDNA, molecules which have various biological functions.[181-184]

The steady-state absorption and emission properties of the 2Ap monomer are known to be dependent on the solvent and the lowest energy absorption band shifts from 305 nm in water to 309 nm in methanol.[16] A redshift similar in magnitude is observed in the emission excitation spectrum when 2Ap is π-stacked with neighboring nucleobases in DNA.[185-186] The dependence of the position of 2Ap in 5'-A_5-3' has revealed that the absorbance band is redshifted approximately 5 nm when 2Ap is on the interior position relative to the end.[45] The same study found the π-stacking dependent energy transfer from A to 2Ap is much more efficient when 2Ap is on the interior of the ssDNA compared to the ends. This finding has been shown to hold true for 2Ap-labeled 5'-A_{30}-3' as well.[46] TA studies of the stacking heterogeneity in A containing ssDNA oligomers of increasing length found the yields of CT states to agree with a nearest neighbor stacking model.[28] Su et al.[28] provide the simple explanation that interior bases which have two neighbors have twice the likelihood to form a base stack than the end bases which have only one neighbor. Here, it is anticipated that 2Ap on the interior of the strands studied here will have a higher stacking propensity, a hypothesis that will be tested by measuring the steady-state absorption spectra and by quantifying the amount of stacking from extensive MD simulations.

The signals observed in TRF experiments of 2Ap-labeled 5'-A_{30}-3',[46] exhibit the characteristic multi-exponential decays of 2Ap-labeled DNA oligomers.[10] While the fastest decay timescales (tens to hundreds of ps) observed in TRF of 2Ap are often attributed to the population of 2Ap which is stacked, the results by Ramreddy et al. show the interior label has a lower amplitude of this component.[46] This suggests the picture of 2Ap quenching being dependent on a static distribution of conformations is insufficient.

This has been noted previously in studies of 2Ap-labeled 5'-AAA-3'[22] and longer oligomers[10, 82], which have suggested the CD of the DNA system affects TRF signals. However, besides the study presented here in Chapter 4,[37] this CD has yet to be quantified independently in a specific system.

To explore the microscopic underpinnings of the CD in these oligomers extensive MD simulations were carried out. Similar RNA oligomers have been studied previously and it was found that MSMs could describe their CD.[142] The experimental observable that Pinamonti et al[142] could compare to was the single relaxation timescale from a temperature jump measurement.[25] Given the MD simulations predict more than a single timescale, the MSM methodology could use an experimental observable sensitive to more of the dynamics, like the TRF of 2Ap, to compare to. Here, simulations of dA_2 and dA_3 were conducted and compared to experiment under the assumption that the 2Ap label behaves similar to A. The simulations on dA_2 reveal the same seven conformational states and comparable stacking kinetics as shown in Chapter 4 for the 2Ap-labeled dinucleotides.[37] Using the seven state MSMs built from the dA_2 simulations and nearest-neighbor stacking rules, a multi-state nearest neighbor stacking model predicts the structures 5'-A_n-3' can adopt. The utility of this stacking model is scrutinized by comparing it to 15.5 μs of 5'-AAA-3'. Ultimately the following explores the ability of both TRF and MSMs from MD simulations to determine the dynamic properties of ssDNA oligomers.

5.2 Methods

5.2.1 Experimental Details

Experimental protocols for the steady-state absorption and emission experiments are described in sections 2.1.1-2.1.2. The TRF measurements are described in section 2.1.3. The samples were prepared as described in Section 2.2.2, but the tris buffer solution had a pH 6.94 for these oligomers. The steady-state and TRF measurements were analyzed by using the methods in section 2.3.

5.2.2 Molecular Dynamics Simulations of Adenine Containing Dinucleotides

Five independent 3 μs MD simulations of the 5'-AA-3' dinucleotide were performed using the procedure described in Sections 2.4.1-2. The six HP and the two χ backbone dihedral angles described in section 2.5.1 were calculated and used as a dataset for MSM construction. The 15 μs data set was clustered using the regular space clustering algorithm with a minimum distance of 0.18 and a stride of 200 ps. The implied timescales for this dataset is shown in Figure 5.1, and based on the plateau at τ_{lag} greater than 500 ps a reversible MSM was estimated with a τ_{lag} of 500 ps. The MSM was then coarse grained into a seven state HMSM and the resultant metastable memberships used to define assign the seven structures to the $\beta\alpha$, $\beta\beta$, $\alpha\alpha$, $\alpha\alpha$*, $\alpha\beta$, $\alpha\beta$*, and Ex conformations. The static occupation probabilities were taken from the equilibrium eigenvector of the reversible seven state HMSM constructed using all five data sets. The transition probability matrix containing the probability of transitioning between the seven states during a 500 ps time frame was converted to a rate matrix by dividing by 500 ps.

Furthermore, the diagonal elements were then set to the negative off-diagonal row sums to conserve population.

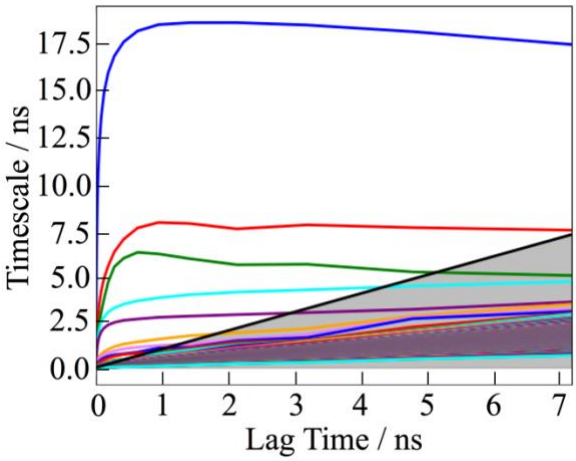

Figure 5.1. The implied timescales for the 5'-AA-3' oligomer demonstrate that the timescales become independent of τ_{lag} at about 500 ps. The black line is y=x with the region below shaded to represent regions where τ_{lag} is greater than the timescale.

5.2.3 Generation of Initial Structures for the Trinucleotide

As described in the results (Section 5.3.3), the seven structures sampled by 5'-AA-3' were used to create a multi-state nearest neighbor stacking model for the structures available to A-containing oligomers of arbitrary length. Initial structures predicted by this model were generated in the following manner. The five 3 μs trajectories of dA_2 were projected onto the cos(θ) and displacement space described in Chapter 4.[37] Structures sampled every 100 ps from the dA_2 trajectories were then sorted into sets of $\beta\alpha$, $\beta\beta$, $\alpha\alpha$, $\alpha\alpha$*, $\alpha\beta$, $\alpha\beta$*, and Ex conformations based on the cos(θ) and displacement values falling into the regions defined in Figure 5.2 and the two χ dihedrals being in the correct *syn*/*anti* pairs described in Chapter 4. Structures that did not simultaneously satisfy both the cos(θ) and displacement and the two χ dihedral criteria were not included in the sets.

Each structure in the dA_2 sets was rotated/translated so that the 3 base axes of the 5' base are always centered at the origin and aligned collinear to the first structure in the set.

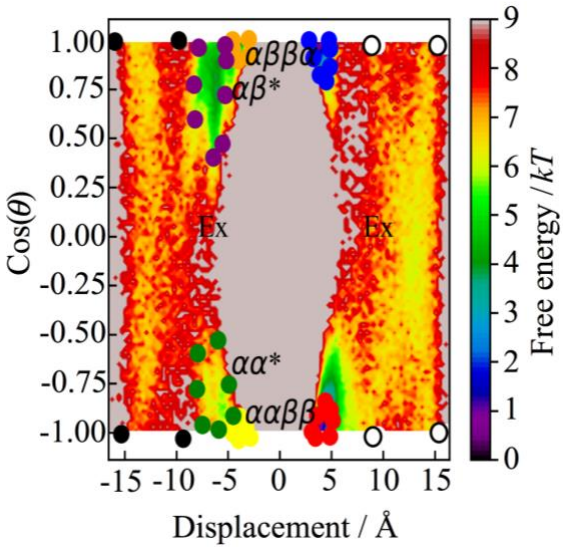

Figure 5.2. The regions of the cos(θ) and displacement space used to partition the dA_2 into sets of $\beta\alpha$, $\beta\beta$, $\alpha\alpha$, $\alpha\alpha$*, $\alpha\beta$, $\alpha\beta$*, and Ex conformations. The Ex conformations were saved if they were in either of the regions defined by the black or white circles.

To build a particular conformation of 5'-A_n-3', say $\beta\alpha\beta\alpha$ for 5'-AAA-3' (see Section 5.3.3 for definition), the set of $\beta\alpha$ structures was searched for a candidate structure that, upon alignment of the 5' base of the candidate structure to the 3' base of another structure in the $\beta\alpha$ set, had a phosphodiester bond length between 1 and 2.5 Å. Once two structures meeting this condition were found, the 3' H atom of the first structure was deleted the structures concatenated and saved. For oligomers longer than 3 bases, the process can be repeated for each n-1 nearest neighbor interactions, where only a single candidate structure is compared with the 3' base of the last nearest neighbor interaction. These initial structures were then solvated, minimized, and thermalized as described in Section 2.4.1-2.

5.2.4 Simulation and Analysis Protocol for the Oligomer Simulations

For 5'-AAA-3', 31 different 500 ns trajectories were run starting from the distinct structures defined by the multi-state nearest neighbor stacking model. The coordinates from each of these trajectories were saved every 10 ps and stripped of water and ions for analysis. For these systems the HPs for each nearest neighbor interaction and the backbone dihedrals were calculated and saved.

To analyze the four dimensional FES for the cos(θ) and displacement space of the 5'-AAA-3' oligomer described in Section 5.3.4, a projection of this FES onto the cos(θ) and displacement space of the 5'-end nearest neighbor interaction was calculated. This projected FES was then partitioned manually using the methods described in Section 2.5.2 and the regions in Figure 5.3.

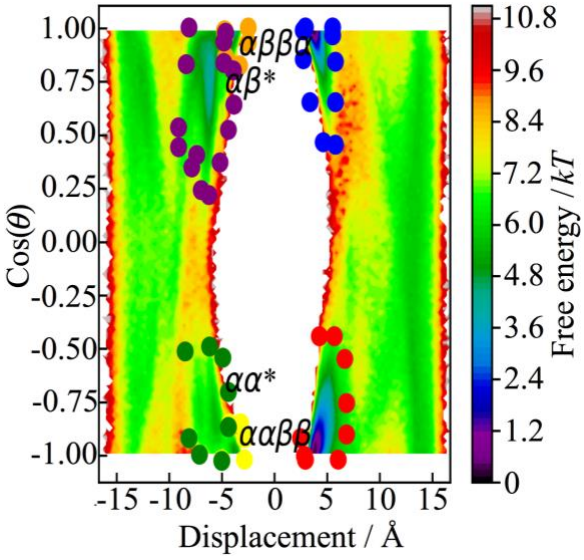

Figure 5.3. The regions used to define the different conformations in the 5'-AAA-3' simulations. If a structure did not fall into one of the polygons defined by the colored circles it was saved as an Ex structure.

Data used to construct MSMs for the trinucleotide included the two sets of three distance HPs and the 24 pairs of x and y-components for the backbone dihedral angles, φ, as x = cos(φ) and y = sin(φ). The dimension of this 54 dimensional dataset was reduced using the TICA technique described in Section 2.5.1. The cumulative variance of 5'-AAA-3' (Figure 5.4) as a function of lag time demonstrate that reducing the dimension of the dataset to 12 still retains 85% of the kinetic variance for the two oligomers respectively when a TICA lag-time of 0.5 ns is used.

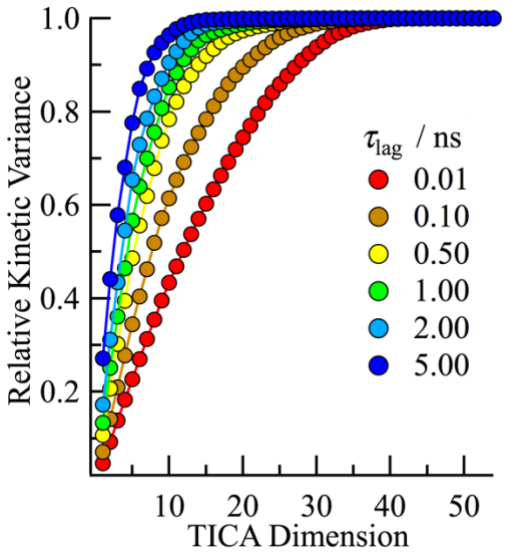

Figure 5.4. The cumulative relative kinetic variance for the TICA data reduction technique applied to the 54 dimension 5'-AAA-3' datasets.

The reduced dimension TICA coordinates were then clustered using *k*-means with 2000 cluster centers and the trajectories discretized. The implied timescales for 5'-AAA-3' are shown in Figure 5.5. Based on the nice plateau at lag times greater than 1 ns, a lag time of 1 ns was chosen for reversible MSM estimation.

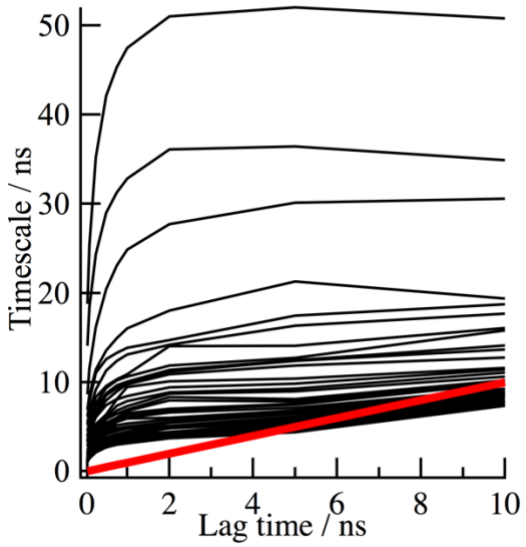

Figure 5.5. The implied timescales for 5'-AAA-3' show that the timescales become independent of the lag time for MSM estimation at about 1 ns. The black curves are the first 40 timescales of the 2000 state MSM while the red curve shows the line $y=x$.

5.3 Results

5.3.1 Steady-State and Time-Resolved Emission

The dinucleotides in aqueous buffer were first characterized with UV-visible absorption and steady-state fluorescence spectroscopy. The room temperature absorption spectra are shown in Figure 5.6A, while the emission spectra obtained with λ_{ex} = 310 nm are shown in Figure 5.6B. The absorbance spectrum for the d2Ap reference compound peaks at 305 nm while its emission peaks at 365 nm. Also, the absorbance spectra for the dinucleotides has significant greater absorbance at 260 nm than d2Ap, which corresponds to the peak absorbance band for A. Within the experimental uncertainty (which is high due to the low sample concentration), the absorbance and emission spectra for 2Ap in the dinucleotides are not shifted appreciably. However, the emission quenching of 2Ap in the dinucleotides is clear in Figure 5.6B as the dinucleotide spectral maximum are

approximately an order of magnitude lower than the absorbance matched d2Ap sample. The Φ for these samples were reported in Chapter 3 for 5′-(2Ap)A-3′and 5′-A(2Ap)-3′as 0.072 ± 0.009 and 0.061 ± 0.012 respectively.

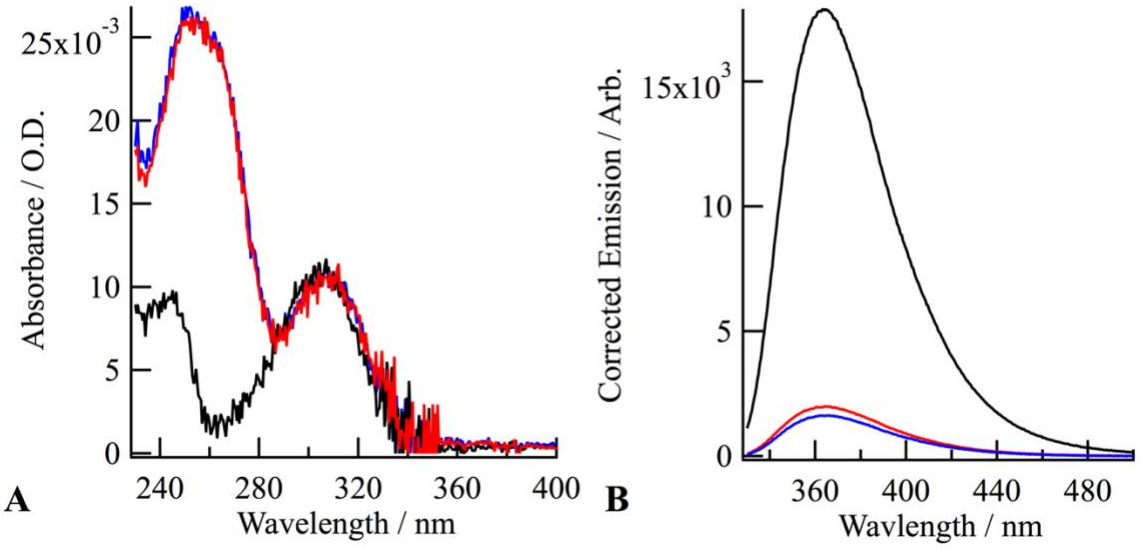

Figure 5.6. The UV-visible absorbance (A) and corrected steady-state fluorescence (B) spectra of d2Ap (black), 5'-(2Ap)A-3' (red), and 5'-A(2Ap)-3' (blue) demonstrates only minor differences between the 2Ap labels absorbance and emission maxima. For the emission spectra, λ_{ex} = 310 nm was used.

Just like the dinucleotides, the 15 base oligomers were studied at room temperature with UV-visible absorption and steady-state fluorescence spectroscopy in aqueous buffer. The absorbance spectra are shown in Figure 5.7A, while the emission spectra obtained with λ_{ex} = 310 nm are shown in Figure 5.7B. Unlike the dinucleotides, the lowest energy absorbance band associated with the 2Ap is redshifted in the oligomers. This shift is 3 nm for 2Ap-labeled on either end of the strand and up to 7 nm when 2Ap is at the interior of the strand. These shifts are comparable to those observed previous in 2Ap-labeled pentanucleotides where an approximate 5 nm shift is observed when 2Ap is

at the 5' or interior position.[45] The emission spectra obtained the two end labels show no noticeable shift, but 2Ap at the interior position is blue shifted 2 nm. The modest changes to the emission maxima have been reported before as well.[187] For these three samples the Φ were found to be 0.087 ± 0.007, 0.094 ± 0.007, and 0.114 ± 0.096 for 5'-(2Ap)A_{14}-3', 5'-A_{14}(2Ap)-3', and 5'-A_7(2Ap)A_7-3' respectively. The Φ suggest the 2Ap on the interior is quenched slightly less than the 2Ap on the two ends which are quenched a similar magnitude.

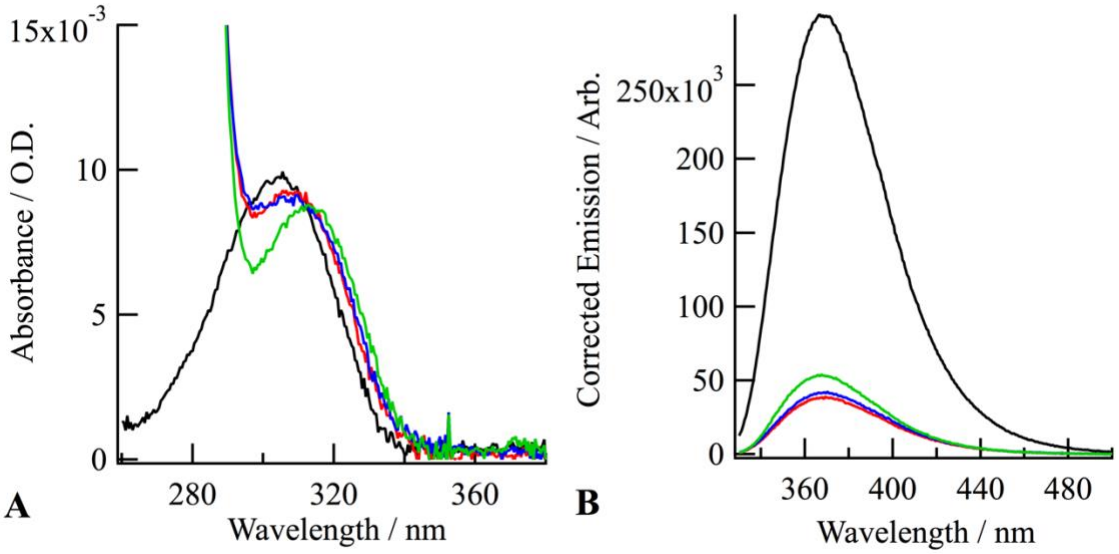

Figure 5.7. The UV-visible absorbance (A) and corrected steady-state fluorescence (B) spectra of d2Ap (black), 5'-(2Ap)A_{14}-3' (red), 5'-A_{14}(2Ap)-3' (blue), and 5'-A_7(2Ap) A_7-3' (green) are shown to reveal the significant shifts in the absorbance and quenching of the emission spectra. For the emission spectra, λ_{ex} = 310 nm was used.

The TRF of all eight 2Ap-labeled ssDNA oligomers in aqueous buffer at 10 °C were measured with the TCSPC technique with λ_{ex} = 310 nm and λ_{em} = 370 nm. The TRF signals from all 8 were found to decay non-monoexponentially upon 310 nm excitation (Figure 5.8). Besides the larger signal observed from 5'-(2Ap)A-3' relative to

the longer 5' labeled oligomers at times greater than 8 ns in Figure 5.8A, the quenching generally occurs slower as the length of the oligomer is increased. When 2Ap is on the 5' or 3' end the slowdown is most significant between the dinucleotides and trinucleotides whereas the trinucleotides and 15 base oligomers behave similarly (Figures 5.8A and C). However, 2Ap at the central position of 15 base oligomer decays significantly slower than the 2Ap at the central position of a trinucleotide (Figure 5.8B).

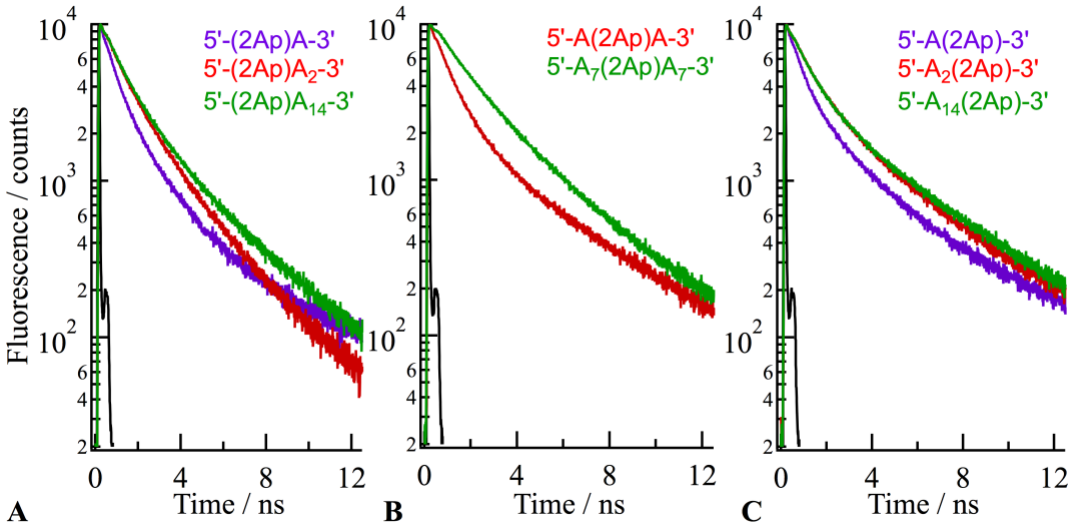

Figure 5.8. The dependence of TRF signals on oligomer length is shown for 2Ap-labeled at the 5' position (A), middle (B), and 3' position (C) of the dinucleotides (purple), trinucleotides (red), and 15 base oligomers (green). Also shown is the IRF (black) scaled to match the peak emission.

In addition to the dependence of the timescales for emission quenching of 2Ap on the length of the oligomers, the quenching also depends on the position of the label as well. This can be seen in Figure 5.8 by comparing the time it takes for a particular colored signal to decay by 90% in each panel. For example, to qualitatively see the influence of position of the 2Ap in a dinucleotide (already discussed in Chapter 3) the purple trace in Figure 5.8A takes a little less than 4 ns to reach 10^3 while the purple trace

in Figure 5.8C takes a little more than 4 ns, so 2Ap at the 3' label is quenched more slowly. The trinucleotides and 15 base oligomers also continue this trend as the same is true for the red and green curves in Figure 5.8A and C. Emission from the interior labeled position appears to be more complicated as the time it takes to reach 90% is slightly faster than the end labels in the trinucleotides but slightly longer in the 15 base oligomers.

To quantify the dependence of the position of the 2Ap label and the length of the oligomer on the emission quenching timescales at 10 °C, sums of three or four exponential decay components were fit to the data. The results of these fits are shown in Table 5.1 with the amplitude weighted mean lifetimes, $<\tau>$, shown in Table 5.2, and calculated with Equation 3.5. The signals from the dinucleotides required four exponentials to model but the longer oligomers only required three exponentials. For all samples the dominant timescale was the intermediate one which ranged from 720-780 ps for the dinucleotides to 1.55-2.95 ns in the 15 base oligomers. Furthermore the shortest timescale for the dinucleotides (240-290 ps) was significantly faster than shortest timescale for the longer oligomers (410-640 ps). These two trends and the increasing mean lifetimes in Table 5.2 are consistent with the observation that the 2Ap is quenched more slowly as the oligomer length increases. While the longest timescale for the dinucleotides was found to be only 2 ns faster than the d2Ap monomer lifetime of 10.4 ns (Table A.1.5), the longest timescale for the longer oligomers was found to be 4 or more ns faster. Additionally the amplitude of the longest timescale also increases with increasing oligomer length. Finally, the mean lifetimes clearly show the 2Ap on the 5' end is quenched faster than the 2Ap on the 3' end.

Table 5.1. Fit parameters for TRF data at 10 °C from the 2Ap-labeled ssDNA oligomers.

Sample	A_1 / %	τ_1 / ns	A_2 / %	τ_2 / ns	A_3 / %	τ_3 / ns	A_4 / %	τ_4 / ns	χ^2
5'-(2Ap)A-3'	18.5	0.24	52.2	0.72	24.5	2.20	4.8	8.34	1.083
5'-A(2Ap)-3'	17.7	0.29	55.7	0.78	19.0	2.74	7.6	8.10	0.989
5'-(2Ap)AA-3'	24.4	0.56	62.5	1.90	13.1	4.77			0.935
5'-A(2Ap)A-3'	34.9	0.64	53.8	1.55	11.3	6.53			1.031
5'-AA(2Ap)-3'	19.0	0.41	64.4	1.82	16.6	6.34			1.094
5'-(2Ap)A_{14}-3'	24.4	0.57	62.3	1.98	13.3	5.78			1.056
5'-A_7(2Ap)A_7-3'	9.0	0.58	66.0	2.95	25.0	6.07			1.06
5'-A_{14}(2Ap)-3	13.5	0.55	67.9	1.70	18.6	6.44			1.136

Table 5.2. Amplitude weighted mean lifetimes at 10 °C from the 2Ap-labeled ssDNA oligomers.

Sample	$<\tau>$ / ns
5'-(2Ap)A-3'	1.36
5'-A(2Ap)-3'	1.62
5'-(2Ap)AA-3'	1.95
5'-A(2Ap)A-3'	1.79
5'-AA(2Ap)-3'	2.31
5'-(2Ap)A_{14}-3'	2.14
5'-A_7(2Ap)A_7-3'	3.52
5'-A_{14}(2Ap)-3	2.42

5.3.2 Molecular Dynamics Simulations of the Adenine Dinucleotide

A total of 15 μs of MD simulations (from five independent three μs runs) for 5'-AA-3' was performed and analyzed following the methodology used in Chapter 4 for the 5'-A(2Ap)-3' and 5'-(2Ap)A-3' dinucleotides. To determine the SH of this system, the FES for the cos(θ) and displacement coordinates was calculated (Figure 5.9). This FES shows that the same seven structures sampled by the 2Ap-labeled dinucleotides, $\beta\alpha$, $\beta\beta$, $\alpha\alpha$, $\alpha\beta$, $\alpha\alpha$*, $\alpha\beta$*, and Ex, are sampled by the A only containing dinucleotide.

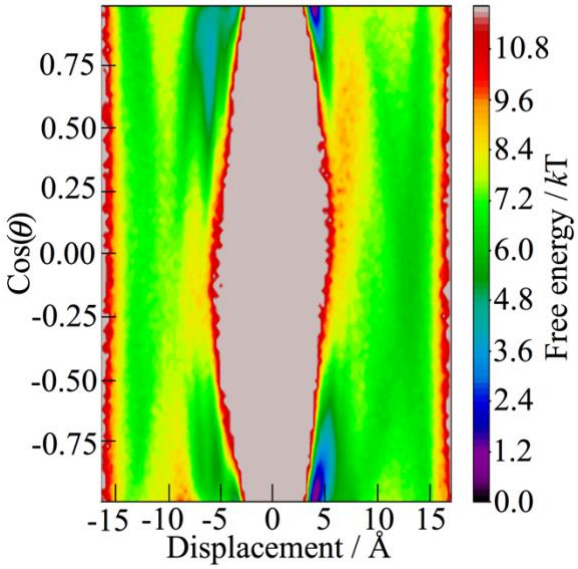

Figure 5.9. FES for the cos(θ) and displacement coordinates of 5'-AA-3' obtained via MD simulations.

Constructing a MSM from the 5'-AA-3' trajectories and coarse graining it into seven states was again achieved similar to the 2Ap-labeled dinucleotide runs (Chapter 4). The equilibrium distribution for the seven conformations are shown in Table 5.3, and are in qualitative agreement with visual inspection of Figure 5.9.

Table 5.3. Equilibrium Distribution for the Seven State HMSM of 5'-AA-3' [a]

Structure	Probability / %
$\beta\alpha$	25.7 (0.9)
$\beta\beta$	44.4 (0.5)
$\alpha\alpha$	0.49 (0.31)
$\alpha\beta$	1.4 (1.1)
$\alpha\alpha$*	8 (1)
$\alpha\beta$*	12.1 (0.8)
Ex	7.8 (0.8)

[a]Values in parentheses are 1σ errors estimated from independent trajectories.

To explore the CD of the 5'-AA-3' dinucleotides, the trajectories were used to construct a HMSM using a lag time of 500 ps. The longest timescale for this model was 18.3 ns (Figure 5.1). The transition rate matrix for this HMSM is shown in Table 5.4. The

same trend observed for the 2Ap-labeled dinucleotides (Chapter 4) is apparent in Table 5.4. Namely, the transitions where the rise and cos(θ) change sign (blue in Table 5.4) occur with higher rates than when only the rise changes sign (red in Table 5.4) or only the cos(θ) changes sign (purple in Table 5.4).

Table 5.4. Transition Rates (in μs^{-1}) for 5'-AA-3'[a,b]

Final / Initial	$\beta\alpha$	$\beta\beta$	$\alpha\alpha$	$\alpha\beta$	$\alpha\alpha$*	$\alpha\beta$*	Ex
$\beta\alpha$	-48.9 (1.6)	7.1 (2.2)	0.77 (1.2)	0 (0)	35.2 (4.5)	1.3 (0.6)	4.6 (0.6)
$\beta\beta$	4.1 (1.5)	-90.8 (6.3)	0 (0)	1.4 (0.8)	4.4 (0.7)	47.6 (5.0)	33.2 (7.9)
$\alpha\alpha$	40 (43)	4 (21)	-202 (52)	93 (52)	64 (16)	0.2 (1.0)	0 (0)
$\alpha\beta$	0 (0)	43 (46)	32 (13)	-174 (40)	5.9 (7.7)	47 (19)	46 (18)
$\alpha\alpha$*	111 (31)	24.1 (6.5)	3.9 (2.1)	1.0 (0.6)	-299 (47)	16.9 (2.3)	142 (12)
$\alpha\beta$*	2.8 (1.3)	175 (11)	0 (0)	5.6 (6.2)	11.4 (1.8)	-295 (24)	100 (19)
Ex	15.0 (2.6)	190 (31)	0 (0)	8.6 (3.6)	147.9 (9.6)	155 (20)	-516 (11)

[a]Values in parentheses are 1σ errors estimated from independent trajectories. [b]With the exception of transitions involving the Ex state, transition rates are color-coded, depending on whether rise (red), cos(θ) (purple), or both rise and cos(θ) change sign (blue).

5.3.3 Multi-State Nearest Neighbor Stacking Model

The intrinsic repetition of a polymers chemical composition can be exploited to build a model for ssDNA. Here, a multi-state nearest neighbor stacking model is defined which can predict the possible structures available to 5'-A$_n$-3'. The seven conformations available to 5'-AA-3' form a basis for a particular nearest neighbor interaction for 5'-A$_n$-3'. Starting from the 5' end, one of the seven conformations for the first two bases can be chosen. This choice restricts the possible conformations available for the second nearest

neighbor interaction because the second base can only expose the face to the third base. For example, if $\beta\beta$ is the conformation of the first two bases, the β-face of the second base is already stacking with the first base so only its α-face is exposed allowing for $\alpha\alpha$, $\alpha\alpha$*, $\alpha\beta$, $\alpha\beta$*, or Ex interactions between the second and third bases. Conversely, if the first two bases are in the $\beta\alpha$ conformation, the second and third bases can only form $\beta\alpha$, $\beta\beta$, or Ex conformations. In this model Ex conformations can always be formed, have both α and β-faces exposed, and can therefore allow the next base to adopt any of the seven conformations. To name a particular conformation, a sequence of the nearest neighbor conformations is concatenated. For example, a three base oligomer containing two $\beta\alpha$ interactions is $\beta\alpha\beta\alpha$. For 5'-AAA-3' this model predicts 31 possible conformations, while for 5'-A_n-3' the number of possible conformations computed numerically are shown in Figure 5.10. Figure 5.10 shows the number of conformations increases exponentially with 10 bases already able to adopt 10^6 conformations. This model was used to predict the relative probabilities of each 5'-AAA-3' structure using statistical mechanics. First the probabilities in Table 5.3 were converted into relative energies. Then, assuming the energy of the oligomer is just the sum of nearest neighbor interaction energies, the probability of each conformation was determined from the Boltzmann factors corresponding to the 31 different structures.

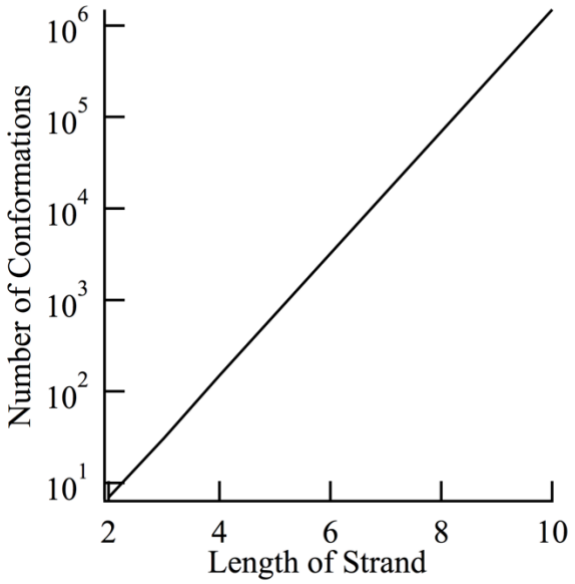

Figure 5.10. The number of conformations predicted by the nearest neighbor stacking model described in the text as a function of strand length.

5.3.4 Molecular Dynamics Simulations of the Adenine Trinucleotide

For the 5'-AAA-3' oligomer 31 MD simulations were conducted where the initial configuration in each run was set to a distinct structure predicted by the multi-state nearest neighbor stacking model. These simulations were each run for 500 ns giving a total simulation time of 15.5 μs. To analyze the SH of this system, two FES for the two pairs of cos(θ) and displacement coordinates are shown in Figure 5.11. The first nearest neighbor interaction has a FES (Figure 5.11A) that closely resembles the FES for the 5'-AA-3' sample (Figure 5.9). However, the second nearest neighbor interaction (Figure 5.11B) has new two new minima with positive displacement values. The $\alpha\beta$ and $\alpha\alpha$ conformations appear to occur more frequently in the second interaction. The extended region is also significantly stabilized in the second interaction.

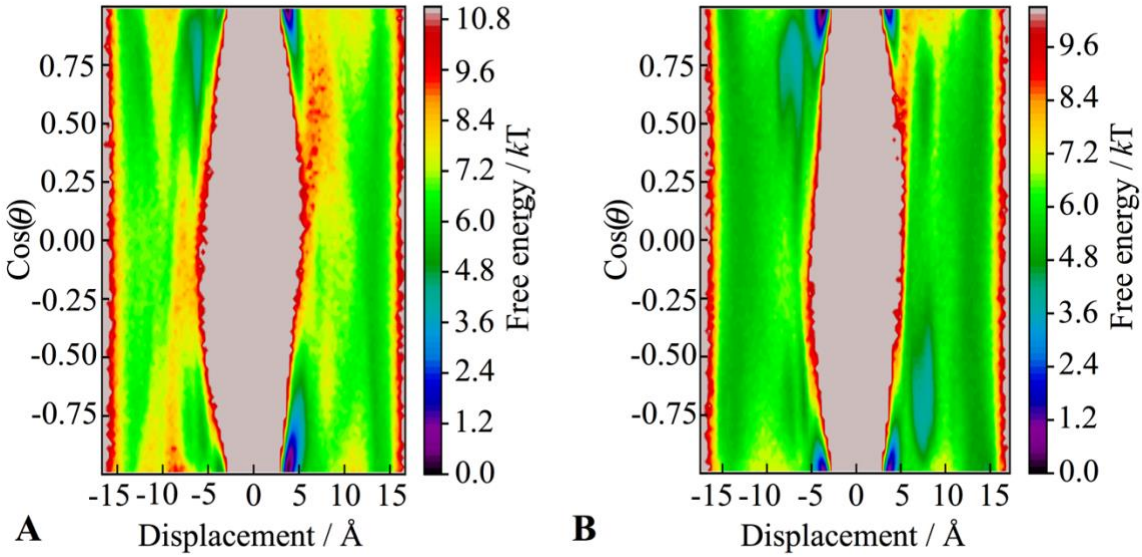

Figure 5.11. The FES for the cos(θ) and displacement coordinates of the 5'-end A and interior A in 5'-AAA-3' (A). The FES for the cos(θ) and displacement coordinates of the interior A and 3'-A in 5'-AAA-3' (B).

To explore how different the SH determined from the MD simulations is from the predicted SH using the multi-state nearest neighbor stacking model (Section 5.3.3), the regions defined in Figure 5.3 were used to partition the four dimensional FES defined by the two sets of cos(θ) and displacement pairs. Integrating the probability over the 31 defined regions gives the relative probability of being in each in of the 31 structures. This is compared to the predicted probabilities from the nearest neighbor stacking model in Figure 5.12. While the presentation in Figure 5.12 appears to show the nearest neighbor model qualitatively tracking the probability for each structure, the correlation is poor (Pearson correlation coefficient = 0.247) as the error scales with predicted probability.

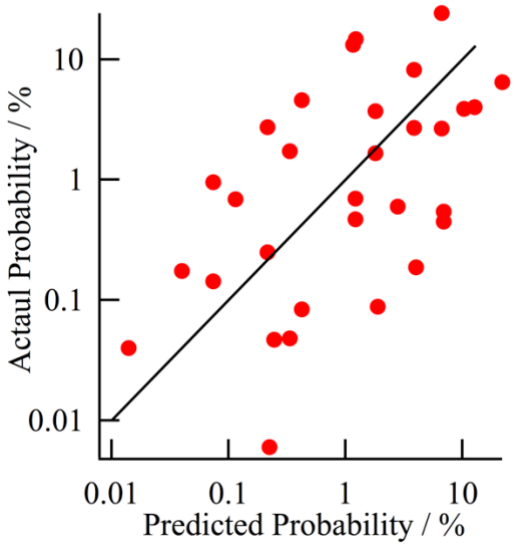

Figure 5.12. Comparison of the probability of 5'-AAA-3' adopting the 31 structures predicted by nearest neighbor stacking model with actual probabilities determined from the MD simulations. Each red dot represents one structure while the black line is the line y=x. The Pearson correlation coefficient is 0.247 for this data set. Note the logarithmic axes.

5.4 Discussion

5.4.1 Implications of Experiments on the Structure and Dynamics of the ssDNA Oligomers

The UV-visible absorption spectra of the d2Ap reference compound (black in Figures 5.6A and 5.7A) show the lowest energy absorption band of 2Ap is located at 305 nm in agreement with literature.[16] In the dinucleotides shifts of this absorption band are not resolvable within the experimental uncertainty. Much more obvious shifts in the spectra are seen when 2Ap is in the 15 base oligomers (Figure 5.7A) indicating 2Ap is more π-stacked in these systems. However this is in contradiction with the steady-state emission measurements if it is assumed that more π-stacking means more quenching. Both the steady-state emission (Figures 5.6B and 5.7B) and emission quantum yields for

the dinucleotides (Φ_s = 0.061-0.072) versus the 15 base oligomers (Φ_s = 0.087-0.114) suggest that 2Ap is more π-stacked in the dinucleotides. Furthermore, the fact that an interior base has two neighbors suggests it is twice as likely to be stacked as an exterior one,[28] but the Φ_s of 2Ap at the interior of the 15 base oligomer is larger than 2Ap on either ends. Instead, the absorbance maxima for 2Ap is a better measure of the degree of stacking at the instant of photoexcitation as 2Ap at the interior position of the 15 base oligomer is more red-shifted than the two end positions (Figure 5.7A). This is likely because absorption only depends on the instantaneous nuclear configuration, while florescence quenching can occur at all times following photoexcitation. The higher stacking of 2Ap at the interior is consistent with measurements by Davis et al. which showed that energy transfer from A to 2Ap is far more efficient when 2Ap is at the interior positions of a five base oligomer relative to the ends.[45] The observation that Φ_s does not track well with the stacked population of 2Ap highlights the limitations of the steady-state emission signals as they cannot differentiate between quenching due to 2Ap π-stacked at the instant of photo-excitation and the 2Ap which are initially unstacked but form π-stacks within 2Ap's excited state lifetime.

To explore the dynamic aspects of the emission quenching, TRF signals of the 2Ap-labeled oligomers were measured. Using the quenching timescales determined from the TRF signals as a proxy for the timescales for unstacked 2Ap to form a π-stack (an approximation that will be more carefully assessed in Section 5.4.3) the observed trends suggest π-stacking occurs at different rates in different length strands and at the interior versus ends. Specifically, this approximation suggests two key results: π-stacking occurs

with higher rates on the 5' end of a strand versus the 3' end, and that π-stacking occurs with higher rates in shorter oligomers relative to longer ones. The position dependent mean lifetimes are consistent with previous reports of 2Ap-labeled 5'-A_{30}-3'.[46] The observation that multiple discrete exponentials are required to model the data (Table 5.1) highlights the π-stacking phenomena may occur over multiple timescales due to the involvement at least three to four different structures.

5.4.2 Implications of MD Simulation on the Structure and Dynamics of the ssDNA Oligomers

An important consideration is that the 5'-AA-3' simulations described here are very similar to the 5'-(2Ap)A-3' and 5'-A(2Ap)-3' simulations shown in Chapter 4. The same seven structures are sampled in all three systems (Figure 5.9), and the conformational dynamics for the three systems is also very similar (Table 5.4). The same trend, that structural transitions which minimize the exposure of the hydrophobic nucleobase faces occur more rapidly, observed for the 2Ap-labeled dinucleotides is seen in 5'-AA-3' (Table 5.4). These considerations suggest that incorporation of the 2Ap label doesn't drastically alter the ssDNA, supporting its use as a fluorescence probe. Furthermore, it justifies the simplification made here that simulations of 5'-AAA-3' suffice for comparison with experiments on the 2Ap-labeled trinucleotides.

Completely quantifying the SH of oligomers longer than the dinucleotides is a significantly more difficult task. This is exemplified by the nearest neighbor stacking model that suggests 31 meta-stable structures for 5'-AAA-3' and billions for 5'-A_{15}-3'. To drive this point home, inspection of Figure 5.11B demonstrates that the nearest neighbor stacking model is actually underestimating the SH in 5'-AAA-3', as the

extended region with positive displacement includes new sheared structures not included in the nearest neighbor stacking model. However, Figure 5.12 shows that roughly half of the structures predicted for 5'-AAA-3' occur with <1% relative probability, and the poor correlation in Figure 5.12 indicates that the nearest neighbor stacking model doesn't get the probability of structures in the longer oligomers correct. Regardless, it is questionable whether a standard MD simulation could ever converge the full SH of 5'-A_{15}-3'. Computing the average rate for interconversions between different structures from Table 5.4 gives a rate of roughly 40 μs^{-1}. This suggests a simulation of 5'-A_{15}-3' would require roughly 25 seconds of simulation time to just visit all conformations once, much less sample them statistically. For this reason, only simulations of 5'-AAA-3' are compared to experiment and simulations of 5'-A_{15}-3' will require advances in simulation methodology,[188] or additional simplifications like performing non-equilibrium simulations on the subset of conformations where the position of 2Ap is initially not stacked.

5.4.3 Insights into Fluorescence Quenching of 2Ap Gained from the Molecular Dynamics Simulation

The SH implicated by the spectroscopic experiments is now compared to the SH quantified by the MD simulations of 5'-AAA-3'. The steady-state absorption and emission measurements have conflicting implications on the position dependence of the likelihood to be stacked in the 2Ap-labeled 15 base oligomers (see Section 5.4.1 for discussion). Using the probabilities for the 31 different structures of 5'-AAA-3' taken from Figure 5.12, the probability of the 5'-A, interior-A, and 3'-A being stacked was determined to be 67 %, 79 %, and 42 % respectively. Given the lowest energy absorbance

of 2Ap on the interior of the 15 base oligomer is most significantly red-shifted (Figure 5.7A) these computed stacking likelihoods are in line with the interpretation that the location of 2Ap's lowest energy absorbance peak is correlated with the probability of the 2Ap probe being π-stacked.

Often, the sum of multiple exponential decay models used to model TRF measurements of 2Ap-labeled ssDNA oligomers are interpreted as the existence of multiple ssDNA structures.[10] This SH is quantified in the MD simulations of 5'-AA-3' and 5'-AAA-3'. Interestingly, the number of components required to fit the TRF signals doesn't agree with the number of structures observed in the simulations, as the dinucleotides require four exponential components while the trinucleotides and 15 base oligomers require only three (Table 5.1). Frequently, the relative amplitude of the shortest decay component in the TRF measurement is taken as the relative probability of 2Ap being stacked.[10] Here, the amplitudes for this component are less than 25% (Table 5.1) and do not correlate well with the position dependent stacking likelihoods taken from the MD simulations (67 %, 79 %, and 42 % for the 5'-A, interior-A, and 3'-A).

As shown above the SH alone doesn't fully explain the observed emission decay heterogeneity, so the influence of the CD predicted from the MD simulations was determined. This was achieved by first calculating a MSM for 5'-AAA-3' which had a longest timescale of 50 ns (Figure 5.5) consistent with the temperature-jump relaxation timescale of 45 ± 9 ns for the RNA form of the same molecule.[25] Unfortunately, constructing a 31 state HMSM to extract transitions rates between the different structures in Figure 5.11 was unfeasible. Instead, an analogous model to the second model used in Chapter 4 was constructed for 5'-AAA-3'. Recall that the key assumption for model 2 in

Chapter 4 was that the emission decay signal arises from 2Ap which are not stacked, and the quenching timescales are the timescales at which the various unstacked conformations form base stacks. To enforce this assumption on the 5'-AAA-3' MSM, each of the 2000 microstates were classified as having either the first nearest neighbor interaction or the second interaction stacked or not based on the COM distance being less than 5 Å for at least 90% of the data points in the microstate. The simple 5 Å cut-off can distinguish the stacked and unstacked conformations in Figure 5.11. Then for a particular location of 2Ap in 5'-AAA-3', the 2000 state rate matrix was condensed into a matrix with size equal to the number of microstates where the A is not stacked with the other bases. The diagonal elements for the rate matrix were adjusted to account for the monomer-like deactivation of 2Ap and the sum of base stacking rates for each microstate. The predicted emission signals, amplitudes, and timescales were computed as described in Section 2.3.2.

The results of the quenching model, which assumes 2Ap is quenched due to the dynamic formation of base stacks, are shown in Figure 5.13. This model predicts multi-exponential emission decays for 2Ap at all three positions in 5'-AAA-3' (Figure 5.13A). The amplitude weighted mean lifetimes predicted by this model for the three positions are 3.1, 3.0, and 4.7 ns for the 5'-end, interior, and 3'-end respectively. These are generally slower than the experimental values of 1.95, 1.79, and 2.31 ns (Table 5.2) for 2Ap the same three positions, but capture the experimental position dependence quite well. The slower values from MD simulations are likely due to the use of a 1 ns lag-time for the MSM construction. This lag-time was required for the timescales in Figure 5.5 to converge, but prohibited the identification of conformational dynamics which occur on

faster timescales. The trend that the 5’ base is quenched faster than the 3’ base seen in experiment (Figure 5.8A and B) is reproduced in the model of stacking kinetics from the MD (red and purple curves in Figure 5.13A). This suggests that the faster quenching of 2Ap on the 5’ end is not due to a static modulation of electron transfer rate, but instead due to the faster stacking rate for bases on that end. Furthermore, the longest two timescales observed experimentally (1.5 - 6.5 ns) overlap with the multiple longest timescales predicted by the model (1.2 - 6.7 ns). The fact that so many timescales have positive amplitudes from the simulation (Figure 5.13 B-D) reinforces the idea that base stacking is a multi-scale phenomena, which then maps onto the emission decay heterogeneity observed from 2Ap-labeled oligomers.

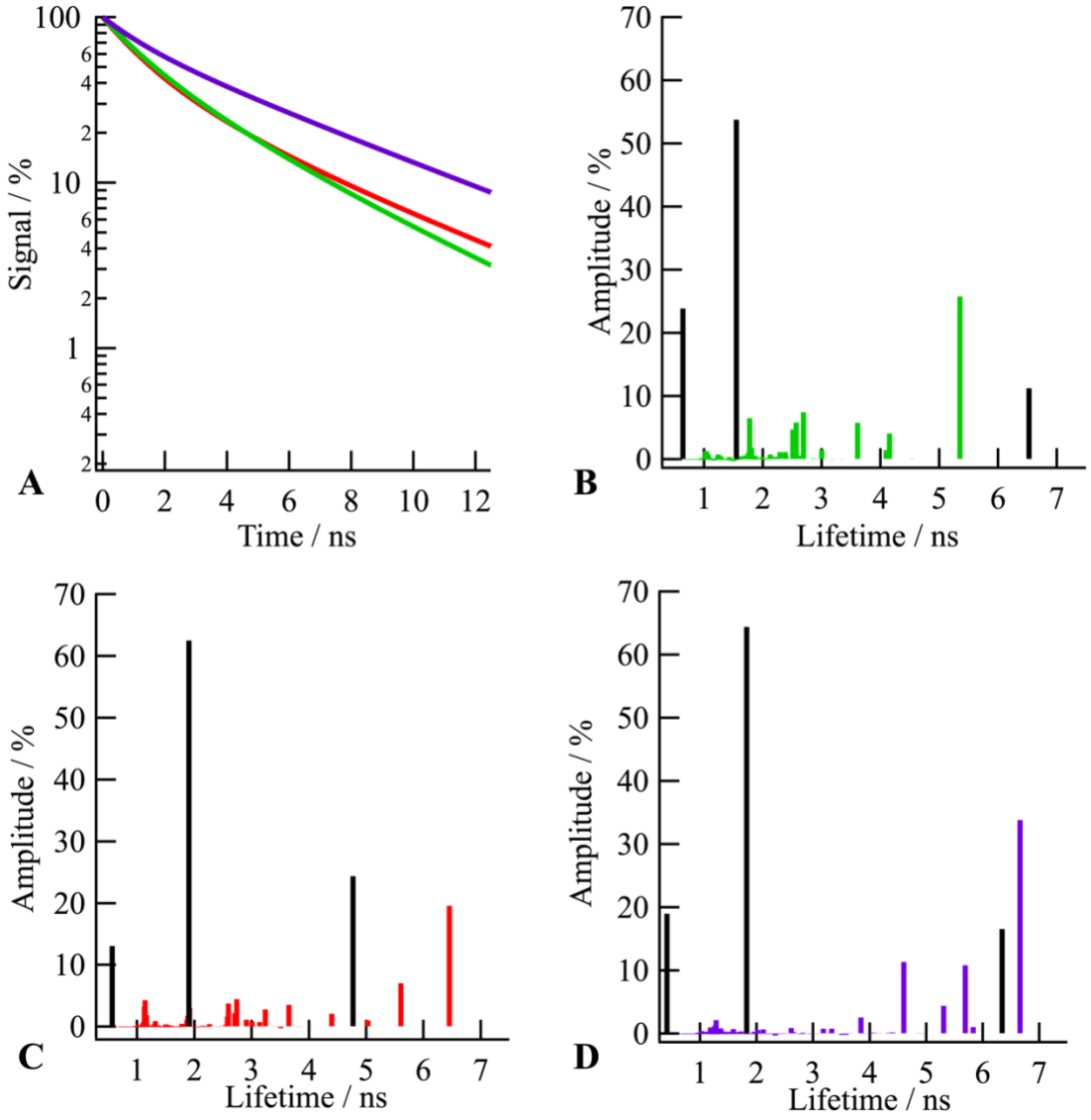

Figure 5.13. The predicted signals (A) with lifetimes and amplitudes (B-C) for the 2Ap quenching model of 5'-AAA-3' described in the text are the colored lines. Experimental lifetimes and amplitudes are the black lines. Predictions for 2Ap at the interior position (green), 5'-end (red), and 3'-end (purple) are shown in panels B,C, and D respectively, while all three are shown in panel A.

The MSM built from the MD simulations of the trinucleotides further enforce the idea that the conformational dynamics occurs on timescales (Figure 5.5) commensurate with the observed excited state deactivation of 2Ap in the trinucleotides (Figure 5.1). Additionally, the MSM demonstrates that already in trinucleotides there are so many structures and associated relaxation times that models of the TRF decays that include rate distributions are justified over purely discrete fits. For examples see Figure 5.5 where at least 40 timescales are converged at a 1 ns lag time, or Figure 5.13 B-D where 10 or more discrete exponential components have positive amplitude. Future work could explore how to broaden the discrete nature of the predicted timescale plots shown in Figure 5.13 B-D to a continuum lifetime distribution comparable to experiment. This is analogous to the convolution of the discrete transition energies estimated from QM calculations with gaussians to mimic the inhomogeneous broadening of absorption and emission line-shapes.[174] However, comparing predicted and experimentally determined lifetime distributions should be treated with caution as qualitatively different lifetime distributions produce similar signals in the time-domain. For example in the trinucleotide the predicted quenching of 2Ap from the MSM (Figure 5.13 B-D) clearly show at least 10 decay timescales with positive amplitudes, yet visually there is no way to tell this from figure 5.13 A. Finally, because the only contribution provided by this model to the quenching observed in the TRF measurements comes from the time dependent formation of stacked bases from initially unstacked ones, the agreement of the predicted position dependence of 2Ap's mean lifetime with experiment, and the emission decay heterogeneity with experiment (Figure 5.13) strongly support the idea that the TRF experiments of 2Ap-labeled oligomers are sensitive to the dynamic stacking of initially unstacked 2Ap.

Additional work could be done to build models analogous to model 3 in Chapter 4, that introduce conformation specific quenching constants for 2Ap in the many different conformations obtained by the MSM, but the agreement of the purely dynamic model used here suggests this may be unessisary.

5.5 Conclusion

Both the SH and CD of adenine containing ssDNA oligomers were explored using experiment and simulation. The MD simulations of 5'-AA-3' are very similar to MD simulations of 2Ap-labeled 5'-AA-3' supporting the idea that 2Ap does not significantly alter the SH and CD and therefore is an effective fluorescence probe. While the MD simulations predict a large number of structures for 5'-A_3-3' and 5'-A_{15}-3', the TRF emission measurements require only a small number of exponentials to model. However, this is likely because the TRF measurement are subject to experimental noise and do not contain enough information to fully resolve the microscopic heterogeneity. For these ssDNA oligomers with many available conformations, it is suggested that a future approach could be to broaden the discrete predicted decay lifetimes into a lifetime distribution for comparison with an experimentally determined one.

Taking a step back and focusing on general trends, the UV-visible absorption measurements and MD simulations suggest that bases are more likely to stacked when they reside in the interior positions the strand. Simply assuming that the emission of 2Ap will be quenched more at the interior positions because of the increased stacking was shown to be inconsistent with the steady-state emission experiments of 2Ap-labeled

5’-A_{15}-3’. The TRF measurements of 2Ap-labeled 5’-A_{15}-3’ also showed slower quenching time scales for 2Ap in the interior corroborating the steady-state emission experiments. However, in the trinucleotides the 2Ap at the interior was found to be quenched faster than 2Ap on either end which, along with MD simulations showing more stacking on the interior, suggest the emission quenching is more complicated than just considerations of the static stacking propensity. This finding agrees with the intuition provided by a previous report[28] that bases on the interior of the strand have two neighbors to stack with instead of one.

To incorporate the dynamic effect of the formation of stacked 2Ap from initially unstacked 2Ap, MSMs were built from extensive MD simulations of 5’-AAA-3’. Adapting this MSM to a quenching model that assumes the TRF signal arises from only the unstacked conformations which are quenched by base stack formation predicts multiple emission timescales commensurate with the experimental ones. The multitude of predicted quenching timescales for 2Ap in this model are entirely due to the disperse base stacking kinetics, while the timescales themselves are slightly faster due to the monomer-like deactivation of 2Ap. The observation that 2Ap is quenched faster on the 5’ end vs 3’ end, seen experimentally for all oligomers, is reproduced by the faster stacking kinetics on the 5’ end verses the 3’ end seen in the MD simulations. Furthermore, the apparent discrepancy that the interior base in the trinucleotide is quenched faster than the ends, is explained by faster stacking kinetics predicted by the MD simulations for this residue. These considerations strongly indicate that the ubiquitous heterogeneous time-resolved emission quenching measurements of 2Ap in ssDNA oligomers are mostly due to the multi-scale kinetics of π-stack formation. This is in agreement with previous studies of

2Ap-labeled trinucleotides that suggest the formation of highly quenched conformations is important, but these studies lacked the microscopic MD simulations used here.[78] The use of TRF and MD simulations analyzed with MSMs represent a powerful toolset to understand the SH and CD of these ssDNA and likely other biopolymers.

5.6 References

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

16. Kervin Evans, D. X., Younsik Kim, Thomas M. Nordlund, 2-Aminopurine Optical Spectra: Solvent, Pentose Ring, and DNA Helix Melting Dependence. *J. Fluoresc.* **1992,** *2*, 209-216.

22. Jean, J. M.; Krueger, B. P., Structural Fluctuations and Excitation Transfer between Adenine and 2-Aminopurine in
Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B.* **2006,** *110*, 2899-2909.

25. Pörschke, D., Dynamics of Nucleic-Acid Single-Strand Conformation Changes - Oligoriboadenylic and Polyriboadenylic Acids. *Eur. J. Biochem.* **1973,** *39*, 117-126.

28. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B.* **2012,** *116*, 10266-10274.

37. Remington, J. M.; McCullagh, M.; Kohler, B., Molecular Dynamics Simulations of 2-Aminopurine-Labeled Dinucleoside Monophosphates Reveal Mutliscale Stacking Kinetics. *J. Phys. Chem. B* **2019,** *123*, 2291-2304.

45. Steven P. Davis, M. M., Andreia Williams, and Thomas M. Nordlund, Position Dependence of 2-Aminopurine Spectra in Adenosine Pentadeoxynucleotides. *J. Fluoresc.* **2003,** *13*, 249-259.

46. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B.* **2007,** *111*, 5757-5766.

78. Jean, J. M.; Hall, K. B., Stacking-Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry* **2004,** *43*, 10277-10284.

82. Melanie A. O'Neill, J. K. B., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

142. Pinamonti, G.; Zhao, J. B.; Condon, D. E.; Paul, F.; Noé, F.; Turner, D. H.; Bussi, G., Predicting the Kinetics of RNA Oligonucleotides Using Markov State Models. *J. Chem. Theory Comput.* **2017,** *13*, 926-934.

174. Callis, P. R., Predicting Fluorescence Lifetimes and Spectra of Biopolymers. In *Methods in Enzymology*, Johnson, M. L.; Brand, L., Eds. Elsevier Academic Press Inc: San Diego, 2011; Vol. 487, pp 1-38.

179. Sakakibara, Y.; Abeysirigunawardena, S. C.; Duc, A.-C. E.; Dremann, D. N.; Chow, C. S., Ligand- and pH-Induced Conformational Changes of RNA Domain Helix 69 Revealed by 2-Aminopurine Fluorescence. *Angew. Chem., Int. Ed.* **2012,** *51*, 12095-12098.

180. Lee, H.-W.; Briggs, K. T.; Marino, J. P., Dissecting Structural Transitions in the HIV-1 Dimerization Initiation Site RNA using 2-Aminopurine Fluorescence. *Methods* **2009,** *49*, 118-127.

181. Liang, C.; Li, X.; Rong, L.; Inouye, P.; Quan, Y.; Kleiman, L.; Wainberg, M. A., The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *Journal of Virology* **1997,** *71*, 5750-5757.

182. Sinsheimer, R. L., A Single-Stranded Deoxyribonucleic Acid from Bacteriophage φX174. *J. Mol. Biol.* **1959,** *1*, 43.

183. Zou, Y.; Liu, Y.; Wu, X.; Shell, S. M., Functions of Human Replication Protein A (RPA): From DNA Replication to DNA Damage and Stress Responses. *Journal of Cellular Physiology* **2006,** *208*, 267-273.

184. Avilov, S. V.; Piemont, E.; Shvadchak, V.; de Rocquigny, H.; Mely, Y., Probing Dynamics of HIV-1 Nucleocapsid Protein/Target Hexanucleotide Complexes by 2-Aminopurine. *Nucleic Acids Res* **2008,** *36*, 885-96.

185. Stivers, J. T., 2-Aminopurine Fluorescence Studies of Base Stacking Interactions at Abasic Sites in DNA: Metal-Ion and Base Sequence Effects. *Nucleic Acids Res* **1998,** *26*, 3837-3844.

186. Nordlund, D.-G. X. a. T. M., Sequence Dependence of Energy Transfer in DNA Oligonucleotides. *Biophys. J.* **2000,** *78*, 1042-1058.

187. Yang, K.; Stanley, R. J., Differential Distortion of Substrate Occurs When It Binds to DNA Photolyase: A 2-Aminopurine Study. *Biochemistry* **2006,** *45*, 11239-11245.

188. Sultan, M. M.; Pande, V. S., tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017,** *13*, 2440-2447.

CHAPTER SIX

SUMMARY AND OUTLOOK

6.1 Summary

The models of 2Ap emission quenching developed herein build upon previous understanding of 2Ap emission quenching[10-11, 46, 50, 82] while also leveraging insight gained from femtosecond pump-probe TA spectroscopy of natural base containing oligomers.[3, 26-27] Specifically, it was shown in Chapter 3 that the population of stacked 2Ap at the instant of photoexcitation was not required to model the TRF signals observed on timescales greater than 65 ps. This strongly contrasts with the common assumption that the relative amplitude of the fastest timescale observed in picosecond TRF measurements is proportional to the relative population of stacked 2Ap.[11, 78 10] Instead, it was argued that both a conformation dependent 2Ap excited state decay rate and a continuum of conformations could reproduce the observed non-exponential emission decays from 2Ap-labeled dinucleotides. Despite the agreement between this model and TRF measurements at a single temperature, it was presumed that there were no interconversions between these structures on the timescale of 2Ap emission (<10 ns), something not supported by the temperature-dependent TRF measurements which show faster decay timescales at higher temperatures.

In Chapter 4, the assumption that emission quenching of 2Ap occurred via a static array of structures was carefully scrutinized by quantifying the SH and CD of the same dinucleotide systems using MD simulations. These simulations revealed that instead of

the two state structural model assumed in Chapter 3, and commonly used for dinucleotides elsewhere,[23-24, 140] the dinucleotides sampled at least seven different conformations. These seven conformations included two sheared conformations where the rate of emission quenching for 2Ap is likely to take on intermediate values between the fully stacked and extended conformations. The estimated timescales for transitions between the seven structures strongly overlapped with the timescales for emission quenching of 2Ap, signifying the purely static model was inappropriate. On the other hand, a model which presumed the TRF signals arise entirely from unstacked 2Ap that quenched 2Ap emission by forming base stacks, predicted mono-exponentially decaying signals for the dinucleotides with 2-4 ns timescales. This further suggested that the CD should be included in the model for quenching of 2Ap. Finally, the model that predicted multiple <10 ns exponential decays that agreed nicely with experiment included both SH, in the form of heuristic conformer-specific quenching rates, and CD, as the structural transition rates from MD simulation. An important feature of this later model is that because the structural interconversion rates determined from the MD simulation were included, the relative amplitudes of specific time constants did not correspond to relative populations of 2Ap in specific structures. This is in contrast to prior analysis of 2Ap fluorescence decays which presume these amplitudes arise from specific conformations.[10-11]

In Chapter 5, MD simulations of 5'-AA-3' were first conducted and found to be similar to 2Ap-labeled ones. This reveals that 2Ap does not drastically affect the SH and CD of the ssDNA it is utilized to quantify. This is an important result as it justifies the use of 2Ap as a non-invasive fluorescence probe. This is in mild contradiction with the

finding that base-pair opening dynamics in DNA duplexes are slightly faster when A is replaced by 2Ap.[189] However, the difference here is probably because there is no base-pairing involved in the 2Ap-labeled ssDNA, and 2Ap and A stacks are similar to A and A stacks.

Furthermore, longer 2Ap-labeled ssDNA oligomers were studied in Chapter 5 with simulation and experiments. First, the steady-state UV-visible absorption and emission measurements reinforced the finding that a static model of stacked and unstacked 2Ap bases does not capture the observed quenching. Instead, CD must be included in the models of fluorescence quenching. As suggested for the dinucleotides, the TRF experiments of the 2Ap-labeled systems are sensitive to the CD. For the trinucleotides the TRF demonstrated that the 2Ap probe is quenched faster on the 5' end versus the 3' ends of the ssDNA. For 2Ap at a particular position (5'-end, 3'-end, or interior) the quenching was found to be slower as the oligomer length was increased. These findings suggest the local SH or CD of the ssDNA at the different positions is altered, but as shown for the dinucleotides, MD simulations should be conducted to independently estimate the CD.

To ensure MD simulations of the 5'-AAA-3' were converged, the seven structures present in 5'-AA-3' were used to build a multi-state nearest neighbor stacking model for 5'-A_n-3' of arbitrary length. This model predicted 31 structures for 5'-AAA-3', and a 500 ns simulation was run from each. The aggregate of the simulations suggest the nearest neighbor stacking model does not predict the probabilities of the longer 5'-AAA-3' well, but all structures predicted by the model were observed to be meta-stable in the simulations. This suggests that the billions of structures predicted for 5'-A_{15}-3' are

unlikely to be sampled in tens or hundreds of μs of MD simulations, and will require more advanced MD simulation techniques.

The experimental signals from 2Ap-labeled 5'-AAA-3' were then predicted using a model where 2Ap was quenched solely due to the dynamic formation of base stacks. These predicted signals included significant decay heterogeneity as they included more than ten discrete exponential decays with positive amplitudes. These predicted signals independently revealed the position dependent trend that 2Ap is quenched faster on the 5' end relative to the 3' end. Thus, the combination of MD and TRF revealed a dynamic feature of ssDNA that the stacking dynamics is more rapid on the 5' end relative to the 3' end. In addition, the dynamic model showed that the stacking kinetics were slower in the trinucleotide relative to the dinucleotides which only showed 2-4 ns timescales. Both of these findings further reinforces the overall theme that using both MD simulations and TRF experiments can provide insight into the mechanisms of SH and CD in ssDNA oligomers.

The primary techniques used in this thesis, picosecond TRF and MD simulations, have proven to be complementary techniques for elucidating the SH and CD present in 2Ap-labeled ssDNA oligomers. The TRF measurements of 2Ap were sensitive to the ensemble averaged CD of the oligomers on timescales ranging from 65 ps to ~50 ns, while MD simulations could probe a single oligomer on 2 ps to 15 μs timescales. For systems which can sample the SH on 15 μs timescales, like di- and trinucleotides, the single molecule simulations were found to be in agreement with the bulk experiments. This agreement was conditional on the construction of emission quenching models that explicitly included the effects of SH and CD. The approach utilized here, to construct

quenching models using independent estimates of SH and CD from MSMs built from MD simulations and then converting to a master equation for the excited state population of 2Ap shows promise to unravel the complexity present in ssDNA and other 2Ap-labeled nucleic acids.

6.2 Outlook

The focus in this thesis was on comparing the SH and CD predicted by MD simulations to the signals obtained from a specific set of experimental techniques and 2Ap-labeled ssDNA systems. However, both different experimental observables and other 2Ap-labeled DNA strands are sensitive to the SH and CD accessible to MD simulations. Another possible experiment would be to use TA measurements to characterize the time dependent excited state population of 2Ap. TA measurements have the advantage of ultrafast (< ps) time resolution and is capable of monitoring the population of dark states like the CT states presumed to result from excitation of 2Ap π-stacked with adenine.[28] Additionally, perturbative methods which monitor the relaxation of the steady emission from 2Ap following the production of a non-equilibrium distribution of conformations (e.g. by using a temperature jump apparatus[23]) are expected to provide a window into the CD of the 2Ap-labeled nucleic acids that is independent of the 10 ns lifetime of 2Ap.[190] However, measurements that assume the time dependence of the amount of steady-state emission quenching of 2AP is directly related to the timescales for π-stacking should be treated with caution as we have argued here the steady-state quantum yield is not well correlated with the probability of the 2Ap being stacked. Instead, the steady-state quantum yield is more correlated with the timescales at which

base stacking occurs. Therefore, measuring the relative steady-state emission of 2Ap following a temperature jump would track how the stacking timescales themselves are influenced by temperature. Given that the peak wavelength of the lowest energy UV-visible absorption band (or excitation emission band) of 2Ap correlates with the degree of stacking, a temperature jump experiment with uv-visible absorption detection may be more suited to determine the position dependent stacking kinetics. Or perhaps the relaxation of a steady state fluorescence excitation spectra would provide a sensitive handle on the structural relaxation timescales. All the above experiments are amendable to the dynamical fingerprint model developed by others[153] which we have adapted here to excited state models in Chapter 4.

While it was found that simulation of the equilibrium structural heterogeneity present in longer 2Ap-labeled and A-containing ssDNA oligomers may be practically unfeasible, there may be pathways to approximate the stacking dynamics relevant to the 2Ap probe in longer oligomers. For example, it was shown in the trinucleotide that the 2Ap probe is sensitive to the local conformational dynamics, i.e. the time it takes to form stacks with the nearest neighbors of 2Ap in the strand. Therefore, for a longer oligomer one could run non-equilibrium simulations where the initial structures are set to structures where the 2Ap is initially unstacked, and the simulation run until the 2Ap stacks. The use of non-equilibrium simulations to construct models analogous to the equilibrium MSMs used here has recently been developed[44, 191-192] and would serve as a powerful analysis tool for this purpose. Furthermore, computation power from the consumers perspective is anticipated to continue growing,[193] GPU accelerated standard MD simulations are only recently gaining steam,[194] and the advanced methods for integrating newtons laws of

motion used by using metadynamics[188, 195] or coarse-graining[196-197] suggests the equilibrium dynamics of much larger DNA strands are possible. Therefore MD simulations on larger 2Ap-labeled oligomers aimed at quantifying the SH and CD are feasible in the future.

6.3 References

3. Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernàndez, C. E.; Kohler, B., DNA Excited-State Dynamics: From Single Bases to the Double Helix. In *Annu. Rev. Phys. Chem.*, Annual Reviews: Palo Alto, 2009; Vol. 60, pp 217-239.

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

11. Jean, J. M.; Hall, K. B., 2-Aminopurine Fluorescence Quenching and Lifetimes: Role of Base Stacking. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 37-41.

23. Pörschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

24. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 2315-2328.

26. Chen, J.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

27. Takaya, T.; Su, C.; de La Harpe, K.; Crespo-Hernàndez, C. E.; Kohler, B., UV Excitation of Single DNA and RNA Strands Produces High Yields of Exciplex States Between Two Stacked Bases. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 10285-10290.

28. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B.* **2012,** *116*, 10266-10274.

44. Husic, B. E.; Pande, V. S., Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018,** *140*, 2386-2396.

46. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B.* **2007,** *111*, 5757-5766.

50. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of Lifetime Distributions from Fluorescence Decays with Application to DNA-Base Analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

78. Jean, J. M.; Hall, K. B., Stacking-Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry* **2004,** *43*, 10277-10284.

82. Melanie A. O'Neill, J. K. B., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

140. Pörschke, D.; Eggers, F., Thermodynamics and Kinetics of Base-Stacking Interactions. *Eur. J. Biochem.* **1972,** *26*, 490-498.

153. Keller, B. G.; Prinz, J. H.; Noé, F., Markov Models and Dynamical Fingerprints: Unraveling the Complexity of Molecular Kinetics. *Chem. Phys.* **2012,** *396*, 92-107.

188. Sultan, M. M.; Pande, V. S., tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017,** *13*, 2440-2447.

189. P.-O.Lycksell, A. G., F.Claesens, L.W.McLaughlin, U.Larsson, and R.Rigler, Base Pair Opening Dynamics of a 2-Aminopurine Substituted Eco RI Restriction Sequence and its Unsubstituted Counterpart in Oligonucleotides. *Nucleic Acids Res* **1987,** *15*, 9011-9025.

190. Ballin, J. D.; Bharill, S.; Fialcowitz-White, E. J.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M., Site-Specific Variations in RNA Folding Thermodynamics Visualized by 2-Aminopurine Fluorescence. *Biochemistry* **2007,** *46*, 13948-13960.

191. Nuske, F.; Keller, B. G.; Pérez-Hernàndez, G.; Mey, A.; Noe, F., Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014,** *10*, 1739-1752.

192. Mardt, A.; Pasquali, L.; Wu, H.; Noe, F., VAMPnets for Deep Learning of Molecular Kinetics. *Nat. Commun.* **2018,** *9*, 11.

193. Pirati, A.; Schoot, J. v.; Troost, K.; Ballegoij, R. v.; Krabbendam, P.; Stoeldraijer, J.; Loopstra, E.; Benschop, J.; Finders, J.; Meiling, H.; Setten, E. v.; Mika, N.; Dredonx, J.; Stamm, U.; Kneer, B.; Thuering, B.; Kaiser, W.; Heil, T.; Migura, S., *The Future of EUV Lithography: Enabling Moore's Law in the Next Decade*. SPIE: 2017; Vol. 10143.

194. Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M., GPU-Accelerated Molecular

Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018,** *58*, 2043-2050.

195. Laio, A.; Parrinello, M., Escaping Free-Energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 12562-12566.

196. Izvekov, S.; Voth, G. A., A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B.* **2005,** *109*, 2469-2473.

197. Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C., The Multiscale Coarse-Graining Method. I. A Rigorous Bridge Between Atomistic and Coarse-Grained Models. *J. Chem. Phys.* **2008,** *128*, 11.

REFERENCES CITED

1. Pecourt, J. M. L.; Peon, J.; Kohler, B., Ultrafast Internal Conversion of Electronically Excited RNA and DNA Nucleosides in Water. *J. Am. Chem. Soc.* **2000,** *122*, 9348-9349.

2. Pecourt, J. M. L.; Peon, J.; Kohler, B., DNA Excited-State Dynamics: Ultrafast Internal Conversion and Vibrational Cooling in a Series of Nucleosides. *J. Am. Chem. Soc.* **2001,** *123*, 10370-10378.

3. Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernàndez, C. E.; Kohler, B., DNA Excited-State Dynamics: From Single Bases to the Double Helix. In *Annu. Rev. Phys. Chem.*, Annual Reviews: Palo Alto, 2009; Vol. 60, pp 217-239.

4. Hahn, S., Structure and Nechanism of the RNA Polymerase II Transcription Machinery. *Nat. Struct. Mol. Biol.* **2004,** *11*, 394-403.

5. Dror, R. O.; Mildorf, T. J.; Hilger, D.; Manglik, A.; Borhani, D. W.; Arlow, D. H.; Philippsen, A.; Villanueva, N.; Yang, Z. Y.; Lerch, M. T.; Hubbell, W. L.; Kobilka, B. K.; Sunahara, R. K.; Shaw, D. E., Structural Basis for Nucleotide Exchange in Heterotrimeric G Proteins. *Science* **2015,** *348*, 1361-1365.

6. Zeng, X. J.; Zhang, L. Y.; Xiao, X. C.; Jiang, Y. Y.; Guo, Y. Z.; Yu, X. Y.; Pu, X. M.; Li, M. L., Unfolding Mechanism of Thrombin-Binding Aptamer Revealed by Molecular Dynamics Simulation and Markov State Model. *Sci Rep* **2016,** *6*, 14.

7. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct Observation of Hierarchical Protein Dynamics. *Science* **2015,** *348*, 578-581.

8. Uversky, V. N., Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **2016,** *291*, 6681-6688.

9. Liao, C.; May, V.; Li, J., PAC1 Receptors: Shapeshifters in Motion. *Journal of molecular neuroscience : MN* **2018**.

10. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

11. Jean, J. M.; Hall, K. B., 2-Aminopurine Fluorescence Quenching and Lifetimes: Role of Base Stacking. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 37-41.

12. Wan, C.; Xia, T.; Becker, H. C.; Zewail, A. H., Ultrafast Unequilibrated Charge Transfer: A New Channel in the Quenching of Fluorescent Biological Probes. *Chem. Phys. Lett.* **2005,** *412*, 158-163.

13. Madhavan Narayanan, G. K., Yangjun Xing, and Robert J. Stanley, Photoinduced Electron Transfer Occurs between 2-Aminopurine and the DNA Nucleic Acid Monophosphates: Results from Cyclic Voltammetry and Fluorescence Quenching. *J. Phys. Chem. B.* **2010,** *114*, 10573-10580.

14. Remington, J. M.; Philip, A. M.; Hariharan, M.; Kohler, B., On the Origin of Multiexponential Fluorescence Decays from 2-Aminopurine-Labeled Dinucleotides. *J. Chem. Phys.* **2016,** *145*, 155101.

15. Evans, K.; Xu, D.; Kim, Y.; Nordlund, T. M., 2-Aminopurine Optical Spectra: Solvent, Pentose Ring, and DNA Helix Melting Dependence. *J. Fluoresc.* **1992,** *2*, 209-216.

16. Evans, K.; Xu, D.; Kim, Y.; Nordlund, T. M., 2-Aminopurine Optical Spectra: Solvent, Pentose Ring, and DNA Helix Melting Dependence. *J. Fluoresc.* **1992,** *2*, 209-216.

17. D. C. Ward, E. R. a. L. S., Fluorescence Studies of Nucleotides and Polynucleotides: I. Formycin, 2-Aminopurine Riboside, 2,6-Diaminopurine Riboside, and Their Derivatives. *J. Biol. Chem.* **1969,** *244*, 1228-1237.

18. Greiner, V. J.; Kovalenko, L.; Humbert, N.; Richert, L.; Birck, C.; Ruff, M.; Zaporozhets, O. A.; Dhe-Paganon, S.; Bronner, C.; Mély, Y., Site-Selective Monitoring of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry* **2015,** *54*, 6012-6020.

19. Voltz, K.; Léonard, J.; Touceda, P. T.; Conyard, J.; Chaker, Z.; Dejaegere, A.; Godet, J.; Mély, Y.; Haacke, S.; Stote, R. H., Quantitative Sampling of Conformational Heterogeneity of a DNA Hairpin Using Molecular Dynamics Simulations and Ultrafast Fluorescence Spectroscopy. *Nucleic Acids Res.* **2016,** *44*, 3408-3419.

20. Tolbert, M.; Morgan, C. E.; Pollum, M.; Crespo-Hernández, C. E.; Li, M. L.; Brewer, G.; Tolbert, B. S., HnRNP Al Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation. *J. Mol. Biol.* **2017,** *429*, 2841-2858.

21. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

22. Jean, J. M.; Krueger, B. P., Structural Fluctuations and Excitation Transfer between Adenine and 2-Aminopurine in
Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B.* **2006,** *110*, 2899-2909.

23. Pörschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

24. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 2315-2328.

25. Pörschke, D., Dynamics of Nucleic-Acid Single-Strand Conformation Changes - Oligoriboadenylic and Polyriboadenylic Acids. *Eur. J. Biochem.* **1973,** *39*, 117-126.

26. Chen, J.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

27. Takaya, T.; Su, C.; de La Harpe, K.; Crespo-Hernàndez, C. E.; Kohler, B., UV Excitation of Single DNA and RNA Strands Produces High Yields of Exciplex States Between Two Stacked Bases. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 10285-10290.

28. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B.* **2012,** *116*, 10266-10274.

29. Zhang, Y. Y.; Li, X. B.; Fleming, A. M.; Dood, J.; Beckstead, A. A.; Orendt, A. M.; Burrows, C. J.; Kohler, B., UV-Induced Proton-Coupled Electron Transfer in Cyclic DNA Miniduplexes. *J. Am. Chem. Soc.* **2016,** *138*, 7395-7401.

30. Somsen, O. J.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Structural Heterogeneity in DNA: Temperature Dependence of 2-Aminopurine Fluorescence in Dinucleotides. *Chemphyschem.* **2005,** *6*, 1622-1627.

31. Kang, H.; Chou, P. J.; Johnson, W. C.; Weller, D.; Huang, S. B.; Summerton, J. E., Stacking Interactions of ApA Analogs with Modified Backbones. *Biopolymers* **1992,** *32*, 1351-1363.

32. Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007,** *92*, 3817-3829.

33. Chen, A. A.; García, A. E., High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *P. Natl. Acad. Sci. USA* **2013,** *110*, 16820-16825.

34. Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J., Reference Simulations of Noncanonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8*, 2506-2520.

35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

36. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, et al *AMBER 2018*, University of California, San Francisco, 2018.

37. Remington, J. M.; McCullagh, M.; Kohler, B., Molecular Dynamics Simulations of 2-Aminopurine-Labeled Dinucleoside Monophosphates Reveal Mutliscale Stacking Kinetics. *J. Phys. Chem. B* **2019,** *123*, 2291-2304.

38. Hayatshahi, H. S.; Henriksen, N. M.; Cheatham, T. E., Consensus Conformations of Dinucleoside Monophosphates Described with Well-Converged Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018,** *14*, 1456-1470.

39. Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R., Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide. *J. Phys. Chem. B.* **2004,** *108* (21), 6582-6594.

40. Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S., Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009,** *131*, 124101.

41. Pande, V. S.; Beauchamp, K.; Bowman, G. R., Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010,** *52*, 99-105.

42. Chodera, J. D.; Noé, F., Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014,** *25*, 135-144.

43. Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F., Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011,** *134*, 174105.

44. Husic, B. E.; Pande, V. S., Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018,** *140*, 2386-2396.

45. Steven P. Davis, M. M., Andreia Williams, and Thomas M. Nordlund, Position Dependence of 2-Aminopurine Spectra in Adenosine Pentadeoxynucleotides. *J. Fluoresc.* **2003,** *13*, 249-259.

46. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B.* **2007,** *111*, 5757-5766.

47. Stark, C. Excited State Processes in Ruthenium(II) Polypyridyl Complexes and Cerium Oxide Nanoparticles. Montana State University, Bozeman, Montana, 2016.

48. Letsinger, R. L.; Wu, T., Use of a Stilbenedicarboxamide Bridge in Stabilizing, Monitoring, and Photochemically Altering Folded Conformations of Oligonucleotides. *J. Am. Chem. Soc.* **1995,** *117*, 7323-7328.

49. Steinfeld, J. I.; Francisco, J. S.; Hase, W. L., *Chemical Kinetics and Dynamics*. Prentice Hall: Upper Saddle River, N.J., 1999.

50. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of Lifetime Distributions from Fluorescence Decays with Application to DNA-Base Analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

51. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. Kluwer Academic/Plenum: New York, 1999.

52. Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham, T. E.; Otyepka, M.; Jurečka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013,** *9*, 2339-2354.

53. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

54. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomicand Molecular Electronic-Structure System. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

55. Gordon, M. S.; Schmidt, M. W., *Advances in Electronic Structure Theory: GAMESS a Decade Later*. Elsevier Science Bv: Amsterdam, 2005; p 1167-1189.

56. Joung, I. S.; Cheatham, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B.* **2008,** *112*, 9020-9041.

57. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

58. Wu, Y. J.; Tepper, H. L.; Voth, G. A., Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006,** *124*, 12.

59. Andrea, T. A.; Swope, W. C.; Andersen, H. C., The Role of Long Ranged Forces in Determining the Structure and Properties of Liquid Water. *J. Chem. Phys.* **1983,** *79*, 4576-4584.

60. Center, O. S., Ohio Supercomputer Center. 1987.

61. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

62. R. J. Gowers, M. L., J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, et al *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulation*, SciPy: Austin, TX, 2016.

63. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.

64. Jones, E.; Oliphant, E.; Peterson, P.; others SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ (accessed 2017-).

65. Blondel, A.; Karplus, M., New Formulation for Derivatives of Torsion Angles and Improper Torsion Angles in Molecular Mechanics: Elimination of Singularities. *J. Comput. Chem.* **1996,** *17*, 1132-1141.

66. Connolly, M. L., Analytical Molecular-Surface Calculation. *J. Appl. Crystallogr.* **1983,** *16*, 548-558.

67. Lu, X. J.; Olson, W. K., 3DNA: a Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic. Acids. Res.* **2003,** *31*, 5108-5121.

68. Diekmann, S., Definitions and Nomenclature of Nucleic-Acid Structure Parameters. *J. Mol. Biol.* **1989,** *205*, 787-791.

69. Lu, X. J.; ElHassan, M. A.; Hunter, C. A., Structure and Conformation of Helical Nucleic Acids: Analysis Program (SCHNAaP). *J. Mol. Biol.* **1997,** *273*, 668-680.

70. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA Dodecamer - Conformation and Dynamics. *P. Natl. Acad. Schi-Biol.* **1981,** *78*, 2179-2183.

71. Schwantes, C. R.; Pande, V. S., Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013,** *9*, 2000-2009.

72. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013,** *139*, 015102.

73. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernàndez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015,** *11*, 5525-5542.

74. S., G.; others *Shapely: Manipulation and Analysis of Geometric Objects*, toblerity.org: 2007—.

75. Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R., Describing Protein Folding Kinetics by Molecular Dynamics Dimulations. 2. Example Spplications to Alanine Dipeptide and Beta-Hairpin Peptide. *J. Phys. Chem. B.* **2004,** *108*, 6582-6594.

76. Gohlke J. R.; Brand L., Fluorescence Probes for Structure. *Annu. Rev. Biochem.* **1972,** *41*, 841-868.

7. T. M. Nordlund, S. A., L. Nilsson, and R. Rigler, A. Graslund, L. W. McLaughlin, Structure and Dynamics of a Fluorescent DNA Oligomer Containing the EcoRI Recognition Sequence: Fluorescence, Molecular Dynamics, and NMR Studies. *Biochemistry* **1989,** *28*, 9095-9103.

78. Jean, J. M.; Hall, K. B., Stacking-Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry* **2004,** *43*, 10277-10284.

79. Bharill, S.; Sarkar, P.; Ballin, J. D.; Gryczynski, I.; Wilson, G. M.; Gryczynski, Z., Fluorescence intensity decays of 2-aminopurine solutions: Lifetime distribution approach. *Anal. Biochem.* **2008,** *377*, 141-149.

80. Torsten Fiebig, C. W., and Ahmed H. Zewail, Femtosecond Charge Transfer Dynamics of a Modified DNA Base: 2-Aminopurine in Complexes with Nucleotides. *Chemphyschem* **2002,** *3*, 781-788.

81. Lobsiger, S.; Blaser, S.; Sinha, R. K.; Frey, H. M.; Leutwyler, S., Switching on the Fluorescence of 2-Aminopurine by Site-Selective Microhydration. *Nat. Chem.* **2014,** *6*, 989-993.

82. Melanie A. O'Neill, J. K. B., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

83. Melanie A. O'Neill, J. K. B., DNA-Mediated Charge Transport Requires Conformational Motion of the DNA Bases: Elimination of Charge Transport in Rigid Glasses at 77 K. *J. Am. Chem. Soc.* **2004,** *126*, 13234-13235.

84. Fogarty, A. C.; Jones, A. C.; Camp, P. J., Extraction of lifetime distributions from fluorescence decays with application to DNA-base analogues. *Phys. Chem. Chem. Phys.* **2011,** *13*, 3819-3830.

85. Siemiarczuk, A.; Wagner, B. D.; Ware, W. R., Comparison of the Maximum-Entropy and Exponential Series Methods for the Recovery of Distributions of Lifetimes From Fluorescence Lifetime Data. *J. Phys. Chem.* **1990,** *94*, 1661-1666.

86. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3 ed.; Springer: Berlin Heidelverg, 2006.

87. Ward, D. C.; Reich, E.; Stryer, L., Fluorescence Studies of Nucleotides and Polynucleotides. I. Formycin, 2-Aminopurineriboside, 2,6-Diaminopurineriboside, and their Derivatives. *J. Biol. Chem.* **1969,** *244*, 1228-1237.

88. Somsen, O. J. G.; Keukens, L. B.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Structural Heterogeneity in DNA: Temperature Dependence of 2-Aminopurine Fluorescence in Dinucleotides. *ChemPhysChem* **2005,** *6*, 1622-1627.

89. Ramreddy, T.; Rao, B. J.; Krishnamoorthy, G., Site-Specific Dynamics of Strands in ss- and dsDNA as Revealed by Time-Domain Fluorescence of 2-Aminopurine. *J. Phys. Chem. B* **2007,** *111*, 5757-5766.

90. Wan, C. Z.; Xia, T. B.; Becker, H. C.; Zewail, A. H., Ultrafast Unequilibrated Charge Transfer: A New Channel in the Quenching of Fluorescent Biological Probes. *Chem. Phys. Lett.* **2005,** *412*, 158-163.

91. Somsen, O. J. G.; Trinkunas, G.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Local Diffusive Dynamics in DNA - A time-Resolved Fluorescence and Molecular-Dynamics Study of Dinucleotides with 2-Aminopurine. *J. Lumin.* **2006,** *119*, 100-104.

92. Becker, W., *Advanced Time-Correlated Single Photon Counting Techniques*. Springer: Berlin Heidelberg, 2005; p 415.

93. Jones, A. C.; Neely, R. K., 2-Aminopurine as a Fluorescent Probe of DNA Conformation and the DNA-Enzyme Interface. *Q. Rev. Biophys.* **2015,** *48*, 244-279.

94. Callis, P. R., Binding Phenomena and Fluorescence Quenching. I: Descriptive Quantum Principles of Fluorescence Quenching Using a Supermolecule Approach. *J. Mol. Struct.* **2014,** *1077*, 14-21.

95. Chen, J. Q.; Zhang, Y. Y.; Kohler, B., Excited States in DNA Strands Investigated by Ultrafast Laser Spectroscopy. In *Photoinduced Phenomena in Nucleic Acids Ii: DNA Fragments and Phenomenological Aspects*, Barbatti, M.; Borin, A. C.; Ullrich, S., Eds. Springer-Verlag Berlin: Berlin, 2015; Vol. 356, pp 39-87.

96. Chen, J. Q.; Kohler, B., Base Stacking in Adenosine Dimers Revealed by Femtosecond Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **2014,** *136*, 6362-6372.

97. Su, C.; Middleton, C. T.; Kohler, B., Base-Stacking Disorder and Excited-State Dynamics in Single-Stranded Adenine Homo-oligonucleotides. *J. Phys. Chem. B* **2012,** *116*, 10266-10274.

98. Narayanan, M.; Kodali, G.; Xing, Y. J.; Stanley, R. J., Photoinduced Electron Transfer Occurs between 2-Aminopurine and the DNA Nucleic Acid Monophosphates: Results from Cyclic Voltammetry and Fluorescence Quenching. *J. Phys. Chem. B* **2010,** *114*, 10573-10580.

99. Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J. K.; Zewail, A. H., Femtosecond Direct Observation of Charge Transfer Between Bases in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97*, 14052-14055.

100. Stuhldreier, M. C.; Temps, F., Ultrafast Photo-Initiated Molecular Quantum Dynamics in the DNA Dinucleotide d(ApG) Revealed by Broadband Transient Absorption Spectroscopy. *Faraday Discuss.* **2013,** *163*, 173-188.

101. Skowron, D. J.; Zhang, Y.; Beckstead, A. A.; Remington, J. M.; Strawn, M.; Kohler, B., Subnanosecond Emission Dynamics of AT DNA Oligonucleotides. *ChemPhysChem* **2016,** *in press*.

102. Liang, J. X.; Nguyen, Q. L.; Matsika, S., Exciplexes and Conical Intersections Lead to Fluorescence Quenching in Pi-Stacked Dimers of 2-Aminopurine With Natural Purine Nucleobases. *Photochem. Photobiol. Sci.* **2013,** *12*, 1387-1400.

103. Kwok, W. M.; Ma, C. S.; Phillips, D. L., Femtosecond Time- and Wavelength-Resolved Fluorescence and Absorption Spectroscopic Study of the Excited States of Adenosine and an Adenine Oligomer. *J. Am. Chem. Soc.* **2006,** *128*, 11894-11905.

104. Stuhldreier, M. C.; Schüler, C.; Kleber, J.; Temps, F., Femtosecond Fluorescence Measurements of the Adenine Dinucleotide: Direct Observation of the Excimer State. In *Ultrafast Phenomena XVII Proceedings of the 17th International Conference,The Silvertree Hotel and Snowmass Conference Center, Snowmass, Colorado, United States, July 18-23, 2010* Chergui, M.; Jonas, D.; Riedle, E.; Schoenlein, R.; Taylor, A., Eds. Oxford University Press: New York, 2011; pp 553-555.

105. Liang, J. X.; Matsika, S., Pathways for Fluorescence Quenching in 2-Aminopurine pi-Stacked with Pyrimidine Nucleobases. *J. Am. Chem. Soc.* **2011,** *133*, 6799-6808.

106. Banyasz, A.; Gustavsson, T.; Onidas, D.; Changenet-Barret, P.; Markovitsi, D.; Improta, R., Multi-Pathway Excited State Relaxation of Adenine Oligomers in Aqueous Solution: A Joint Theoretical and Experimental Study. *Chem. Eur. J.* **2013,** *19*, 3762-3774.

107. Bonnist, E. Y. M.; Jones, A. C., Long-Wavelength Fluorescence from 2-Aminopurine-Nucleobase Dimers in DNA. *ChemPhysChem* **2008,** *9*, 1121-1129.

108. Rist, M.; Wagenknecht, H.-A.; Fiebig, T., Exciton and excimer formation in DNA at room temperature. *ChemPhysChem* **2002,** *3*, 704-707.

109. Kuzmin, M. G.; Soboleva, I. V.; Dolotova, E. V., Transient Exciplex Formation Electron Transfer Mechanism. *Adv. Phys. Chem.* **2011,** *2011*, 813987.

110. Kang, H.; Chou, P. J.; Johnson, W. C.; Weller, D.; Huang, S. B.; Summerton, J. E., Stacking Interactions of ApA Analogs With Modified Backbones. *Biopolymers* **1992,** *32*, 1351-1363.

111. Fornasiero, D.; Kurucsev, T., Vibronic Exciton Interactions - Resolution and Interpretation of the Temperature-Dependent Circular-Dichroism and Absorption-Spectra of ApA and of dApdA. *Eur. J. of Biochem.* **1984,** *143*, 1-7.

112. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Thoery Comput.* **2015,** *11*, 2315-2328.

113. Jafilan, S.; Klein, L.; Hyun, C.; Florian, J., Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B* **2012,** *116*, 3613-3618.

114. Dallmann, A.; Dehmel, L.; Peters, T.; Mugge, C.; Griesinger, C.; Tuma, J.; Ernsting, N. P., 2-Aminopurine Incorporation Perturbs the Dynamics and Structure of DNA. *Angew. Chem. Int. Edit.* **2010,** *49*, 5989-5992.

115. Voltz, K.; Leonard, J.; Touceda, P. T.; Conyard, J.; Chaker, Z.; Dejaegere, A.; Godet, J.; Mely, Y.; Haacke, S.; Stote, R. H., Quantitative Sampling of Conformational Heterogeneity of a DNA Hairpin Using Molecular Dynamics Simulations and Ultrafast Fluorescence Spectroscopy. *Nucleic Acids Res.* **2016,** *44*, 3408-3419.

116. Chodera, J. D.; Noe, F., Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struc. Biol.* **2014,** *25*, 135-144.

117. Albery, W. J.; Bartlett, P. N.; Wilde, C. P.; Darwent, J. R., A General-Model for Dispersed Kinetics in Heterogeneous Systems. *J. Am. Chem. Soc.* **1985,** *107*, 1854-1858.

118. Vix, A.; Lami, H., Protein Fluorescence Decay - Discrete Components or Distribution of Lifetimes - Really No Way Out of the Dilemma. *Biophys. J.* **1995,** *68*, 1145-1151.

119. Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L., Nature OF Biological Electron-Transfer. *Nature* **1992,** *355*, 796-802.

120. Arkin, M. R.; Stemp, E. D. A.; Holmlin, R. E.; Barton, J. K.; Hormann, A.; Olson, E. J. C.; Barbara, P. F., Rates of DNA-mediated Electron Transfer Between Metallointercalators. *Science* **1996,** *273*, 475-480.

121. Murphy, C. J.; Arkin, M. R.; Jenkins, Y.; Ghatlia, N. D.; Bossmann, S. H.; Turro, N. J.; Barton, J. K., Long-Range Photoinduced Electron-Transfer Through a DNA Helix. *Science* **1993,** *262*, 1025-1029.

122. Lewis, F. D.; Wu, T. F.; Zhang, Y. F.; Letsinger, R. L.; Greenfield, S. R.; Wasielewski, M. R., Distance-Dependent Electron Transfer in DNA Hairpins. *Science* **1997,** *277*, 673-676.

123. Tezcan, F. A.; Crane, B. R.; Winkler, J. R.; Gray, H. B., Electron Tunneling in Protein Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 5002-5006.

124. Langen, R.; Chang, I. J.; Germanas, J. P.; Richards, J. H.; Winkler, J. R.; Gray, H. B., Electron-Tunneling in Proteins - Coupling Through a Beta-Strand. *Science* **1995,** *268*, 1733-1735.

125. Lin, J. P.; Balabin, I. A.; Beratan, D. N., The Nature of Aqueous Tunneling Pathways Between Electron-Transfer Proteins. *Science* **2005,** *310*, 1311-1313.

126. Ponce, A.; Gray, H. B.; Winkler, J. R., Electron Tunneling Through Water: Oxidative Quenching of Electronically Excited Ru(tpy)(2)(2+) (tpy=2,2 ': 6,2 "-terpyridine) by Ferric Ions in Aqueous Glasses at 77 K. *J. Am. Chem. Soc.* **2000,** *122*, 8187-8191.

127. Tavernier, H. L.; Fayer, M. D., Distance Dependence of Electron Transfer in DNA: The Role Of The Reorganization Energy and Free Energy. *J. Phys. Chem. B* **2000,** *104*, 11541-11550.

128. Norberg, J.; Nilsson, L., Solvent Influence on Base Stacking. *Biophys. J.* **1998,** *74*, 394-402.

129. O'Neill, M. A.; Dohno, C.; Barton, J. K., Direct Chemical Evidence for Charge Transfer between Photoexcited 2-Aminopurine and Guanine in Duplex DNA. *J. Am. Chem. Soc.* **2004,** *126*, 1316-1317.

130. O'Neill, M. A.; Barton, J. K., DNA-Mediated Charge Transport Requires Conformational Motion of the DNA Bases: Elimination of Charge Transport in Rigid Glasses at 77 K. *J. Am. Chem. Soc.* **2004,** *126*, 13234-13235.

131. O'Neill, M. A.; Barton, J. K., DNA Charge Transport: Conformationally Gated Hopping through Stacked Domains. *J. Am. Chem. Soc.* **2004,** *126*, 11471-11483.

132. Porschke, D., Molecular States in Single-Stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978,** *17*, 315-323.

133. Heyn, M. P.; Nicola, C. U.; Schwarz, G., Kinetics of Base-Stacking Reaction of N6,N6-Dimethyladenosine - Ultrasonic-Absorption and Dispersion Study. *J. Phys. Chem.* **1977,** *81*, 1611-1617.

134. Jean, J. M.; Krueger, B. P., Structural Fluctuations And Excitation Transfer Between Adenine and 2-Aminopurine in Single-Stranded Deoxytrinucleotides. *J. Phys. Chem. B* **2006,** *110*, 2899-2909.

135. Lakowicz, J. R.; Wiczk, W.; Gryczynski, I.; Szmacinski, H.; Johnson, M. L., Influence Of End-To-End Diffusion On Intramolecular Energy-Transfer As Observed By Frequency-Domain Fluorometry. *Biophys. Chem.* **1990,** *38*, 99-109.

136. Rachofsky, E. L.; Osman, R.; Ross, A. J. B., Probing Structure and Dynamics of DNA with 2-Aminopurine: Effects of Local Environment on Fluorescence. *Biochemistry* **2001,** *40*, 946-956.

137. Larsen, O. F. A.; van Stokkum, I. H. M.; Gobets, B.; van Grondelle, R.; van Amerongen, H., Probing the Structure and Dynamics of a DNA Hairpin by Ultrafast Quenching and Fluorescence Depolarization. *Biophys. J.* **2001,** *81*, 1115-1126.

138. Rai, P.; Cole, T. D.; Thompson, E.; Millar, D. P.; Linn, S., Steady-State and Time-Resolved Fluorescence Studies Indicate an Unusual Conformation of 2-Aminopurine Within ATAT and TATA Duplex DNA Sequences. *Nucleic Acids Res. Acids Research* **2003,** *31*, 2323-2332.

139. Somsen, O. J.; van Hoek, A.; van Amerongen, H., Fluorescence Quenching of 2-Aminopurine in Dinucleotides. *Chem. Phys. Lett.* **2005,** *402*, 61-65.

140. Pörschke, D.; Eggers, F., Thermodynamics and Kinetics of Base-Stacking Interactions. *Eur. J. Biochem.* **1972,** *26*, 490-498.

141. Heyn, M. P.; Nicola, C. U.; Schwarz, G., Kinetics of Base-Stacking Reaction of N6,N6-Dimethyladenosine - Ultrasonic-Absorption and Dispersion Study. *J. Phys. Chem.* **1977,** *81*, 1611-1617.

142. Pinamonti, G.; Zhao, J. B.; Condon, D. E.; Paul, F.; Noé, F.; Turner, D. H.; Bussi, G., Predicting the Kinetics of RNA Oligonucleotides Using Markov State Models. *J. Chem. Theory Comput.* **2017,** *13*, 926-934.

143. van Mourik, T.; Hogan, S. W. L., DNA Base Stacking Involving Adenine and 2-Aminopurine. *Struct. Chem.* **2016,** *27*, 145-158.

144. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C., Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007,** *126*, 155101.

145. Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. In *Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Bowman, G. R.; Pande, V. S.; Noé, F., Eds. Springer-Verlag Berlin, Heidelberger Platz 3, D-14197 Berlin, Germany: 2014; Vol. 797.

146. Wales, D. J., *Energy Landscapes: With Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press: The Edinburgh Building, Cambridge CB2 2RU, UK, 2003; p 671.

147. Rao, F.; Caflisch, A., The Protein Folding Network. *J. Mol. Biol.* **2004,** *342*, 299-306.

148. Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C., Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.* **2007,** *126*, 155102.

149. Wales, D. J.; Bogdan, T. V., Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B.* **2006,** *110*, 20765-20776.

150. Noé, F.; Wu, H.; Prinz, J. H.; Plattner, N., Projected and Hidden Markov Models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* **2013,** *139*, 184114.

151. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. Recommendations 1982. *Eur. J. Biochem.* **1983,** *131*, 9-15.

152. Kube, S.; Weber, M., A Coarse Graining Method for the Identification of Transition Rates Between Molecular Conformations. *J. Chem. Phys.* **2007,** *126*, 024103.

153. Keller, B. G.; Prinz, J. H.; Noé, F., Markov Models and Dynamical Fingerprints: Unraveling the Complexity of Molecular Kinetics. *Chem. Phys.* **2012,** *396*, 92-107.

154. Rose, I. A.; Hanson, K. R.; Wilkinson, K. D.; Wimmer, M. J., A Suggestion for Naming Faces of Ring Compounds. *P. Natl. Acad. Schi-Biol.* **1980,** *77*, 2439-2441.

155. Sundaralingam, M.; Westhof, E., *The Nature of the Mobility of the Sugar and its Effects on the Dynamics and Functions of RNA and DNA*. 1981; p 301-326.

156. Jafilan, S.; Klein, L.; Hyun, C.; Florián, J., Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B.* **2012,** *116*, 3613-3618.

157. Fornasiero, D.; Kurucsev, T., Vibronic Exciton Interactions - Resolution and Interpretation of the Temperature-Dependent Circular-Dichroism and Absorption-Spectra of ApA and of dApdA. *Eur. J. Biochem.* **1984,** *143*, 1-7.

158. Neidle, S., *Principles of Nucleic Acid Structure*. Elsevier Academic Press Inc: San Diego, 2008; p 1-289.

159. Newcomb, L. F.; Gellman, S. H., Aromatic Stacking Interactions in Aqueous-Solution - Evidence that Neither Classical Hydrophobic Effects nor Dispersion Forces are Important. *J. Am. Chem. Soc.* **1994,** *116*, 4993-4994.

160. Šponer, J.; Riley, K. E.; Hobza, P., Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases. *Physical Chemistry Chemical Physics* **2008,** *10*, 2595-2610.

161. Mak, C. H., Unraveling Base Stacking Driving Forces in DNA. *J. Phys. Chem. B.* **2016,** *120*, 6010-6020.

162. Brennan, M.; Kustin, K., Ultrasonic Attenuation in Basic Purine Solutions. *J. Phys. Chem.* **1972,** *76*, 2838-2841.

163. Engel, J. D.; Hippel, P. H. V., Effects of Methylation on Stability of Nucleic-Acid Conformations - Studies at Monomer Level. *Biochemistry* **1974,** *13*, 4143-4158.

164. Garland, F.; Christian, S. D., Thermodynamic and Kinetic-Model of Sequential Nucleoside Base Aggregation in Aqueous-Solution. *J. Phys. Chem.* **1975,** *79*, 1247-1252.

165. Hemmes, P. R.; Oppenheimer, L.; Jordan, F., Ultrasonic Relaxation Evaluation of Thermodynamics of Syn-Anti Glycosidic Isomerization in Adenosine. *J. Am. Chem. Soc.* **1974,** *96*, 6023-6026.

166. Rhodes, L. M.; Schimmel, P. R., Nanosecond Relaxation Processes in Aqueous Mononucleoside Solutions. *Biochemistry* **1971,** *10*, 4423-4433.

167. Polacek, R.; Buckin, V. A.; Eggers, F.; Kaatze, U., Kinetics of Base-Stacking Interactions and Proton Exchange of 6-Methylpurine Aqueous Solutions. *J. Phys. Chem. A.* **2004,** *108*, 1867-1872.

168. Hoffman, G. W., Nanosecond Temperature-Jump Apparatus. *Rev. Sci. Instrum.* **1971,** *42*, 1643-1647.

169. Warshaw, M. M.; Tinoco, I., Optical Properties of 16 Dinucleoside Phosphates. *J. Mol. Biol.* **1966,** *20*, 29-38.

170. Lee, B.; Richards, F. M., Interpretation of Protein Structures - Estimation of Static Accessibility. *J. Mol. Biol.* **1971,** *55*, 379-400.

171. Wallqvist, A.; Berne, B. J., Molecular-Dynamics Study of the Dependence of Water Solvation Free-Energy on Solute Curvature and Surface-Area. *J. Phys. Chem.* **1995,** *99*, 2885-2892.

172. Somsen, O. J. G.; Trinkunas, G.; de Keijzer, M. N.; van Hoek, A.; van Amerongen, H., Local Diffusive Dynamics in DNA - A Time-Resolved Fluorescence and Molecular-Dynamics Study of Dinucleotides with 2-Aminopurine. *J. Lumin.* **2006,** *119*, 100-104.

173. Edholm, O.; Blomberg, C., Stretched Exponentials and Barrier Distributions. *Chem. Phys.* **2000,** *252*, 221-225.

174. Callis, P. R., Predicting Fluorescence Lifetimes and Spectra of Biopolymers. In *Methods in Enzymology*, Johnson, M. L.; Brand, L., Eds. Elsevier Academic Press Inc: San Diego, 2011; Vol. 487, pp 1-38.

175. Blumberger, J., Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **2015,** *115*, 11191-11238.

176. Burshtein, A. I., Non-Markovian Theories of Transfer Reactions in Luminescence and Chemiluminescence and Photo- and Electrochemistry. *Adv. Chem. Phys.* **2004,** *129*, 105-418.

177. Min, W.; Luo, G. B.; Cherayil, B. J.; Kou, S. C.; Xie, X. S., Observation of a Power-Law Memory Kernel for Fluctuations Within a Single Protein Molecule. *Phys. Rev. Lett.* **2005,** *94*, 198302.

178. Suárez, E.; Adelman, J. L.; Zuckerman, D. M., Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. *J. Chem. Theory Comput.* **2016,** *12*, 3473-3481.

179. Sakakibara, Y.; Abeysirigunawardena, S. C.; Duc, A.-C. E.; Dremann, D. N.; Chow, C. S., Ligand- and pH-Induced Conformational Changes of RNA Domain Helix 69 Revealed by 2-Aminopurine Fluorescence. *Angew. Chem., Int. Ed.* **2012,** *51*, 12095-12098.

180. Lee, H.-W.; Briggs, K. T.; Marino, J. P., Dissecting Structural Transitions in the HIV-1 Dimerization Initiation Site RNA using 2-Aminopurine Fluorescence. *Methods* **2009,** *49*, 118-127.

181. Liang, C.; Li, X.; Rong, L.; Inouye, P.; Quan, Y.; Kleiman, L.; Wainberg, M. A., The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *Journal of Virology* **1997,** *71*, 5750-5757.

182. Sinsheimer, R. L., A Single-Stranded Deoxyribonucleic Acid from Bacteriophage φX174. *J. Mol. Biol.* **1959,** *1*, 43.

183. Zou, Y.; Liu, Y.; Wu, X.; Shell, S. M., Functions of Human Replication Protein A (RPA): From DNA Replication to DNA Damage and Stress Responses. *J. Cell. Physiol.* **2006,** *208*, 267-273.

184. Avilov, S. V.; Piemont, E.; Shvadchak, V.; de Rocquigny, H.; Mely, Y., Probing Dynamics of HIV-1 Nucleocapsid Protein/Target Hexanucleotide Complexes by 2-Aminopurine. *Nucleic Acids Res.* **2008,** *36*, 885-96.

185. Stivers, J. T., 2-Aminopurine Fluorescence Studies of Base Stacking Interactions at Abasic Sites in DNA: Metal-Ion and Base Sequence Effects. *Nucleic Acids Res.* **1998,** *26*, 3837-3844.

186. Nordlund, D.-G. X. a. T. M., Sequence Dependence of Energy Transfer in DNA Oligonucleotides. *Biophys. J.* **2000,** *78*, 1042-1058.

187. Yang, K.; Stanley, R. J., Differential Distortion of Substrate Occurs When It Binds to DNA Photolyase: A 2-Aminopurine Study. *Biochemistry* **2006,** *45*, 11239-11245.

188. Sultan, M. M.; Pande, V. S., tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017,** *13*, 2440-2447.

189. P.-O.Lycksell, A. G., F.Claesens, L.W.McLaughlin, U.Larsson, and R.Rigler, Base Pair Opening Dynamics of a 2-Aminopurine Substituted Eco RI Restriction Sequence and its Unsubstituted Counterpart in Oligonucleotides. *Nucleic Acids Res.* **1987,** *15*, 9011-9025.

190. Ballin, J. D.; Bharill, S.; Fialcowitz-White, E. J.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M., Site-Specific Variations in RNA Folding Thermodynamics Visualized by 2-Aminopurine Fluorescence. *Biochemistry* **2007,** *46*, 13948-13960.

191. Nuske, F.; Keller, B. G.; Pérez-Hernàndez, G.; Mey, A.; Noe, F., Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014,** *10*, 1739-1752.

192. Mardt, A.; Pasquali, L.; Wu, H.; Noe, F., VAMPnets for Deep Learning of Molecular Kinetics. *Nat. Commun.* **2018,** *9*, 11.

193. Pirati, A.; Schoot, J. v.; Troost, K.; Ballegoij, R. v.; Krabbendam, P.; Stoeldraijer, J.; Loopstra, E.; Benschop, J.; Finders, J.; Meiling, H.; Setten, E. v.; Mika, N.; Dredonx, J.; Stamm, U.; Kneer, B.; Thuering, B.; Kaiser, W.; Heil, T.; Migura, S., *The Future of EUV Lithography: Enabling Moore's Law in the Next Decade*. SPIE: 2017; Vol. 10143.

194. Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M., GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018,** *58*, 2043-2050.

195. Laio, A.; Parrinello, M., Escaping Free-Energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99*, 12562-12566.

196. Izvekov, S.; Voth, G. A., A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B.* **2005,** *109*, 2469-2473.

197. Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C., The Multiscale Coarse-Graining Method. I. A Rigorous Bridge Between Atomistic and Coarse-Grained Models. *J. Chem. Phys.* **2008,** *128*, 11.

APPENDICES

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER THREE, ON THE ORIGIN OF MULTIEXPONENTIAL FLUORESCENCE DECAYS FROM 2-AMINOPURINE LABELED DINUCLEOTIDES

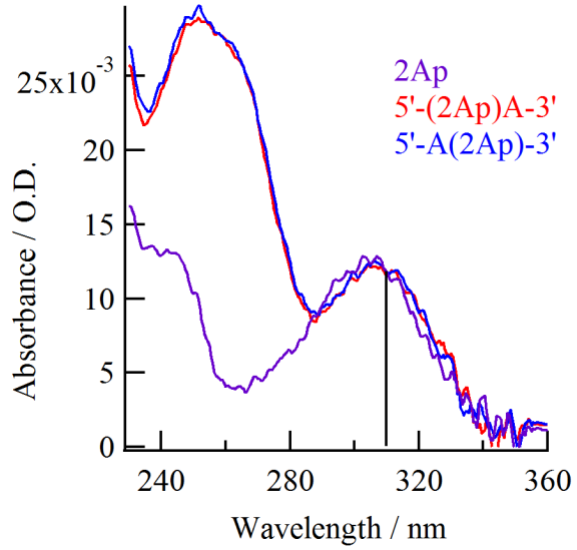

Figure A.1.1**.** UV-visible absorption spectra of absorbance matched solutions of 5'-(2Ap)A-3', 5'-A(2Ap)-3', and 2Ap. The main difference between the dinucleotides and 2Ap is apparent at 260 nm where A strongly absorbs. While the peak of 2Ap near 305 nm appears to be a bit higher than for the dinucleotide samples, the slight redshift of the absorption by the latter brings the absorbance at 310 nm, vertical line, to equivalent values.

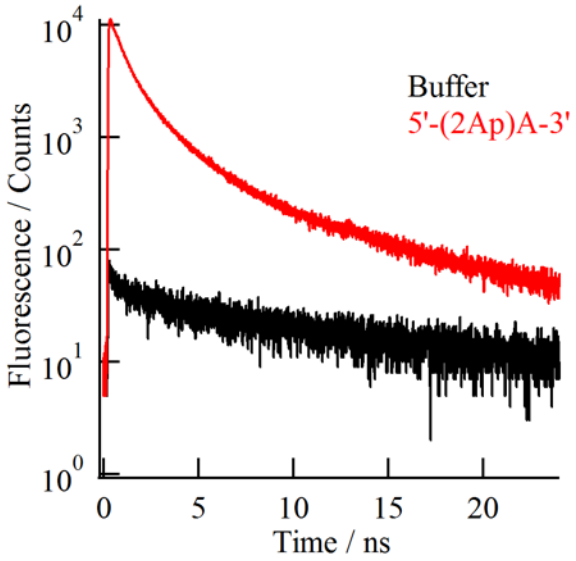

Figure A.1.2. The signal from the tris-HCl buffer (bottom) is compared with the signal from 5'-(2Ap)A-3' (top) with λ_{ex} = 310 nm and λ_{em} = 370 nm. The absence of a short lifetime component (τ < 100 ps) in the buffer-only signal is an indication that scattered pump photons contribute negligibly to the final decay signals.

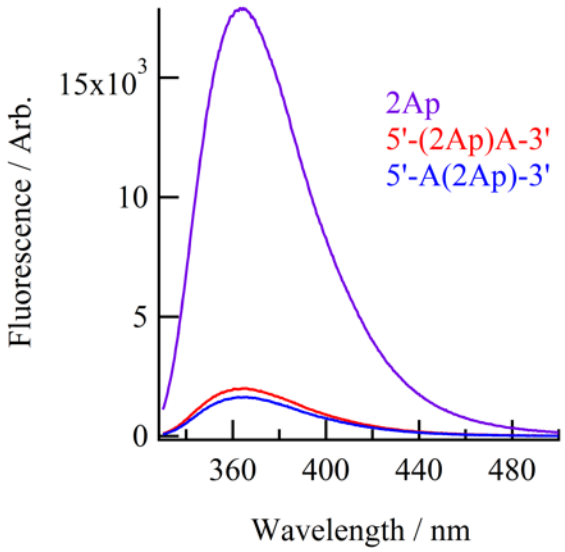

Figure A.1.3. The corrected steady-state fluorescence spectra of the three samples recorded with λ_{ex} = 310 nm. The area under each curve was used to estimate the steady-state quantum yields of the dinucleotides. The ordering on the legend corresponds to the relative peak intensity.

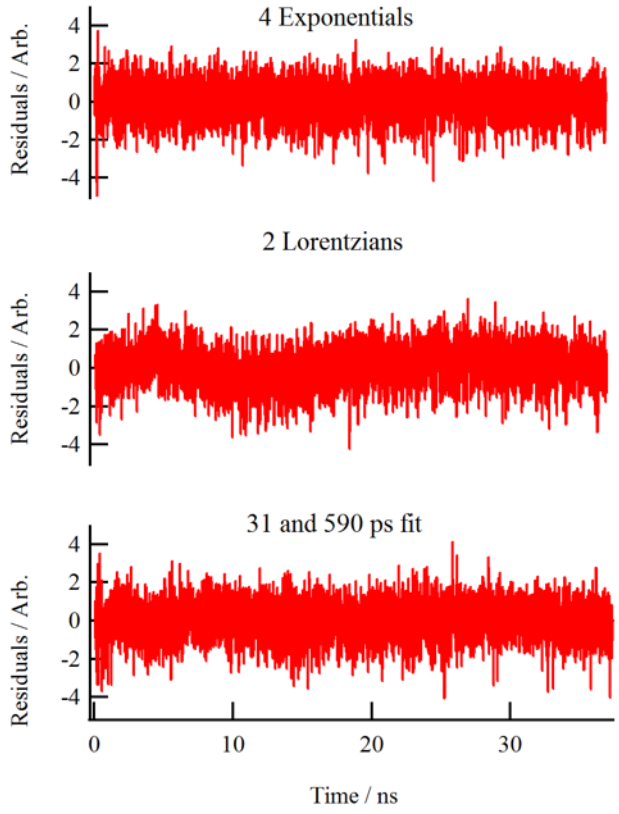

Figure A.1.4. The weighted residuals (equation 5) for a discrete four exponential model vs. the bi-Lorentzian rate distribution model from fits to 5'-(2Ap)A-3' at 20 °C. The bi-Lorentzian model was globally fit between the 5' and 3'-labeled dinucleotides with the μ and Δ of the major peak as global parameters. In the bottom panel, the data was fit using a four exponential model with two of the lifetimes fixed at 31 ps and 590 ps lifetimes, the values reported by O. J. G. Somsen et al. (*ChemPhysChem* **2005,** *6*, 1622). The slightly non-random residuals in the first few ns indicate that the short 31 ps lifetime may not be present in our data; instead we find a shortest decay component of 320 ps.

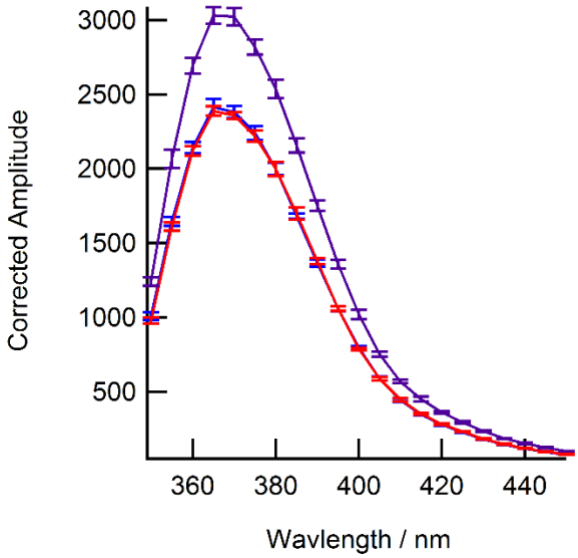

Figure A.1.5. The deconvoluted signal ("Corrected Amplitude") obtained by evaluating Eq. (3) in the main text at $t = 0$ for fits made with a bi-Lorentzian lifetime distribution. Results are shown for 2Ap (top), 5'-(2Ap)A-3' (middle), and 5'-A(2Ap)-3' (bottom, obscured by the 5'-(2Ap)A-3') curve), samples are shown in purple, red, and blue, respectively. The error bars are the 95% confidence intervals obtained from support plane analysis. Values at 370 nm were used to calculate the equilibrium constants using eq. 10 in the main text.

Table A.1.1. Temperature-dependent fitting parameters for the 5'-(2Ap)A-3' sample are shown for the two Lorentzian rate distribution model (eqs. 3 and 4). The μ_1 and Δ_1 were globally fit for the 3' and 5' labels due to the similarity in the short time (< 5 ns) kinetics. The observed narrowing can be seen in the value of Δ_1 at higher temperatures. Fit uncertainties were used to round all table values to an appropriate number of significant figures, except for the Δ_2 and μ_2 for which errors via support plane method were not found. As can be seen from the χ^2 values, the fits are slightly worse at high temperatures, indicating the lifetime distribution picture may be inappropriate at higher temperatures.

T / °C	A_1	μ_1 / ns	Δ_1 / ns	A_2	μ_2 / ns	Δ_2 / ns	χ^2
10	13900	0.72	0.77	4600	0.79	23.8	1.28
15	13900	0.71	0.76	4499	0.74	26.2	1.40
20	14000	0.70	0.76	4300	0.76	29.0	1.47
25	14200	0.71	0.72	4200	0.81	30.8	1.62
30	13800	0.73	0.70	3800	0.73	31.7	1.87
35	13900	0.74	0.64	4000	0.74	34.3	2.09
40	13600	0.75	0.61	4000	0.77	37.0	2.21
45	13100	0.79	0.55	4200	0.73	37.9	2.35
50	13200	0.78	0.49	4200	0.74	38.8	2.40
55	12500	0.82	0.44	4100	0.77	37.5	2.41
60	12200	0.84	0.38	4300	0.74	38.1	2.54

Table A.1.2. Temperature-dependent fitting parameters for the 5'-A(2Ap)-3' sample are shown for the two Lorentzian rate distribution model. The μ_1 and Δ_1 were globally fit for the 3' and 5' labels due to the similarity in the short time (< 5 ns) kinetics. Fit uncertainties were used to round all table values to an appropriate number of significant figures, except for the Δ_2 and μ_2 for which errors via support plane method were not found. As can be seen from the χ^2 values, the fits becomes slightly worse with increasing temperature, indicating that the lifetime distribution picture may be inappropriate at higher temperatures.

T / °C	A_1	μ_1 / ns	Δ_1 / ns	A_2	μ_2 / ns	Δ_2 / ns	χ^2
10	12600	0.72	0.77	6500	1.17	15.86	1.28
15	12700	0.71	0.76	6300	1.11	14.74	1.40
20	12700	0.70	0.76	5800	1.03	15.01	1.47
25	12900	0.71	0.72	5700	1.00	14.38	1.62
30	12900	0.73	0.70	5300	0.95	14.41	1.87
35	12700	0.74	0.64	5100	0.92	13.46	2.09
40	12500	0.75	0.61	4700	0.92	12.92	2.21
45	12200	0.79	0.55	4500	0.85	11.90	2.35
50	12200	0.78	0.49	4500	1.80	11.66	2.40
55	12000	0.82	0.44	4400	0.77	12.17	2.41
60	11600	0.84	0.38	4300	0.72	11.22	2.54

Table A.1.3. Temperature-dependent fitting parameters for multiexponential fits (Eq. 1 and 2) to the 5'-(2Ap)A-3' fluorescence decays. At lower temperatures, four exponential components were required, while the higher temperature fits required just three components.

T / °C	A_1	τ_1 / ns	A_2	τ_2 / ns	A_3	τ_3 / ns	A_4	τ_4 / ns	χ^2
10	3000	0.26	8500	0.84	3000	2.71	560	9.8	0.9739
15	2500	0.25	8400	0.76	3410	2.49	510	10.0	0.9945
20	3700	0.32	7500	0.80	3470	2.40	460	10.4	0.9692
25	----	----	9307	0.56	4830	1.94	520	9.4	1.0189
30	----	----	8700	0.52	5400	1.77	500	9.3	0.9660
35	----	----	8400	0.49	6040	1.62	510	8.9	0.9895
40	----	----	8100	0.47	6380	1.54	480	8.8	0.9919
45	----	----	7400	0.45	7220	1.40	480	8.6	0.9573
50	----	----	6900	0.43	7820	1.32	470	8.4	0.9550
55	----	----	6300	0.43	8100	1.26	440	8.2	0.9625
60	----	----	5700	0.39	9000	1.18	430	7.9	0.9511

Table A.1.4**.** Temperature-dependent fitting parameters for the four exponential model applied to the 5'-A(2Ap)-3' sample.

T / °C	A_1	τ_1 / ns	A_2	τ_2 / ns	A_3	τ_3 / ns	A_4	τ_4 / ns	χ^2_{red}
10	3500	0.46	7900	0.81	3220	2.85	1130	8.26	1.0340
15	8300	0.55	3800	1.11	3020	3.18	840	9.1	1.0307
20	8200	0.51	3800	1.04	3350	2.87	880	8.7	1.0125
25	8800	0.49	4000	1.22	3180	3.04	690	9.3	1.0081
30	8900	0.47	3800	1.19	3550	2.77	680	9.2	0.9866
35	8500	0.45	4000	1.17	3620	2.62	640	9.2	0.9948
40	8500	0.45	4700	1.33	2940	2.73	540	9.6	1.0028
45	8200	0.41	4300	1.06	4550	2.21	630	8.8	1.0068
50	8200	0.40	4900	1.14	4070	2.21	580	8.8	0.9587
55	7100	0.38	5300	1.02	4600	2.04	570	8.7	1.0308
60	7900	0.40	4700	1.12	5000	1.86	610	8.1	1.0373

Table A.1.5. Temperature-dependent lifetimes from single exponential fits to fluorescence decays of the 2Ap monomer (λ_{exc} = 310 nm, λ_{em} = 370 nm). The value of reduced chi-squared (χ^2) for each fit is also shown.

T / °C	τ / *ns*	χ^2
10	10.43	1.0868
15	10.27	1.0688
20	10.12	1.1231
25	9.89	1.033
30	9.81	1.0045
35	9.57	1.0466
40	9.35	1.0092
45	9.11	1.0607
50	8.86	1.0077
55	8.60	1.0391
60	8.30	1.0369

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER FOUR, MOLECULAR DYNAMICS SIMULATIONS OF 2-AMINOPURINE-LABELED DINUCLEOSIDE MONOPHOSPHATES REVEAL MULTISCALE STACKING KINETICS

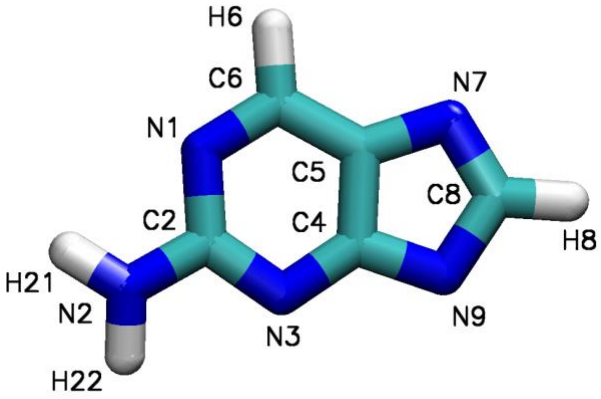

Figure A.2.1. Atom names for the 2Ap residue used in the simulations.

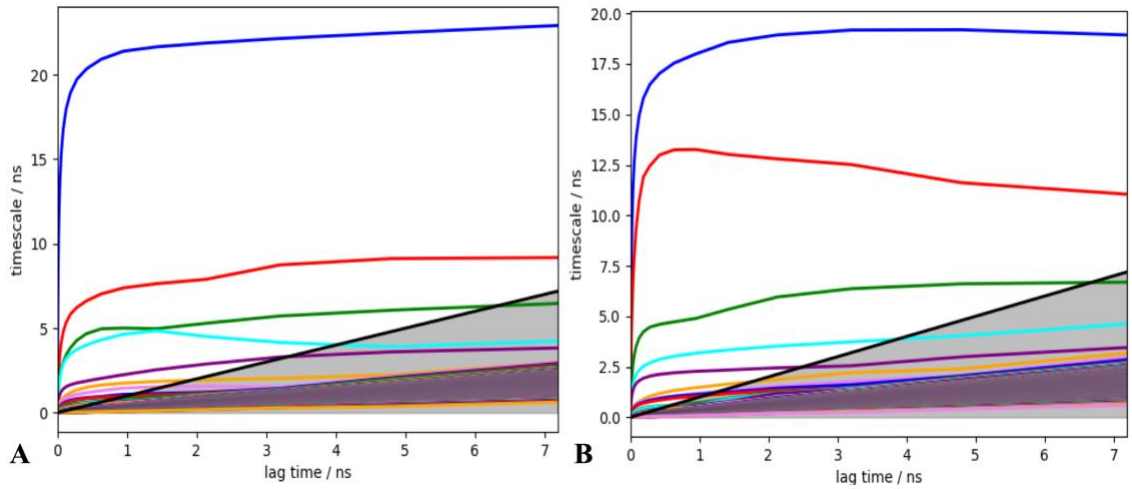

Figure A.2.2. Implied timescales for validation of MSM construction on ~2000 microstates in the 8 dimensional coordinate space. Figures A and B are the 5'-A(Ap)-3' and 5'-(Ap)A-3' samples, respectively. These figures show the existence of at least 6 timescales at a 500 ps lag time, and relatively little dependence of the timescales on lag time after that. The line $y = x$ is also shown to clearly distinguish time scales which are not significant due to having a time constant shorter than the lag time.

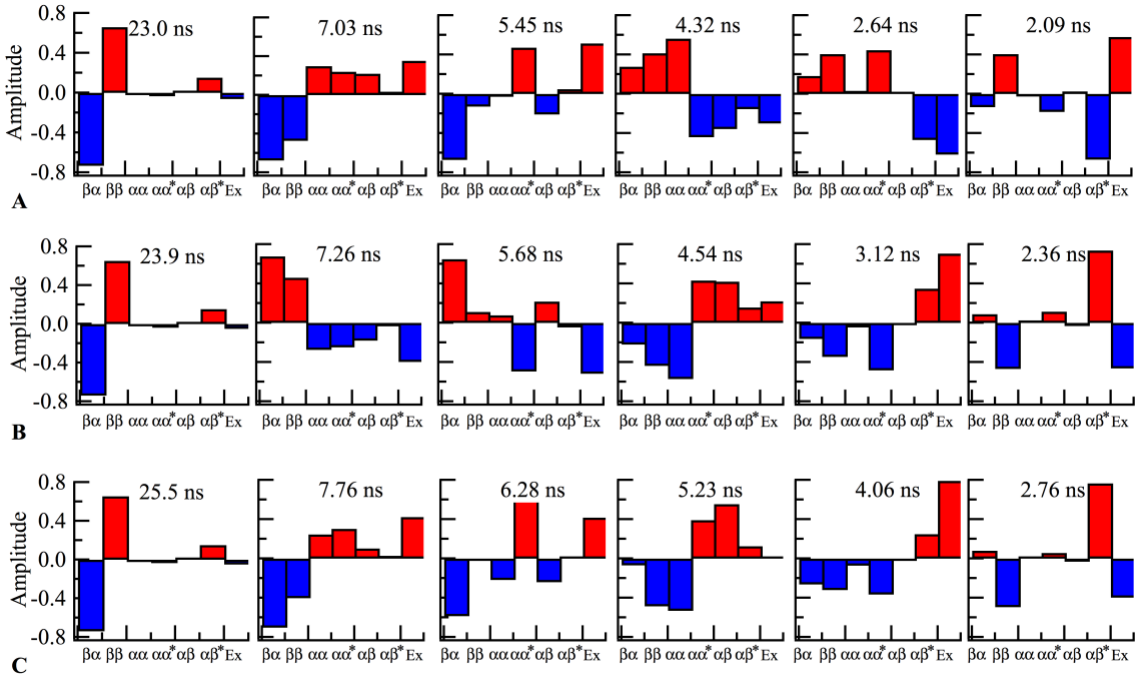

Figure A.2.3 Comparison of the 7 state HMSMs constructed for the 5'-A(2Ap)-3' sample at increasing lag times, 0.50, 0.75, and 1.25 ns from A to C. The amplitudes of each eigenvector for each model are plotted with their associated timescales, which lengthen slightly with increasing lag time. Besides irrelevant changes in sign, the eigenvectors are remarkably similar with only slight differences in the eigenvectors corresponding to the short (< 6 ns) timescales when the longer 0.75 and 1.25 ns lag times are used.

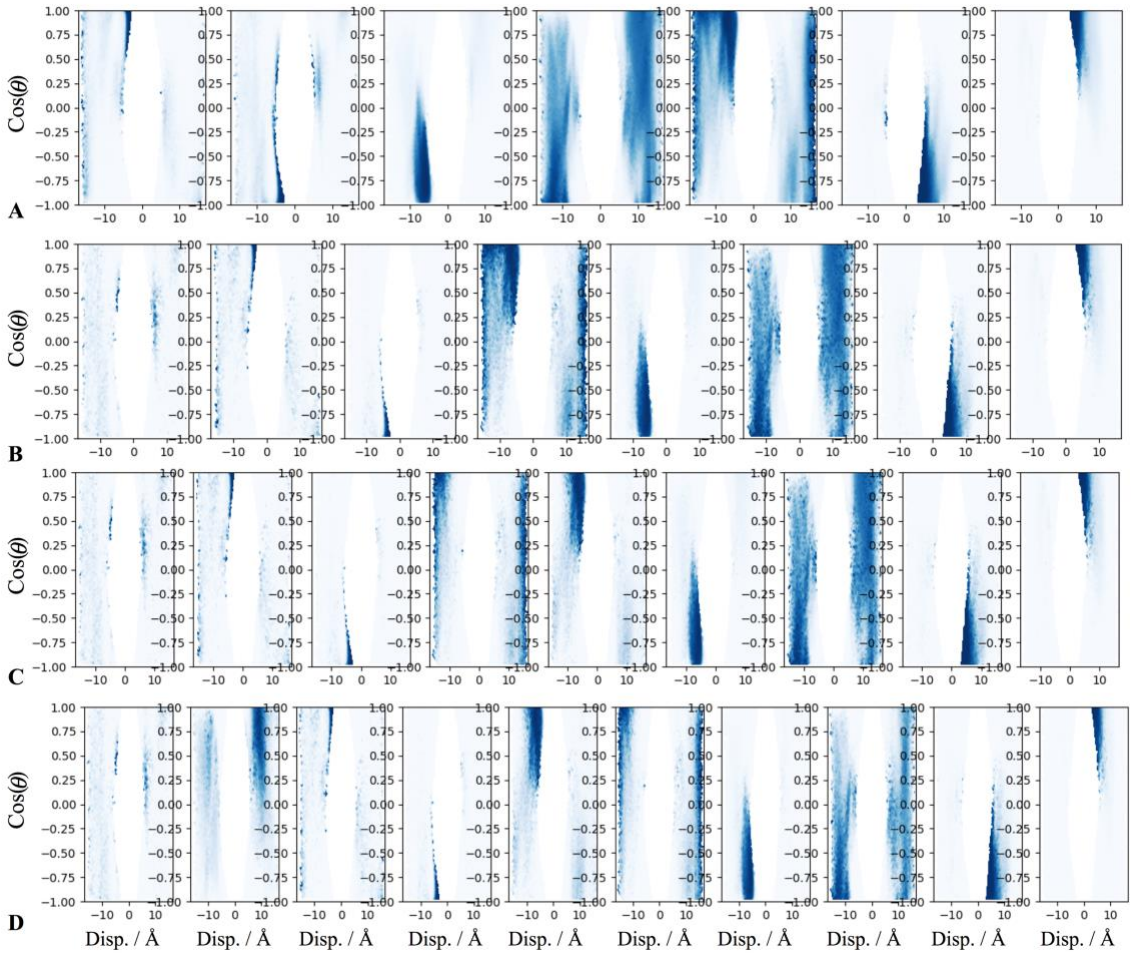

Figure A.2.4. HMSM fuzzy memberships for the 5'-A(2Ap)-3' system with increasing number of states from A to D (7, 8, 9, and 10). States beyond the first 7 obtained from the coarse graining algorithm partition the unstacked region more finely and a T-shaped conformation is identified in the 10 state model.

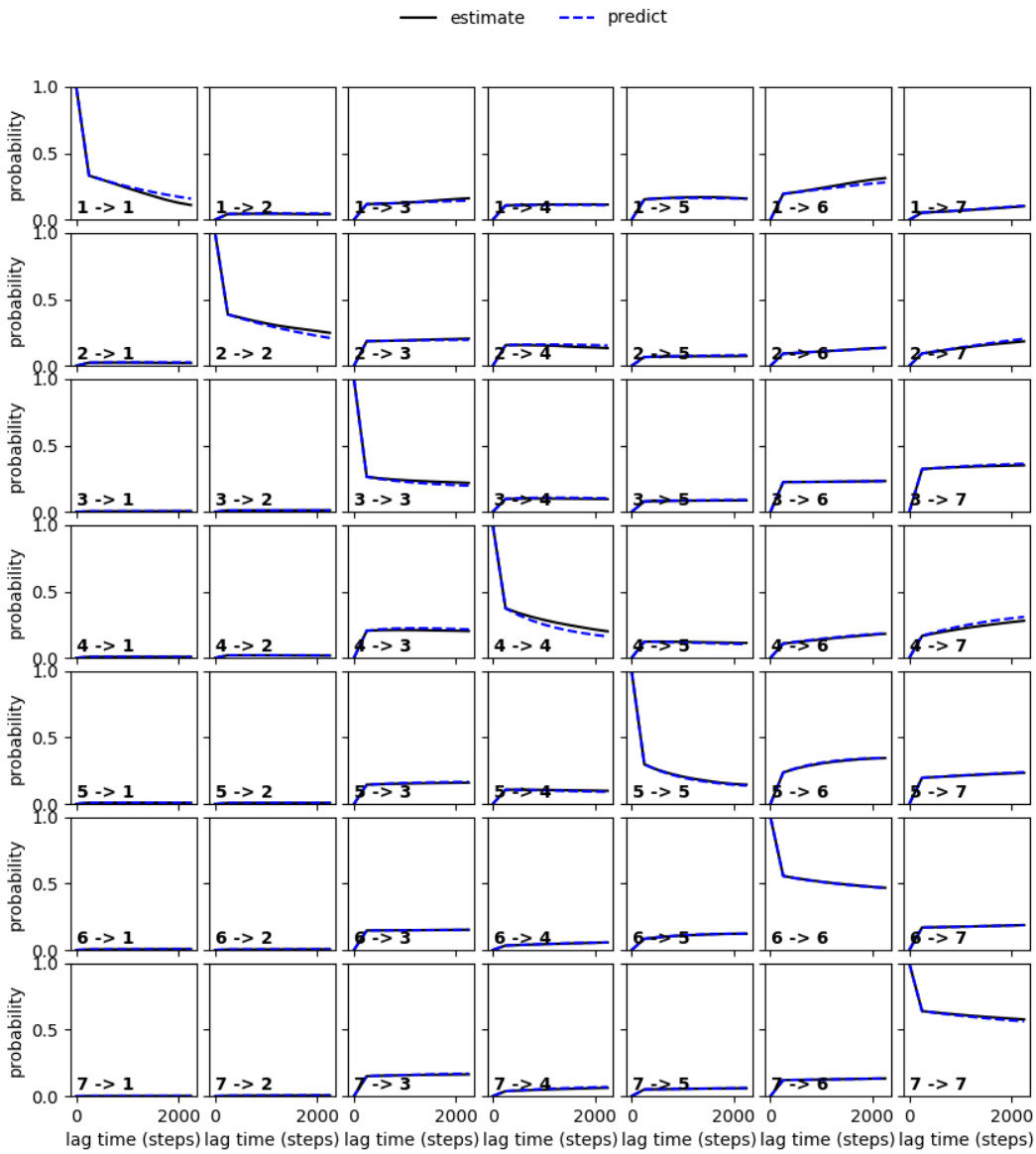

Figure A.2.5. The Chapman-Kolmogorov test for coarse graining the ~2000 state MSM into a 7 state HMSM for the 5'-A(2Ap)-3' system. The curves labeled $i \rightarrow j$ are the conditional probabilities that an initial state i is found in state j at a later time. The blue curves are predicted by the HMSM with a 500 ps (250 step) lag time, while the black curves are estimated with longer lag times. The similarities between these curves show that the 7 HMSM is indeed a good model for the system.

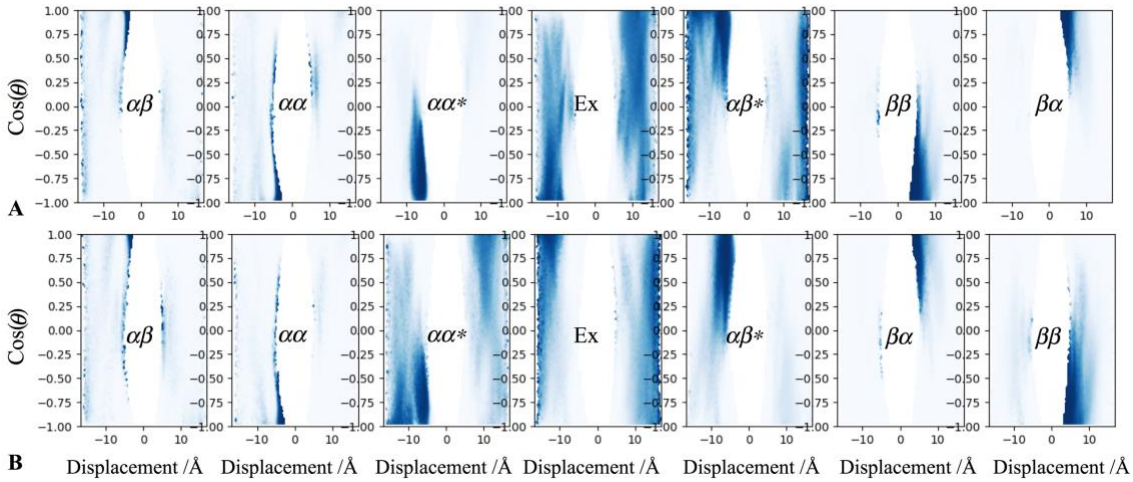

Figure A.2.6. The fuzzy memberships for the 7 state HMSMs constructed for 5'-A(Ap)-3' (top) and 5'-(Ap)A-3' (bottom) projected onto the displacement and cos(θ) space, showing how many of the ~2000 total microstates are represented by the 7 states in the HMSM. There is some 'leakage' of the sheared conformations into the extended region, and, similarly, a few of the rare T-shaped structures (small distance, small angle) are classified as rare stacked states. This is likely due to the 500 ps lag time, and it suggests that distinct states that are combined can interconvert on timescales faster than 500 ps.

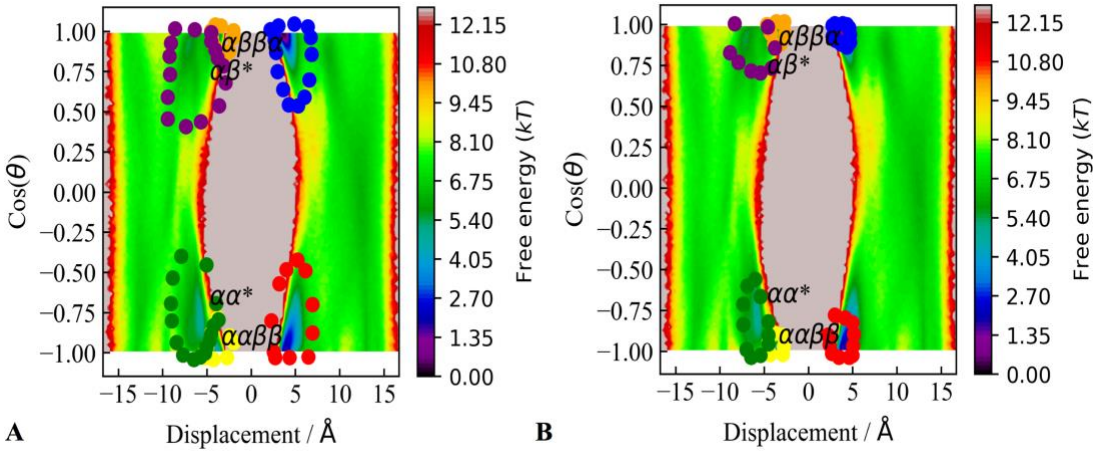

Figure A.2.7. Two different sets of regions (panels A and B) in the displacement cos(θ) space for 5'-A(2Ap)-3' used to define the 7 metastable conformations. The extended conformation (Ex) is defined by the remaining regions.

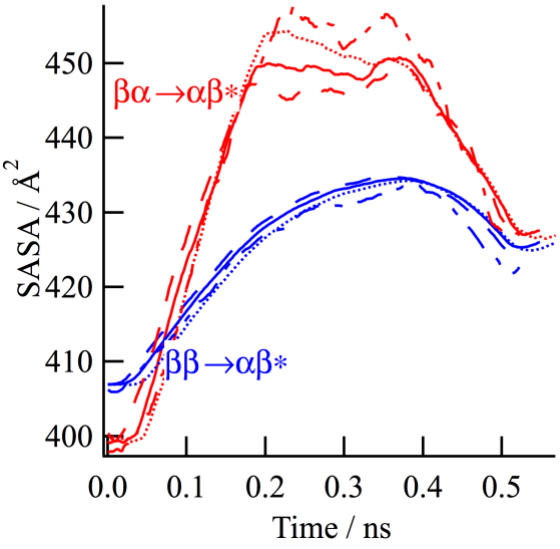

Figure A.2.8. The SASA averaged over 500 ps transtion paths obtained by requiring that the initial/final states defined in Figure A.2.7 panel A were visited for 80% of a 30 ps padding time (dashed), 75% of a 25 ps padding time (solid), and 70% of a 70 ps padding time (dotted). The same calculation was repeated for the regions defined in Figure S7 panel B requiring the transitions spent 75% of a 25 ps padding time (mixed dashed). These curves show that the choice of parameters and regions does not affect the conclusion that $\beta\alpha \rightarrow \alpha\beta^*$ transitions expose more surface area to the solvent than $\beta\beta \rightarrow \alpha\beta^*$ transitions.

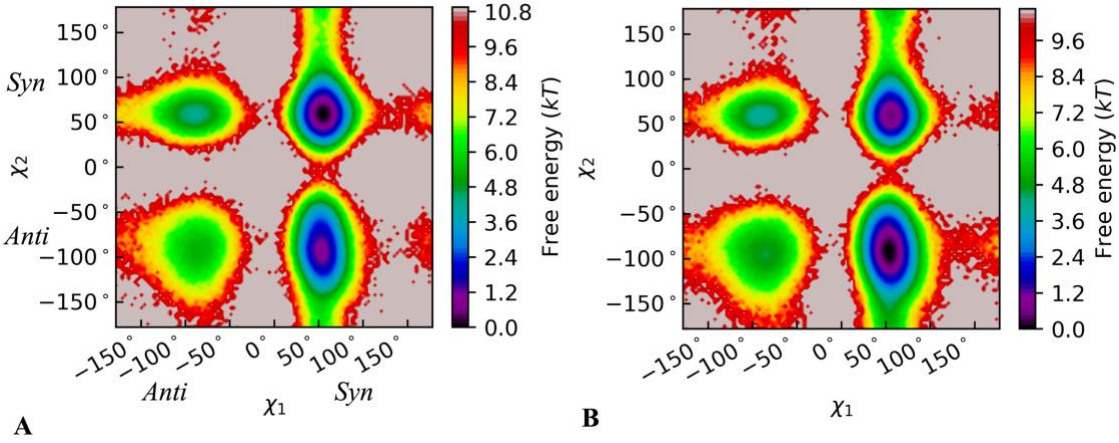

Figure A.2.9. The free energy surface for χ_1 and χ_2 computed for 5'-A(Ap)-3' (A) and 5'-(Ap)A-3' (B). The figure shows that positive χ_1 angles are favored, corresponding to the $\beta\alpha$, $\beta\beta$, $\alpha\beta$*, and $\alpha\alpha$* states. Regions of *anti* (-180, 0) and *syn* (0, 180) for χ are also shown in panel A for clarity.

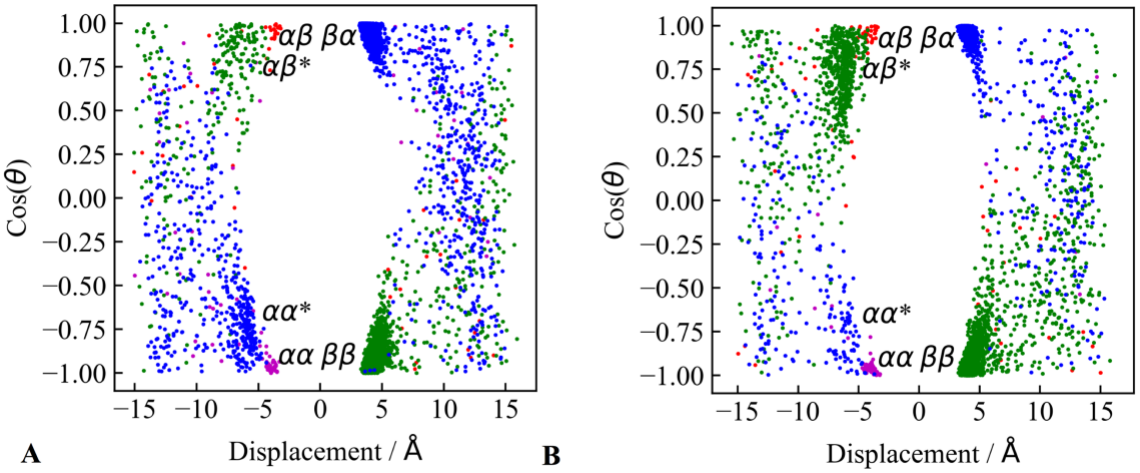

Figure A.2.10. A scatter plot of the displacement and cos(θ) values from 5000 frames taken in 2 ns intervals is shown to demonstrate how χ_1 and χ_2 partition the phase space of 5'-A(Ap)-3' (A) and 5'-(Ap)A-3' (B). The colors are coded based on the values of χ_1 and χ_2. In particular, blue, green, red, and magenta are used to identify conformations having (χ_1, χ_2) pairs that are (*syn*,*syn*), (*syn*,*anti*), (*anti*,*anti*), and (*anti*,*syn*), respectively. These combinations correspond to the four minima in Figure S9.

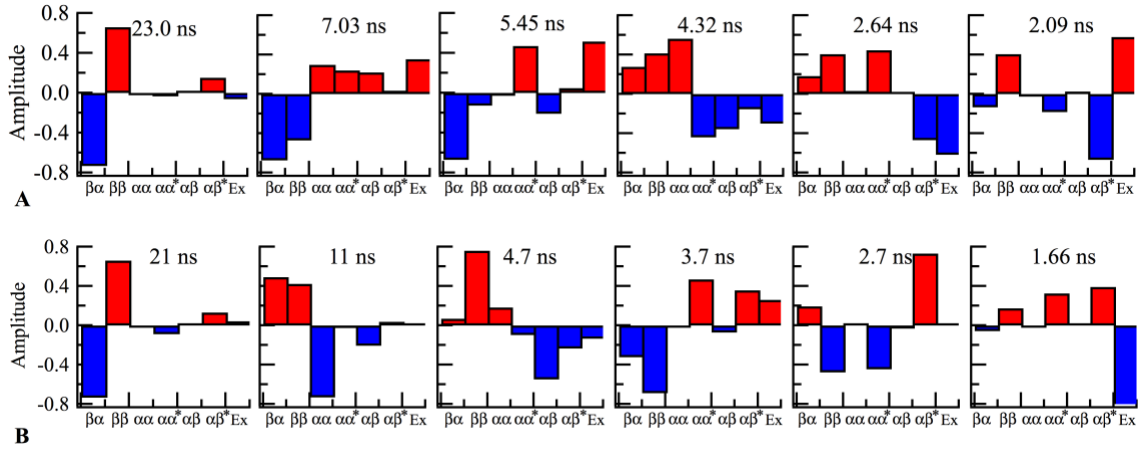

Figure A.2.11. Amplitudes of the eigenvectors of the transition rate matrices for 5'-A(2Ap)-3' (top) and 5'-(2Ap)A-3' (bottom) constructed from HMSMs with lag times of 500 ps. Positive values of the components of the eigenvectors are red while negative values are blue. The timescales associated with each eigenvector are shown in each frame.

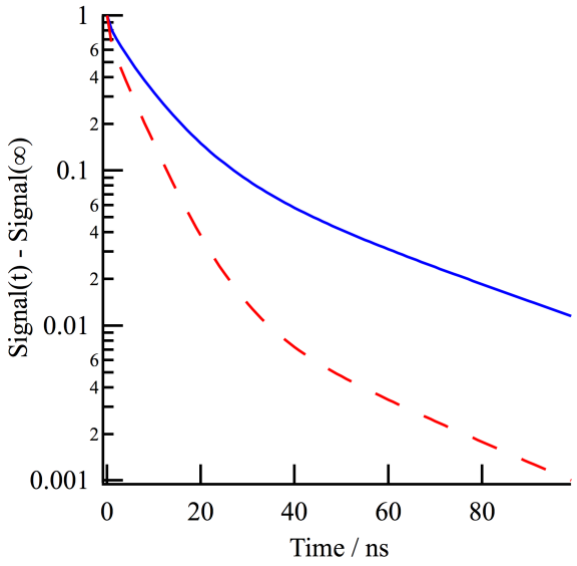

Figure A.2.12. The predicted signals for a simulated 2.5 K temperature jump measurement with UV-visible detection as described in the main text. The results for 5'-A(2Ap)-3' are shown by the blue/solid curve, while the 5'-(2Ap)A-3' sample is shown by the red/dashed curve. The slowest six amplitudes and lifetimes for these curves are shown in Table A.2.3.

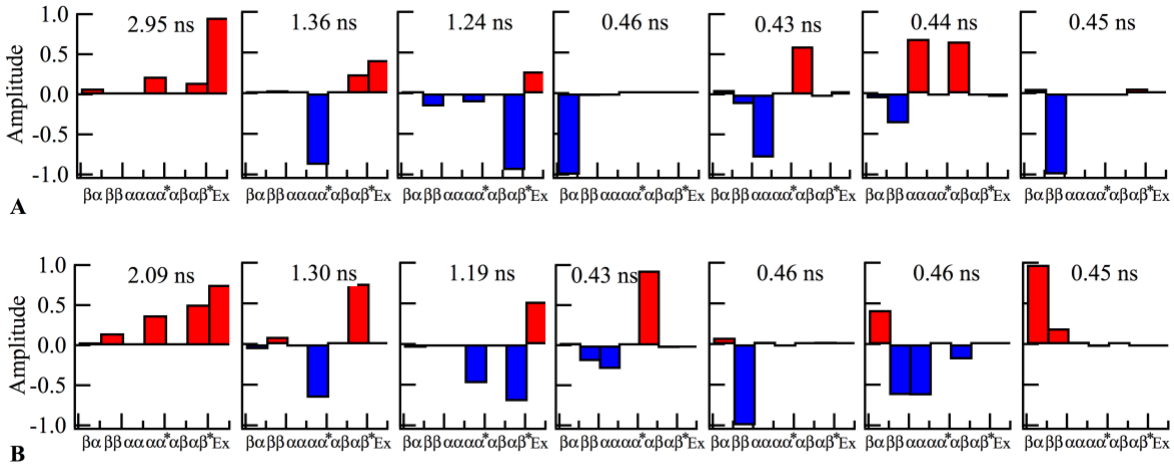

Figure S13. The amplitudes of the eigenvectors of rate matrices which regulate the excited state population of 2Ap* predicted by the hybrid static/dynamic model 3 described in the main text. Panels A and B show the results for 5'-A(2Ap)-3' and 5'-(2Ap)A-3' respectively. These eigenvectors are ordered the same as the timescales/amplitudes in Table A.2.5 for clear comparison.

Modifications to the dynamical fingerprint model

Here master equations are built for time-resolved measurements that are sensitive to the ground and/or excited state populations of a system with *n* distinct conformations that interconvert. In the following, it is assumed that quenching is irreversible and the re-formation of excited states is neglected. A two electronic state system is considered with a ground and excited state, where N_i is the population in the ground state with conformation *i* and N_i^* is the population in the excited state with conformation *i*. The rate of decay of an excited state in conformation *i* to the ground state is given by the sum of radiative, $k_{\mathrm{r},i}$ and non radiative $k_{\mathrm{nr},i}$ rates. The conformational transition rates from conformation i to conformation j are $k_{i,j}$. Then, assuming that transition rates are independent of electronic excitation, the master equation controlling the population of N_i and N_i^* is given by eq. S1,

$$\frac{d}{dt}\begin{bmatrix} N_1 \\ \vdots \\ N_n \\ N_1^* \\ \vdots \\ N_n^* \end{bmatrix} = \begin{bmatrix} -\sum_{j\neq 1} k_{1j} & \cdots & k_{n,1} & k_{\mathrm{dec},1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k_{1,n} & \cdots & -\sum_{j\neq n} k_{nj} & 0 & \cdots & k_{\mathrm{dec},n} \\ 0 & \cdots & 0 & -k_{\mathrm{r},1} - k_{\mathrm{nr},1} - \sum_{j\neq 1} k_{1j} & \cdots & k_{n,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & k_{1,n} & \cdots & -k_{\mathrm{r},1} - k_{\mathrm{nr},1} - \sum_{j\neq n} k_{nj} \end{bmatrix} \begin{bmatrix} N_1 \\ \vdots \\ N_n \\ N_1^* \\ \vdots \\ N_n^* \end{bmatrix} \quad \text{(S1)}$$

The proof that equation 1 satisfies detailed balance ($\pi_i \boldsymbol{K}_{ij} = \pi_j \boldsymbol{K}_{ji}$ where $\vec{\pi}$ is the equilibrium vector) is straightforward: First, it is noted that $\vec{\pi}$ is 0 for excited states and non-zero for the ground state entries with the value of the ground state equilibrium,

because the ground state is an absorbing state. If *i* and *j* are both ground states, then detailed balance is satisfied because the ground state master equation is assumed to satisfy detailed balance. If *i* is a ground state and *j* is an excited state then $\boldsymbol{K}_{ij} = 0$ and π_j=0 so detailed balance holds. If *i* is an excited state and *j* is a ground state then $\pi_i = 0$ and $\boldsymbol{K}_{ji} = 0$ so detailed balance holds. Finally, if *i* is an excited state and *j* is a excited state then $\pi_i = 0$ and $\pi_j = 0$ so detailed balance holds. Therefore, Equation 1 satisfies detailed balance for all *i* and *j*. This equation can be solved via standard eigenvalue decomposition techniques.

Table A.2.1 Atomic charges and assigned atom types for 2Ap used in the simulations.

Atom Name	Atom Type	Charge / eV
N9	N*	-0.047936
C8	C2	0.150546
H8	H5	0.165748
N7	NB	-0.598370
C5	CB	0.156676
C6	CA	0.157719
H6	H4	0.150049
N1	NC	-0.674915
C2	CQ	0.879792
N2	N2	-0.930086
H21	H	0.394869
H22	H	0.394869
N3	NC	-0.682870
C4	CB	0.378610
Methyl Cap	----	0.105299

Table A.2.2. Average values of the $\chi_{5'}$ and $\chi_{3'}$ dihedral angles along with the twist helical parameter for the six regions defined in Figure A.2.7 panel A. Averages for these angular quantities were obtained by first averaging their x and y components on the unit circle and then using the inverse tangent function taking into account the proper quadrant.

	5'-A(2Ap)-3'			5'-(2Ap)A-3'		
State	$\chi_{5'}$	$\chi_{3'}$	Twist	$\chi_{5'}$	$\chi_{3'}$	Twist
$\beta\alpha$	56.3	60.4	36.3	56.8	60.5	39.3
$\beta\beta$	53.7	-92.8	155	54.4	-93.0	149
$\alpha\alpha$	-91.5	62.5	63.5	-90.7	62.4	61.7
$\alpha\alpha$*	51.6	67.5	79.5	50.0	70.0	81.0
$\alpha\beta$	-88.2	-102	-44.6	-83.9	-99.9	-35.3
$\alpha\beta$*	50.2	-76.3	-27.2	51.9	-81.7	-34.3

Table A.2.3. The slowest six relaxation times and their relative amplitudes for a simulated 2.5 K temperature jump experiment, assuming that the percentage hypochromicity of stacked conformations is 10%. The signals from these curves can be seen in Figure A.2.12.

	5'-A(2Ap)-3'		5'-(2Ap)A-3'	
Index	τ / ns	A / %	τ / ns	A / %
1	20.6	14	17.3	2
2	6.8	11	13.2	1
3	4.9	0	4.6	6
4	4.1	62	2.9	88
5	1.9	13	2.2	-3
6	1.6	0	1.2	6

Table A.2.4. Rate matrices for the master equations regulating the 2Ap excited state population with rates in units of ns^{-1} for (A) 5'-A(2Ap)-3' and (B) 5'-(2Ap)A-3'. These tables are for the second model where the stacked conformations are assumed to decay instantaneously, and the signal is due to the unstacked populations. All diagonal elements were altered to account for a 1/9.89 ns^{-1} monomer-like radiative decay rate.

(A)

Final / Initial	αα*	αβ*	Ex
αα*	-0.363	0.015	0.183
αβ*	0.01	-0.473	0.097
Ex	0.081	0.066	-0.400

(B)

Final / Initial	αα*	αβ*	Ex
αα*	-0.416	0.0336	0.117
αβ*	0.0207	-0.418	0.096
Ex	0.119	0.157	-0.608

Table A.2.5. Rate matrices for the master equations regulating the 2Ap excited state population with rates in units of ns^{-1} for 5'-A(2Ap)-3' (A) and 5'-(2Ap)A-3' (B). The tables are for the third model where the stacked and partial stacks are subject to nonradiative quenching rates of 1/500 ps^{-1} and 1/3 ns^{-1}, respectively. All the diagonal elements were also altered to account for a 1/9.89 ns^{-1} monomer-like radiative decay rate.

(A)

Final / Initial	βα	ββ	αα	αα*	αβ	αβ*	Ex
βα	-2.1517 (0.0069)	0.00226 (0.00078)	0.00070 (0.00050)	0.007 (0.019)	0 (0)	0.0026 (0.0014)	0.038 (0.021)
ββ	0.0038 (0.0027)	-2.1973 (0.0065)	0.00011 (0.00067)	0.0017 (0.0029)	0.0017 (0.0011)	0.0833 (0.0086)	0.0056 (0.0026)
αα	0.020 (0.022)	0.0018 (0.0052)	-2.293 (0.086)	0.0175 (0.089)	0.032 (0.037)	0.0052 (0.0080)	0.116 (0.070)
αα*	0.051 (0.077)	0.0074 (0.0069)	0.0047 (0.0051)	-0.696 (0.089)	0 (0)	0.015 (0.025)	0.1834 (0.082)
αβ	0 (0)	0.039 (0.011)	0.044 (0.031)	0 (0)	-2.286 (0.080)	0.094 (0.048)	0.008 (0.084)
αβ*	0.0126 (0.0051)	0.240 (0.048)	0.0009 (0.0016)	0.010 (0.034)	0.01185 (0.0040)	-0.806 (0.041)	0.097 (0.036)
Ex	0.127 (0.061)	0.0110 (0.0057)	0.0136 (0.0076)	0.081 (0.035)	0.001 (0.054)	0.066 (0.056)	-0.401 (0.051)

(B)

Final / Initial	βα	ββ	αα	αα*	αβ	αβ*	Ex
βα	-2.1613 (0.0066)	0.0076 (0.0018)	0.00105 (0.00072)	0.044 (0.017)	0 (0)	0.00161 (0.00097)	0.006 (0.023)
ββ	0.0037 (0.0014)	-2.1899 (0.0059)	0 (0)	0.0030 (0.0015)	0.001 (0.016)	0.0491 (0.0062)	0.032 (0.014)
αα	0.017 (0.026)	0.007 (0.020)	-2.201 (0.038)	0.0417 (0.0032)	0.034 (0.024)	0 (0)	0 (0)
αα*	0.137 (0.056)	0.019 (0.014)	0.00778 (0.00038)	-0.749 (0.014)	0.0006 (0.0055)	0.033 (0.012)	0.117 (0.066)
αβ	0 (0)	0.041 (0.087)	0.037 (0.011)	0.004 (0.011)	-2.30 (0.22)	0.068 (0.062)	0.056 (0.088)
αβ*	0.0031 (0.0012)	0.190 (0.031)	0 (0)	0.0207 (0.0034)	0.007 (0.048)	-0.751 (0.025)	0.096 (0.035)
Ex	0.018 (0.050)	0.20 (0.12)	0 (0)	0.119 (0.024)	0.010 (0.073)	0.157 (0.087)	-0.60 (0.11)

Table A.2.6. All seven time constants and amplitudes predicted by model 3. The time constants are the negative inverse eigenvalues of the rate matrices in Table A.2.5.

5’-A(2Ap)-3’ Amplitude	5’-A(2Ap)-3’ Time constant / ns	5’-(2Ap)A-3’ Amplitude	5’-(2Ap)A-3’ Time constant / ns
0.277	2.95	0.419	2.09
-0.007	1.36	0.010	1.30
0.084	1.24	0.077	1.19
0.444	0.46	-0.035	0.43
0.001	0.43	0.371	0.46
-0.091	0.44	0.197	0.46
0.205	0.45	0.370	0.45