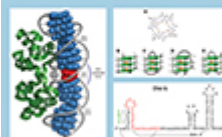
ACS SYMPOSIUM SERIES 1062

Frontiers in Nucleic Acids



EDITED BY
Richard D. Sheardy
and Stephen A. Winkle

Frontiers in Nucleic Acids

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.fw001

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.fw001

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

ACS SYMPOSIUM SERIES **1082**

Frontiers in Nucleic Acids

Richard D. Sheardy, Editor
Department of Chemistry and Physics
Texas Woman's University

Stephen A. Winkle, Editor
Department of Chemistry and Biochemistry
Florida International University

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.fw001

Sponsored by the
ACS Division of Biological Chemistry

American Chemical Society, Washington, DC

Distributed in print by Oxford University Press, Inc.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Library of Congress Cataloging-in-Publication Data

Library of Congress Cataloging-in-Publication Data

Frontiers in nucleic acids / Richard D. Sheardy, editor, Stephen A. Winkle, editor.
p. cm. -- (ACS symposium series ; 1082)
"Sponsored by the ACS Division of Biological Chemistry, American Chemical Society, Washington DC."
Includes bibliographical references and index.
ISBN 978-0-8412-2623-4 (alk. paper)
1. Nucleic acids--Congresses. I. Sheardy, Richard Dean. II. Winkle, Stephen A. III. American Chemical Society. Division of Biological Chemistry.
QP620.F76 2011
612′.0157--dc23

2011041734

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.fw001

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48n1984.

Copyright © 2011 American Chemical Society

Distributed in print by Oxford University Press, Inc.

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of $40.25 plus $0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.fw001

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Chapter 1

Inhibition of G-Quadruplex Helicase Activity As Determined by Surface Plasmon Resonance

Matthew Decker,[1] Bhupendra Mishra,[2] Dalip Kumar,[2] and Wendi M. David*,[1]

[1]Department of Chemistry and Biochemistry, Texas State University-San Marcos, San Marcos, Texas 78666
[2]Chemistry Group, Birla Institute of Technology and Science, Pilani 333031, Rajasthan, India
***E-mail: wdavid@txstate.edu**

The ability of G-quadruplex interactive small molecules to inhibit SV40 Large T-Antigen G-quadruplex helicase activity was determined using a surface plasmon resonance based assay. Unwinding of a physiologically relevant intramolecular G-quadruplex DNA substrate was completely inhibited by one novel porphyrin derivative under these conditions. This SPR-based method represents an important strategy for comparing helicase inhibition effectiveness.

Introduction

A number of processes including replication, repair, and transcription, involve DNA helicases. Besides unwinding of duplex DNA, it is increasingly evident that manipulation of G-quadruplex-forming regions of G/C genomic tracts is necessary *in vivo*. G-Quadruplex DNA stability is potentially important in the inhibition of telomerase (*1*), modulation of the promotor regions of oncogenes such as c-myc (*2*), and for correct helicase functioning in genome maintenance (*3*). Development of selective G-quadruplex stabilizing small molecules is thus of interest (*4–6*).

We used a model helicase to characterize inhibition of quadruplex DNA helicase activity by a set of G-quadruplex stabilizing small molecules. SV40 large T-antigen (T-ag), essential for replication of the SV40 genome, unwinds viral DNA in a bidirectional manner after recruiting host cellular factors to initiate replication (*8*). In addition to duplex unwinding, T-ag has been demonstrated

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

to unwind G-quadruplex substrates (*9*), perhaps because quadruplex-forming regions of the SV40 genome must be resolved for efficient replication to occur. We have previously demonstrated the use of surface plasmon resonance (SPR) for investigating T-ag helicase activity (*10*, *11*) and extend that method in the current study for comparing potential G-quadruplex helicase inhibitors.

The strategy for our SPR-based determination of G-quadruplex helicase activity and inhibition is shown in Figure 1. An oligonucleotide containing an intramolecular G-quadruplex forming region is immobilized on a streptavidin derivatized SPR sensor chip surface through a 5′-biotin moiety. *In situ* hybridization of a short complementary oligonucleotide downstream to the quadruplex folded region generates an immobilized intramolecular quadruplex substrate with a free 3′-end and a short forked duplex region (Figure 1, top). Loading of an active T-ag hexamer onto the free 3′-end of the quadruplex folded substrate (Figure 1, second from top) and subsequent unwinding of the intramolecular quadruplex generates a forked duplex (Figure 1, second from bottom). Continued translocation of T-ag unwinds the duplex region, liberating the hybridized complementary oligonucleotide from the sensor surface (Figure 1, bottom). Loss of the complement from the sensor surface results in a decreased SPR signal, indicating helicase activity in a proportional manner.

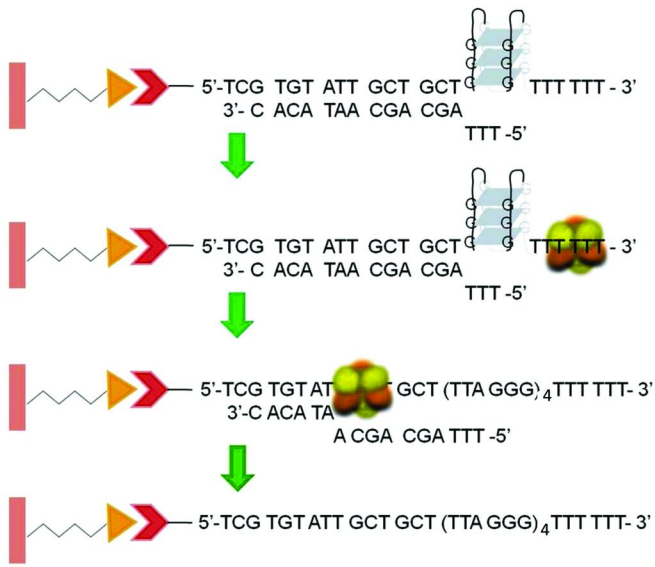


Figure 1. Strategy for observing quadruplex helicase activity by SPR.

The ability of a novel set of porphyrin analogs (**BM042**, **BM043**, and **BM044**), **TMPyP4** and **360** (Figure 2) to inhibit the G-quadruplex helicase activity of T-ag was characterized using our SPR-based helicase assay. Inhibition of helicase activity observed in this manner would result in a decrease of the net loss of signal from the sensor surface after exposure of the DNA substrate to T-ag. Since the DNA substrate contains an intramolecular quadruplex forming

region, inhibition of the T-ag helicase activity represents inhibition of unwinding the intramolecular quadruplex. T-ag does not bind 5′-ends of synthetic substrates (*12*) and does not unwind synthetic oligonucleotide substrates in our assay that do not contain a free 3′-end.

Both TMPyP4 and 360 are known G-quadruplex interactive compounds (*13*, *14*) and TMPyP4 has previously been shown to inhibit T-ag duplex and G-quadruplex helicase activity (*15*). The porphyrin analogs shown in Figure 2 are novel G-quadruplex interactive compounds and their synthesis and characterization will be reported elsewhere.

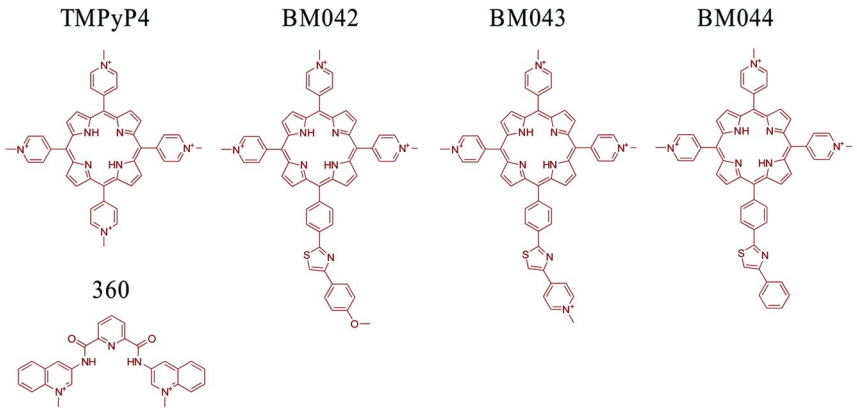


Figure 2. Structure of small molecule G-quadruplex DNA interactive compounds.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Experimental

Chemicals and Reagents

All reagents for buffer solutions were obtained from Sigma-Aldrich (St. Louis, MO) or GE Healthcare (Piscataway, NJ) unless otherwise noted. Potassium chloride was obtained from EM Science (Gibbstown, NJ) and magnesium chloride was obtained from ACROS Organics (Geel, Belgium). A working stock solution of 10X concentrated **HBS-EP Buffer** (0.1 M HEPES, 30 mM EDTA) was used to prepare the following running buffers: **HBS-EP-Mg^{2+} Buffer** (0.01 M HEPES, pH 7.4, 0.15 M KCl, 3 mM EDTA, 10 mM $MgCl_2$, and 0.005% v/v P20 surfactant); **HBS-EP-K^+ Buffer** (0.01 M HEPES, pH 7.4, 0.15 M KCl, 3 mM EDTA, and 0.005% P20 surfactant); and **HBS-EP-Li^+** (0.01 M HEPES, pH 7.4, 0.01 M $LiCl_2$, 3 mM EDTA, and 0.005% P20 surfactant). Buffers were passed through 0.2 mm filters (Nalgene) and degassed for a minimum of ten minutes prior to use.

The following HPLC-purified oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA): 5′-BioTEG-TC GTG TAT TGC TGC TTT TTT (TTA GGG)$_4$ TT-3′ (**imG4**) and 3′-CAC ATA ACG ACG ATT T-5′ (**imG4comp2**). Oligonucleotides were stored at -20 °C and diluted with buffer prior to use.

SV40 T-ag was obtained from CHIMERx (Milwaukee, WI), where it was isolated from cultured insect cells and stored in 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1.0 mM EDTA, 1.0 mM dithiothreitol, and 50% (v/v) glycerol. Aliquots of T-ag were stored at -80 °C and diluted with HBS-EP-Mg^{2+} buffer prior to each use. Adenosine 5′-triphosphate disodium salt (ATP) was obtained from Sigma-Aldrich and stored at -20 °C. ATP was added to T-ag injection solutions just prior to injection.

TMPyP4 was obtained from Sigma-Aldrich. The small molecules BM042, BM043, BM044 were obtained as a gift from Dr. Dalip Kumar and Bhupendra Mishra. The synthesis of these novel porphyrin analogs will be reported elsewhere. The compound 360 has been previously reported and was a gift from Dr. Sean Kerwin (The University of Texas, Austin, TX). Each of these compounds was diluted in HPLC grade dimethylsulfoxide (DMSO) and protected from light during storage at 4 °C.

Surface plasmon resonance experiments were performed using a Biacore® X instrument and streptavidin-derivatized (SA) sensor chips from GE Healthcare. Each sensor chip was stored at 4 °C prior to use. All experiments were performed at 25 °C.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Immobilization and Preparation of DNA Substrates

SA sensor chips were pre-conditioned prior to immobilization of the desired nucleotide sequence by performing a series of three 20 μL injections of 1 M NaCl/50 mM NaOH in dual channel mode at 50 μL/min. HBS-EP-K^+ buffer was used as a running buffer and also served as the diluent for the DNA sequences to be immobilized. Upon dilution, biotinylated **imG4** DNA was injected in 10-20 μL increments in single channel mode at 10 μL/min until the desired immobilization level of ~100-500 response units (RUs) was achieved. In order to ensure maximum formation of the intramolecular quadruplex structure contained in the **imG4** oligonucleotide, the immobilized substrate was equilibrated in HBS-EP-K^+ buffer overnight prior to any subsequent experiments. In all cases, one flowcell contained immobilized oligonucleotide while the other flowcell was used as a reference cell to correct for nonspecific binding.

Upon equilibration of the immobilized oligonucleotide, a solution of partially complimentary DNA, **imG4comp2** was injected until complete hybridization was observed. Since hybridization of **imG4comp2** to the immobilized **imG4** oligonucleotide is a 1:1 interation, the SPR response indicating maximum hybridization was easily calculated based on the RUs resulting from immobilized **imG4** multiplied by the ratio of molecular weights for the two oligonucleotides (MW imG4comp2/ MW imG4).

Binding of Small Molecules to Immobilized Oligonucleotide

Dilutions (500 nM – 5 μM) of the respective compound were prepared in HBS-EP-K^+ buffer immediately prior to each binding assay. Solutions were injected over a reference subtracted flowcell immobilized with **imG4** at 20 μL/min for 180 seconds with a delayed wash of 180 seconds. The sensorgram

was allowed to proceed for a period of seven minutes after each injection had completely finished, including the wash, before regenerating the chip with either 1 M KCl/1 mM NaOH or 1 M KCl/50 mM NaOH at 50 μL/min flow rate. Reference subtracted sensorgrams for varying concentrations of individual compounds were analyzed by BiaEvaluation software (v. 2.0; GE Healthcare); however, in the case of BM042-BM044, the observed binding appeared to involve either multiple binding modes or aggregation of compound, precluding accurate determination of binding constants. Therefore, for clarity, the binding of each compound at 1 μM to the immobilized DNA substrate was used to qualitatively compare binding between the different small molecules and calculated binding constants are not reported.

In order to assess the extent of binding that might occur for each compound to the unwound single-stranded immobilized **imG4**, dilutions of the compounds were also prepared in HBS-EP-Li^+ buffer and binding was determined in HBS-EP-Li^+ running buffer. The immobilized SA sensorchip was heated to 40 °C while equilibrating with HBS-EP-Li^+ running buffer. The sensorchip was slowly cooled to 25 °C prior to binding assays. Regeneration was accomplished by injecting 1 M $LiCl_2$/50 mM NaOH at 50 μL/min.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Determination of T-ag G-Quadruplex Helicase Activity and Inhibition

Immobilization of **imG4** oligonucleotide and hybridization of the partially complementary oligonucleotide **imG4comp2** was performed as described above. The resulting DNA substrate was equilibrated in HBS-EP-K^+ buffer to ensure proper folding of the intramolecular quadruplex forming region. Varying concentrations of T-ag were then injected in HBS-EP-Mg^{2+} buffer containing 15 mM ATP at 20 μL/min for 40 seconds. After completion of the injection, the response (RU) originating from removal of **imG4comp2** was determined. The ratio of this response divided by the RUs originating from hybridization of **imG4comp2** multiplied by 100 is equal to the % T-ag activity. The chip was rehybridized with complement prior to each successive injection.

Inhibition of T-ag G-quadruplex helicase activity was characterized for each of the small molecule G-quadruplex interactive compounds mentioned above. These compounds were prepared and stored under light sensitive conditions to prevent possible induction of nucleic acid photocleavage properties. Dilutions for injection were prepared immediately prior to each assay and immobilized DNA substrate was equilibrated as mentioned previously. After equilibration, the running buffer was changed to HBS-EP-Mg^{2+} buffer for characterization of T-ag activity and inhibition. Next, 40 μL of the respective small molecule was injected using manual inject mode at 20 μL/min for 120 seconds. This injection was paused, followed by an immediate injection of T-ag in HBS-EP-Mg^{2+} buffer containing 15 mM ATP and the manual injection was resumed. The injection was stopped after the completion of the T-ag injection and the amount of **imG4comp2** removed from the sensor surface was determined. The flow rate was then increased to 50 μL/min and the sensor chip was regenerated using two 30 μL injections of 1 M KCl/50 mM NaOH. The % T-ag activity was calculated and

normalized relative to % T-ag activity in the absence of any added inhibitor. This normalized value was reported as % inhibition.

Results and Discussion

Previous investigations detailed the optimization of our SPR-based assay for determining the duplex and G-quadruplex helicase activity of T-ag (*10*, *11*). The current study focused on utilizing our assay to compare the ability of several G-quadruplex interactive small molecules (Figure 2) to inhibit unwinding of an intramolecular G-quadruplex DNA substrate by T-ag.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Binding of G-Quadruplex Interactive Compounds to the Immobilized DNA Substrate

The ability of small molecules TMPyP4, BM042-BM044, and 360 to bind to the immobilized DNA substrate was compared prior to assessing inhibition of the enzymatic G-quadruplex helicase activity of T-ag. A concentration series of each compound was injected in HBS-EP-K^+ buffer over the immobilized **imG4** substrate, which contains an intramolecular G-quadruplex folding region based on the human telomeric repeat sequence TTA GGG (*16*). The amount of immobilized substrate for these binding experiments was kept low to minimize the effects of mass transfer.

A representative sensorgram showing binding of BM044 to immobilized **imG4** substrate is shown in Figure 3. The response was concentration dependent with potential cooperative binding or aggregation evident at higher concentrations. The *on* rate of binding appeared slow, as did the *off* rate. Complete removal of compound from the DNA substrate did not occur in the time frame examined without regeneration solution, indicating persistent binding.

A comparison of TMPyP4, BM042, BM043, and BM044 binding to the immobilized **imG4** substrate is shown in Figure 4. All of the BM0 compounds exhibited slower association and dissociation profiles in comparison to TMPyP4. SPR-determined parameters for binding of TMPyP4 to quadruplex DNA have been previously reported (*15*) and the binding profile displayed in Figure 3 is similar to published data. Of particular note compared to the other compounds is the relatively fast association and dissociation of TMPyP4 from the quadruplex substrate.

BM042 appeared to associate most slowly with the G-quadruplex DNA substrate and displayed evidence of cooperative binding or aggregation in the concentration series (data not shown). It is possible that the methoxy-substituted phenyl substituent appended to one of the 'arms' of the planar structure may initially restrict access of the compound to the quadruplex structure, but it likely interacts with a groove of the quadruplex structure after initial binding is established. The dissociation of the compound was also the slowest among the group, indicating that upon interaction with the G-quadruplex substrate the compound forms a tight interaction.

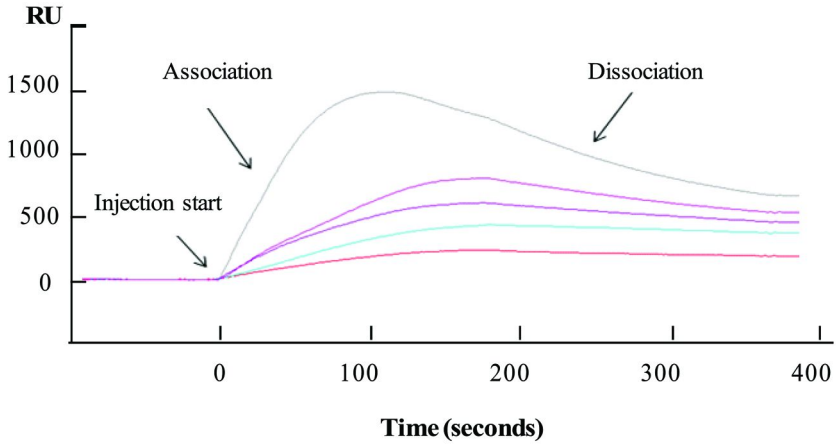


Figure 3. ***Binding of BM044 to immobilized imG4 substrate.*** *BM044 at 2 μM (top, grey trace), 1.7 μM (2nd from top, magenta trace), 1.2 μM (3rd from top, dark purple trace), 900 nM (4th from top, cyan trace), and 750 nM (bottom, red trace) was injected at 20 μL/min in HBS-EP-K+ buffer. Each reference subtracted sensorgram represents binding of BM044 to ~150 RUs of immobilized imG4 equilibrated in HBS-EP-K+ buffer prior to injections.*

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

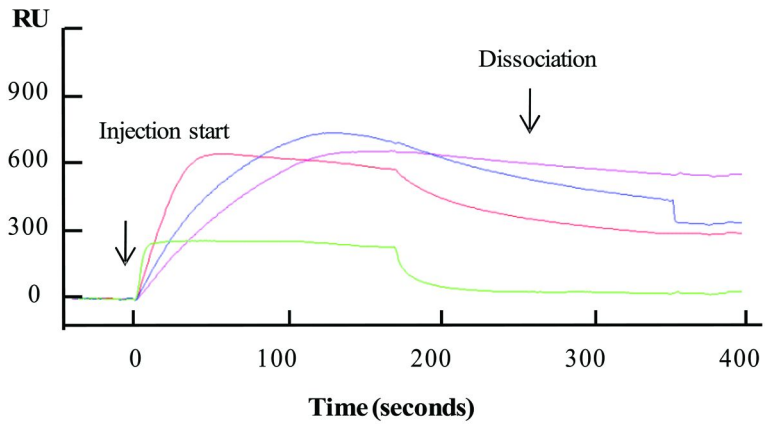


Figure 4. ***Comparison of binding profiles at 1 μM small molecule concentration.*** *A solution of BM042 (magenta trace, top line in dissociation region), BM043 (red trace, 3rd from top in dissociation region), BM044 (blue trace, 2nd from top in dissociation region), and TMPyP4 (light green trace, bottom line in dissociation region) was injected at 20 μL/min in HBS-EP-K+ buffer over a sensor surface immobilized with ~150 RUs of imG4. Each sensorgram was reference subtracted.*

BM043 associated with the **imG4** substrate nearly as fast as TMPyP4 (as demonstrated by the sharp increase in RUs upon compound injection). The fast *on* rate may be attributed to charge: with four cationic charges the electrostatic attraction to the DNA should be high. The dissociation of the compound was also rapid, with much of the compound releasing rapidly upon completion of the injection. However, some interaction with the quadruplex persisted.

The compound 360 is a known selective G-quadruplex interactive agent (*14*). Structurally, the crescent shape of 360 is more similar to Distamycin A, than to the porphyrin analogs TMPyP4, BM042, BM043, and BM044. Distamycin A, has been previously reported to bind G-quadruplex structures (*17*). Binding of 360 to immobilized imG4 substrate is shown in Figure 5. A very fast association rate was evident from the sensorgram. At concentrations greater than 1 μM some aggregation may have occurred but dissociation was fairly rapid after the end of the injection.

In agreement with the reported selectively of 360 for binding G-quadruplex structures and maintaining a tight interaction at low concentrations, evaluation of 360 binding parameters using BiaEvaluation software yielded a dissociation constant, K_D = 3.96e-10 M, indicating the tightest binding of all compounds compared in this study (fitted data not shown).

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

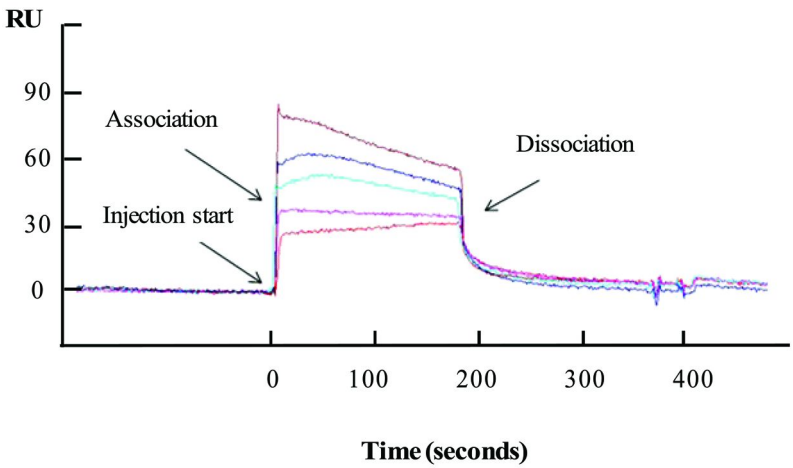


Figure 5. ***Binding of 360 to immobilized imG4 substrate.*** *The compound 360 at 5 μM (top, maroon trace), 3 μM (2nd from top, blue trace), 2 μM (3rd from top, cyan trace), 1 μM (2nd from bottom, magenta trace), and 500 nM (bottom, red trace) was injected at 20 μL/min in HBS-EP-K+ over the sensor surface immobilized with ~375 RUs of imG4. Each sensorgram was reference subtracted.*

In order to assess the selectivity of the compounds BM042, BM043, and BM044 for the intramolecular quadruplex forming region of immobilized imG4 the binding of these compounds to the DNA substrate was characterized in HBS-EP-Li+ buffer. We have previously demonstrated the substantially decreased

amount of immobilized quadruplex formation under these buffer conditions (*11*). A direct comparison of these compounds' ability to bind primarily unfolded **imG4** substrate is shown in Figure 6. Each compound appeared to slowly associate and remain bound to the substrate. The low amount of persistent binding is likely electrostatic, non-specific binding, indicating a relatively low affinity for the unfolded substrate. In contrast, TMPyP4 showed a more substantial amount of binding to the single-stranded immobilized oligonucleotide. It is possible that TMPyP4 facilitated formation of the folded intramolecular quadruplex region under these conditions.

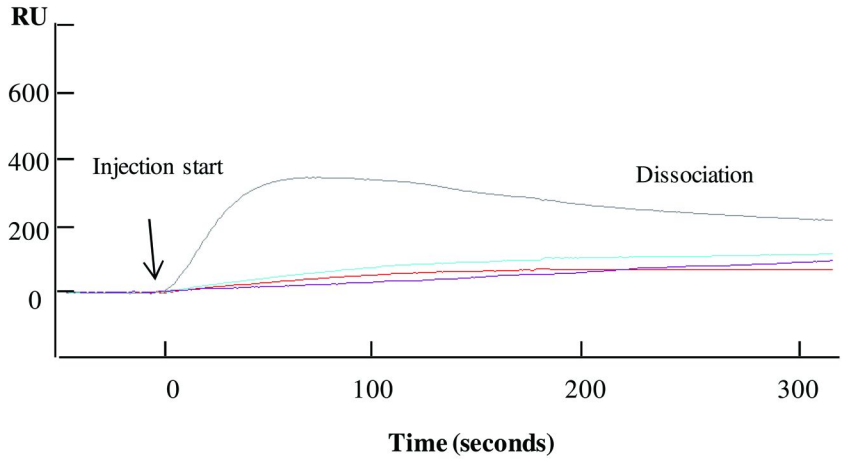


Figure 6. ***Comparison of binding at 500 nM small molecule concentration to the unfolded immobilized imG4 substrate.*** *BM044 (bottom, purple trace), BM042 (2nd from bottom, red trace), BM043 (2nd from top, light blue trace), and TMPyP4 (top, gray trace) were injected in HBS-EP-Li+ buffer at 20 µL/min over the sensor surface immobilized with ~150 RUs of imG4. Each sensorgram was reference subtracted.*

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Inhibition of T-ag G-Quadruplex Helicase Activity

In order to easily observe inhibition and normalize results between the different compounds, assay conditions were adjusted to determine ~70% unwinding of an intramolecular G-quadruplex substrate in the absence of any added inhibitor. By immobilizing **imG4** DNA to the SA sensorchip and introducing a partially complementary oligonucleotide that would *only be removed* by T-ag *if the intramolecular quadruplex structure was unwound*, we constructed a probe for determination of T-ag enzymatic activity. For all experiments using this approach, the partially complementary oligonucleotide contained a free 5′-end to minimize the potential for helicase loading onto the reporter oligonucleotide strand, ensuring that T-ag was forming a complex on the free 3′-end of the immobilized **imG4** G-quadruplex DNA substrate.

A sensorgram depicting unwinding of the **imG4** substrate by T-ag in the absence of any added inhibitor is shown in Figure 7. The immobilization of **imG4** and hybridization of **imG4comp2** is not shown. Unwinding experiments were performed in HBS-EP-K^+ buffer with 430 RU of immobilized **imG4** DNA, fully hybridized with partially complementary **imG4comp2** DNA (127 RUs). ATP concentration (15 mM) was held constant, while enzyme concentration was varied. The concentration of T-ag hexamer required to generate ~70% unwinding was 1.0 nM.

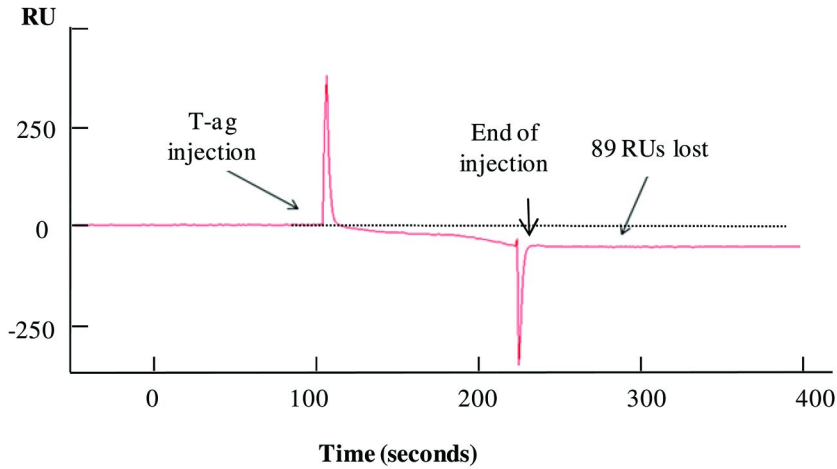


Figure 7. ***Unwinding of immobilized DNA substrate by T-ag.*** *The reference subtracted sensorgram represents the removal of 89 RUs of hybridized imG4comp2 DNA following T-ag injection in HBS-EP-Mg^{2+} buffer containing 15 mM ATP. The total amount of hybridized imG4comp2 DNA prior to helicase activity was 127 RUs.*

A representative sensorgram for inhibition of T-ag G-quadruplex helicase activity with BM043 is shown in Figure 8. The DNA substrate was equilibrated as described prior to injection of 1 μM BM043 and subsequent injection of active T-ag. Note that after the T-ag injection was complete the sensorgram indicated no removal of **imG4comp2** due to substrate unwinding. Complete inhibition of T-ag helicase activity was observed in this assay.

Characterization of the ability of the compounds in this study to inhibit T-ag unwinding of the immobilized DNA substrate revealed the following % inhibition values: BM043, ~100%; BM044, 81%; 360, 75%; TMPyP4, 54%, and BM042, 33%. Compound 360 demonstrated high binding affinity for the DNA substrate and high inhibition of T-ag quadruplex helicase activity. Although the binding affinity of BM043 for the DNA substrate is fairly low, the compound does display a slow *on* and *off* rate, as well as some selectivity for the folded immobilized substrate. In addition, it completely inhibits the T-ag quadruplex helicase activity in this assay.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

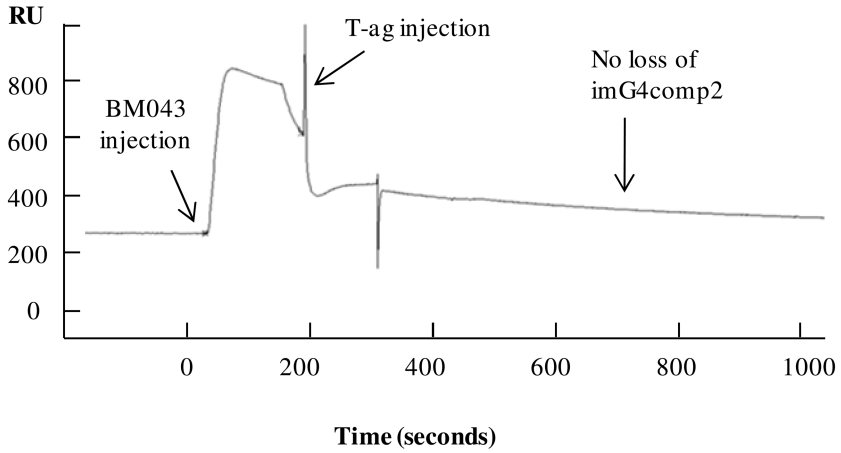


Figure 8. ***Inhibition of T-ag helicase activity with BM043.*** *A 1 μM concentration of BM043 was injected at 20 μL/min in HBS-EP-Mg² buffer in manual inject mode for 120 seconds, immediately followed by injection of 1 nM T-ag/15 mM ATP. The sensorgram is reference subtracted.*

We have demonstrated effective inhibition of intramolecular G-quadruplex helicase activity using our SPR-based technique. Since it is likely that intramolecular G-quadruplex formation (and not just bimolecular or tetramolecular quadruplex formation) is physiologically relevant (*18*), our method for assessing quadruplex helicase inhibition represents an important strategy for determining potential effectiveness of designed helicase inhibitors.

Acknowledgments

Support from the Welch Foundation (AI0045 to Texas State Department of Chemistry and Biochemistry) is gratefully acknowledged. The Department of Science & Technology, New Delhi is also gratefully acknowledged.

References

1. Hahn, W.; Weinberg, R. *Nat. Rev. Cancer* **2002**, *2*, 331–341.
2. Freyer, M.; Buscaglia, R.; Kaplan, K.; Cashman, D.; Hurley, L.; Lewis, E. *Biophys. J.* **2007**, *92*, 2007–2015.
3. Wu, Y.; Shin-ya, K.; Brosh, R. M. *Mol. Cell. Biol.* **2008**, *28*, 4116–4128.
4. Dash, J.; Waller, Z. A. E.; Pantos, G. D.; Balasubramanian, S. *Chem. Eur. J.* **2011**, *17*, 4571–4581.
5. Wu, Y.; Brosh, R. M. *FEBS J.* **2010**, *277*, 3470–3488.
6. Chaires, J. B.; Mergny, J.-L. *Biochimie* **2008**, *90*, 973–975.
7. Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441–471.
8. Simmons, D. T. *Adv. Virus Res.* **2000**, *55*, 75–133.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

9. Baran, N.; Pucshansky, L.; Marco, Y.; Benjamin, S.; Manor, H. *Nucleic Acids Res.* **1997**, *25*, 297–303.
10. Plyler, J.; Jasheway, K.; Tuesuwan, B.; Karr, J.; Brennan, J.; Kerwin, S. M.; David, W. M. *Cell Biochem. Biophys.* **2009**, *53*, 43–52.
11. Plyler, J. R.; Sanjar, F.; Howard, R.; Araki, N.; David, W. M. *J. Biotechnol. Res.* **2010**, *2*, 56–66.
12. SenGupta, D. J.; Borowiec, J. A. *Science* **1992**, *256*, 1656–1661.
13. Li, J. L.; Harrison, R. J.; Reszka, A. P.; Brosh, R. M.; Bohr, V. A.; Neidle, S.; Kickson, I. D. *Biochem.* **2001**, *40*, 15194–15202.
14. De Cian, A.; Mergny, J.-L. *Nucleic Acids Res.* **2007**, *35*, 2483–2493.
15. Tuesuwan, B.; Kern, J. T.; Thomas, P. W.; Rodriguez, M.; Li, J.; David, W. M.; Kerwin, S. M. *Biochemistry* **2008**, *47*, 1896–1909.
16. Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6622–6626.
17. Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; De Tito, S.; Romagnoli, T.; Pagano, B.; Limongelli, V.; Giancola, C.; Baraldi, P. G.; Mayol, L.; Novellino, E.; Randazzo, A. *J. Am. Chem. Soc.* **2010**, *132*, 6425–6433.
18. Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2008**, *36*, 2700–2704.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch001

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Chapter 2

A Fluorescence-Based G-Quadruplex DNA Cleavage Assay

Michelle Schoonover[1] and Sean M. Kerwin[*,1,2]

[1]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, 78712 USA
[2]Division of Medicinal Chemistry, College of Pharmacy, University of Texas at Austin, Austin, TX, 78712 USA
[*]E-mail: skerwin@mail.utexas.edu

We describe a fluorescence-based G-quadruplex DNA cleavage assay based on intramolecular G-quadruplex-forming 5′,3′-dual labeled oligonucleotides. By utilizing the difference in fluorescence emission spectra between an intact and cleaved quadruplex, the extent of G-quadruplex cleavage can be determined for a range of agents including hydroxyl radical, photo-activated porphyrin TMPyP4, and the oxo-metalloporphyrin Mn(=O)TMPyP4. This assay has been adapted to 96-well format for rapid screening, and the cleavage reaction samples can be subsequently directly analyzed by PAGE with fluorescence visualization. Agents that selectively cleave G-quadruplexes may serve as valuable probes for elucidating the formation and resolution of these structures in cells or as therapeutic agents. This assay should aid in identifying G-quadruplex cleavage agents for these various applications.

Small molecular compounds capable of initiating the cleavage of proteins or nucleic acids represent potential molecular probes and therapeutic agents. For example, DNA-cleaving agents isolated from nature (*1*) or derived from synthesis (*2*) are currently used to treat cancer. One major limitation of these agents is their lack of selectivity, which leads to toxicity limiting their effectiveness.

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Guanosine-rich DNA sequences are proposed to transiently form stable secondary structures known as G-quadruplexes in species ranging from bacteria to humans (*3–6*). The building blocks of G-quadruplex DNA are tetrads made of four hydrogen bonded guanine bases (G) that can stack to give rise to highly polymorphic structures, adopting multiple molecularities and topologies based on both DNA sequence and environmental factors (*7*, *8*) (Figure 1). *In vitro*, G-rich single-stranded and duplex oligonucleotides have been shown to form highly stable G-quadruplex structures (*9–11*), that have been characterized by a wide range of techniques, such as circular dichroism (CD) (*12*), nuclear magnetic resonance (NMR) (*13*) and X-ray crystallography (*14*).

These G-rich sequences have been hypothesized to form G-quadruplex structures involved in the regulation of cancer-related genes (*15*, *16*), telomere function (*17–19*), and various other cellular processes (*20–22*). It is estimated that there are over three hundred thousand potential G-quadruplex DNA forming sequences in the human genome (*23*, *24*). However, the cellular detection of G-quadruplexes has been elusive (*25–27*).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

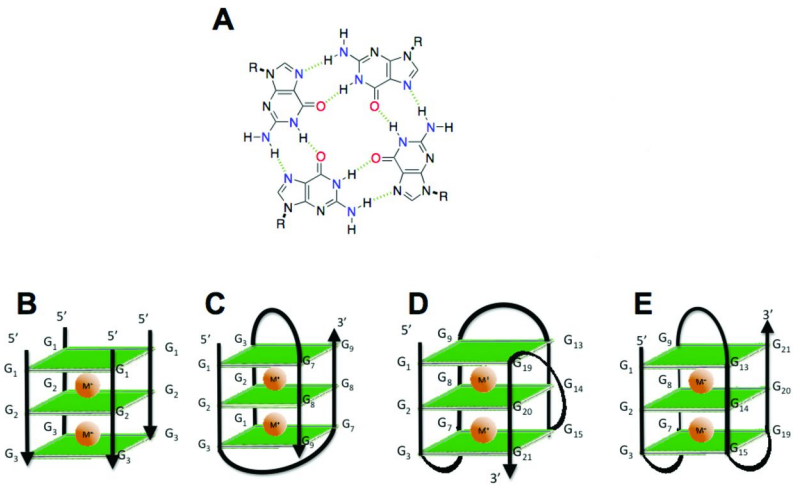


Figure 1. (A) Structure of a G-tetrad (B) Model of the topology of a termolecular, parallel stranded G-quadruplex. (C) Model of the topology of a bimolecular G-qaudruplex. (D). Model of the topology of the intramolecular "3+1" mixed parallel /anti parallel G-quadruplex formed by the human telomeric sequence. (E) Model of the topology of the intramolecular antiparallel G-quadruplex formed by the human telomeric sequence.

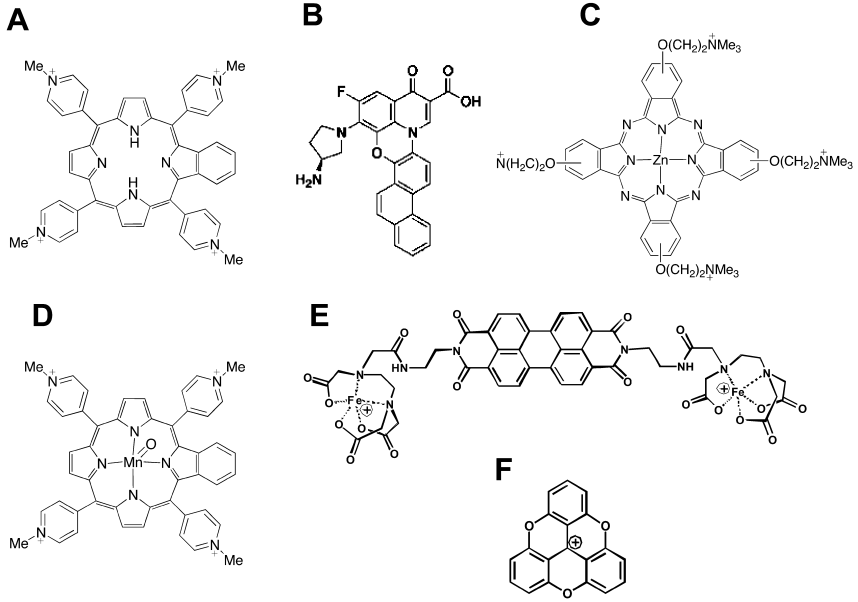


Figure 2. Known G-quadruplex DNA cleavage agents (A) 5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine (TmPyP4). (B) A fluoroquinoanthroxazine. (C) Tetrakis(2-trimethylaminoethoxy)phthalocyaninato zinc (Zn-TTAPc). (D) Mn(=O)TMyP4. (E). Perylene-EDTA. (F). Trioxotriangulenium ion (TOTA).

Small molecular agents capable of effecting the specific cleavage of G-quadruplex DNA could play an important role as molecular probes for these structure *in vitro* and *in vivo*. These probes could be used to map the regions of the genome where G-quadruplexes form. Importantly, photoactivated G-quadruplex cleavage agents have the unique potential to probe the dynamic nature of G-quadruplex structures, allowing studies of their formation and dissolution during various cellular events. By targeting specific G-quadruplex DNA structures, G-quadruplex DNA cleavage agents can effect highly sequence specific DNA cleavage. In this way, these cleavage agents may be more selective that the currently used, relatively non-sequeunce selective DNA cleaving anti-cancer agents. Currently, very few compounds have been reported to cleave G-quadruplexes (*28–35*) (Figure 2), and no photo-activated cleavage agent with high selectivity for G-quadruplex DNA is currently known. Importantly, while progress has been made in identifying G-quadruplex ligands with increased binding selectivity for quadruplex versus duplex DNA (*36*), there has been very little work devoted to understanding and exploiting the differences that may exist in DNA cleavage chemistry of G-quadruplex versus duplex DNA that could facilitate the discovery of such highly selective cleavage agents.

This work describes a new assay developed to rapidly identify G-quadruplex specific cleavage agents. Standard assays for duplex DNA cleavage exist, such as supercoiled plasmid nicking methods (*37*), or more recent solution methods based on fluorescently-labeled oligonucleotides (*38*) or cationic polymers (*39*).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

However, the only reported studies of G-quadruplex DNA cleavage have relied on laborious PAGE-based methods. Our approach is inspired by fluorescent-based "stop-light" duplex DNA cleavage assay first used in screening for enediyne natural products (*40*). However, unique features of G-quadruplex DNA complicate this approach. Here we describe a fluorescence-based G-quadruplex DNA cleavage assay that overcomes these difficulties and is suitable for use in a 96-well plate format. We describe the results of this assay in analyzing the G-quadruplex DNA cleavage by hydroxyl radical, generated by Fe•EDTA, and two model G-quadruplex DNA cleavage agents: photochemically activated *meso*-5,10,15,20-tetrakis-(N-methyl-4-pyridyl)porphine (TMPyP4) (*30*, *31*) and oxo-Mn-TMPyP4 (*34*). We show that this assay revealed both photochemical and oxidative G-quadruplex DNA cleavage events with good sensitivity, demonstrating that this assay could be a valuable tool in screening and identifying highly selective G-quadruplex cleavage compounds.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Material and Methods

DNA Purification and Quadruplex Formation

The oligonucleotide, HT22, (dAGGG(TTAGGG)$_3$) was synthesized in house on a DNA synthesizer (Expedite 8900) and purified by HPLC. The fluorescent dual labeled analog, F21T, (5′-*FAM*-dGGG(TTAGGG)$_3$-*TAM*-3′; *FAM*: 6-carboxyfluorescein; *TAM*: tetramethylrhodamine) (*41*) was purchased from Integrated DNA Technologies and used without further purification.

The F21T probe was diluted from water stock solution to 400 nM in 50 mM potassium cacodylate pH 7.4 and annealed by heating to 95 °C for 5 minutes in a water bath, followed by slowly cooling to room temperature. FRET melting experiments confirmed a T_m of 54 °C under these conditions, which corresponds to the literature value (*42*). Solutions of HT22 were annealed under the same conditions but at 5 μM strand concentration. Authentic cleaved F21T samples were prepared by incubating 5 pmol F21T in 180 μL 50 mM potassium cacodylate containing 20 mM mM $MgCl_2$, pH 7.4 with 10 units of S1 nuclease at 37 °C for 24 hours. The reaction was stopped by adding 20 μL of 250 mM EDTA. Where indicated, 200 μL of this solution was added to a well on each plate as an internal standard.

Hydroxyl Radical Cleavage

Stock solutions of 100 mM $Fe(NH_4)(SO_4)_2$ (Sigma Aldrich) and 100 mM sodium ascorbate (Sigma Aldrich) were stored in small aliquots away from light at -70 °C. Stock solutions of 200 mM Na_2-EDTA (Sigma Aldrich) and 30% H_2O_2 (Fisher) were stored at room temperature and 4 °C respectively. Working solutions of 1-10 mM Fe•EDTA were prepared immediately before use by diluting the $Fe(NH_3)(SO_4)_2$ to the nominal concentration and adding a two-fold excess of Na_2-EDTA. Working solutions of 10-100 mM sodium ascorbate and 0.6% H_2O_2 were also prepared immediately before use. For each reaction, the concentration

of sodium ascorbate was 10-times higher that that of the Fe•EDTA, while the H_2O_2 remained constant at 0.05%.

The hydroxyl radical cleavage reactions were initiated by adding 15 μL of each working solution (0-10 mM Fe EDTA, 0-100 mM sodium ascorbate and 0.6% H_2O_2) in triplicate individual wells of a 96-well black polystyrene plate (Whatman) containing 5 pmol of annealed F21T in 135 μL of buffer (final concentrations: 0-833 μM Fe•EDTA, 0-8.3 mM sodium ascorbate, 0.05% H_2O_2, 27 nM F21T, 50 mM potassium cacodylate pH 7.4). The change in FAM fluorescence of F21T was monitored over time using a Perkin Elmer Victor ^{3}V plate reader (fluorescence filters: ex 485 ±13 nm, em 535± 25nm). The plate was shaken between reads to ensure homogenous samples throughout the reaction. After 15 minutes the reactions were quenched by the addition of 20 μL of 50% DMSO (Fisher).

A control reaction, monitoring the oxidative bleaching of fluorescein isothiocyanate (Baltimore Biological Laboratory) by hydroxyl radical was performed using the same experimental procedures as mentioned above but substituting 5 pmol free fluorophore for F21T.

Heat Denaturation and Reannealing

A portion of the F21T cleavage samples were heated and cooled to room temperature to reform the G-quadruplex of intact sequences. To accomplish this, the 96-well plate was affixed with an airtight seal (0.05 mm adhesive polyester seal- Uniseal, Whatman) to prevent evaporation and heated for 30 minutes in a 85 °C oven. The plate was allowed to cool to room temperature. This time and temperature was chosen ensure complete G-quadruplex denaturation while preventing the thermal distortion of the polystrene 96-well plate. After cooling, the plate was spun in a swinging bucket rotor centrifuge (Eppendorf, 5810R) for 5 minutes at 3000 rpm to remove condensation on the seal. The FAM fluorescence was recorded as before.

For PAGE analysis, 1.5 or 5 pmol F21T/F21T cleavage products were resuspended in 1x denaturing buffer (formamide, 5 mM EDTA), heated to 95 °C for 5 minutes and subjected to electrophoresis on a 20% denaturing polyacrylamide gel (1.5 mm thick). Standard Maxim Gilbert sequencing was performed on F21T for gel analysis (*43*). The fluorescent gel was visualized by a Typhoon Trio and quantified by GelQuant.Net (BiochemLabSolutions). The percent cleavage for each reaction mixture was calculated using equations 1 and 2:

$$FC_x = C_x/Tot_x \quad (1)$$

$$\text{percent cleavage} = [(FC_x\text{-}FC_o)/(100\text{-}FC_o)] \times 100 \quad (2)$$

where FC_x is the raw fraction cleaved for a given cleavage reaction, C_x is the sum of the integrated band intensity for all cleavage bands in a given lane, Tot_x is the sum of band intensity for all bands in a given lane, and FC_o is the raw fraction

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

cleaved determined for a control sample of F21T not subjected to the cleavage reaction conditions.

Synthesis of Mn-TMPyP4

The metallporphyrin was synthesized as previously described (*43*) with slight modification to the purification process. Briefly, meso-5,10,15,20-tetrakis-(N-methyl-4-pyridyl)porphine chloride (Sigma) (1 mmol) and 16 equivalents of $MnCl_2$ (Fisher) were mixed in 5 mL of 3:2 water/methanol with catalytic 2,6 lutidine at room temperature. The reaction was monitored by UV-visible spectroscopy (Varian Cary100) and judged complete after 24 hours by the disappearance of TMPyP4 Soret and Q bands and appearance of red-shifted Soret (462 nm) and Q bands (560 nm) (*44*). Excess Mn^{2+} was removed by adding 16 meq of analytical grade Chelex100 resin (sodium salt, BioRad), stirring for 1 hour at room temperature, and removing the supernatant. Metalloporphyrin concentration was determined spectroscopically: $\varepsilon_{462\ nm} = 101{,}000\ M^{-1}\ cm^{-1}$ (*44*).

Photochemical Cleavage of F21T

Being careful to avoid background photochemical reaction by working under subdued lighting, 20 μL reaction mixtures of 250 nM of annealed F21T and 0-2.5 μM TMPyP4 tosylate (Sigma) in 50 mM potassium cacodylate, pH 7.4 were incubated at room temperature for 30 minutes. The reaction mixtures were then placed in a 96-well fluorescent plate. Dark wells were blocked from light by a sheet of aluminum foil on top of the well. The plate was covered with a glass filter to eliminate UV light under 300 nm. The plate was then placed in a Luzchem photoreactor equipped with eight RPR-4190 lamps (emission centered at 420 nm) (Rayonet) for 0, 30, 60 or 120 minutes. After irradiation, 180 μL of a dissociation buffer (5.55 mM potassium cacodylate pH 7.4 and 10 μg calf thymus DNA per well) was added, and the wells were covered with foil to prevent further photoreactions. The plates were subjected to heat denaturation /reannealing as described above. The FAM fluorescence intensity was converted to apparent percent cleavage by applying equation 3:

$$\text{Apparent percent cleavage} = (F_x - F_{min})/(F_{max} - F_{min}) \times 100\% \quad (3)$$

where F_x is the observed fluorescent corrected for any background fluorescence observed for TMPyP4 in the absence of F21T; F_{min} is the average fluorescence of the corresponding samples that had not been irradiated (dark control); and F_{max} is the average fluorescence of a completely cleaved F21T construct (positive control). The apparent percent cleavage for a given reaction was determined by averaging the values for triplicate wells, and this together with the standard deviation, is reported in the Figures.

Selected reaction samples were also subjected to PAGE analysis. Prior to the required heat denaturation step, various samples were subjected to strand breakage without ethanol precipitation by directly adding 22 μL piperidine (10% final

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

concentration) and performing the heat denaturation step as above. Alternatively, piperidine heat treatment (10% v/v for 30 minutes at 90 °C) was performed on selected samples directly prior to PAGE analysis. For PAGE analysis, 1.5 to 5 pmol of dual labeled DNA cleavage products were lypholized to dryness. Piperidine treated samples were further subjected to repeated addition of water and lyophilization to remove traces of piperidine. DNA was then resuspended in 1x denaturing buffer (formamide, 5 mM EDTA), heated to 95 °C for 5 minutes and subjected to electrophoresis on a 20% denaturing polyacrylamide gel (1.5 mm thickness).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Mn(=O)TMPyP4 G-Quadruplex DNA Cleavage

Reaction mixtures of 16 μL containing 313 nM of F21T and 0-1.6 μM Mn•TMPyP4 in 62.5 mM potassium cacodylate, pH 7.4 were placed in individual wells of a 96 well fluorescent plate and preincubated for 30 minutes at room temperature away from ambient light to prevent background photocleavage. Cleavage reactions where initiated by introducing 4 μL of freshly prepared 500 μM $KHSO_5$ (Sigma) in water to the wells. Reactions were stopped between 0-15 minutes by the addition of 180 μL dissociation buffer (5.5 mM potassium cacodylate pH 7.4 and 10 μg calf thymus DNA/sample). Stopped samples were also covered with foil to prevent photoreaction. Analysis of these plates was carried out by fluorescence before and after heating/cooling as described above. Selected samples were also subjected to PAGE analysis before and after piperidine-heat treatment as described above.

TMPyP4 Photocleavage of Unlabeled Human Telomeric DNA

For each reaction, 10 μL reactions containing 150 nM of HT22, annealed as previously described, was mixed with 0-1.5 μM TmPyP4 in 50 mM potassium cacodylate, pH 7.4. Photoreactions were performed as previously described. Reactions were stopped by the addition of 40 μL dissociation buffer (final buffer concentration: 10 mM potassium cacodylate pH 7.4 and 10 μg calf thymus DNA per well). Stopped samples were also covered with foil to prevent further photoreactions.

For PAGE analysis, the DNA samples were resuspended in 1x denaturing buffer (formamide, 5 mM EDTA), heated to 95 °C for 5 minutes, ran on a 20% denaturing polyacrylamide gel, stained with SYBR Gold (Invitrogen) and visualized by UV illumination (BioRAD) and quantified by GelQuant.Net (BiochemLabSolutions).

Results and Discussion

To identify a G-quadruplex specific cleavage agent, a fluorescent-based assay was developed exploiting the diminished quenching by energy transfer of a dual donor-acceptor labeled G-quadruplex deoxyoligonucleotide. These dual-labeled probes have been previously used in FRET-based melting

temperature determinations of G-quadruplex DNA, and as a means to screen potential G-quadruplex ligands for their ability to stabilize these structures (*41*, *42*, *46*). However, these probes have not been previously explored as a means to screen compounds for their ability to cleave G-quadruplex DNA. In exploring this application of these dual labeled probes, certain critical features had to be addressed. For example, while some cleavage events might be expected to lead to spontaneous separation of the fluorophores (e.g., cleavage on the base to which the fluorophore is attached), other cleavage events would in essence simply convert the intramolecular G-quadruplex into a bimolecular G-quadruplex. If the resulting bimolecular quadruplex were stable under the cleavage reaction conditions, no change in fluorescence would result. Thus, a separate heating/reannealing step might be required in order to reveal all cleavage events through the essentially irreversible G-quadruplex denaturation of these bimolecular quadruplexes. This leads to a second consideration in cases where the G-quadruplex cleavage is the result of the interaction of a ligand binding to the quadruplex. If such a ligand stabilizes the nicked G-quadruplexes such that they do not undergo dissociation even when heated, these cleavage events would not give rise to increased fluorescence. Thus, one also has to consider a means to dissociate any ligands from the quadruplex prior to the heat denaturation/reannealing procedure. In the sections below we describe how each of these issues were successfully resolved to afford an assay that can be carried out quickly, inexpensively, and in 96-well plate format.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Fe•EDTA Cleavage of G-Quadruplex DNA

The Fenton reaction is widely used in nucleic acid research to footprint binding sites of compounds, proteins and other secondary structures on DNA (*47*, *48*):

$$\text{Ascorbate} + \text{EDTA}\bullet\text{Fe}^{3+} \rightarrow \text{Oxidized ascorbate} + \text{EDTA}\bullet\text{Fe}^{2+}$$
$$\text{EDTA}\bullet\text{Fe}^{2+} + H_2O_2 \rightarrow \text{EDTA}\bullet\text{Fe}^{3+} + \ \bullet OH + OH^-$$

The hydroxyl radical is a small diffusible oxidant which produces a variety of cleavage products, including frank strand breakage due to hydrogen abstraction from the deoxyribose and other nucleobase modifications in single and double stranded DNA (*49*, *50*). DNA cleavage by hydroxyl radical is largely independent of base identity, but is affected by the deoxyribose hydrogen atoms' accessibility, yet hydroxyl radical cleavage is not widely used in studying G-quadruplex DNA (*51*). Notably, a perylene diimide –EDTA•Fe^{2+} ligand was found to selectively cleave quadruplex DNA in the presence of H_2O_2 at DNA sites proximal to the perylene diimide's preferred binding site (*32*). Due to its well-established mode of DNA cleavage and its known ability to cleave structured DNA, Fe•EDTA-generated hydroxyl radical was selected to for initial studies to determine the feasibility of monitoring G-quadruplex cleavage by fluorescence.

We employed a dual-labeled oligonucleotide probe, F21T, consisting of 21 bases of human telomeric GGGTTA repeats labled at the 5′-end with 6-carboxyfluorescein (FAM) and at the 3′-end by tetramethylrhodamine (TAM).

Previous work has established the similarity in fold and stability of the quadruplex form of F21T to the unmodified human telomeric G-quadruplex structures (*41*), and the utility of F21T for screening ligands for their abiltiy to bind to and stabilize this quadruplex (*46*).

When exposed to Fenton reaction conditions (8.3–83 μM Fe•EDTA, 10 equiv sodium ascorbate, 0.05% H_2O_2) in potassium cacodylate buffer (50 mM, pH 7.4), the FAM fluorescence of F21T (27 nM) increases as a function of time, finally reaching a plateau (data not shown). This increase in FAM fluorescence is accompanied by a decrease in the FRET signal observed when exciting the FAM fluorophore and monitoring the TAM emmission intensity. There is an increase in FAM fluorescence with increasing Fe•EDTA concentration, with a plateau in fluorescence intensity at higher Fe•EDTA concentrations (Figure 3A). The overall increase in final FAM fluorescence at high Fe•EDTA concentration is ~3-fold.

Two possibilities were considered to explain the relatively small change in FAM fluorescence. One possibility was that the hydroxyl radical cleavage of the compact G-quadruplex structure might be relatively inefficient. Alternatively, cleavage events might not be revealed under these conditions due to the stability of bimolecular quadruplexes resulting from cleavage events.

To explore the second possibility, an additional step was added to irreversibly denature any bimolecular quadruplexes formed during hydroxyl radical cleavage. This involved carefully heating the 96-well plate followed by slow cooling in order to allow the intact oligonucleotide to refold (Figure 4). The low concentration of DNA (nM range) minimizes the rate of intermolecular G-quadruplex reformation. In order to avoid any additional cleavage due to hydroxyl radical generation during this step, DMSO (5 % final concentration) was added to each well. In control experiments, this concentration of DMSO was sufficient to prevent any bleaching of free fluorescein due to hydroxyl radical produced under these conditions. Additional FRET melting experiments demonstrated that this concentration of DMSO does not significantly affect the stability of the G-quadruplex formed by F21T (data not shown).

The FAM fluorescence of F21T treated with Fe•EDTA followed by this heat/cool cycle was increased at all concentrations of Fe•EDTA (Figure 3A). We attribute this increase to the dissociation of nicked G-quadruplexes that are stable under the reaction conditions, but which undergo dissociation during the heat/reannealing step.

In order to confirm that the increased FAM fluorescence corresponds to G-quadruplex cleavage, these samples were also analyzed by PAGE and visualized by fluorescence (Figure 3B). Bands corresponding to nonsequence-dependent cleavage were observed for all samples treated with Fe•EDTA, and the intensity of these bands increased with increasing Fe•EDTA concentration. Two additional bands running just below intact F21T were also observed in all samples treated with Fe•EDTA (arrow in Figure 3B). Because these bands were not seen when the gel was visualized using the TAM fluorescence, these were attributed to products in which the TAM, or the TAM plus *N*-(5-aminopentyl)acrylamide linker, are cleaved from the 3′ end of the oligonucleotide. The correlation between the percent cleavage determined by PAGE analysis and the FAM fluorescence

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

intensity (Figure 3C) demonstrates that the intensity of the FAM fluorescence signal is directly related to the amount of cleaved F21T.

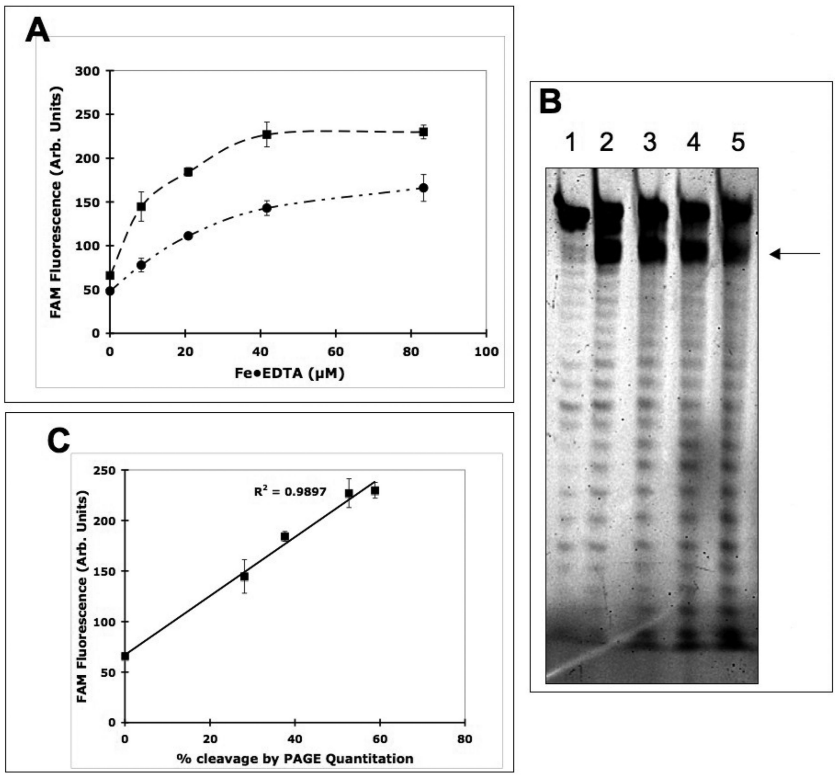


Figure 3. (A) Solution fluorescence of F21T treated with increasing concentrations of Fe•EDTA in the presence of sodium ascorbate and H2O2 for 15 min at room temperature before (circles) or after (squares) subjecting the samples to heating and slow cooling. (B) Fluorescent image (FAM excitation/emission) of a 20% denaturing PAGE analysis of the F21T samples subjected to hydroxyl radical cleavage in the presence of 0 μM (lane 1), 8.3 μM (lane 2), 20.8 μM (lane 3), 41.6 μM (lane 4), or 83.3 μM (lane 5) Fe•EDTA after the heat/cool cycle. The arrow indicates bands corresponding to cleavage of the TAM or the TAM plus N-(5-aminopentyl)acrylamide linker from the DNA. (C) A plot of the percent cleavage of F21T treated with different concentrations of Fe•EDTA as determined by PAGE analysis shown in (B) versus the fluorescence intensity shown by the squares in (A).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

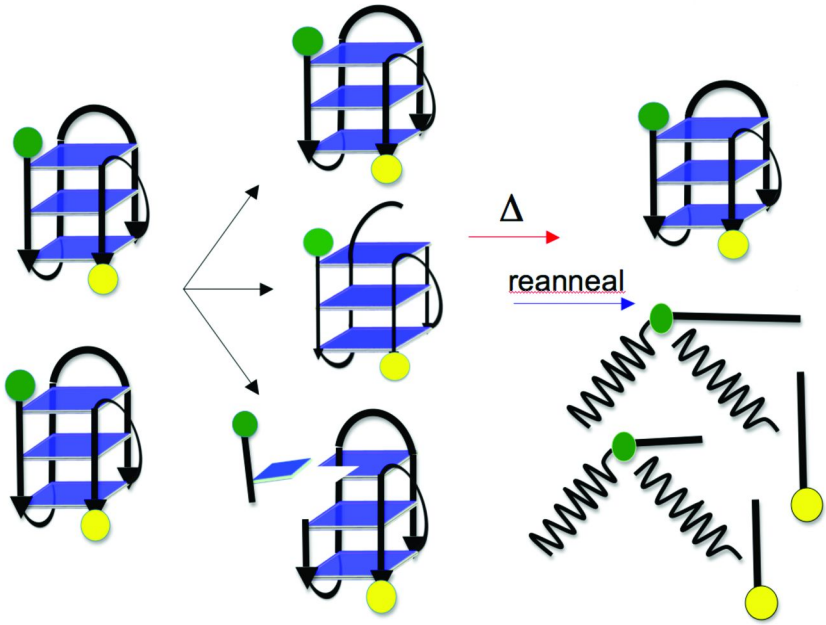


Figure 4. Cartoon depicting G-qaudruplex DNA cleavage events leading to either spontaneous nucleotide dissociation or a stable biomoleclar quadruplex. Heat denaturation of a bimolecular G-quadruplex formed upon cleavage is essentially irreversible under the reaction conditions.

TMPyP4 Photocleavage of F21T

The photochemical cleavage of G-quadruplex DNA by TMPyP4 (*30*, *31*) was used to further test this fluorescence-based quadruplex cleavage assay. Although the mechanism of photocleavage of G-quadruplex DNA by TMPyP4 is unknown, porphyrin-based duplex DNA photocleavage has been proposed to occur primarily via a singlet oxygen (*52*, *53*), although there is also evidence for cleavage due to hydrogen abstraction from the deoxyribose (*54*).

Samples of 250 nM F21T were pre-incubated with various concentrations of TMPyP4 for 30 minutes in a 96-well plate. The 96-well plate was covered with a glass plate to filter light below 300 nm and to prevent sample evaporation during irradiation. The plate was then placed in a photoreactor equipped with lamps with a maximal emission at 420 nm. After various irradiation times, a dissociation buffer consisting of 10 μg calf thymus DNA/well, final $[K^+] = 10$ mM) was added to each well. The excess duplex DNA was added to sequester TMPyP4 from the quadruplex in order to prevent porphyrin-mediated quenching of the oligonucleotide fluorophores and any photocleavage due to adventitious light. The final lower salt concentration allowed subsequent direct PAGE analysis without adversely affecting the stability of the G-quadruplex DNA. After the addition of the dissociation buffer, the wells were covered with foil.

After carefully subjecting the plate to the same heat/reannealing step as described above, the FAM fluorescence was measured and converted to apparent percent cleavage as described in the experimental section (equation 3).

As shown in Figure 5, the apparent percent cleavage of F21T by photoactivated TMPyP4 determined by this assay shows a concentration dependence up to approximately 750 nM TMPyP4 (i.e., at 3:1 ratio of TMPyP4 to F21T). At higher TMPyP4 concentrations, there is little apparent additional cleavage. At the highest TMPyP4 concentration (2.5 μM) and irradiation time (120 min) the apparent cleavage of F21T is approximately 60%.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

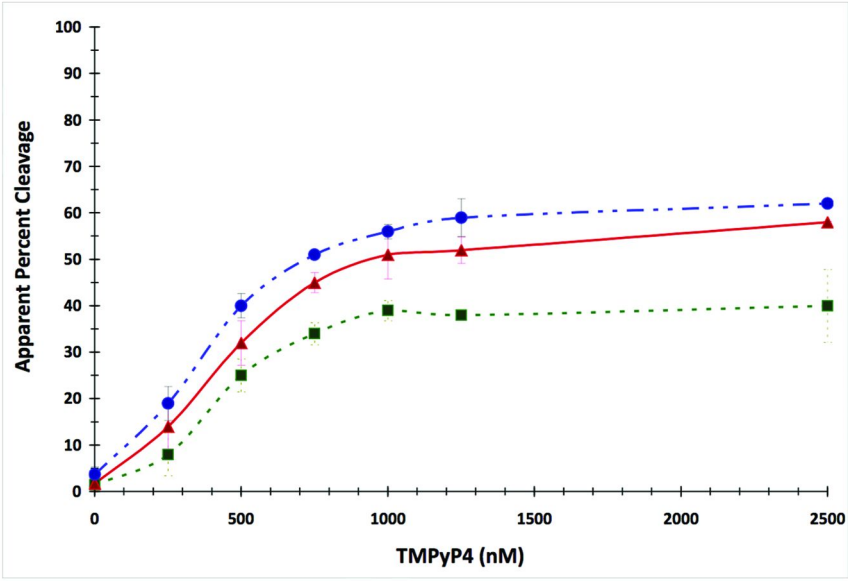


Figure 5. Apparent percent cleavage determined from fluorescence change of F21T using equation 3 (y-axis) after photolysis for 30 (squares), 60 (triangles) or 120 mins (circles) in the presence of various concentrations of TMPyP4 (x-axis). Error bars represent the standard deviation in apparent precent cleavage for each condition (n=3).

PAGE analysis was also carried out on the TMPyP4 photochemical F21T cleavage reactions (Figure 6). Samples that were subjected to PAGE prior to piperidine-heat treatment revealed multiple bands that were not well resolved. At the highest TMPyP4 concentrations, the bands are very faint, presumably due to photobleaching of the FAM of F21T at these TMPyP4 concentrations. After piperidine-heat treatment, the photocleavage bands are better resolved and indicate site selective G cleavage at the external tetrads. Previous studies of TMPyP4 photocleavage of G-quadruplex DNA have been limited to duplex-quadruplex (*30*) and termolecular G-quadruplex (*31*) structures. In both instances, the observed photocleavage has been attributed to stacking of TMPyP4 on terminal G-tetrads leading to preferential cleavage of the Gs comprising these tetrads. The

photocleavage pattern that we observe here is commensurate with these previous studies. In addition, a similar photocleavage pattern has recently been reported for the quadruplex formed by a 3′-FAM-labeled human telomeric oligonucleotide in the presence of a zinc-phthalocyanine (*35*).

The extent of photocleavage of F21T by TMPyP4 was quantified by the PAGE analysis as described for the Fe•EDTA cleavage studies. Since there was clearly extensive loss of the FAM signal at the highest TMPyP4 concentration, only the lanes corresponding to lower TMPyP4 concentrations were quantified. Figure 7 demonstrates the excellent agreement between the solution fluorescence data and the extent of photocleavage determined by PAGE. Furthermore, an unlabeled oligonucleotide HT22 was subjected to the same photocleavage conditions and the products analyzed by PAGE with SYBR Gold visualization of the bands (data not shown). The photocleavage was quantified by the decrease in the bands for full-length DNA as a function of TMPyP4 concentraion, giving results similar to those shown in Figure 7. This correlation between solution and PAGE analysis demonstrates the fluorescence-based assay can be used to accurately quantify G-quadruplex DNA photocleavage under conditions where both extensive cleavage (i.e., multiple "hit" regime) and photobleaching/oxidation fluorophores are avoided.

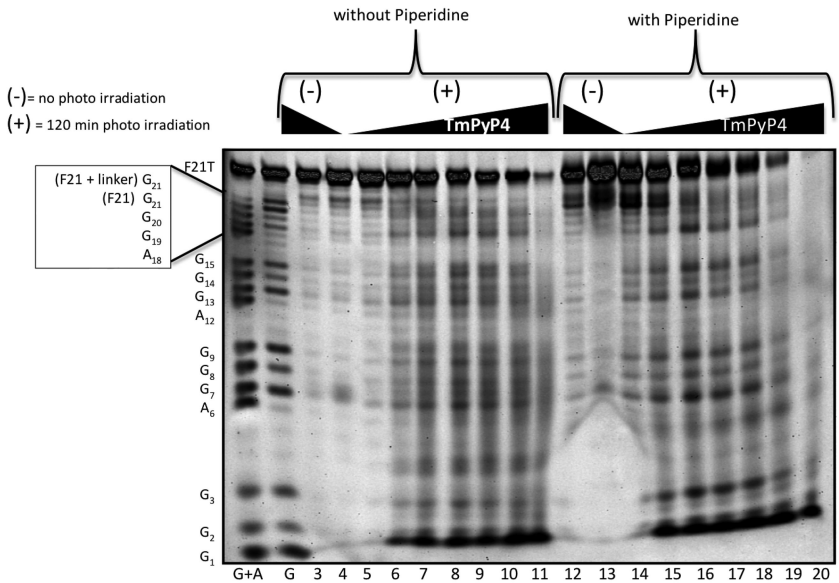


Figure 6. Fluorescent image (FAM excitation/emission) of a 20% denaturing PAGE analysis of the F21T samples irradiated for 120 min in the presence of various concentrations of TMPyP4 before and after piperidine/heat treatment. Reactions contained 0 nM (lanes 5,14), 250 nM (lanes 6,15), 500 nM (lanes 7,16), 750 nM (lanes 8,17), 1 μM (lanes 9,18), 1.25 μM (lanes 10,19) or 2.5 μM (lanes 11, 20) of TMPyP4. Control samples that were not irradiated contained 0 μM (lanes 4, 13) or 2.5 μM TMPyP4 (lanes 3, 12).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

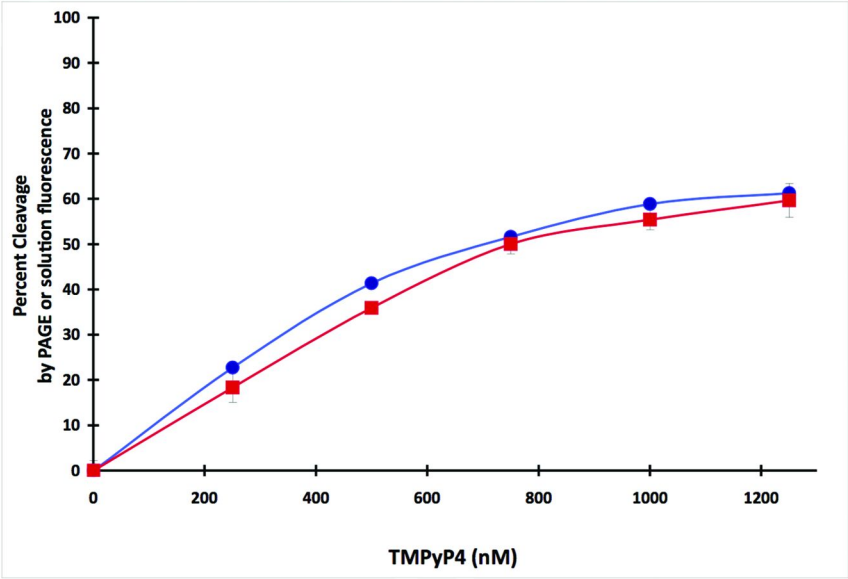


Figure 7. Comparison of apparent percent cleavage as determined by the solution fluorescence intensity using equation 3 (squares) and the percent cleavage determined by subsequent PAGE analysis using equation 2 (circles) for samples of F21T incubated with various concentrations of TMPyP4 and subjected to photolysis for 120 min. Error bars represent the standard deviation in apparent precent cleavage or percent cleavage for each condition (n=3).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Mn(=O)TMPyP4 Oxidative Cleavage of F21T

The oxo-metalloporphyrin Mn(=O)TMPyP4, generated *in situ* by oxidation of Mn•TMPyP4, has been reported to cleave the G-quadruplex formed by the human telomeric sequence (*34*). We find that F21T is cleaved very efficiently when treated with Mn•TMPyP4 and potassium peroxysulfate. F21T was pre-incubated with various concentrations of Mn•TMPyP4, and cleavage initiated by addition of potassium peroxysulfate (100 μM). The cleavage reactions were stopped by the addition of dissociation buffer (10 μg calf thymus DNA/sample) and the FAM fluorescence was evaluated after heating and re-annealing and converted to apparent percent cleavage as shown above. As shown in Figure 8, even at very short reaction times, there is significant apparent cleavage of F21T, as evidenced by increased FAM fluorescence. Maximum cleavage of approximately 75 % occurred at 750 nM of Mn•TMPyP4 after just 15 minutes, the longest reaction time employed. At higher Mn•TMPyP4 concentrations, the apparent cleavage determined by FAM fluorescence is reduced. Control reactions in which F21T was incubated with $KHSO_5$ in the absence of Mn•TMPyP4 did not show any changes in FAM fluorescence.

To further explore this oxidative cleavage of F21T by Mn(=O)TMPyP4, the cleavage reactions mixtures were subjected to PAGE analysis. Cleavage reaction mixtures exposed to various concentrations of Mn•TMPyP4 and $KHSO_5$ for 15 min were analyzed before and after piperidine/heat treatment (Figure 9A). Samples that were not treated with piperide/heat prior to PAGE showed multiple diffuse bands of higher mobility than full-length F21T. In addition, bands that migrate slower than full-length F21T, including material that remained in the loading well, were also observed (arrows, Figure 9A). These low-mobility bands are also observed when the gel is visualized by TAM fluorescence (data not shown), indicating that these are due to adducts of the full length DNA. At lower MnTMPyP4 concentrations, the intensity of both the higher- and lower-mobility bands increased; however, the total signal for all bands in each lane decreased at higher Mn•TMPyP4 concentration, most likely due to bleaching of the FAM fluorophore under the reaction conditions.

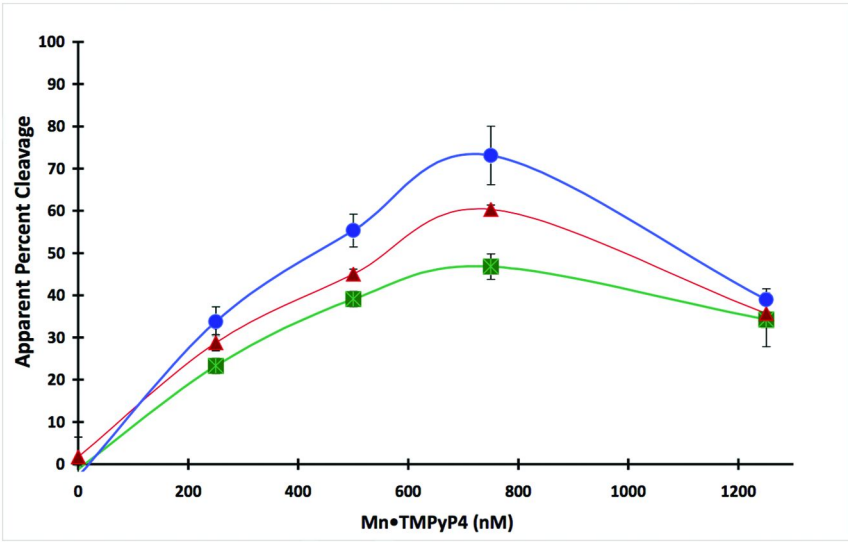


Figure 8. Apparent percent cleavage calculated from the fluorescence change according to equation 3 of F21T treated with various concentrations of Mn•TMPyP4 in the presence of 100 μM $KHSO_5$ for 1 (squares), 5 (triangles) or 15 min (circles). Error bars represent the standard deviation in apparent precent cleavage for each condition (n=3).

After piperidine-heat treatment, the bands were better resolved (Figure 9A). While weak bands remained at the loading well for reaction mixtures containing Mn•TMPyP4, most of the lower mobility DNA appeared as slightly diffuse bands that migrate just above intact F21T. There is a noticeable decrease in the intensity of the intact F21T band with increasing Mn•TMPyP4 concentration. Bands corresponding to cleavage of the F21T at the second T of each TTA loop as well as cleavage of the two G's comprising the 5′ terminal G-tetrad are

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

observed (Figure 9A). The location of these major cleavage sites mapped on to the potassium-stabilized hybrid 3+1 fold of the human telomeric G-quadruplex (*55*) is shown in Figure 9B. This cleavage pattern is similar to that previously reported for Mn(=O)TMPyP4 with a different human telomeric G-quadruplex substrate (*45*) in that the main cleavage occurs at G residues of a single tetrad and the T residues in loop regions.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

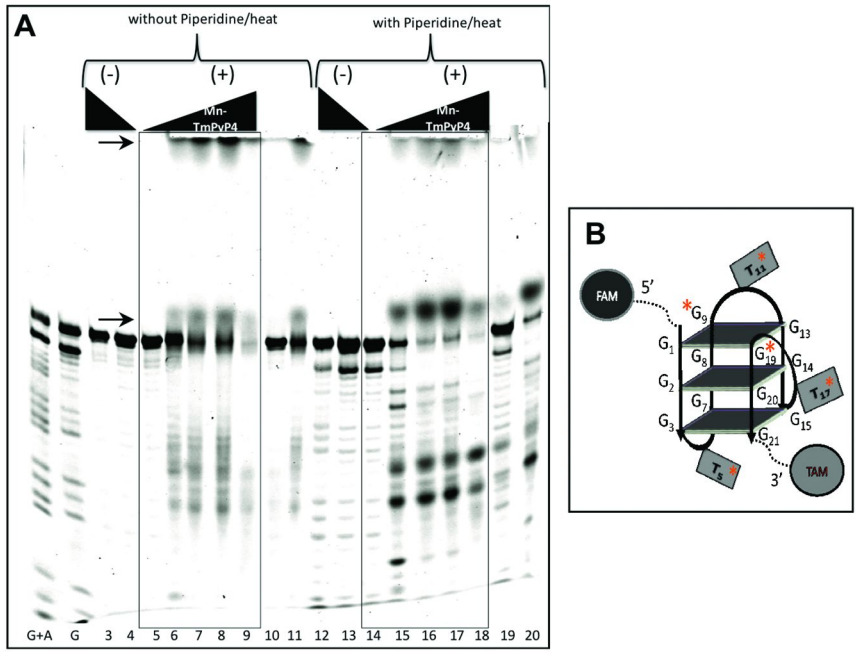


Figure 9. (A) Fluorescent image (FAM excitation/emission) of a 20% denaturing PAGE analysis of the F21T samples treated for 15 min with various concentrations of Mn•TMPyP4 in the presence of 100 μM $KHSO_5$ before and after piperidine/heat treatment. Reactions contained 0 nM (lanes 5,14), 250 nM (lanes 6,10,15,19), 500 nM (lanes 7,16), 750 nM (lanes 8,17), 1.25 μM (lanes 9,18) of Mn•TMPyP4. Samples in lanes 10 and 19 did not undergo the heat/reanneal cycle. Reactions in lanes 11 and 20 were treated with 500 nM Mn•TMPyP4 in the presence of 100 μM $KHS0_5$ for only 1 min. Control samples contained F21T with either 0 nM (lanes 4 and 13) or 2500 nM (lanes 3 and 12) Mn•TMPyP4 in the absence of $KHSO_5$. (B) Cartoon of antiparallel/parallel quadruplex structure formed by $GGG(TTAGGG)_3$ in the presence of potassium. Sites of Mn(=O)TMPyP4 cleavage, as determined by PAGE analysis are marked with a "".*

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

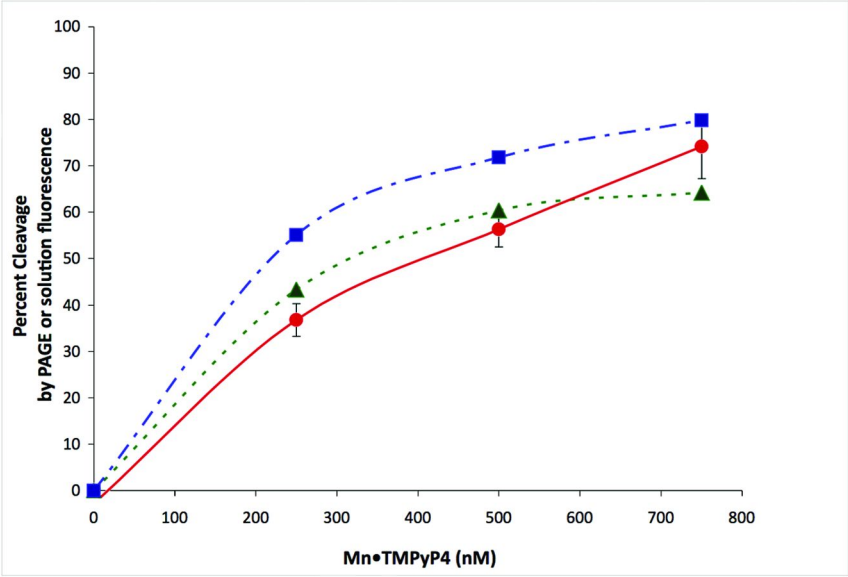


Figure 10. Comparison of apparent percent cleavage as calculated from the solution fluorescence according to equation 3 (circles) and the percent cleavage determined by subsequent PAGE analysis using equation 2 before (triangles) or after (squares) piperidine/heat treatment for samples of F21T incubated with various concentrations of Mn•TMPyP4 in the presence of 100 μM $KHSO_5$ for 15 min. Error bars represent the standard deviation in apparent precent cleavage for each condition (n=3).

Quantification of the PAGE bands for intact F21T (including adduct bands) as a fraction of all bands versus Mn•TMPyP4 concentration in the cleavage reactions is shown in Figure 10, plotted along with the apparent percent cleavage determined by solution fluorescence intensity. There is good overall agreement with the solution and PAGE results at the lower Mn•TMPyP4 concentrations prior to piperidine/heat treatment. However, there is clearly more cleavage after piperidine/heat treatment at all Mn•TMPyP4 concentrations. Even when the solution fluorescence data for the highest concentrations of Mn•TMPyP4 are excluded due to the bleaching of the FAM fluorophore, it is apparent that certain covalent modifications of the F21T are only revealed as cleavage events after piperidine/heat treatment. While this requirement for piperidine/heat treatment for G-quadruplex cleavage by Mn(=O)TMPyP4 has been attributed to C1′ hydrogen atom abstraction (*45*), here it appears that adduct formation may also play a role in the G-quadruplex cleavage chemistry of Mn(=O)TMPyP4.

Conclusions and Future Prospects

Here we show that dual-labeled G-quadruplex DNA-forming oligonucleotide probes are useful tools for probing G-quadruplex DNA cleavage. Specifically, despite some limitations of the particular fluorophores and linkers, the F21T probe

that we employed was adequate for revealing cleavage resulting from a variety of chemistries including hydroxyl radical cleavage, photochemical cleavage due to TMPyP4, and oxidative cleavage by Mn(=O)TMPyP4. Although some cleavage events give rise to an immediate increase in FAM fluorescence, many cleavage events are "silent" until the probe undergoes heating/reannealing during which stable bimolecular G-quadruplex structures formed by cleavage events are essentially irreversibly denatured. There are some limitations associated with the use of this specific F21T probe, such oxidative bleaching of the FAM fluorophore and liability of the TAM linker. However, these limitations are generally restricted to conditions where extensive cleavage has already occurred and could be addressed in the future with FRET pairs that are more robust. Still, even with these limitations, the correlation of the increase in FAM fluorescence with cleavage, as verified by PAGE analysis, serves to validate the utility of this assay under a range of cleavage conditions. Only in the case of covalent modifications of G-quadruplex DNA that require post-reaction processing, such as piperidine/heat treatment of Mn(=O)TMPyP4-modified DNA, does this solution assay underestimate the covalent modifications leading to G-quadruplex DNA cleavage.

Agents that selectively cleave G-quadruplexes will be critical tools for establishing the existence of these structures in cells. More importantly, photochemically activated G-quadruplex cleavage agents will provide an essential ability to probe for both the temporal and spatial (e.g., DNA sequence) aspects of G-quadruplex formation and resolution. Given the propensity of G-rich DNA sequences to form these structures in vitro, studies employing these photocleavage agents will be critical in ruling out artifacts related to the *post facto* formation of these structures by correlating their formation with the various biological processes that they have been proposed to mediate.

There are also potential clinical applications for G-quadruplex cleavage agents. Existing DNA cleavage agents have proven effectiveness as anticancer drugs, but lack selectivity. By targeting only those DNA sequences that form G-quadruplex structures, these cleavage agents could be much more selective, particularly if these G-qaudruplex structures are over-represented in diseased cells. In an ideal situation, such agents would target the unique G-quadruplex structure(s) formed by a DNA sequence (e.g., mutation in tumor cells or proviral DNA in HIV-infected cells) that only exists in the diseased cell. Of course, such an approach is not limited to DNA quadruplexes, and RNA quadruplexes could also serve as targets for such highly selective cleavage agents. The compact nature and structural diversity of G-quadruplexes make this approach plausible, in which a *small* molecule effectively targets a unique DNA (or RNA) sequence of more than 15 bases. We believe that the assay we have presented will be a valuable tool in identifying G-quadruplex cleavage agents for these various applications.

Acknowledgments

We are grateful for support from the Robert Welch Foundation (F1298).

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

References

1. Galm, U.; Hager, M. H.; Van Lanen, S. G.; Ju, J. H.; Thorson, J. S.; Shen, B. *Chem. Rev.* **2005**, *105*, 739–758.
2. Guirouilh-Barbat, J.; Zhang, Y. W.; Pommier, Y. *Mol. Cancer Ther.* **2009**, *8*, 1985–1994.
3. Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
4. Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L. S.; Johnson, F. B. *Nucleic Acids Res.* **2008**, *36*, 144–156.
5. Matilla, I.; Alfonso, C.; Rivas, G.; Bolt, E. L.; de la Cruz, F.; Cabezon, E. *J. Biol. Chem.* **2010**, *285*, 17537–17544.
6. Cahoon, L. A.; Seifert, H. S. *Science* **2009**, *325*, 764–767.
7. Parkinson, G. N. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; The Royal Society of Chemistry: Cambridge, U.K., 2006; pp 1–30.
8. Rachwal, P. A.; Brown, T.; Fox, K. R. *Biochemistry* **2007**, *46*, 3036–3044.
9. Olsen, C. M.; Gmeiner, W. H.; Marky, L. A. *J. Phys. Chem. B* **2006**, *110*, 6962–6969.
10. Zheng, K. W.; Chen, Z.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2010**, *38*, 327–338.
11. Kumar, N.; Maiti, S. *Nucleic Acids Res.* **2008**, *36*, 5610–5622.
12. Paramasivan, S.; Rujan, I.; Bolton, P. H. *Methods* **2007**, *43*, 324–331.
13. Dai, J. X.; Carver, M.; Yang, D. Z. *Biochimie* **2008**, *90*, 1172–1183.
14. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
15. Qin, Y.; Hurley, L. H. *Biochimie* **2008**, *90*, 1149–1171.
16. Kendrick, S.; Hurley, L. H. *Pure Appl. Chem.* **2010**, *82*, 1609–1621.
17. Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441–471.
18. Paeschke, K.; McDonald, K. R.; Zakian, V. A. *FEBS Lett.* **2010**, *584*, 3760–3772.
19. Yang, D. Z.; Okamoto, K. *Future Med. Chem.* **2010**, *2*, 619–646.
20. Johnson, J. E.; Smith, J. S.; Kozak, M. L.; Johnson, F. B. *Biochimie* **2008**, *90*, 1250–1263.
21. Parker, S. C. J.; Hansen, L.; Abaan, H. O.; Tullius, T. D.; Margulies, E. H. *Science* **2009**, *324*, 389–392.
22. Maizels, N. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055–1059.
23. Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901–2907.
24. Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
25. Lipps, H. J.; Rhodes, D. *Trends Cell Biol.* **2009**, *19*, 414–422.
26. Huppert, J. L. *FEBS J.* **2010**, *277*, 3452–3458.
27. Muller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. *Nat. Chem.* **2010**, *2*, 1095–1098.
28. Kim, M. Y.; Duan, W. H.; Gleason-Guzman, M.; Hurley, L. H. *J. Med. Chem.* **2003**, *46*, 571–583.
29. Duan, W. H.; Rangan, A.; Vankayalapati, H.; Kim, M. Y.; Zeng, Q. P.; Sun, D. K.; Han, H. Y.; Fedoroff, O. Y.; Nishioka, D.; Rha, S. Y.; Izbicka, E.; Von Hoff, D. D.; Hurley, L. H. *Mol. Cancer Ther.* **2001**, *1*, 103–120.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

30. Han, F. X. G.; Wheelhouse, R. T.; Hurley, L. H. *J. Am. Chem. Soc.* **1999**, *121*, 3561–3570.
31. Han, H. Y.; Langley, D. R.; Rangan, A.; Hurley, L. H. *J. Am. Chem. Soc.* **2001**, *123*, 8902–8913.
32. Tuntiwechapikul, W.; Salazar, M. *Biochemistry* **2001**, *40*, 13652–13658.
33. Pothukuchy, A.; Mazzitelli, C. L.; Rodriguez, M. L.; Tuesuwan, B.; Salazar, M.; Brodbelt, J. S.; Kerwin, S. M. *Biochemistry* **2005**, *44*, 2163–2172.
34. Vialas, C.; Pratviel, G.; Meunier, B. *Biochemistry* **2000**, *39*, 9514–9522.
35. Zheng, K. W.; Zhang, D.; Zhang, L. X.; Hao, Y. H.; Zhou, X. A.; Tan, Z. *J. Am. Chem. Soc.* **2011**, *133*, 1475–1483.
36. Monchaud, D.; Teulade-Fichou, M. P. *Org. Biomol. Chem.* **2008**, *6*, 627–636.
37. Tuntiwechapikul, W.; David, W. M.; Kumar, D.; Salazar, M.; Kerwin, S. M. *Biochemistry* **2002**, *41*, 5283–5290.
38. Lee, S. P.; Han, M. K. *Methods Enzymol.* **1997**, *278*, 343–363.
39. Pu, F.; Hu, D.; Ren, J. S.; Wang, S.; Qu, X. G. *Langmuir* **2010**, *26*, 4540–4545.
40. Biggins, J. B.; Prudent, J. R.; Marshall, D. J.; Ruppen, M.; Thorson, J. S. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 13537–13542.
41. Mergny, J. L.; Maurizot, J. C. *ChemBioChem* **2001**, *2*, 124–132.
42. Guyen, B.; Schultes, C. M.; Hazel, P.; Mann, J.; Neidle, S. *Org. Biomol. Chem.* **2004**, *2*, 981–988.
43. Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499–560.
44. Borovkov, V. V.; Lintuluoto, J. M.; Inoue, Y. *Synlett* **1999**, *61*–62.
45. Bernadou, J.; Pratviel, G.; Bennis, F.; Girardet, M.; Meunier, B. *Biochemistry* **1989**, *28*, 7268–7275.
46. De Cian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teulade-Fichou, M. P.; Lacroix, L.; Mergny, J. L. *Methods* **2007**, *42*, 183–195.
47. Tullius, T. D. *Nature* **1988**, *332*, 663–664.
48. Jain, S. S.; Tullius, T. D. *Nature Protocol.* **2008**, *3*, 1092–1100.
49. Pogozelski, W. K.; McNeese, T. J.; Tullius, T. D. *J. Am. Chem. Soc.* **1995**, *117*, 11673–11679.
50. Pogozelski, W. K.; Tullius, T. D. *Chem. Rev.* **1998**, *98*, 1089–1107.
51. Paramasivan, S.; Bolton, P. H. *J. Biomol. Struct. Dyn.* **2007**, *24*, 709.
52. Aoki, K.; Ishikawa, Y.; Oyama, M.; Tomisugi, Y.; Uno, T. *Chem. Pharm. Bull.* **2003**, *51*, 899–903.
53. Ishikawa, Y.; Yamakawa, N.; Uno, T. *Bioorg. Med. Chem.* **2007**, *15*, 5230–5238.
54. Croke, D. T.; Perrouault, L.; Sari, M. A.; Battioni, J. P.; Mansuy, D.; Helene, C.; Trung, L. D. *J. Photochem. Photobiol., B* **1993**, *18*, 41–50.
55. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch002

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Chapter 3

Biophysical Studies of the Structure, Stability, and Ligand Binding Properties of G-Quadruplex DNA: Thoughts and Comparisons of the K-ras, c-MYC, and Bcl-2 Oncogene Promoter Sequence Quadruplexes

Jamie M. Dettler and Edwin A. Lewis*

Department of Chemistry, Mississippi State University, Mississippi State, MS 39762
***E-mail: elewis@chemistry.msstate.edu**

Expression of the K-ras proto-oncogene is a well known hallmark of pancreatic and other cancers. We report here the results of a biophysical characterization of model human K-ras promoter sequence constructs having either the wild type, WT, 30-mer sequence or four mutant 26-mer sequences in which three G→T mutations were introduced in the second, third, forth, or fifth G-run starting from the 5′ end of the WT construct. The model WT and mutant K-ras constructs were studied using analytical ultracentrifugation (AUC), circular dichroism spectroscopy (CD), differential scanning calorimetry (DSC), isothermal titration Calorimetry (ITC), and DMS footprinting. Analytical ultracentrifugation experiments demonstrated that the WT K-ras promoter sequence folds into a compact structure with a sedimentation coefficient of 2.4 S. This result is consistent with the calculated sedimentation coefficient for the proposed K-ras G-quadruplex structure and similar to the sedimentation coefficients found for c-MYC and Bcl-2 quadruplex structures. The CD and DMS footprinting results indicate that the WT and two mutant (Mut1-2-3-4 and Mut1-2-3-5) K-ras promoter sequences fold into intramolecular, parallel-stranded, G-quadruplexes. The DMS cleavage patterns observed for the WT sequence are most similar to those for

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Mut1-2-3-5 sequence indicating that the ensemble of K-ras WT G-quadruplex motifs includes a quadruplex incorporating a kinked out base (T7) and a large lateral or end loop having as many as 11 unstructured bases. The two mutant sequences (Mut1-3-4-5 and Mut1-2-4-5) exhibit CD and DMS footprinting spectra that indicate essentially no G-quadruplex formation. DSC experiments demonstrate that the WT K-ras sequence folds into a mixture of at least two thermodynamically unique conformers with the two overlapping melting transitions having *T*m values of 58.5 °C and 71.2 °C. In comparison to simpler G-quadruplex forming sequences, e.g. c-MYC or even Bcl-2, formation of stable K-ras promoter sequence G-quadruplexes requires the incorporation of large lateral or end loops and/or a corner backbone kinked out base.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Introduction

Pancreatic Cancer is the 4th leading cause of cancer death in the United States. In 2006 there were 33,730 new diagnoses of pancreatic cancer, and 32,300 deaths resulting from pancreatic cancer (*1*). The prevalence of recurrence of pancreatic tumors after surgery is extremely high and the lack of efficacy of current anti-pancreatic cancer therapeutics highlights the importance of developing new molecular targets for pancreatic cancer treatment (*2*). The K-ras proto-oncogene is mutated and overexpressed in more than 90% of pancreatic tumors making it an attractive target for anticancer drug design (*3–6*). The overexpression of the K-ras oncogene is associated with cancers that are more aggressive and have shorter survival rates such as lung, breast, prostate, colorectal, skin, and most importantly pancreatic cancer (*3–6*). The one year survival rate of a patient with pancreatic cancer is less than 20% and five year survival rate less than 5% (*1*).

The drug targeting of unique higher-order DNA structures, e.g. G-quadruplex motifs represent alternatives to DNA sequence recognition. Binding small molecules to quadruplex DNA has been demonstrated to downregulate the expression of oncogenes such as c-MYC and Bcl-2 *in vivo* (*7*). The G-quadruplex DNA target has shown promise in increasing binding specificity for anti-cancer agents over small molecules that bind in a sequence specific manner to duplex DNA (*8–11*). Recent studies have suggested that the *K-ras* oncogene contains a nuclease hypersensitivity element upstream from its promoter region that is capable of forming a G-quadruplex (*12*). At this point it is uncertain whether or not these sequences can form stable unimolecular structures. Although the human K-ras promoter sequence is rich in guanine, G-quadruplexes formed by this sequence would necessarily incorporate large destabilizing lateral loop regions and/or "*kinks*" in the corner backbone between the stacked G-tetrads. Because of the role this gene plays in pancreatic cancer, it is essential to have a clear understanding of whether or not this promoter forms a G-quadruplex *in vivo*. The impact of these large loops on quadruplex stability, as well as the potential for

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

forming a physiologically relevant quadruplex in the *K-ras* NHE is the focus of this study.

In this study we have performed a biophysical characterization of the WT and four mutant K-ras P1 promoter sequences using AUC, CD, DSC, ITC, and DMS footprinting techniques. The initial focus of this study was to determine whether the WT K-ras promoter sequence can form a stable G-quadruplex. In comparison to the simpler and more widely studied c-MYC and Bcl-2 promoter sequences G-quadruplexes, a K-ras WT promoter sequence must incorporate one or more large lateral loops and/or a corner "kinked out" base. We have found that the WT K-ras polypurine sequence forms an ensemble of intramolecular G-quadruplex motifs in dilute solution. We also found that the predominant WT G-quadruplex motif formed includes both a corner kink and a very large (11 base) loop. Four mutant K-ras sequences, created to probe the folding topology of the WT sequence, clearly demonstrated that G-runs 1, 2, and 3 are employed in all stable K-ras promoter sequence quadruplexes and that there is very little difference in the stability of quadruplexes having one base (T7) in a kinked conformation and either four or 11 bases in a large lateral/end loop. Finally, the K-ras WT promoter sequence G-quadruplex structures we report here include the earlier structure proposed by Xodo et al on the basis of their DMS footprinting results (*12*).

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Background

It has been proposed that focusing on structural selectivity rather than sequence selectivity, when targeting DNA replication or transcription, is the more promising area of drug design (*13*). G-quadruplex and i-Motif higher order DNA structures have been the focus of recent strategies for oncogene regulation (*13–17*). Approximately 40% of all human genes have been found to have i-Motif forming purine and G-quadruplex forming pyrimidine-rich sequences within 1Kb of their start sequences (*13*, *18*). Of the known human oncogenes, more than half have been shown to have G-quadruplex and i-Motif forming sequences in their promoter regions (*19*). One of the most significant problems, and an area of much debate, when designing drugs for oncogene (or gene) regulation, is the uncertainty of whether or not the G-quadruplex and i-Motif higher order structures are biologically relevant or exist at all under human physiological conditions (*13–18*). However, it has been established that binding a drug with high affinity for G-quadruplex drives the folding structural equilibrium in the direction of G-quadruplex formation (*13*, *18*). Hurley et al. reported on the use of topoisomerase II inhibiters that interact with G-quadruplexes as well as subsequent studies that show TMPyP4 is able to further stabilize the G-quadruplex structure and suppress any further c-MYC transcriptional activation (*7*, *14*, *23*, *24*).

Model oncogene promoter region sequences have shown the potential to fold into multiple unique G-quadruplex conformations (*13–17*, *19–21*, *23*, *24*). The structural diversity and conformational equilibrium exhibited by these sequences is the source of the second most significant problem in drug targeting G-quadruplex (and/or i-Motif) structures, i.e. what is the most relevant target structure? In

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

order to design a small molecule having very high affinity and selectivity for the K-ras promoter region, we must know the structures of each of the receptor G-quadruplexes or at least the structure of the most significant target conformation.

Materials and Methods

The native (WT) 30-mer and mutant (Mut) 26-mer K-ras oligonucleotide constructs were obtained from Operon (Huntsville, AL) and their sequences are shown in Scheme 1. The runs of guanines assumed to be involved in G-quadruplex formation are underlined while the G→T mutations are outlined.

Oligo		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
WT	5'-	G	G	G	C	G	G	T	G	T	G	G	G	A	A	G	A	G	G	G	A	A	G	A	G	G	G	G	G	A	G	3'
Mut 1-3-4-5	5'-	G	G	G	C	T	T	T	T	T	G	G	G	A	A	G	A	G	G	G	A	A	G	A	G	G	G	3'				
Mut 1-2-4-5	5'-	G	G	G	C	G	G	T	G	T	T	T	T	A	A	G	A	G	G	G	A	A	G	A	G	G	G	3'				
Mut 1-2-3-5	5'-	G	G	G	C	G	G	T	G	T	G	G	G	A	A	G	A	T	T	T	A	A	G	A	G	G	G	3'				
Mut 1-2-3-4	5'-	G	G	G	C	G	G	T	G	T	G	G	G	A	A	G	A	G	G	G	A	A	G	A	T	T	T	3'				

Scheme 1. Model WT and mutant K-ras promoter sequences.

Within the WT sequence studied here there are 5 runs of guanines that can be used or involved in G-quadruplex formation, for a general naming system we assigned a number, Run 1 through Run 5, to each run of guanines starting from the 5′ end. Based upon this assignment the WT can be described as: Run 1 including G1-G3, Run 2 including G5-G8 (including T7 as a kinked out base), Run 3 including G10-G12, Run 4 including G-17-G-19, and Run5 including G-24-G28. The G→T mutations were made in an attempt to simplify the number of folded species in dilute solution by selectively eliminating the involvement of one of the last four guanine runs in potential quadruplex formation. For example Mut1-3-4-5 has the guanines in Run 2 mutated to thymines with G-runs 1, 3, 4, and 5having their native sequence.

All stock solutions were prepared by dissolution of weighed amounts of lyophilized oligonucleotide into [K^+] BPES buffer at a single salt concentration of 130 mM [K^+] and pH 7.0 Approximately 1 mL of the oligonucleotide was exhaustively dialyzed (1000 molecular-weight cutoff membrane) with two changes of buffer solution (1 L, 24h each) at 4°C. The concentrations of all DNA solutions were verified using ultraviolet-visible spectrophotometry (UV-vis) and molar extinction coefficients determined by using a nearest-neighbor method for single stranded DNA (ε_{260}); WT (3.174×10^5), Mut1-3-4-5 (2.661×10^5), Mut1-2-4-5 (2.687×10^5), Mut1-2-3-5 (2.687×10^5), and Mut1-2-3-4 (2.677×10^5). See *Oligonucleotide MW and Extinction Coefficient Calculator*. Available from: http://www.ambion.com/techlib/misc/oligo_calculator.html.

TMPyP4 was obtained from Frontier Scientific (Logan, UT). All TMPyP4 solutions were prepared by dissolution of a known amount of TMPyP4 into a measured volume of the dialysate buffer solution from the corresponding oligonucleotide. All TMPyP4 concentrations were verified using UV-Vis with molar extinction coefficient of ε_{424}=2.26 X 10^5 $M^{-1}cm^{-1}$ (*13*, *20*).

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Sedimentation velocity experiments were performed on a Beckman-Coulter Proteome Lab XL-A analytical ultracentrifuge. Experiments were performed at 20 ºC with a rotor speed of 60,000 rpm. DNA concentrations ranged from 0.82 to 3.4 μM. Typically 35 to 40 scans were collected for each sample per experiment. Sedimentation velocity experiments were analyzed with SEDFIT using a partial specific volume (0.56 cm3/g) calculated from the oligonucleotide sequences as described by Hatters et al (*22*). For comparative purposes, the sedimentation coefficients were estimated for the WT K-ras (30-mer), WT c-MYC (22-mer and 36-mer) and the WT Bcl-2 39-mer using HydroPro.

CD measurements were performed using an OLIS DSM-20 CD spectropolarimeter (OLIS, Bogart, GA). Spectra were collected over a wavelength range of 200-305 nm. Oligonucleotide samples were prepared at a nominal concentration of 3.0 μM and were annealed prior to CD experiments. All measurements were done at 25°C and all experiments done as described in references (*13*) and (*20*).

Differential Scanning Calorimetry (DSC) experiments were performed using a Microcal VP-DSC instrument (Microcal, Northampton, MA). The experimental scan rate was set at 90°C/h with a temperature range from 10° to 110° C. The thermograms were deconvoluted using Origin 7.1 software into the minimum number of "two-state" transitions needed to fit the thermogram within expected experimental error as previously described (*13*, *20*).

Isothermal Titration Calorimetry (ITC) experiments were performed using a Microcal VP-ITC (Microcal, Northampton, MA). ITC titrations were performed by addition of the 25×10 μL injections of 1.5 mM TMPyP4 into1.5 mL of the annealed oligonucleotide solution ([DNA] ≈ 50 μM). All titrations were done at 25°C and at a single supporting electrolyte concentration of 130 mM [K^+] (pH=7.0 BPES buffer). Three replicate titration experiments were performed on each of the four model oligonucleotides. The nonlinear regression fitting was done using Mathematica 8.0 and a two independent sites model algorithm developed in our lab. The ITC experimental protocols and data fitting procedures have been previously described in detail (*13*, *20*).

Gel electrophoresis experiments were performed using 20% non-denaturing polyacrylamide gels obtained from Bio-Rad. A 10 bp and 20 bp DNA step ladder was obtained from Promega for molecular weight referencing and the DNA stained with Stains-All ® (Sigma Aldrich) for visualization. All samples were prepared with the standard 130 mM [K+] BPES pH 7.0 with 10% TBE buffer was used as the running buffer. All samples were prepared at nominally 100 μM and were annealed at 95 °C for 10 minutes then flash cooled in ice at ~4 °C for approximately 10 minutes before loading. All PAGE experiments were run for approximately 45 minutes at 150 V. Gels were then stained with EtBr and visualized using a UV-lamp.

Dimethyl Sulfate (DMS) Footprinting experiments were performed using 3′-end-labeled WT and mutant constructs that were diluted with the 130 mM [K^+] BPES buffer at pH 7.0 to a final concentration of 10 μM. 40 μL of 3′ 32P-labeled DNA (20 000 cpm; ca. 5 ng) was annealed at 95 °C for 10 min, and snap-cooled to room temperature. 2 μL of Calf-Thymus DNA and 2 μL of 10% DMS in 50% EtOH (25 μL EtOH, 5 μL DMS, 20 μL ddH_2O) was added to the DNA

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

solution mixture and incubated at room temperature for 10 minutes. The reaction was stopped/quenched by adding 3.5 μL of 98% *â*-mercaptoethanol and 10 μL of glycerol (20%) and then subjected to preparative gel electrophoresis. The DNA samples were then subjected to ethanol precipitation, piperidine treatment, and 16% polyacrylamide gel electrophoresis.

Results and Discussion

We report here one of the first studies to establish that the WT K-ras oncogene promoter region purine rich sequence is capable of forming stable G-quadruplex structures. We have previously reported on the deconvolution of the WT c-MYC and Bcl-2 promoter sequences using the same biophysical techniques reported here for the deconvolution of the K-ras promoter region G-quadruplex motifs. The guanine rich region of the K-ras oncogene promoter exhibits some sequence similarities to the guanine rich regions of the c-MYC and Bcl-2 oncogene promoter sequences. However the K-ras purine rich region includes additional guanines not located in G-runs, five guanine runs including one run of three guanines that can be formed by "kinking-out" a thymine (T7), and longer runs of non-guanine bases between G-runs. The structures formed by the WT K-ras sequence exhibit thermal stabilities, spectroscopic signatures, and TMPyP4 binding characteristics that are similar to those exhibited by well characterized c-MYC and Bcl-2 oncogene promoter region G-quadruplexes (*13*, *20*).

Non-denaturing polyacrylamide gel electrophoresis experiments and analytical ultracentrifugation experiments demonstrate that unannealed WT K-ras like unannealed c-MYC and Bcl-2 is polydisperse. A PAGE gel is shown in Figure 1 for unannealed WT K-ras vs. annealed WT K-ras along with a standard DNA ladder containing ds-DNAs having from 10 to 100 base pairs (as labeled). The unannealed K-ras sample shows a number of higher molecular weight bands, in comparison to the annealed K-ras sample.

Annealed K-ras samples exhibit sedimentation and electrophoretic properties that are similar to those reported for the WT c-MYC and Bcl-2 G-quadruplexes. The annealed WT K-ras construct having a MW of 9,612 Dalton's has two principle species having sedimentation coefficients of 2.4 and 4.9 S. The large 2.4 S peak observed in the C(S) vs. S plot is consistent with a compact quadruplex species having a molecular weight of approximately 10,000 Daltons (see Figure 2). The analytical ultracentrifuge data shown in Figure 2 for two well characterized oncogene promoter G-quadruplexes, c-MYC and Bcl-2, in addition to the WT K-ras construct are remarkably similar in that all three exhibit monomeric and dimeric sedimenting species.

The experimental sedimentation coefficients for all three monomeric G-quadruplexes are in excellent agreement with the sedimentation coefficients calculated for these structures using the HydroPro program (see Table 1).

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

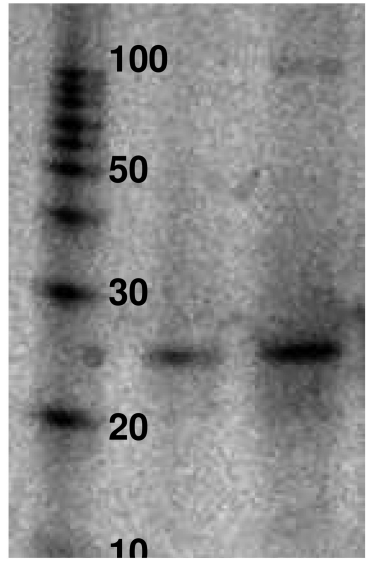


Figure 1. PAGE mobilities for the unnanealed K-ras WT (30-mer) promoter sequence (right lane) and annealed WT K-ras (30-mer) promoter sequence (center lane) in a 20% polyacrylamide gel. The intramolecular (monomer) quadruplex band for the annealed WT K-ras oligo lies between the 20 and 30 bp DNA step bands. The quadruplex mobility is comparable to a duplex structure containing approximately twice as many bases (e.g. 25 bp for the ladder DNA vs. 30 bases for the WT K-ras promoter sequence quadruplex.

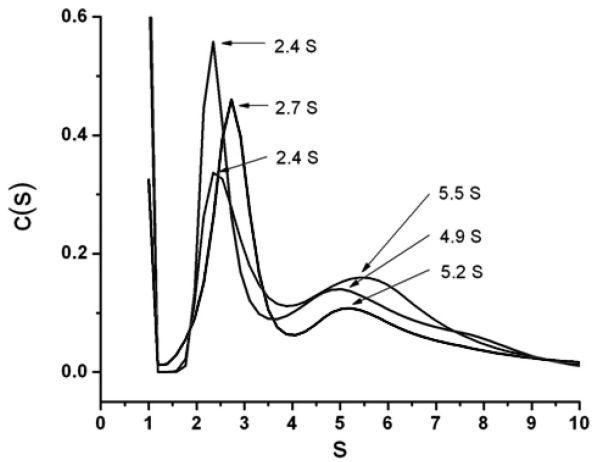


Figure 2. Superimposed C(S) vs. S ultracentrifuge data for the WT K-ras 30-mer (2.4 S), WT c-MYC 36-mer (2.4 S), and the WT Bcl-2 39-mer (2.7 S).

Table 1. Molecular weights, monomer and dimer measured sedimentation coefficients, and HydroPro monomer G-quadruplex estimated sedimentation coefficients for the WT K-ras (30-mer), WT c-MYC (22-mer and 36-mer), and the WT Bcl-2 (39-mer) sequences in 130 mM [K^+] BPES pH 7.0 buffer

Promoter	*Mol. Wt.*	*Sed. Coeff.*[1] *(measured)*	*Sed. Coeff.*[2] *(measured)*	*Sed. Coeff.*[3] *(HydroPro)*
K-ras	9,612	2.4 S	4.9 S	2.2 S
c-MYC	6,991	na	Na	1.6 S
c-MYC	11,519	2.4 S	5.2 S	2.6 S
Bcl-2	12,431	2.7 S	5.5 S	2.9 S

[1] Measured sedimentation coefficients are consistent with a monomeric species and correspond to the maxima of the first peak in the C(S) vs. S plot (Figure 2). [2] Measured sedimentation coefficients are consistent with a dimeric species and correspond to the maxima of the second peak in the C(S) vs. S plot (Figure 2). [3] The calculated sedimentation coefficient for the c-MYC 22-mer (Mw 6,991 Da) was obtained from the published NMR solution structure for this quadruplex (NDB ID: 1XAV) using the appropriate values for the solution density and the partial specific volume for the DNA. Sedimentation coefficients listed for the other G-quadruplexes were estimated assuming that the shape factor is unchanged from the 22-mer standard.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

The CD spectra for the K-ras WT and four mutant constructs in 130 mM [K+] BPES at pH 7.0 are shown in Figure 3. The spectrum for the WT K-ras construct exhibits a maximum molar ellipticity at 265 nm and a minimum in molar ellipticity at 244 nm that is consistent with previously published spectra for "classical" intramolecular parallel stranded G-quadruplex DNA (*13*, *20*). The CD spectra for Mut 1-2-3-5 and Mut 1-2-3-4 also demonstrate maximum and minimum molar ellipticity values that are similar to those observed for the WT. It must however be noted that a slight shoulder appears at 290 nm in both the Mut 1-2-3-5 and Mut 1-2-3-4 that is not observed in the WT spectrum and the 265 nm peak is attenuated in Mut 1-2-3-4 in comparison to the WT sequence. The 290 nm shoulder indicates the presence of anti-parallel species or quadruplexes having mixed topology in these mutant sequence solutions. The CD results suggest that these two mutants, that must include guanine Runs 2 (including the *kink* at T7) and 3, fold into G-quadruplex structures that are similar to those formed by the WT K-ras sequence. On the other hand, mutants 1-3-4-5 and 1-2-4-5, do not exhibit "classical" G-quadruplex CD signatures. CD spectra for Mut 1-3-4-5 and Mut 1-2-4-5 are neither consistent with the formation of an inter/intramolecular G-quadruplex nor with single stranded DNA.

DSC thermal melting profiles for the WT (30-mer) and the four of the K-ras mutant (26mer) purine rich K-ras promoter sequences are shown in Figures 4 and 5 respectively. All four of the thermograms are comprised of either a single symmetric peak (Mut1-3-4-5) or a broad asymmetric peak (Mut1-2-4-5, Mut1-2-3-5, and Mut1-2-3-4) that has been best fit for one or two independent "two-state" melting transitions. The thermogram for Mut1-3-4-5, which has Run 2 of guanines

(including the G8 kink) mutated, is comprised of a single symmetric peak that has been fit with a single "two state" model with a *T*m value of 53.0 °C. With the mutation of Run 3 of guanines, the thermogram for Mut 1-2-4-5 is comprised of a broad asymmetric peak that has been fit with two independent "two state" model with calculated *T*m values of 39.4 °C and 53.0 °C, these Tm values are the lowest observed for all of the K-ras constructs studied here (Table2). The Tm values for Mut1-2-3-5 and Mut1-2-3-4 are similar however the overall distribution of folded species observed is different which is demonstrated by the DSC thermal melting profiles. Mut 1-2-3-5 and Mut1-2-3-4 thermograms are best fit with a two independent "two-state" model with Tm values of 1) 55.6 °C and 2) 69.7 °C and 1) 57.2 °C and 2) 69.9 °C respectively.

The results of the DSC and CD experiments present a consistent picture regarding the structure and stability of the model K-ras WT as well as the four mutant promoter oligonucleotides. We have found that the WT sequence folds into at least two unique structural quadruplexes with different thermal stabilities. The WT thermogram is comprised of an asymmetric peak that has been fit for two overlapping independent "two state" transitions exhibiting *T*m values of 58.97 °C and 71.59 °C (Table 2). The WT raw excess heat capacity exhibited a slight shoulder with a Tm of approximately 83 °C. The observed asymmetric shoulder was attributed to the presence of a small percentage of intramolecular G-quadruplex structures that can be described as an equilibrium between possible dimers, tetramers, or G-wires which are all combinations of more than one G-rich sequence.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

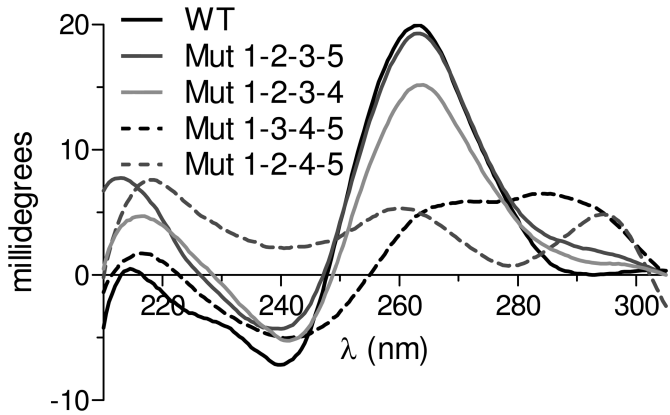


Figure 3. Circular dichroism spectra for the model K-ras WT (30-mer) and four mutant (26-mer) promoter sequences in 130 mM [K⁺] BPES buffer at pH 7.0. The WT, Mut 1-2-3-5, and Mut 1-2-3-4 sequences exhibit CD signatures typical for parallel G-quadruplexes. The Mut 1-3-4-5 and Mut 1-2-4-5 sequences do not appear to form G-quadruplex motifs.

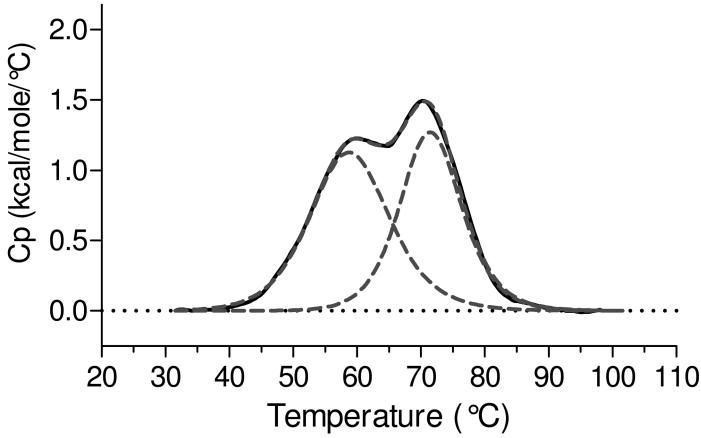


Figure 4. DSC thermogram for the thermal denaturation of the WT K-ras 30-mer construct in 130 mM [K+] BPES buffer at pH 7.0. The raw excess heat capacity has been deconvoluted into two independent "two state" processes. The solid black curve is for the raw excess heat capacity superimposed with the composite curves (dashed grey lines) for the two overlapping "two-state" transitions. The Tm values for the two melting transitions are 58.97 °C and 71.59 °C.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Typical ITC data obtained for the titration of the K-ras WT 30mer construct with TMPyP4 at 25 °C in 130 mM [K+] BPES buffer at pH 7.0 are shown in Figure 6. After making corrections for blank heat effects, baseline, and titrant dilution, the correct heat data were fit using a non-linear regression algorithm to obtain the best fit parameters for the equilibrium constant(s), K, enthalpy change(s), ΔH, and reaction stoichiometry, *n*, for the formation of the G-quadruplex DNA/TMPyP4 complex. The binding thermograms for the Interaction of the K-ras WT 30mer construct and TMPyP4 require a complicated model that exhibits two binding modes with each binding mode exhibiting two equivalent sites. At saturation the binding stoichiometry of TMPyP4 to G-quadruplex DNA is 4:1.

The best fit parameters for the binding of TMPyP4 to the K-ras WT 30mer G-quadruplex are listed in Table 3 along with the best fit parameters for TMPyP4 binding to the WT c-MYC (36mer) and the WT Bcl-2 (39mer) promoter sequences. The values listed for the interaction of TMPyP4 with all three of these WT promoter region G-quadruplexes is similar and demonstrates that the WT K-ras construct is capable of binding TMPyP4 with similar binding modes and affinities.

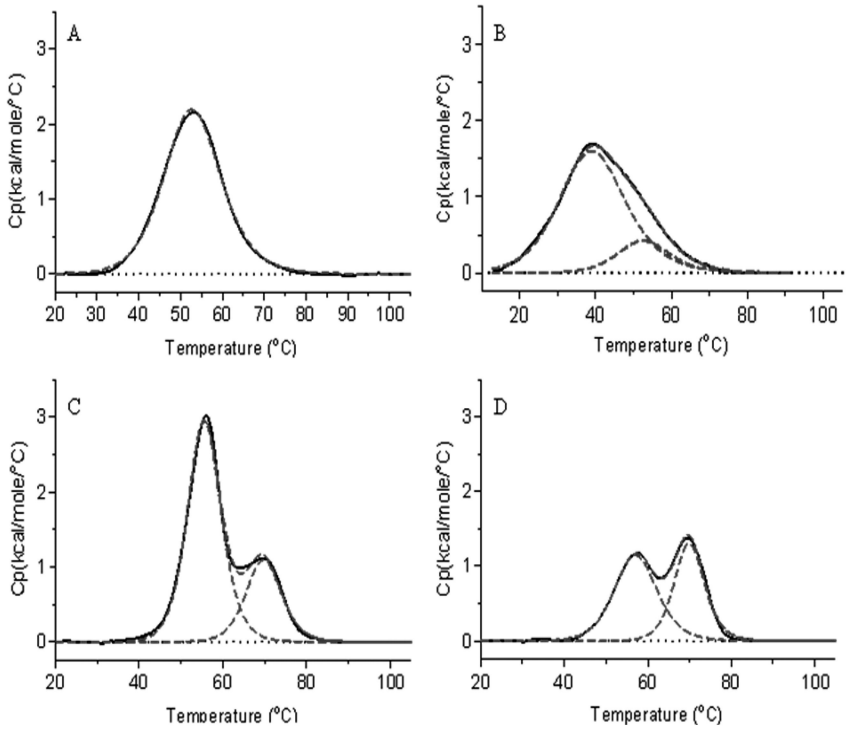


Figure 5. DSC thermograms for the denaturation of K-ras mutant G-quadruplex constructs in 130 mM K+ BPES at pH 7.0. The raw excess heat capacity for all four sequences has been deconvoluted into either one or two "two-state" processes. The uppermost curve (black) is for the raw thermogram superimposed with the composite curve for the two overlapping "two-state" transitions (dashed grey lines). The two curves, the raw and the fit, are virtually indistinguishable. The Tm values for the two overlapping but independent melting transitions are listed in Table 2. A) Mut 1-3-4-5, B) Mut 1-2-4-5, C) Mut 1-2-3-5, D) Mut 1-2-3-4.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Table 2. The thermal denaturation temperatures, Tm's, were determined from the fits of the thermograms shown in Figures 4 and 5. The thermograms were fit for either one or two independent overlapping "two-state" melting transitions. The CD spectra shown in Figure 3 were used to establish the presence of a G-quadruplex and the folding topology. DMS footprinting experiments were used to verify formation of a G-quadruplex having protected bases and the presence of a kink in the quadruplex backbone was predicted from the model sequences.

	DSC		*CD*		*DMS*	*Sequence*
Oligo	*Tm1*	*Tm2*	*Q-Plex*	*Topology*	*Q-Plex*	*Kink*
WT	59.0	71.6	Yes	Parallel	Yes	???
Mut1-3-4-5	53.0	-	No	Mixed	No	No
Mut1-2-4-5	39.4	53.0	No	Mixed	No	No
Mut1-2-3-5	55.6	69.7	Yes	Parallel	Yes	Yes
Mut1-2-3-4	57.2	69.9	Yes	Parallel	Yes	Yes

Table 3. ITC-derived thermodynamic parameters (ΔG_i, ΔH_i and $-T\Delta S_i$) for the binding of TMPyP4 to the K-ras WT 30mer G-quadruplex construct in 130 mM [K+] BPES buffer at pH 7.0. The data for c-MYC (*13*) and Bcl-2 (*20*) are listed for comparison. All parameters are given in kcal/mol of TMPyP4 bound.

Oligo	ΔG_1	ΔH_1	$-T\Delta S_1$	ΔG_2	ΔH_2	$-T\Delta S_2$
K-ras	-10.4±0.4	-3.6 ± 0.2	-6.8	-7.5±0.9	-7.6 ± 0.8	0.2
c-MYC	-9.2±0.2	-2.6 ± 0.1	-6.6	-7.3±1.0	-9.1 ± 1.7	1.8
Bcl-2	-10.4±0.5	-1.8 ± 0.1	-8.6	-7.6±0.5	-6.1 ± 0.1	-1.5

DMS footprinting was used to determine which guanines in the purine rich region of the WT K-ras promoter sequence were involved in the formation of the g-tetrads, the formation of a backbone *kink*, or were involved in the lateral/end loop region. In the absence of KCl; (0 mM [K+]) buffer the formation of stable G-quadruplexes does not occur, therefore a cleavage pattern is observed for each of the guanines in the sequence. Upon the addition of 130 mM [K+] the formation of a stable G-quadruplex protects the guanines that are involved in Hoogsteen Hydrogen bonding within the G-tetrads from being cleaved. As shown in Figure 7, the WT K-ras sequence demonstrates a clear pattern of protected guanines. G1, G2, G3, G5, and G6 are completely protected, G8 shows significant protection,

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

G10, G11, and G12 are again completely protected, and G25, G26, and G27 are again at least partially protected. This pattern is consistent with the formation of one or more quadruplex structures employing the following four G-runs: (G1-G3), (G5, G6, and G8), (G10-G12), and (G25-G27).

The four mutant constructs were also analyzed with DMS footprinting in an effort to verify quadruplex formation by the mutant sequences. Cleavage patterns for both Mut1-3-4-5 and Mut1-2-4-5 in the presence of 130 mM [K+] are similar to the cleavage patterns observed for these sequences in the absence of KCl. The almost complete cleavage of all of the guanines in these mutants is indicates that these constructs do not form stable G-quadruplex structures. In comparison, the Mut1-2-3-5 and Mut1-2-3-4 sequence exhibit protection and cleavage patterns that are very similar to those obtained for the WT K-ras construct. In both the Mut1-2-3-5 and Mut1-2-3-4, where Run 4 or Run 5 are mutated, Runs 1-3 are almost completely protected and minimal cleavage is observed for Run 4 or Run 5 in the Mut 1-2-3-4 or Mut 1-2-3-5 sequences respectively. G-quadruplexes formed by either the K-ras WT or the Mut 1-2-3-4 or Mut 1-2-3-5 sequences are shown in Figure 8.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

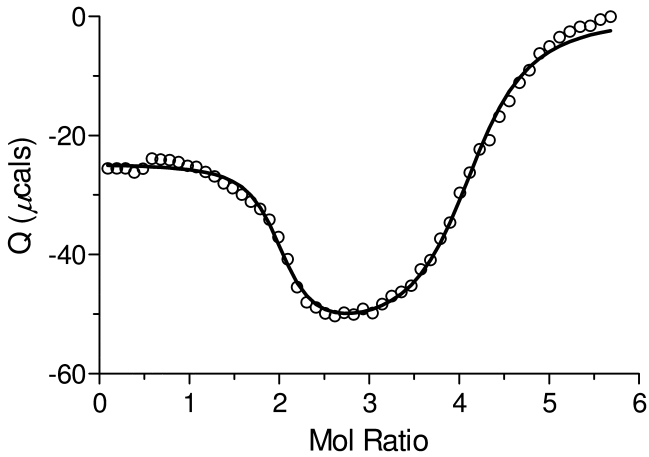


Figure 6. ITC data shown are for the addition of TMPyP4 to the K-ras WT 30mer construct at pH 7.0 in 130mM [K+] BPES. The integrated heat data (points) are for a single titration experiment. The best fit nonlinear regression line shown (—) is for the thermodynamic parameters obtained using a two independent sites model.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

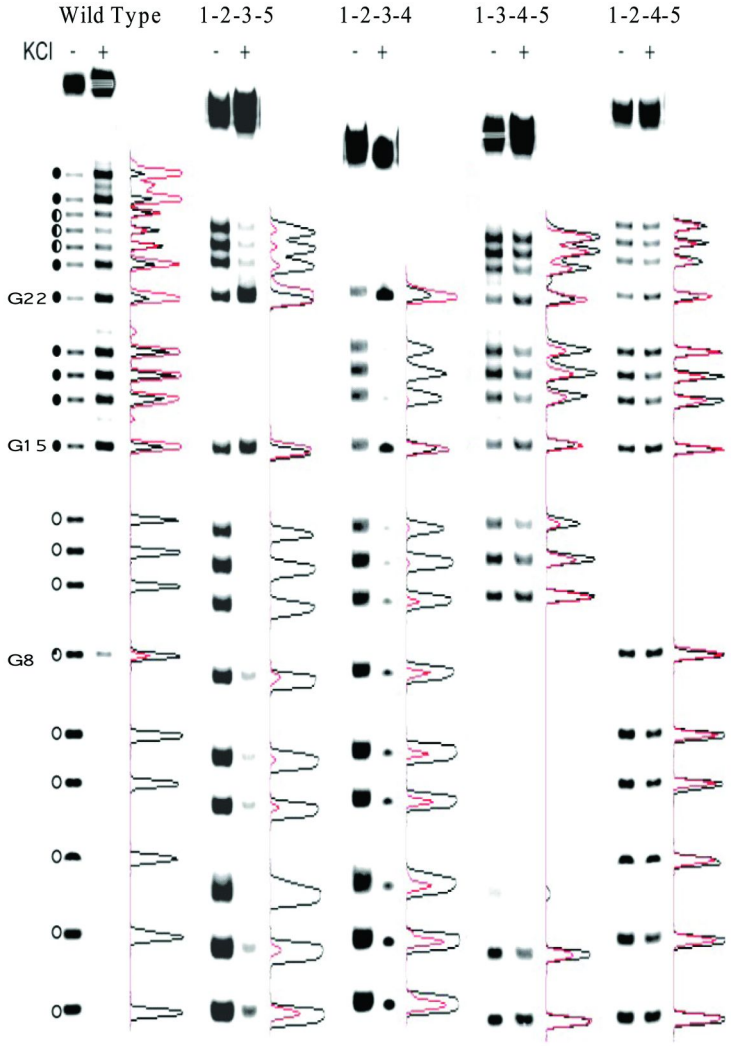


Figure 7. DMS footprinting results (cleavage patterns) are shown for the WT K-ras 30-mer sequence and the four mutant 26-mer sequences. Cleavage at unprotected guanines is indicated by a band on the gel and a bar on the histogram. Protected guanines (in the presence of 130 mM [K^+] result in clear spots in the gel. The lanes and – or + KCl are labeled at the top of the gel.

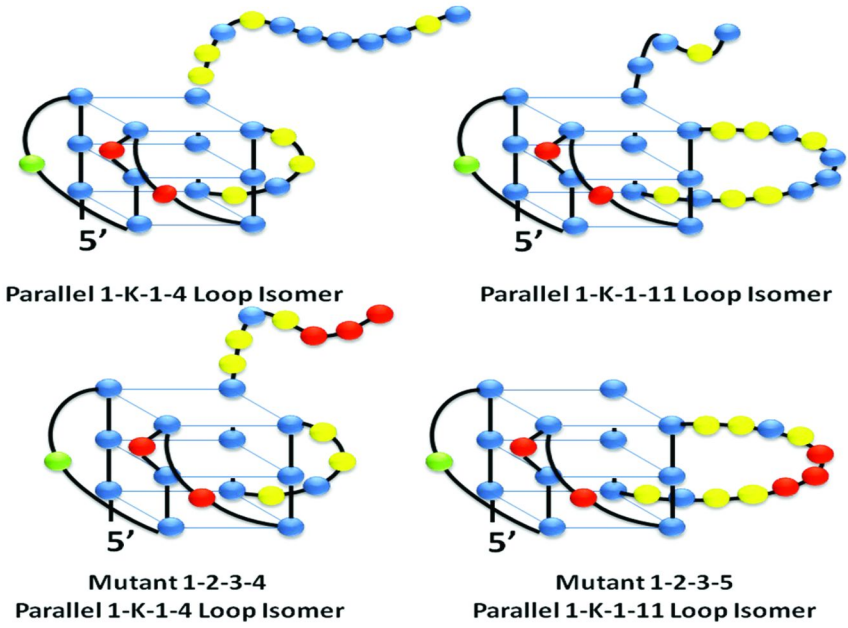


Figure 8. The folding topology is shown for four K-ras WT and mutant G-quadruplexes. The bases are represented as colored balls, blue for guanine, green for cytosine, red for thymine, and yellow for adenine. Multiple G-quadruplexes exist by mixing and matching syn- or anti- base configuations in each of the three stacked G-tetrads. (see color insert)

What have we learned in these studies about K-ras promoter sequence quadruplexes specifically and about G-quadruplex structure and stability in general? Obviously, we have shown that the K-ras promoter sequence forms one or more stable G-quadruplexes that are similar to the quadruplexes formed by the c-MYC and Bcl-2 promoter sequences, at least in terms of their stability and the thermodynamics of their interactions with a quadruplex probe ligand, TMPyP4. The K-ras quadruplexes melt at slightly lower temperatures than the c-MYC or Bcl-2 structures (*13*, *20*), but not by as much as one would think considering the large destabilizing lateral or end loop, 11 bases for the Mut 1-2-3-5 vs. a maximum of 6 unstructured loop bases in the lower melting 1:6:1 c-MYC quadruplex, and the backbone *kink*. Perhaps most interesting is that we only found quadruplexes that contained the T7 kink and that the only K-ras quadruplexes found employed G-runs 1,2,and 3 in combination with either G-run 4 or 5. A K-ras quadruplex formed using G-runs 1, 3, 4, and 5 would necessarily have three larger loops (containing 6, 4, and 4 unstructured bases) while the quadruplex formed using G-runs 1, 2, 4, and 5 would have two larger loops (containing 8 and 4 unstructured bases) and a kink in G-run 2. Apparently a kink is less destabilizing than multiple larger loops. In the quadruplexes utilizing G-runs 1-2-3-4 or 1-2-3-5, the first two loops contain a single base and the third loop either 4 or 11 bases.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Clearly the demonstration of stable quadruplexes with both kinks and large loops suggest that quadruplex forming sequences may be even more prevalent than previously thought and new search algorithms may be needed to find quadruplex forming sequences outside the accepted box of four runs of three or more guanines separated by from 1-3 non guanine bases.

Summary

To date, almost all studies of G-quadruplex DNA have been limited to short single-stranded sequences as models for the structure and stability of G-quadruplexes that may form *in vivo*. We have had some success in modeling NHEs by pairing a quadruplex forming g-rich sequence with a shorter partially complimentary strand to produce a G-quadruplex with flanking duplex regions. Probably the only system that is reasonably well modeled with short single stranded g-rich sequences is the telomere single strand overhang. It might seem that since the equilibrium in systems containing both g-rich quadruplex forming sequences and the complimentary c-rich strand is so far toward formation of B-DNA that studies like the one reported here would not be interesting. However, it is my belief that in the absence of the complimentary strand, quadruplex chemistry comes into its own. The self stacking of guanine bases is clearly a major force in quadruplex formation. Strand separation during DNA replication or in single stranded RNA must lead to (at least some) quadruplex formation. Even the existence of a small population of DNA or RNA quadruplex motifs could result in drugable target concentrations that could be exploited in novel disease therapies.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Acknowledgments

The authors would like to thank Dr. Jonathan “Brad” Chaires at the University of Louisville for help with the analytical ultracentrifuge experiments and Dr. Laurence Hurley at the University of Arizona Cancer center for help with the DMS footprinting experiments.

References

1. Wang, W. W.; Chunyou; Dong, J.; Zhao, G.; Chen, X.; Zhang, M. Detection of K-ras gene point mutation’s style in human pancreatic cancer cell line PANC-1 by PCR-SSP. *Chin.-Ger. J. Clin. Oncol.* **2006**, *5*, 46–48.
2. Barugola, G.; Falconi, M.; Bettini, R.; Boninsegna, L.; Casarotto, A.; Salvia, R.; Bassi, C; Pederzoli, P. The determinant factors of recurrence following resection for ductal pancreatic cancer. *J. Pancreatol.* **2007**, *8*, 132–140.
3. Sarkar, F. H.; Banerjee, S.; Li, Y. Pancreatic cancer: Pathogenesis, prevention and treatment. *Toxicol. Appl. Pharmacol.* **2007**, *224*, 326–336.
4. Maitra, A.; Kern, S. E.; Hruban, R. H. Molecular pathogenesis of pancreatic cancer. *Best Pract. Res., Clin. Gastroenterol.* **2006**, *20*, 211–226.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

5. Talar-Wojnarowska, R.; Malecka-Panas, E. Molecular pathogenesis of pancreatic adenocarcinoma: Potential clinical implications. *Med. Sci. Monit.* **2006**, *12*, RA186–193.
6. Zhao, Y.; Miyashita, K.; Ando, T.; Kakeji, Y.; Yamanaka, T.; Taguchi, K.; Ushijima, T.; Oda, S.; Maehara, Y. Exclusive KRAS mutation in microsatellite-unstable human colorectal carcinomas with sequence alterations in the DNA mismatch repair gene, MLH1. *Gene* **2008**, *423*, 188–193.
7. Hurley, L. H.; Von Hoff, D. D.; Siddiqui-Jain, A.; Yang, D. Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Semin. Oncol.* **2006**, *33*, 498–512.
8. Ladame, S.; Schouten, J. A.; Roldan, J.; Redman, J. E.; Neidle, S.; Balasubramanian, S. Exploring the recognition of quadruplex DNA by an engineered Cys2-His2 zinc finger protein. *Biochemistry* **2006**, *45*, 1393–1399.
9. Kerwin, S. M. G-quadruplex DNA as a target for drug design. *Curr. Pharm. Des.* **2000**, *6*, 441–478.
10. Han, H.; Hurley, L. H. G-quadruplex DNA: A potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **2000**, *21*, 136–142.
11. Neidle, S.; Read, M. A. G-quadruplexes as therapeutic targets. *Biopolymers* **2000**, *56*, 195–208.
12. Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006**, *34*, 2536–2549.
13. Dettler, J. M.; Buscaglia, R.; Le, V. H.; Lewis, E. A. DSC deconvolution of the structural complexity of c-MYC P1 promoter G-quadruplexes. *Biophys. J.* **2011**, *100*, 1517–1525.
14. Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the Bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–5415.
15. Reed, J. E.; Arnal, A. A.; Vilar, R. Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel(II) complexes. *J. Am. Chem. Soc.* **2006**, *128*, 5992–5993.
16. Kaushik, M.; Bansal, A.; Kukreti Possibility of an antiparallel (tetramer) quadruplex exhibited by the double repeat of the human telomere. *Biochemistry* **2007**, *46*, 7119–7131.
17. Kumar, P.; Verma, A.; Chowdhury, S. Tetraplex DNA transitions within the human C-Myc promoter detected by multivariate curve resolution of fluorescence resonance energy transfer. *Biochemistry* **2005**, *44*, 16426–16434.
18. Huppert, J. L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413.
19. Hurley, L. H. Secondary DNA structures as molecular targets for cancer therapeutics. *Biochem. Soc. Trans.* **2001**, *29*, 692–696.
20. Nagesh, N.; Buscaglia, R.; Dettler, J. M.; Lewis, E. A. Studies on the site and mode of TMPyP4 interactions with Bcl-2 promoter sequence G-Quadruplexes. *Biophys. J.* **2010**, *98*, 2628–2633.

21. Cashman, D. J.; Buscaglia, R.; Freyer, M. W.; Dettler, J. M.; Hurley, L. H.; Lewis, E. A. Molecular modeling and biophysical analysis of the C-MYC NHE-III$_1$ silencer element. *J. Mol. Model.* **2008**, *14*, 93–101.
22. Hatters, D. M.; Wilson, L.; Atcliffe, B. W.; Mulhern, T. D.; Guzzo-Pernell, N.; Howlett, G. J. Sedimentation analysis of novel DNA structures formed by homo-oligonucleotides. *Biophys. J.* **2001**, *81*, 371–381.
23. Yang, D.; Hurley, L. H. Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides, Nucleotides Nucleic Acids* **2006**, *25*, 951–968.
24. Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch003

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Chapter 4

A CD Spectroscopic Investigation of Intermolecular and Intramolecular DNA Quadruplexes

Brenna A. Tucker, Sahmla Gabriel, and Richard D. Sheardy*

Department of Chemistry and Physics, Texas Woman's University, Denton, TX 76204
***E-mail: rsheardy@twu.edu**

To gain insight into the fundamental factors that affect the structure, stability and ligand binding properties of the intramolecular DNA quadruplex formed from the human telomere sequence $(TTAGGG)_4$, a library of DNA sequences, $(XXXGGG)_y$ where X = T or A and y = 1, 2, 4 or 8, was generated and characterized by circular dichroism (CD). The wild type sequence folds into a distinct conformation with particular strand and loop orientations. The GGG bases form a G-tetrad while the second, third and fourth TTA segments are found in the loops joining the tetrads. Our initial findings from CD analyses indicate that the conformations of these oligomers are dependent upon sequence context (identity of X), number of repeats (value of y) and cation (Na^+ or K^+) present.

Introduction

Although the Watson-Crick model of DNA as a right-handed double helical structure stands as the primary conformation in how we think about DNA, we now know that DNA is highly polymorphic: (A) DNA can exist as a single strand, duplex, triplex, quadruplex or even multiplex; (B) the duplex can either be right handed of left handed; and, (C), DNA can be sculpted into unusual higher order structures. Ultimately, the conformation and associated conformational properties of a segment of DNA is determined not only by its sequence context but also by the environmental conditions (i.e., temperature, pH, identity of counter ions and their concentrations, etc.) under which it is prepared. Of particular interest

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

recently are the structures formed from G-rich DNAs designated as quadruplexes. DNA sequences that have islands of G_{2-4} separated by 1 to 4 A or T bases can form a rich library of secondary structures with different molecularities, strand orientations, and guanine base conformations (i.e., *syn* or *anti*) (*1–13*). The actual conformation is highly dependent upon the sequence context and buffer conditions. The stabilities of these structures are also both sequence context and cation dependent (*14–18*). We have demonstrated that the thermally induced unfolding of the quadruplex formed from the human telomere sequence $(TTAGGG)_4$ proceeds through a three state mechanism (*19*). This observation has led to new questions. What is the nature of the intermediate formed in the unfolding mechanism? How will sequence context of the loops influence the conformation and stability of the folded structure? Can we use sequence context effects to delineate the unfolding mechanism?

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Background

DNA is a highly polymorphic molecule whose conformation is influenced by both environment and sequence context. Among the various forms, the quadruplex conformers have received a lot of attention recently. Self-assembly of tandem G-rich sequences produces quadruplexes through stacking of G-quartets (tetrads). These units are composed of four coplanar guanine bases arranged in a Hoogsteen base pairing motif (*1*). The Hoogsteen base pairing motif facilitates the creation of a central pore, which can coordinate monovalent cations with the following selectivity $K^+ > Na^+ > Cs^+ > Li^+$ (*1*). In addition, divalent cations (Mg^{2+}, Sr^{2+}, Ca^{2+}, Ba^{2+}) also induce quadruplex formation (*2*, *3*). Quadruplexes can be created with varying strand stoichiometries yielding unique folding topologies, alternate loop arrangements, strand polarity and base orientation (*1–3*, *5–14*).

Perhaps the most interesting biologically relevant quadruplex forming structure is the human telomeric repeat sequence $(TTAGGG)_n$. The telomere contains kilobases of the G-rich tandemly repeated non-coding sequence located at human chromosomal ends. The double stranded telomeric sequences are associated with proteins like TRF1 and TRF2. TRF2 is necessary to prevent end to end fusions of chromosomes, leading to genetic instability (*20*). Unstable telomeric ends could be interpreted by cellular machinery as double stranded breaks, signaling cellular senescence (*20*).

To better understand the local structure of a telomeric sequence, model oligomers of 8 to 24 bases have been synthesized and characterized. The solution NMR structure of $AG_3(TTAGGG)_3$ in Na^+ determined by Wang and Patel contains two lateral loops and one diagonal loop, more commonly known as an intramolecular basket type antiparallel quadruplex (*5*). The oligomer $AG_3(TTAGGG)_3$ in K^+ forms a mixture of conformations and the addition of flanking ends to create the sequence $A_3G_3(TTAGGG)_3A_2$ favors the formation of a single intramolecular quadruplex characterized by a single edgewise loop or propeller loop and two lateral loops, known as the hybrid parallel-antiparallel quadruplex as depicted in Figure 1 (*6*). Patel, *et al* have investigated the structure of a similar sequence, $TTAGGG(TTAGGG)_3A$ in K^+ solution, which also

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

produces a conformation with a single propeller loop and a diagonal loop (*7*). Circular dichroism (CD) spectroscopy is an outstanding method for examining the conformation and conformational changes in DNA secondary structure. In particular, CD can be used to distinguish the various structures of G-quadruplexes (*8*, *18*, *19*, *21–24*) For example, CD studies have indicated the spectra of $A_3G_3(TTAGGG)_3A_2$ and $(TTAGGG)_3A_2$ in K^+ are due predominantly (>95%) to a single intramolecular hybrid quadruplex conformer typified by a peak at 290 nm, a shoulder at 268 nm and a small trough at 240 nm (*8*).

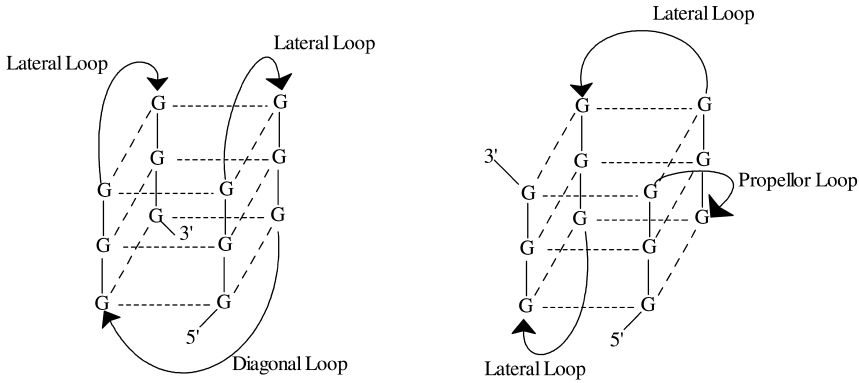


Figure 1. Two different unimolecular folding motifs as discussed above. All three loops in each motif have the sequence 5′-TTA-3′.

We have previously reported structural and thermodynamic studies on the human telomeric repeat $(TTAGGG)_4$ in the presence of potassium ion (*19*). Circular dichroism (CD) studies indicated the presence of a single intramolecular hybrid quadruplex prior to heat induced unfolding. Singular value decomposition of the CD spectra collected at different temperatures indicated three significant spectral species are most likely present in the pathway from the folded to unfolded state. Differential scanning calorimetry confirmed two distinct transitions in the unfolding process. A model was presented for the quadruplex to random coil transition, via formation of an unknown intramolecular intermediate, with exclusion of K^+ during each transition.

In the folded structure, there are three distinct loops each having the sequence 5′-TTA-3′ (See Figure 2). Starting at the 5′ end, the first loop is a propeller loop, while the second and third loops are lateral loops giving rise to a parallel-antiparallel-parallel-antiparallel strand orientation. Hence the orientation of the loops determines the orientation of the strands. The role of the loops in other quadruplexes has been investigated (*15*, *17*, *25–28*). These studies indicate that both loop sequence and length play a role in the secondary structure, stability and ligand binding of the folded quadruplex.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

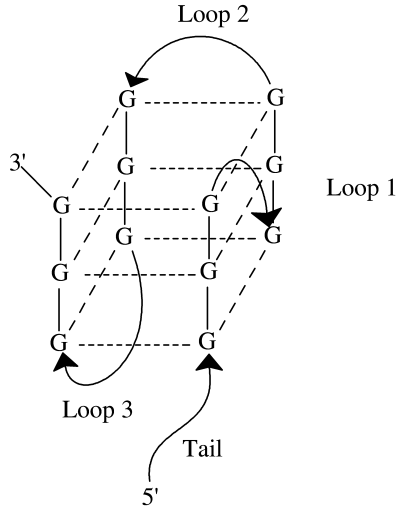


Figure 2. Schematic representation of the folded structure of $(TTAGGG)_4$. Each arrow has the sequence TTA and runs 5' to 3' and the dashed lines represent the Hoogsteen type hydrogen bonds holding the G-tetrads together. As can be seen, Loop 1 is a propeller loop while Loops 2 and 3 are lateral loops. The first segment of TTA is designated as the tail.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Here we report a spectroscopic study of a library of DNA oligomers related to the human telomere sequence. The general sequence of these oligomers is $(XXXGGG)_y$ where X = A or T and y =1, 2, 4 or 8. The goal of this project is to understand the effect of loop sequence (identity of X) , number of repeats (value of Y), and cations (Na^+ or K^+) present on the structure and stability of these sequences.

Materials and Methods

All HPLC-purified oligomers were purchased from Bio-Synthesis, Inc. (Lewisville, TX) and used without further purification. Standard phosphate buffers were prepared using either KH_2PO_4/K_2HPO_4 or NaH_2PO_4/Na_2HPO_4 to 10 mM total phosphate, 0.1 mM EDTA, pH 7.0 and KCl or NaCl added, respectively, to obtain a final $[K^+]$ or $[Na^+]$ of 150 mM. DNA samples were reconstituted in standard phosphate buffer and heated to 95 °C for 5 minutes before slowly cooling to room temperature. After sitting at room temperature for 2 hours, the samples were incubated at 5 °C for 24 hours. UV/Vis studies to determine concentrations were performed with a Varian Cary 100 Bio spectrophotometer (Palo Alto, CA). Samples were run from 320 – 220 nm at 25 °C and 95 °C and concentrations were calculated using the ε_{260} values found in Table 1 as provided by the vendor.

Using an Olis RSM 1000 spectrophotometer (Athens, GA), CD studies were performed in 0.1 cm circular quartz cuvettes between 25 °C and 95 °C at 5 °C

intervals. Data were collected every 1.0 nm over the range of 320 nm – 240 nm with an integration time of 3 seconds. Data were analyzed using Olis Global Works.

Table 1. DNA Oligomers Investigated in this Study

Loops Sequence Effects		*Number of Repeats Effects*	
Oligomer	*ε (mol^{-1} bases cm^{-1})*	*Oligomer*	*ε (mol^{-1} bases cm^{-1})*
$(TTAGGG)_4$	243,718	$(TTAGGG)_1$	60,929
$(TATGGG)_4$	243.718	$(TTAGGG)_2$	121,858
$(ATTGGG)_4$	243,717	$(TTAGGG)_4$	243,718
$(TTTGGG)_4$	220,114	$(TTAGGG)_8$	487,435
$(AATGGG)_4$	267,320		
$(ATAGGG)_4$	267,320		
$(TAAGGG)_4$	267,320		
$(AAAGGG)_4$	290,923		

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Results and Discussion

Effect of Loop Sequence

All DNA oligomers of the sequence $(XXXGGG)_4$ were characterized by both UV/Vis and CD spectroscopic determinations. Figure 3 shows typical UV/Vis (Panel A) and CD (Panel B) spectra of one of the members of this library, $(ATTGGG)_4$, at 25 °C and 95 °C. Examination of the UV/Vis spectra reveals little difference between 25 °C and 95 °C. The lack of significant hyperchromism upon denaturation observed in the UV/Vis spectra suggests that the stacking of the G-tetrads is such that the transition dipole moments interact in a fashion quite different from typical double stranded DNA. On the other hand, a considerable difference is observed for the corresponding CD spectra. Similar results were observed for all members of this library (data not shown) – little difference in the UV/Vis spectra but significant difference in the CD spectra for the two temperatures studied. This is an important observation when consideration of carrying out optical melting studies. Clearly, the CD spectra indicate that all oligomers are fully folded at 25 °C but are totally denatured and single stranded at 95 °C.

The CD spectra for all $(XXXGGG)_4$ oligomers at 25 °C are shown in Figure 4. For clarity, the spectra are grouped by which base occupies the third position of the XXX segment. In panel A are the spectra for that third base being A while the spectra for that third base being T are found in Panel B. Table 2 lists the significant spectral characteristics (wavelength, λ, of trough, peak or shoulder and molar ellipticity, θ) for these oligomers.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

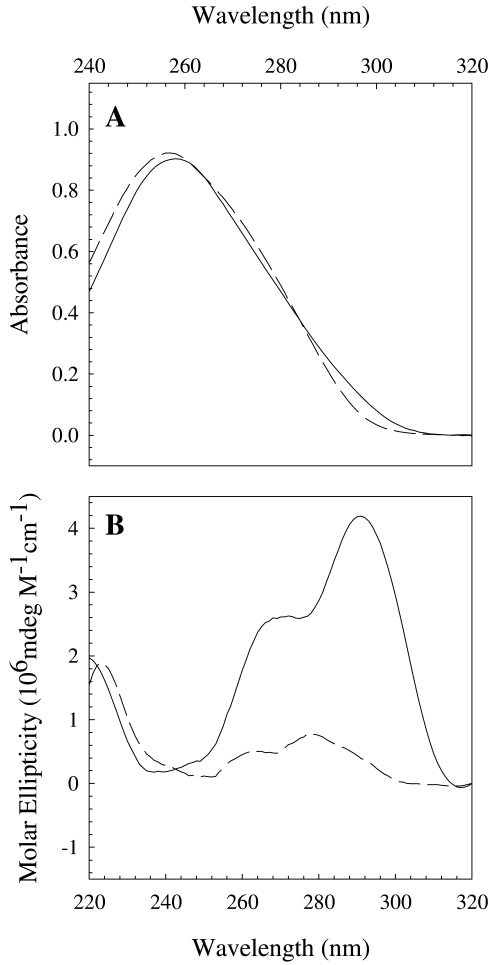


Figure 3. (A) The UV/Vis spectra of $(ATTGGG)_4$ in 10 mM phosphate buffer, pH 7.0 and 150 mM K^+ at 25 °C (solid line) and 95 °C (dashed line). (B) The CD spectra of $(AATGGG)_4$ in 10 mM phosphate buffer, pH 7.0 and 150 mM K^+ at 25 °C (solid line) and 95 °C (dashed line). Similar results were observed for all the members of the $(XXXGGG)_4$ libraries of oligomers.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

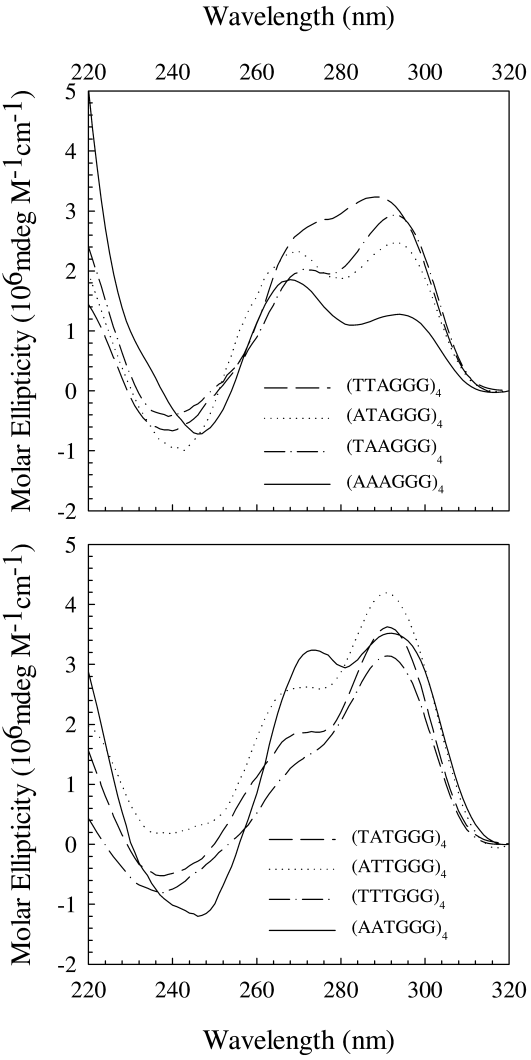


Figure 4. CD spectra of the $(XXXGGG)_4$ oligomers in 10 mM phosphate buffer, pH 7.0 and 150 mM K^+ at 25 °C. The upper panel contains the spectra for those oligomers with an A in the third position of the XXX segment and the lower panel contains the spectra of those oligomers with a T in the third position of the XXX segment.

Table 2. CD Spectral Characteristics of the $(XXXGGG)_4$ Oligomers

XXX	λ_1 *(nm)*	θ_1	λ_2 *(nm)*	θ_2	λ_3 *(nm)*	θ_3
TTA	237	-0.693	271	2.68	288	3.23
ATA	241	-0.945	270	2.32	293	2.47
TAA	238	-0.394	271	2.02	293	2.93
AAA	246	-0.721	268	1.86	294	1.28
TAT	236	-0.514	268	1.79	290	2.63
ATT	240	0.183	268	2.59	290	4.18
TTT	235	-0.768	268	1.26	290	3.12
AAT	244	-0.1.15	272	3.21	291	3.51

Molar ellipticities (θ) are reported as 10^6 mdeg mol^{-1} cm^{-1}.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Several interesting observations can be made from the data presented in Figure 4 and summarized in Table 2. First, each oligomer has a shallow trough in the range of 235 to 244 nm, a peak or shoulder at between 268-271 nm and a second peak at 288 to 293 nm. Oligomers with at least two As have a more well defined peak centered at 268 nm while oligomers with at least two 2 Ts have more of a shoulder at 268-270 nm. Conversely, oligomers with at least two Ts have a more pronounced peak at 288-293 nm.

In a recent study, Gray et al examined the effect of the polarities of stacked G-quartets with respect to the hydrogen bond donors and acceptors (*22*). A quadruplex with all strands oriented in a parallel fashion will necessarily have a parallel orientation for the hydrogen bond donors and acceptors within the G-quartet stacks. Such quadruplexes are characterized by CD peaks at around 260 nm and troughs at 239 nm. On the other hand, quadruplexes with a mixture of parallel and anti-parallel strand orientations will have a mixture of hydrogen bonding orientations with in the stacks. A quadruplex with all adjacent anti parallel hydrogen bonding G-quartet stacks will be characterized by CD peaks at 291 nm and 237 nm and troughs at 264 nm. As can be seen from the CD spectra in Figure 4, all quadruplexes have peaks in the range of 291 nm and 260 nm suggesting that these folded structure have both parallel and anti-parallel hydrogen bonding orientations. Thus, the orientations of the stacking of the G-quartets are the same for all oligomers studied.

If the stacking of the G-quartets is the same for all folded structures as discussed above, the slight variations in CD spectra must arise from the XXX segments. The first XXX segment is the tail of the quadruplex while the second, third and fourth XXX segments of these sequences comprise the loops in the folded structure (Figure 2). For these oligomers, the tail and each loop has the same sequence. Hence, the differences in the CD spectra can be attributed to: 1) different base content of the tail and the loops; 2) how the bases within the tail or within a loop stack with each other; and 3) how the bases within the tail or within a loop stack with the G-tetrads

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Although detailed calorimetric studies will be carried out on these sequences to gain insight into their thermodynamic stabilities, CD optical melts were carried out to generate a preliminary stability order. For these studies, the CD spectrum for each oligomer was recorded every 5 degrees from 25 °C to 95 °C. A plot of the NSR (Normalized Spectral Response) at 288 nm vs temperature yields an optical melting profile as shown in Figure 5.

Determination of the midpoint of each transition plotted in Figure 5 reveals the so called melting temperature, T_m. These values are reported in Table 3. As can be seen most oligomers have T_m values between 73-76 °C, while $(ATTGGG)_4$ and $(TTTGGG)_4$ have T_m values between 65 and 67 °C and, surprisingly, $(AAAGGG)_4$ has a T_m of only about 44 °C. These differences must be attributed to sequence context of the tails and loops. There is much published data on the effect of dangling unpaired bases (i.e., overhangs) on the stability of oligomeric DNA duplexes as well as on the presence of loops and bulges. Typically, the degree of stabilization an overhang offers a duplex is dependent upon the length of the overhang and its sequence context. Overhangs of 1 to 3 bases are more stabilizing than longer overhangs and A is more stabilizing than T (*29–32*). If the tail of the quadruplex is analogous to a 5′ overhang of a DNA duplex, it's presence should be stabilizing and an A adjacent to the G-quartet should be more stabilizing than a T, opposite to what is observed for the all A tail. Thus, it is likely that the lower T_m of $(AAAGGG)_4$ is influenced by the loops. In the absence of any solution structure, the source of this anomaly cannot be determined at this time.

It must be kept in mind, however, that these T_m values are based on a two state transition (folded to unfolded), which may not be the case. We previously reported that the unfolding of the wild type sequence $(TTAGGG)_4$, proceeded through an intermediate in a three state fashion (*19*). The advantage of recording the full spectrum at a particular temperature throughout the melting process allows for analysis of the transition using Singular Value Decomposition (SVD). SVD allows determination of the number of significant spectral species throughout any transition. The determination of the number of significant spectral species present during a transition is based upon the magnitudes of the singular values, the magnitudes of the values from the U matrix and V matrix autocorrelations and the randomness of residual plots (*19*, *33*). The OLIS 100 has software built in to carry out this analysis upon completion of the melting experiment. However, the OLIS program analysis reports weighting factors, U matrix autocorrelation values and basis spectra. Analysis of the data presented in Table 4 indicates that the values of the weighting factors and U matrix correlation fall off dramatically after the third component. Hence our suggestion that the transition from the folded state to the unfolded state for each oligomer studied is follows a three state transition. Hence the T_m values reported can only be used to assess relative stabilities. Actual thermodynamic parameters must be obtained calorimetrically. Further, the DSC results can be used to determine the number of states in the transitions.

The differences in the observed T_m values can be attributed to the same reasons for the differences in the CD spectral characteristics: 1) tail and loop sequence context; 2) stacking of the bases with each other within the tail and within a loop; and 3) stacking of the bases within the tail or within a loop with the G-tetrads. In

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

consideration of sequence context of the XXX segments and its contribution to overall stability, it is difficult to see any particular trends from the data in Table 3. Calorimetric studies are required to further address sequence context issues. None the less, the spectral and melting data all clearly demonstrate that even small changes in sequence context of the loops can have dramatic effects on the structure and stability of the folded quadruplex.

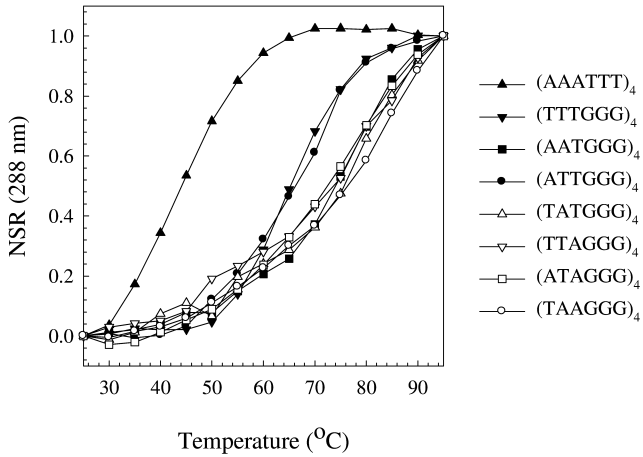


Figure 5. CD optical melting profiles for the $(XXXGGG)_4$ oligomers. The NSR is defined as the fraction of the single strands throughout the transition and is calculated by $(\theta_{25} - \theta_T)/(\theta_{25} - \theta_{95})$ where θ_{25} is the molar ellipticity at 25 °C, θ_{95} is the molar ellipticity at 95 °C and θ_T is the molar ellipticity at temperature T. The midpoint of the transition is defined as T_m.

Table 3. Melting Data for the $(XXXGGG)_4$ Oligomers

Oligomer	*"T_m" (°C)*	*SVD Spectral Species*
$(TTAGGG)_4$	73.5	3
$(TATGGG)_4$	75.8	3
$(AATGGG)_4$	66.5	3
$(TTTGGG)_4$	65.2	3
$(AATGGG)_4$	74.2	3
$(ATAGGG)_4$	73.9	3
$(TAAGGG)_4$	74.7	3
$(AAAGGG)_4$	44.2	3

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Effect of Number of Repeats

For this aspect of the study, DNA oligomers of general sequence $(TTAGGG)_y$, where y = 1, 2, 4, or 8 were used to study the effect of the number of repeats of (TTAGGG) on the conformation and stability of the resultant structure in the presence of either Na^+ or K^+. The results of CD determinations are shown in Figure 6.

Table 4. SVD Analysis Results for the $(XXXGGG)_4$ Oligomers

Oligomer	*Component*	*Weighting Factor*	*U Matrix Auto Correlation Value*
$(TTAGGG)_4$	1	1.00	0.932
	2	0.453	0.807
	3	0.188	0.564
	4	0.037	-0.216
$(TATGGG)_4$	1	1.000	0.937
	2	0.834	0.768
	3	0.756	0.668
	4	0.113	-0.095
$(ATTGGG)_4$	1	1.000	0.935
	2	0.117	0.868
	3	0.060	0.821
	4	0.018	-0.522
$(TTTGGG)_4$	1	1.000	0.949
	2	0.421	0.802
	3	0.116	0.818
	4	0.024	0.027
$(AATGGG)_4$	1	1.000	0.922
	2	0.363	0.857
	3	0.085	0.808
	4	0.030	0.527
$(ATAGGG)_4$	1	1.000	0.918
	2	0.502	0.846
	3	0.175	0.699
	4	0.042	-0.020
$(TAAGGG)_4$	1	1.000	0.935

Continued on next page.

Table 4. (Continued). SVD Analysis Results for the $(XXXGGG)_4$ Oligomers

Oligomer	*Component*	*Weighting Factor*	*U Matrix Auto Correlation Value*
	2	0.560	0.820
	3	0.24	0.345
	4	0.047	-0.114
$(AAAGGG)_4$	1	1.000	0.940
	2	0.331	0.807
	3	0.041	0.561
	4	0.033	-0.092

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

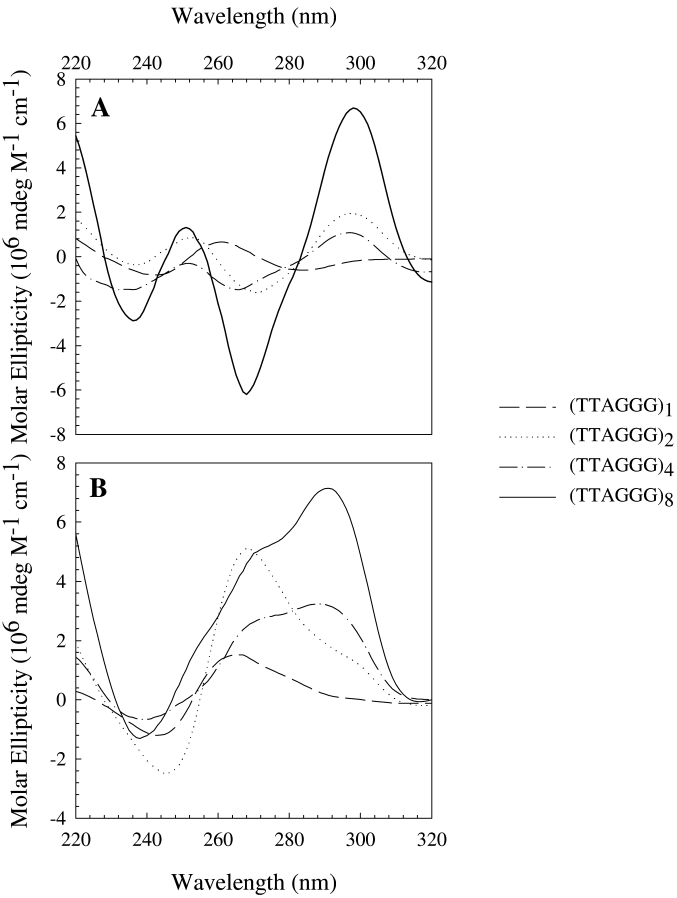


Figure 6. The CD spectra of the $(TTAGGG)_y$ oligomers(where y = 1, 2, 4 or 8) in 10 mM phosphate buffer, pH 7.0 at 25°C with either (A) 150 mM Na^+ or (B) 150 mM K^+.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

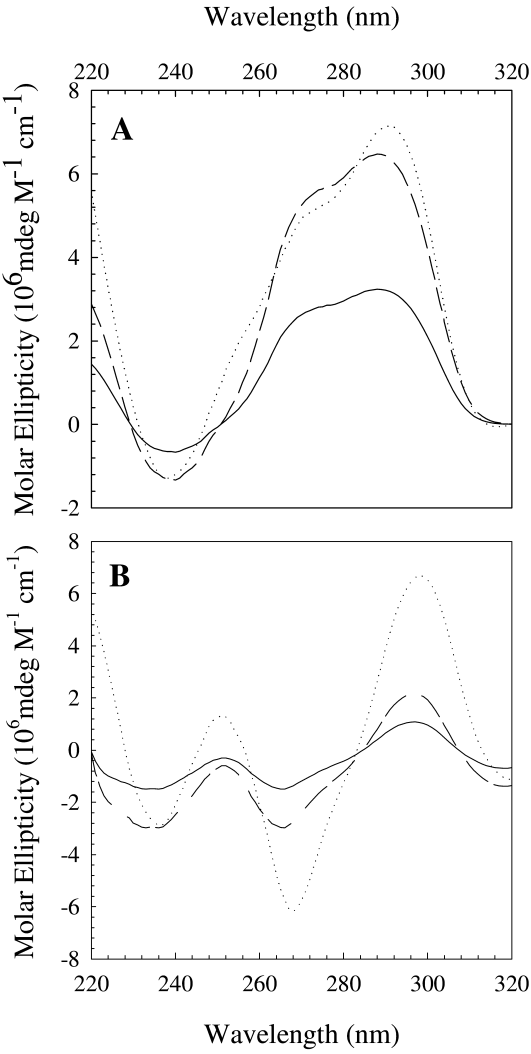


Figure 7. The CD spectra of $(TTAGGG)_4$ (solid line), $(TTAGGG)_8$ (dotted line) and $(TTAGGG)_4$ + $(TTAGGG)_4$ (dashed line) in standard phosphate buffer, pH 7.0, at 25 °C with (A) 150 mM Na^+ or (B) 150 mM K^+.

These oligomers were studied to not only investigate the influence of the number of repeats on the conformation of the resultant quadruplex but also to determine the molecularities of those quadruplexes. Inspection of the CD spectra clearly shows that the CD signature, and hence conformation, of the quadruplex is highly dependent upon both the cation used and the number of repeats, except for the case of only one repeat. Regardless of cation used, the CD spectrum of $(TTAGGG)_1$, with a peak centered at 261 nm and a trough at 242 nm, is consistent

with a parallel stranded quadruplex with a molecularity of four (*21*). The CD spectra of the oligomers with 2, 4 or 8 repeats are very similar in Na^+, with peaks centered at 296 nm and 251 nm and troughs at 268 nm. These spectra are consistent with a quadruplex formed from a bimolecular association. However, $(TTAGGG)_4$ and $(TTAGGG)_8$ could also be forming unimolecular chair type structures with lateral loops with $(TTAGGG)_8$ possessing two such structures linked together. The differences in CD spectral characteristics are just not fine enough to make the distinction between the two possible forms. In K^+, the CD spectrum of $(TTAGGG)_2$ is very reminiscent of that for a bimolecular propeller type structure with external loops determined by the slight shoulder at 297 nm as observed previously (*21*). The CD spectra of $(TTAGGG)_4$ and $(TTAGGG)_8$ are very similar in K^+ with peaks at 288 nm, shoulders at 271 nm and troughs at 237 nm, albeit with different intensities. The sequence $(TTAGGG)_8$ was included in this study to see if it could form a double quadruplex. Figure 7A shows that the CD spectrum of $(TTAGGG)_8$ in K^+ is simply the sum of $(TTAGGG)_4$ with itself. Hence, $(TTAGGG)_8$ is a double quadruplex with each individual quadruplex having the same structure as $(TTAGGG)_4$. Similar results have been reported for $(TTAGGG)_4$ added to $(TTAGGG)_4TT$ to yield $(TTAGGG)_8TT$ (*34*). However, the same cannot be said of $(TTAGGG)_8$ in Na^+ (Figure 7B). Clearly the CD spectrum of $(TTAGGG)_8$ is not the sum of $(TTAGGG)_4$ with itself. Assuming that each individual quadruplex in $(TTAGGG)_8$ have the same conformation in either Na^+ or K^+, the differences in additivity must arise from how the quadruplexes stack with each other. Of course, one cannot rule out the possibility that the two quadruplexes in $(TTAGGG)_8$ have different conformations in Na^+.

Summary

A thorough CD spectroscopic investigation of DNA oligomers of general sequence $(XXXGGG)_y$ (where X = A or T and y = 1, 2, 4, or 8) reveals that the conformations of the quadruplexes formed from these sequence are highly dependent upon sequence context, number of repeats and the identity of the cation present. The different conformations arise from different molecularities (4, 2 or 1), loop orientations (propeller or lateral) and strand orientations (parallel, antiparallel or combinations of parallel and antiparallel). A thorough calorimetric studied is now underway to determine how sequence effects, length effects and identities of cation affect thermodynamic stabilities of the sequences.

Acknowledgments

The authors wish to thank the Robert A Welch Foundation for the departmental research grant supporting this research and the National Science Foundation (NSF# 0820845) for the acquisition of the CD spectrometer.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

References

1. Williamson, J. R. G-Tetrad structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730.
2. Dai, T.; Marotta, S. P.; Sheardy, R. D. Self-assembly of DNA oligomers into high molecular weight species. *Biochemistry* **1995**, *34*, 3655–3662.
3. Venczel, E. A.; Sen, D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* **1993**, *32*, 6220–6228.
4. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
5. Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure* **1993**, *1*, 263–282.
6. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
7. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. Structure of the human telomer in K^+ solution: an intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
8. Xu, Y.; Noguchi, Y.; Sugiyama, H. The new models of the human telomere d[$AGGG(TTAGGG)_3$] in K^+ solution. *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591.
9. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by the natural human telomere sequence in K^+ solution. *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
10. Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Kim, N.; Phan, A. T. Structure of the human telomere in K^+ solution: A stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
11. Kuryavyi, V.; Patel, D. Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure* **2010**, *18*, 73–82.
12. Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K^+ solution. *Nucleic Acids Res.* **2010**, *38*, 1009–1021.
13. Viglasky, V.; Bauer, L.; Tluckova, K. Structural features of intra- and intermolecular G-quadruplexes derived from the human telomere. *Biochemistry* **2010**, *49*, 2110–2120.
14. Rachwal, P. A.; Fox, K. R. Quadruplex melting. *Methods* **2007**, *43*, 291–301.
15. Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M. T.; De Pauw, E.; Gabelica, V. G-quadruplex DNA assemblies: Loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, *130*, 10208–10216.
16. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 5482–5515.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

17. Balkwill, G. D.; Garner, T. P.; Searle, M. S. Folding of single-stranded DNA quadruplexes containing an autonomously stable mini-hairpin loop. *Mol. BioSys.* **2009**, *5*, 542–547.
18. Olsen, C. M.; Marky, L. Monitoring the temperature unfolding of G-quadruplexes by UV and circular dichroism spectroscopies and calorimetry techniques. *Methods Mol. Biol.* **2010**, *608*, 147–158.
19. Antonacci, C.; Chaires, J. B.; Sheardy, R. D Biophysical characterization of the human telomeric repeat $(TTAGGG)_4$ in potassium solution. *Biochemistry* **2007**, *46*, 4654–4660.
20. van Steensel, B.; Smogorzewska, A.; de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **1998**, *92*, 401–413.
21. Paramasivian, S.; Rujan, I.; Bolton, P. Circular dichroism of quadruplex DNAs: Applications to structures, cation effects and ligand binding. *Methods* **2006**, *43*, 324–331.
22. Gray, D. M.; Wen, J.-D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **2008**, *20*, 431–440.
23. Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
24. Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, S.; Perone, R.; Randazzo, A.; Spada, G. P. A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.* **2010**, *8*, 2683–2692.
25. Risitano, A.; Fox, K. R. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **2004**, *32*, 2598–2606.
26. Kumar, N.; Sahoo, B.; Varun, K. A. S.; Maiti, S.; Maiti, S. Effect of loop variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.* **2008**, *36*, 4433–4442.
27. Balkwill, G. D.; Garner, T. P.; Williams, H. E. L.; Searle, M. S. Folding topology of a bimolecular quadruplex containing a stable mini-hairpin motif within the diagonal loop. *J. Mol. Biol.* **2009**, *385*, 1600–1615.
28. Fugimoto, T.; Miyoshi, D.; Tateishi-Karimata, H.; Sugimoto, N. Thermal stability and hydration state of DNA G-quadruplex regulated by loop regions. *Nucleic Acids Symp. Ser.* **2009**, *53*, 237–238.
29. Senior, M.; Jones, R. A.; Breslauer, K. J. Influence of dangling thymne residues on the stability and structure of two DNA duplexes. *Biochemistry* **1988**, *27*, 3879–3885.
30. Marotta, S. P.; Sheardy, R. D. Conformational properties of Z-forming DNA oligomers bearing termijnal unpaired bases. *Biophys. J.* **1996**, *71*, 3361–3369.
31. Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acds Res.* **2000**, *28*, 1929–1934.
32. Ricelli, P. V.; Mandell, K. E.; Benight, A. S. Melting studies of dangling-ended DNA hairpins: Effects of end length, loop sequence and biotinylation. *Nucleic Acids Res.* **2002**, *30*, 4022–4093.

33. Haq, I.; Chowdhry, B. Z.; Chaires, J. B. Singular value decomposition of 3-D DNA melting curves reveals complexity in the melting process. *Eur. Biophys. J.* **1997**, *26*, 419–426.
34. Petraccone, L.; Trent, J. O.; Chaires, J. B. The tail of the telomere. *J. Am. Chem. Soc.* **2008**, *130*, 16530–16532.

Downloaded by UNIV OF GUELPH LIBRARY on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch004

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Chapter 5

Nucleosomal DNA: Kinked, Not Kinked, or Self-Healing Material?

Rajib Mukherjee and Thomas C. Bishop*

Departments of Chemistry and Physics, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272, U.S.A.
***E-mail: bishop@latech.edu. Tel.:(318)257-5209. Fax: (318)257-3823.**

We use all atom molecular dynamics simulations to investigate the structure and dynamics of nucleosomal DNA. Our primary focus is characterization of DNA kinking as a function of position and base pair type within the nucleosome. For this purpose we have constructed 16 fully solvated all atom models of the nucleosome and subjected each to molecular dynamics simulations. Each system consists of over 200,000 atoms and is simulated for 16 ns. Combined the systems represent over 3.2 million atoms and 256 ns of nucleosome dynamics. Each system has a different sequence of DNA. The sequences chosen allow us to investigate each base pair type at each of the 147 positions in the nucleosome, as well as, each of the 16 dinucleotide steps at each of the 146 positions in the nucleosome. Within the limits of this approach we demonstrate that threading different DNA sequences onto a given histone core allows for successful initiation of molecular modeling. We also show that location within the nucleosome affects the structure and dynamics of a given base pair step type, i.e. the histones alter the DNA. However this effect is not so strong as to impose the same structure and dynamics on DNA regardless of sequence. Analysis of interaction energies demonstrates that the histone-DNA interactions vary more as a function of position than as a function of sequence. Moreover, the relative contribution from van der Waals interactions and from Coulomb interactions to the total histone-DNA interaction energy also varies with position. Thus the magnitude and the nature of the histone-DNA interactions vary with location. To

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

analyze DNA kinking in the nucleosome we employed three definitions of kinks, included are a weak and a strong kink criteria. No single kink, by any definition, persisted for an entire 16 ns simulation. The longest continuous lifetime for an individual kink was approximately 6 ns. However, picosecond lifetimes are more typical. On nanosecond time scales kinks are self-healing, repeatedly appearing and disappearing at some locations. Kinking is not determined only by location or only by DNA sequence. Rather kinking results from a combination of the stresses imposed on DNA by the histone core and the sequence specific material properties of DNA. Both are decidedly non-uniform.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

1. Introduction

Nucleosomes consists of 147 base pair of DNA wrapped approximately 1.7 times around a core of eight histones, see Figure 1. The histones prefer some sequences of DNA over others, but biologic functionality demands the histones be capable of binding to any sequence. *In vitro* histones occupy preferred locations on lengths of DNA greater than 147 base pair (positioning) and exhibit preferential binding in mixtures containing different 147 base pair long fragments (affinity). The basis for such sequence preferences has long been attributed to the material properties of DNA (*1*). DNA sequences which more easily conform to the superhelix structure because of their intrinsic flexibility or intrinsic shape will more readily form nucleosomes. This implies that variations in DNA self-interactions dominate over histone-DNA, histone self-interactions and solvent interactions. *In vivo* nucleosome positioning has been reported for several whole genomes, but the interpretation of results and the role of experimental bias is debated, as evidenced by a recent exchange of letters (*2–5*). *In vivo* additional factors can also alter nucleosome stability, including histone modifications, linker DNA constraints, ion concentration and species effects.

The physical basis for nucleosome stability and relationships between positioning and affinity remain unclear. In fact several contradictory sequence motifs have been proposed as nucleosome positioning sequences. One proposes that a well positioned nucleosome will have A·A, T·T or T·A steps at locations where the minor groove faces in towards the histone core and G·C steps at the minor groove out locations (*6*). Another motif requires strong base pair stacking interactions at the minor groove in and weak stacking interactions at the minor groove out locations (*7*). This results in C·G steps at minor groove in and A·T steps at minor groove out locations. Yet a third suggests A·T steps at minor groove in and C·G steps at minor groove out locations (*8*).

Models based on elastic deformations of DNA coupled with simple approximations for solvent and electrostatic interactions do not provide clear insight (*9*). Crick and Klug (*10*) proposed DNA kinking as a means of reducing the energy barrier associated with superhelix formation. By their definition, kinking is a non-elastic deformation that corresponds to a transition from second

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

order to first order variation of energy as a function of deformation. Kinks are identified in the highest resolution X-ray crystal structure of nucleosome (*11*), but the definition employed is based on structure rather than energetics. In (*11*) a kink requires a large negative Roll angle between consecutive base pairs. Negative Roll opens the major groove and thus kinks, by this definition, occur at locations where the minor groove faces the histone core. Analysis of all available nucleosome structures indicates that such kinks are not required to reconstruct the nucleosome superhelix at atomic resolution (*12*).

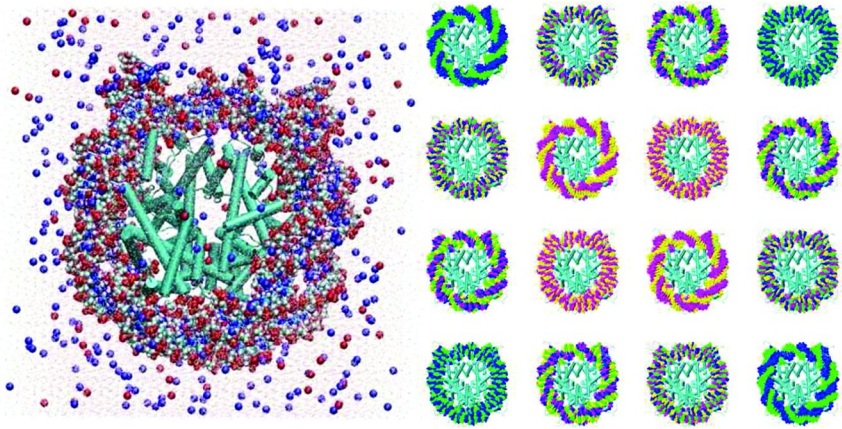


Figure 1. Systems Modeled. Left: A molecular graphics image of one of the fully solvated systems. Ions are colored by charge and are indicated as spheres. Water appears only as dots, and only the heavy atoms in DNA are shown as van der Waals spheres for clarity. All histones are represented in one color as a simple cartoon. The unstructured histone tails are not visible Right: The 16 individual systems simulated are displayed without any water or ions as an array of images. Each nucleosome was simulated as a single system with explicit solvent, in the left image. The homopolymer systems appear on the diagonal as AA, CC, GG, and TT. Dinucleotide repeats along each row: top row AA, AC, AG, AT, 2nd row CA, CC, CG, CT, etc… Bases are colored as follows: A = blue, C= yellow, G= magenta, T= green. As molecular entities the DNA in AA and TT are equivalent but the orientation of each base pair is flipped with respect to the histone core. Thus in the top left nucleosome, AA, blue crosses over green at the dyad, top-center of image, while in the bottom right nucleosome, TT, green crosses over blue. (see color insert)

There is no consensus on what defines a DNA kink or the structural and dynamical properties associated with DNA kinking. Here, we utilize all atom molecular dynamics simulations to study the structure and dynamics of nucleosomal DNA and to investigate DNA kinks in particular. We have designed 16 all atom models of the nucleosome, each with a different sequence of DNA. The sequences represent four homopolymers and twelve oligomers composed of alternating dinucleotides. These sequences allow us to investigate both base pairs (A:T and C:G) and their alternate orientations (T:A and G:C) at all 147 possible

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

base pair positions in a nucleosome. These sequences also place each of the 16 possible dimer steps (A·A, A·C, … T·G, T·T) at each of the 146 possible step locations in a nucleosome

We analyze the histone-DNA interaction energies for each base pair type and conduct a structural analysis based on the inter base pair helical step parameters: Tilt, Roll, Twist, Shift, Slide, and Rise. Different definitions of kinking are considered and are employed to investigate the formation of DNA kinks, their location within the nucleosome and lifetime properties.

2. Methods

In this section we provide details of how the 16 systems were created, the molecular dynamics simulation protocols employed and the methods used for data extraction and analysis. The definition of a DNA kink is included in this section along with expectations based on statistical arguments for how often a kink should be observed in our simulations.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Simulation Methods

System Setup

All models were built using protein databank entry **1kx5** as the initial input. We shortened the histone tails in **1kx5** in order to minimize the amount of solvent required to fully solvate the nucleosome. The amino acid composition of each histone was as follows: H2A, 10-128; H2B, 10-122; H3, 30-135; and H4, 1-102 (numbering corresponds to the numbering in **1kx5**). This enabled us to keep the total system size near 200,000 atoms. Actual system sizes ranged from 205,950 to 206,097 atoms and varied only in DNA content.

A cubic periodic cell of 152 Å × 158 Å × 95 Å, containing 60,548 waters, was sufficient to provide a 12 Å solvent layer for our modified nucleosome. Enough monovalent ions were located within this solvent box to both neutralize the system (170 Na^+ ions) and provide for a bulk ion concentration of 150 mM (155 Na^+ ; 155 Cl^-). The histones and solvent were saved as AMBER library files and then combined with different DNA molecules constructed as detailed below to create each initial model. This methodology was employed to eliminate, in so far as possible, variations in the initial system configuration that might affect the sampling observed during our simulations. Thus, except for the DNA sequence and the corresponding atomic model the starting configuration of each system was the same, see images in Figure 1.

We created 16 different 147 nucleotide long sequences of DNA and folded each sequence into an all atom model using 3DNA (*13*). Each initial model of DNA possessed the same set of intra and inter base pair step helical parameters as the DNA in crystallographic structure **1kx5**. Thus, in terms of helical parameters, the initial conformations of DNA in all systems were identical. We note that if the Cartesian coordinates of DNA in **1kx5** are converted to helical parameters and then the helical parameter description is converted back to Cartesian coordinates

then the root mean square deviation between the x-ray and converted coordinates is less than 1 Å rmsd (*12*).

We docked each modeled sequence onto the x-ray structure by root mean square fitting the 294 phosphorous atoms of the model to those in the **1kx5** structure. Once the DNA was oriented, we combined its coordinates with our histone library and our nucleosome solvent library using AMBER's tleap module to create the initial parameter and coordinate files. All simulations used the parm99 force-field parameters (*14*) with parmbsc0 corrections for DNA (*15*), TIP3P water, sodium and chlorine parameters as distributed with AMBER 10 (*16*). NAMD version 2.7 (*17*) was used for all minimization, equilibration and dynamics calculations.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Choice of DNA Sequences

The 16 sequences chosen allow us to investigate the structure and dynamics of each of the 16 base pair step types (A·A, A·C, … T·G, T·T) at each of the 146 different locations within a nucleosome. Only 10 of the 16 are unique types for free DNA, but once placed in the context of the nucleosome these symmetries are removed. For example as a molecular entity a T·T step is equivalent to an A·A step, but within the context of the nucleosome a T·T step at a given location is not equivalent to a A·A step. The orientation is reversed and the strand of DNA closest to the histone core becomes the one farthest away. The solvent environment and opportunities for contacts with histone may differ significantly with orientation. We therefore consider all 16 base pair step types. This can be achieved with 16 sequences. The sequences represent four homopolymers, i.e. repeats of A, C, G, or T, and twelve oligomers composed of dinucleotide repeats, i.e. repeats of AC, AG, … We label each simulation according to sequence. The label *AC* represents the system which contained the shorted histones, solvent, ions, and DNA with repeated A·C dinucleotide steps. Note for this system every other step is a C·A step.

The central base pair (dyad) for any sequence *XY* is a $Y{:}\tilde{Y}$ pair and the sequence has *X*:*X* pairs at both the 5' and 3' ends. (Here a tilde denotes the complementary base pair.) The sequences are not palindromic. In terms of individual base pairs, the simulations represent each base pair at each of the 147 positions eight times. For example, the systems *AA*, *CA*, *GA* and *TA* each have A:T base pairs located at the dyad and at all even numbered steps from the dyad, while the systems *TT*, *GT*, *CT* and *AT* have T:A pairs at the dyad and all even numbered steps. The latter set represents the same base pair, but its' orientation is inverted compared to the first set.

Simulation Protocol

We utilized the same simulation protocol as in our previous nucleosome studies (*18*, *19*). Briefly the systems were minimized with a cutoff of 12 Å and a maximum move of 0.1 Å per step for 1000 steps, a time step of 1 fs

and shake applied to all hydrogen. Minimization beyond this point was not a necessary prerequisite for equilibration suggesting that the initial fitting and solvating procedure were acceptable. The system was then equilibrated in two stages. Both stages utilized NAMD's Berendsen pressure coupling scheme (relaxation time 500 fs, target pressure 1.01325 atm, compressibility 4.57E-5), PME electrostatic calculations at every step (grid size 128 Å × 128 Å × 128 Å, 4th order interpolation, Ewald coefficient 0.26, tolerance 1E-6, max spacing 1.5 Å), NAMD's C1 switching from 10 Å to 12 Å, pair list cutoff of 14 Å, 1-4scaling of 0.8333, the SETTLE algorithm applied to all hydrogen, and a time step of 2 fs. In stage one, 150 ps were computed with a temperature reassignment every picoseconds that began at 0 K, was incremented by 50 K for each reassignment, and achieved a maximum reassignment temperature of 300 K. During the second stage 150 ps were conducted with NAMD's temperature coupling to a 300 K bath and temperature reassignment turned off. After equilibration, 16 ns of NTP dynamics were computed with a pressure relaxation time of 5 ps and Langevin damping coefficient of 5 ps. Only the 16 ns production runs are considered for any of the analysis.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Simulation Throughput

Our workflow strategy was inspired by NAMD-G (*20*) but was structured so as to take advantage of Louisiana Optical Network Initiative (LONI) resources. LONI consists of five supercomputing installations located across the state of Louisiana connected by a dedicated fiber optic network that allows each supercomputer fast access to a distributed data network called PetaShare (*21*). We used four of LONI's 128 node (512 cores) Linux clusters and PetaShare to accomplish our simulations with a high throughput high performance computing workflow strategy.

Each simulation utilized 64 cores (16 nodes) and all tasks were run in 1 ns increments, requiring approximately 16 hours of run time per nanosecond of simulation. In total there were 16 × 16 = 256 tasks to be completed. Each task began by retrieving the necessary inputs (extended system, coordinate, velocity and parameter files) from PetaShare and ended by depositing the outputs (extended system, coordinate, velocity, and trajectory files) into PetaShare. During run time, trajectory data was also copied on a daily basis from PetaShare to local computing resources for the purpose of analysis and visualization. This avoided a lengthy post-production trajectory transfer. In this manner we were able to complete all simulation tasks during two separate 10 day periods. The first period produced 8 ns for each of the 16 systems, and the second period produced the final 8 ns for each system. We have formalized the work flow tools pioneered for this effort and expanded their capabilities (*22*). These tools are available as Python scripts that can be downloaded by following the ManyJobs link at http://dna.engr.latech.edu.

Analysis Techniques

Trajectory Sampling and Standard Analyses

During run time, simulation snapshots were saved every 1 ps, producing 16,000 snapshots for each simulation. For some analyses the snapshots were reduced to a 10 ps sampling (1600 snapshots) and for others to a 100 ps sampling (160 snapshots).

RMSD values for the protein backbone and DNA heavy atoms were calculated using the 1 ps sampling. RMSD values for the entire histone complex exhibited long time scale variations and anomalously high values (> 4.0Å), data not shown. This result is entirely due to the unstructured histone tails. For the well-structured histone core, identified as H2A, 15-116; H2B, 26-121; H3, 39-133; and H4, 23-100 (numbering corresponds to **1kx5**), the rmsd for all heavy atoms remained below 2.75 Å for the duration of each simulation. RMSD values obtained from consideration of DNA heavy atoms never exceeded 3.5 Å; 2.5 Å was a more typical value. B-factors were calculated for the histone CA atoms using AMBER's ptraj utility and the 10 ps sampling. Only the histone core domain was used to calculate protein-DNA interaction energies as described below.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Energy Analysis

The NAMD energy plug-in for VMD (*23*) was used to calculate Coulomb and van der Waals interaction energies between the histone core and DNA for the entire length of DNA and separately for each base pair. A total of 160 equally spaced snapshots from each 16 ns trajectory were evaluated. For this analysis the water and ions were stripped from the snapshots and a dielectric constant of 80 was used as an *ad hoc* method of approximating solvent screening. For all energy calculations a cutoff of 400 Å was used with switching that began 300 Å and no periodic boundaries. This provided a complete accounting of the long range Coulomb interactions in our analysis. Note that during the simulations PME with periodic boundaries and a 10-12 Å switching function was utilized to determine long range interactions. The simulations contained explicit solvent and ions and therefore used a dielectric constant of 1.

Base Pair Step Analysis

Local base pair step parameter values were obtained for 16,000 equally spaced snapshots from each simulation using 3DNA (*13*). The base pair step parameters are Shift, Slide, Rise, Tilt, Roll, and Twist (*24*, *25*). These parameters measure the relative translation and rotation, respectively, from one base pair plane to the next.

A single "findpair" descriptor in 3DNA that identified all bases as paired was used for each trajectory. In this manner each snapshot reported 147 base pairs and a set of 146 helical step parameters regardless of whether or not the bases exhibited full hydrogen bonding. Statistical analysis included determination of the mean,

max, min, range, standard deviation and normality for each of the 16 x 146 sets of 16,000 inter base pair helical parameters. We utilized the Jarque-Bera test (*26*) to categorize the distributions as normal. Since some base pair steps are known to exhibit non-normal behavior even when free in solution (*27*), we report the normality using confidence levels of 0.80, 0.90, 0.95, and 0.99. A score of zero is assigned to any step that fails the minimum normality test criteria of 0.80. In some instances the mean and standard deviation values were observed to be anomalously high for individual base pair step parameters. This observation provides reason to investigate the following metrics for DNA kinking.

Kink Analysis

To date, there is no consensus on how to define a kink. Conceptually a kink can be defined as a deviation of any of the base pair step parameters beyond some specified range of values. This is a geometric definition rather than a definition based on energetics. By such definition, even DNA free in solution is expected to spontaneously kink due to thermal fluctuations. For our kink analysis we utilize sequence specific DNA helical parameter values reported by the Ascona B-DNA Consortium (ABC) (*27*). These values were obtained from simulations of B-form DNA free in solution.

We recently assessed DNA helical parameters for all available nucleosome structures in the protein data bank (*12*). The Richmond-Davey kinks, described below, are not required to achieve the overall superhelix geometry. We also characterized the DNA superhelix mathematically as having properties of a shear helix and torsion helix. It is best characterized as a Roll-Slide-Twist helix (*12*, *28*). We thus focused on deformations in Roll, Slide and Twist as the criteria for kinking instead of using all six helical parameters. The same three helical parameters where used to identify kinks in an x-ray crystal structure of the nucleosome (*11*). For the x-ray structure, a step was defined as kinked if Slide > 1.5 Å, Roll <-18° and Twist > 40°. We refer to this definition as a Richmond-Davey kink and denote it K_{rd}. The particular values utilized correspond to approximately three standard deviations from the mean values obtained from simulations of DNA free in solution (*27*). This definition is biased such that kinks occur only where the minor groove faces inwards (negative Roll), and the definition does not explicitly allow for sequence specific properties of DNA.

We utilize a more general definition. Our definition of kink is not biased towards specific locations within the nucleosome. It also accounts for the sequence specific flexibility of DNA and variations in the intrinsic conformation of individual base pair step types. We define a base pair step as kinked when the following criteria are met:

$$|Ro - \langle Ro \rangle| > n\sigma_{ro} \text{ and } |Sl - \langle Sl \rangle| > n\sigma_{sl} \text{ and } |Tw - \langle Tw \rangle| > n\sigma_{tw} \quad (1)$$

where *Ro* denotes the helical parameter Roll, *Sl* denotes Slide and *Tw* denotes Twist. A similar definition was considered in (*29*). To determine if a given base pair step is kinked or not at any given instant of time we compare the helical

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

parameter values obtained from the snapshot to the mean (denoted by angle brackets) and standard deviation (denoted by σ) values obtained from the ABC simulations of DNA free in solution (*27*). In all cases we use sequence specific values, e.g. the á*Ro*ñ and σ_{ro} values used for T·A and C·C steps differ. The step is identified as kinked only if all three inequalities are satisfied. With this definition we further identify two types of kinks. A strong kink, denoted K_s, is defined by $n = 3$, i.e. all three helical parameters must be greater than three standard deviations away from the conformation expected for DNA free in solution. A weak kink, denoted by K_w, is defined by one standard deviation, $n = 1$.

Since our definition of kink is based on sequence specific conformation and thermal fluctuations we can predict how often kinks should occur. Assuming the helical parameter values exhibit normal distributions we expect any single helical parameter, e.g. Roll, to exceed the 1σ threshold in approximately 32 % of the snapshots and the 3σ threshold in less than 0.27% of the snapshots. If Roll, Slide and Twist are tightly coupled, then we expect all three to exceed the threshold simultaneously. In this case, the probabilities of a kink are approximately 1 of every 3 snapshots and 1 of 370 snapshots for K_w and K_s respectively. If the three helical parameters are not tightly coupled but rather deform independently then the expectations are $(0.32)^3$ and $(0.0027)^3$ or approximately 1 of every 31 snapshots and 1 of every 51,000,000 for K_w and K_s respectively. If the helical parameters are completely independent then we should not observe any K_s kinks in our simulations. We also consider the lifetime of a kink, denoted λ, as the longest number of consecutive snapshots for which a kink persists.

3. Results

B-Factor Analysis

B-factors for the histone core are highly consistent from system to system, data not shown. This suggests that the dynamics of the protein core is not significantly affected by the sequence of DNA. In contrast, the unstructured histone tails exhibit considerable variations between simulations and in molecular graphics images appear to be surprisingly mobile. On the time scale of our simulations the histone tails are expected to exhibit an unstructured conformation and stochastic motion that by its' very nature leads to overall asymmetries in the nucleosome. On longer time scales it is expected that the conformational space swept out by each type of histone tail would be similar. However, at any instant of time the conformation of two matching tails in a nucleosome array or in an ensemble of nucleosomes will not be equivalent, producing an inherent asymmetry in the nucleosome.

Energy Analysis

In this section we consider the histone-DNA interactions. For this analysis we consider the interaction between DNA and the histone core. In this manner we avoid the sampling problems associated with the unstructured tail regions, described above. We first consider the total interaction energy as a function of

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

time and then the distribution of histone-base pair interactions as a function of position in the nucleosome.

Histone-DNA Interaction Energy vs. Time

The data presented in Figure 2 represents the total Coulomb and total van der Waals interaction energies versus time for 160 equally spaced snapshots from each simulation. The values for each simulation have been smoothed by a running average with a window size of 10 frames. As described in Methods, the Coulomb energy is calculated with a dielectric constant of 80 because the solvent, present in the original simulation, was removed for analysis. This provided an *ad hoc* method of scaling the Coulomb interactions to represent the solvent screening there were explicitly represented during the simulations. A more sophisticated treatment using a Generalized Born model to analyze the snapshots did not qualitatively change the findings reported here. Our emphasis here is not on the magnitude of the energies or sequence details but rather on the time evolutions. The van der Waals energies converge rapidly for all systems while the Coulomb energies may require on the order of 4 ns to converge depending on the system. This data is presented instead of RMSD values to access convergence since it provides more insight. The observation of these two time scales is consistent with the fact that Coulomb interactions are long range, global interactions, while van der Waals interactions are short range, local interactions. The short ranger interactions are expected to equilibrate faster. Time variations present in Figure 3 span the entire 16 ns and may indicate that the systems have not fully equilibrated. An alternative interpretation is that there are long time scale structural variations in the nucleosome that are not properly sampled on a 16 ns time scale. Such limitations exist in all molecular dynamics simulations.

The raw data indicate that fluctuations for any simulation span the entire range of values observed for all simulations. The relative difference between best and worst total histone-DNA Coulomb interaction energies is approximately 1%. In case of van der Waals interactions the relative difference between best and worst is on the order of 10%. We conclude that sequence differences are difficult to discern with these metrics. Note that *AA* and *TT* contain the same molecule of double stranded DNA, but the rotational phasing of the DNA with respect to the histone-core differs by 180°. They are completely out of phase. Corresponding locations in *AA* and *TT* contain the same base pairs, but their orientations are flipped. The same is true for the other two homopolymers *CC* and *GG* and for the pairs of systems, e.g. *AC* and *TG*, *AG* and *TC*, etc... It may seem surprising that rotational changes of 180° do not have a stronger effect. However, histone contacts with the highly charged and conserved DNA backbone will be largely unaffected by a rotation of 180°.

We conclude that rotational changes of 180° do no strongly affect the histone-DNA interaction energies but that the relative contribution of van der Waals interactions and electrostatic interactions to the total histone-DNA interaction varies with sequence. The latter suggests that the effects of ion

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

concentration and species will vary non-uniformly with sequence. Ionic strength effects should therefore have a non-trivial relation to DNA sequence.

Histone-Base Pair Interaction Energy vs. Position

In this section we consider the interactions between individual base pairs and the histone core. For comparison, if the total interaction energies presented above were uniformly distributed then the energy per base pair would range from -13.6 to -13.8 kcal/mol · bp for Coulomb interactions and from -5.2 to -5.8 kcal/mol · bp for van der Waals interactions, depending upon sequence. In Figure 3 the Coulomb and van der Waals interactions energies between the histone core and individual base pairs are plotted as a function of base pair position in the nucleosome. It is immediately obvious that the distribution of interaction energies tends to be symmetric with respect to the dyad and that variations in the interaction between histone and DNA as a function of position far exceed variations attributable to DNA sequence identity.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

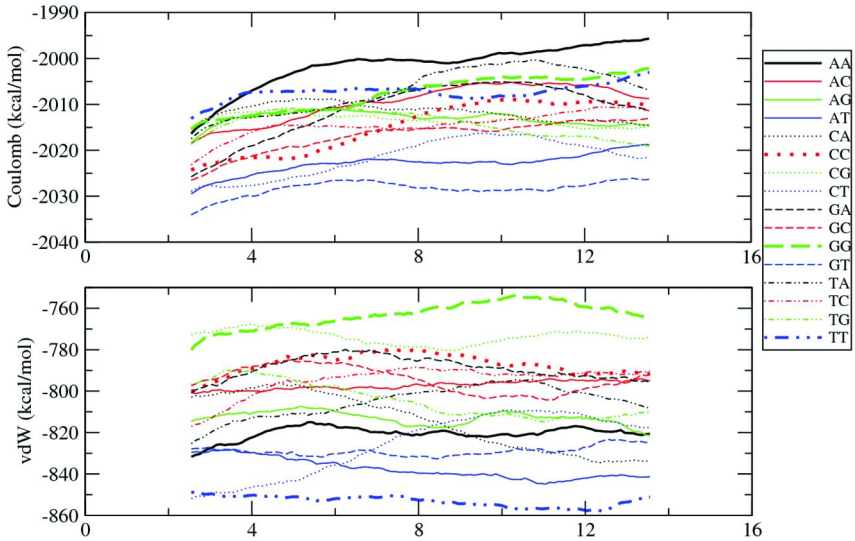


Figure 2. Histone-DNA Interaction Energy versus Time The Coulomb (top) and van der Walls (bottom) interaction energies are plotted as function of time (ns) for the duration of each simulation. Color scheme and line style are as indicated in legend. The data have been smoothed with a moving average that spans 5 ns of data. The Coulomb interactions require more time to equilibrate than the van der Waals interactions. Ranking sequences as most to least favorable by electrostatics and by van der Waals provides different results. The relative difference between most and least favorable is 1% for electrostatics and 10% for van der Waals interactions. (see color insert)

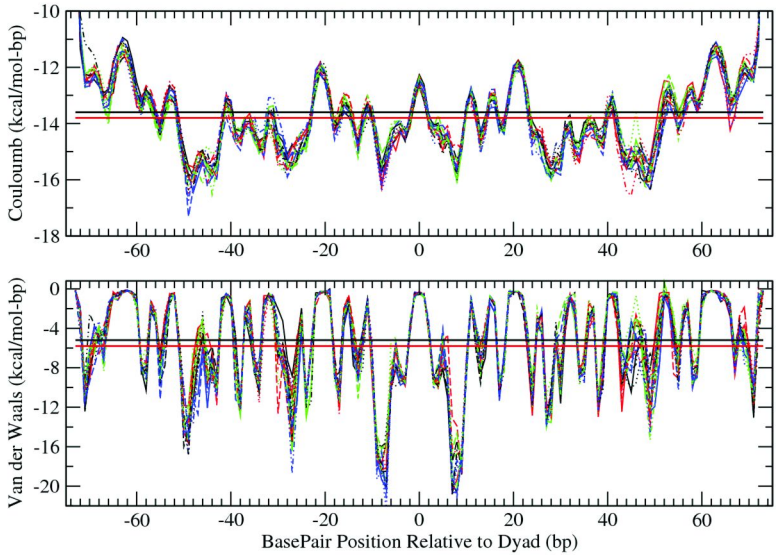


Figure 3. Histone-Base Pair Interaction Energies versus Position Coulomb (top) and van der Waals (bottom) interaction energies are plotted as a function of position relative to the dyad for each of the 16 systems investigated. The color scheme is the same as Figure2. Energies are extracted from our MD simulations as described in Methods. Two horizontal lines indicate the range of mean values obtained for the 16 systems. Overall symmetry with respect to the dyad is clear. Variation as a function of position is much greater than variation as a function of sequence. At this resolution, the individual lines are indistinguishable at most locations. (see color insert)

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

We first consider the overall symmetry. The energy landscapes tend to be symmetry with respect to the dyad, even though each of the individual sequences considered in this study is antisymmetric. For example the orientations of the base pairs at the negative positions of system *AA* are equivalent to the orientations of the base pairs corresponding to the positive positions of *TT*. This is true of each "complementary" sequence, e.g. *CA* and *TG*, or *GT* and *AC*, etc… This result agrees with the invariance of interactions due to rotations of 180°.

As a function of position, the Coulomb interactions varied by more than 5 kcal/mol · bp but less than 1 kcal/mol · bp as a function of sequence. The van der Waals interactions varied by approximately 20 kcal/mol · bp as a function of position but less than 2 kcal/mol · bp as a function of sequence. Thus position rather than sequence identity most strongly influences the interaction between a given base pair and the histone octamer. This result is not surprising if one considers that orientation of a base pair with respect to the histone core changes by approximately 34° between adjacent base pairs. However if the rotational phasing of a given base pair was the sole determinant of variations in the interaction

energies then the energy landscape should tend to vary sinusoidally. It does not. Instead there are pronounced extrema in both the Coulomb interactions and the van der Waals interactions. Of interest are local minimum that occur in both the Coulomb and van der Waals interactions near base pair positions ±10, ±26, and ±49. These are regions of particularly strong histone-DNA interaction. The interactions near ±49 represent the first set of strong histone-DNA interactions that would be encountered while traversing along DNA towards the interior of a nucleosome.

The interaction energy landscape exhibits a similarity to energy landscapes obtained from a recent atomic force study in which DNA was forcefully removed from nucleosomes (*30*), to the analysis of histone-DNA interaction energies in the nucleosome (*31*), and to our earlier observation of local extrema in the histone-DNA interaction potential resulting from the interaction of a simple elastic rod interacting with a cylindrical core particle (*32*).

The positioning ability of a given sequence of DNA is related to the change in total energy that occurs when 147 adjacent base pairs are dragged across such an energy landscape. Some base pairs will achieve more favorable interactions while others will experience less favorable interactions. The extrema represent locations where an individual base pair will gain or lose significant amounts of histone-DNA interaction energy as an entire sequence is repositioned. The missing component of in the energy landscape presented here is the self-energy of DNA which must also change as each base adopts a new conformation concomitant with its new position in the nucleosome. Elastic rod models, as in (*9*), attempt to capture this energy landscape. The total energy of the system is a combination of the histone-DNA interaction energies and the DNA self energy. Both landscapes appear to exhibit rugged variations. If collectively moving all 147 base pairs in either direction increases the total energy then the sequence is positioned. If motion in one direction yields a decrease in the total energy then the nucleosome should favor relocating in the indicated direction. If motion in either direction is energetically favorable or flat the nucleosome position will be unstable or fuzzy.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Structure Analysis

In this section we analyze the DNA helical parameter data as a function of position. Helical parameter values are obtained for 16,000 snapshots for each of the 16 simulations.

Statistical Measures

Mean values of the DNA helical parameters as a function of position indicate that Roll is highly conserved in all systems. Roll exhibits a regular variation as reported in (*11*) and (*12*). The other helical parameters exhibit much greater variability between the different simulations. The conserved Roll-Slide-Twist relations can only be obtained by invoking the Fourier filtering procedure presented in (*12*). Consideration of the unfiltered data indicate several locations

in which the conformation of DNA, as determined by helical parameters Slide and Twist, is restricted, data not shown. At these locations the specific Slide and Twist values exhibit limited variation between sequences. This suggests there are discrete locations where a specific conformation of DNA is required, regardless of sequence. By comparison other locations allow a relative conformational freedom as a function of sequence. Thus any sequence of DNA can achieve the requisite global superhelix conformation while still optimizing the local conformational preferences dictated by sequence specific properties of DNA. This agrees with our earlier observation that instead of being tightly affixed to the histone core, there is considerable slack in DNA, i.e. local conformational freedom (*18*).

Comparison of values of the standard deviation of DNA helical parameters obtained from our simulations to those obtained from an extensive molecular dynamics study of DNA free in solution (*27*) indicate that nucleosomal DNA tends to be stiffer than free DNA. This is true for all of the helical parameters except Tilt. The sequence-averaged values of standard deviations obtained in our simulations were as follows: Shift, 0.48 to 0.64 Å; Slide, 0.46 to 0.58 Å; Rise, 0.30 to 0.33 Å; Tilt, 4.3 to 6.5°; Roll, 4.9 to 5.9°; and Twist, 4.4 to 6.8 °. For simulations of DNA free in solution the observed standard deviations are Shift, 0.76 Å; Slide, 0.68 Å; Rise, 0.37 Å; Tilt, 4.6° Roll, 7.2°, and Twist, 7.3° (*27*).

In Figure 4 we consider the results of the normality test described in Methods. The overall trend is that the helical parameters are normally distributed with a 99% confidence interval. The helical parameter Rise exhibits the fewest number of deviations from normality. Tilt and Roll exhibit the greatest number of non-normal distributions, indicated by a zero, and the greatest tendency for weaker confidence criteria. There are no obvious patterns that can be related to a structural feature of the nucleosome or a sequence effect. The general trend towards normal distributions provides quantitative support for our expectations of kinking frequencies presented in Methods. Such explicit normality testing has not been reported elsewhere.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Kink Analysis

Finally, we analyze kinks in nucleosomal DNA using the definition presented in Methods. For simplicity of interpretation and presentation we consider the time evolution, frequency and lifetimes of kinks only for the homopolymer sequences *AA*, *CC*, *GG*, and *TT*. We then provide a complete map of kinking behavior for all types of base pair steps at all 146 possible positions in the nucleosome.

We first consider kinks as defined by Richmond and Davey (*11*), K_{rd}, see Figure 5. Kinking at each position is shown as a function of time. Figure 5 also includes the percentage of time that kinks appear at a given position and the lifetime of the kink. Kinks appeared at positions other than those identified in the x-ray crystal structure, irrespective of the step type. The kinks are dynamic in nature, appear intermittently and then disappear. In this regard they are self-healing. For example at position -27 of *TT*, the kink appeared in the first three nanoseconds and then disappeared. At position -16 of *TT* a kink appeared after five nanoseconds

but healed after eleven nanoseconds. At other positions, e.g. position 46 of *CC*, kinks appeared and disappeared intermittently for the duration of simulation. The maximum number of times a kink formed in one of the homopolymers is at position -56 of *AA*. This particular kink appeared for 30 % of the total simulation time. At all other locations kinks appeared less than 4 % of the time including the positions identified by Richmond and Davey. At some positions a kink appears to persist even though the percentage of snapshots is relatively low. An example is position -46 of *CC* which appears to persist for the entire simulation even though its' lifetime was 3 ps and only about 1% of the snapshots registered as kinked. In contrast, the kink at position -47 of *AA* persists for approximately 3 ns but has a lifetime of 12 ps with 3% of the snapshots registering as kinked.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

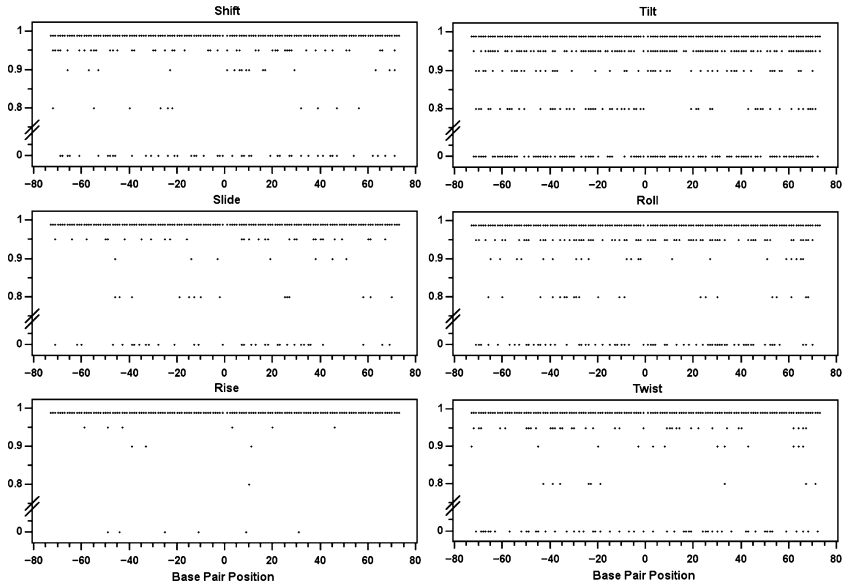


Figure 4. Normality of DNA Helical Parameter Data versus Position The Jarque-Bera test is used to assess the normality of the distributions obtained from the 146 x 6 sets of 16,000 helical parameter data. Normality is determined with different levels of confidence (0.80, 0.90, 0.95, 0.99% confidence intervals) for each distribution. A value of zero is assigned for any distribution that failed all four normality tests. Rise is the most normally distributed regardless of sequence or position. Tilt and Roll exhibit the largest number of non-normal distributions and the lowest confidence scores. There are no simple patterns relating to sequence or position.

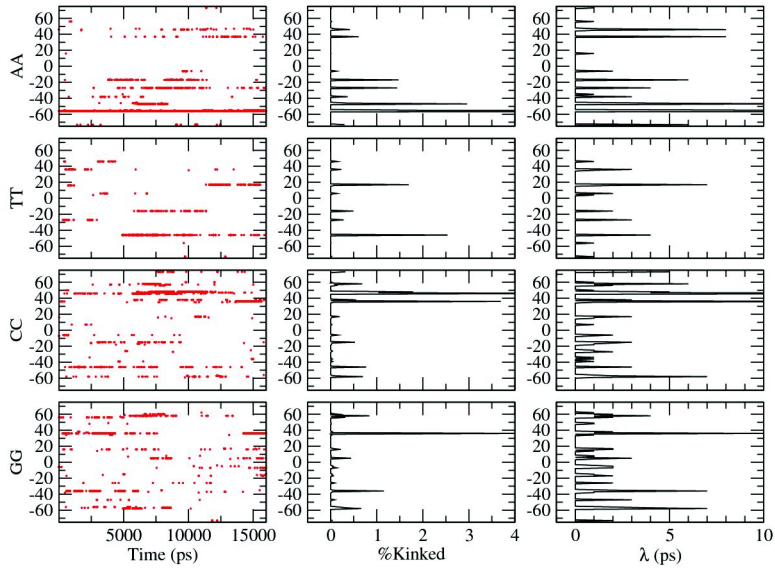


Figure 5. K_{rd} versus time, percentage of appearance and lifetime for homopolymers Left: Kinks identified by the Richmond and Davey criterion versus position and time for the homopolymer systems A·A, T·T, C·C and G·G. Middle: The percentage of snapshots registered as kinked versus position. Right: The maximum consecutive kink lifetime in picoseconds versus position.

Figure 6 presents results based on the weak kink criterion, K_w. Kinks identified by this criterion are expected to be common because weak kinks require all three helical parameters to simultaneously deviate by only one standard deviation from the sequence specific conformation of free DNA. By this criterion, if the three helical parameters are totally dependent, then we expect kinking 32% of the time even for DNA free in solution. If the helical parameters are totally independent then we expect kinking about 3.3% of the time. At some positions kinking far exceeds 32%. The maximum percentage kinking is observed at position -56 of *AA*. This kink appeared in more than 95% of the 16,000 snapshots. This is atypical. Even though the plot of K_w versus time suggests continuous kinking at all locations, the kinks are actually dynamic in nature with only a few weak kinks having a lifetime exceeding 250 ps. More typical is a weak kink lifetime of 10 to 25 ps. Surprisingly, there are positions that exhibited weak kinks less than 3.3% of the time. At these positions even fluctuations in the base pair step conformation are within one standard deviation of the sequence specific conformation of free DNA more often than is expected for free DNA. The histones, rather than deforming DNA, actually support the conformation of DNA free in solution. The absence of weak kinks is more important than their presence.

Figure 7 presents results based on the strong kink criterion, K_s. Kinks identified by this criterion are expected to be rare because strong kinks require all

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

three helical parameters to simultaneously deviate by more than three standard deviations from the sequence specific conformation of free DNA. If the three helical parameters are totally independent of one another, then the kink should never appear in our simulations, as marked by the line at 0%. If the helical parameters are strongly coupled then kinks should appear about 0.27% of the time. Any location that exceeds the 0.27% threshold is decidedly kinked. We observe a few such locations. The lifetimes for K_s kinks are typically less than 1 ps. As with K_{rd} kinks, the K_s kinks appeared intermittently and were self-healing. Noted exceptions are kinks which appeared 30% of the time at position -56 of *AA* and at position -4 of *TT*. The K_s criterion is the most intuitive definition of a kink. It is the most similar to K_{rd}, but it is not biased by position.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

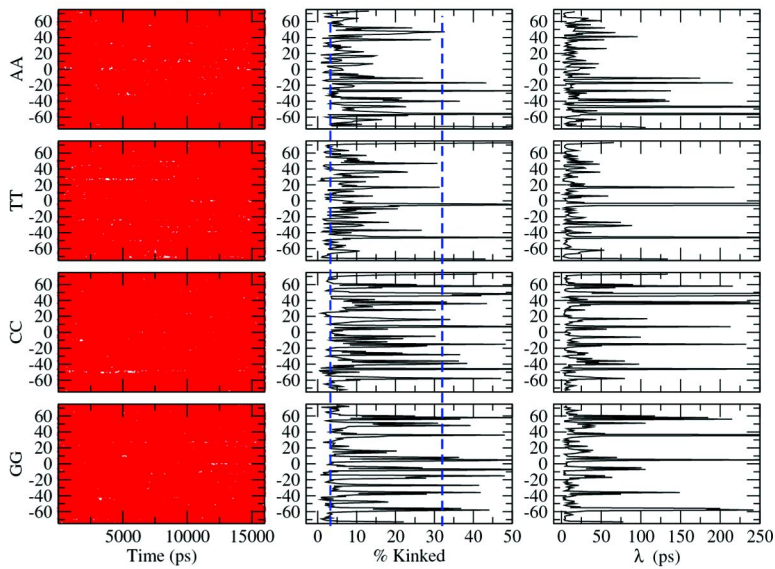


Figure 6. K_w versus time, percentage of appearance and lifetime for homopolymers Left: Kinks identified by weak kink criterion, K_w, versus position and time for the homopolymer systems A·A, T·T, C·C and G·G. Middle: The percentage of snapshots registered as kinked versus position. Right: The maximum consecutive kink lifetime in picoseconds versus position. The weak kink criteria requires Roll, Slide and Twist to simultaneously deviate by more than one standard deviation from the sequence specific values associated with free DNA. Such kinks should appear in a simulation of free DNA in 3.2% to 32% of the snapshots, as indicated by the vertical dotted lines in the Middle plot. The lack of weak kinks is surprising and suggests the conformation in the nucleosome closely resembles free DNA.

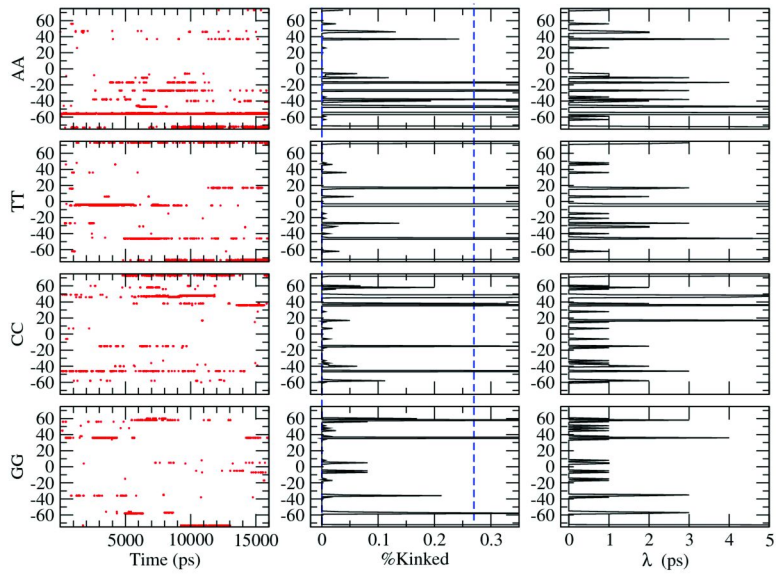


Figure 7. K_s versus time, percentage of appearance and lifetime for homopolymers Left: Kinks identified by strong kink criterion, K_s, versus position and time for the homopolymer systems A·A, T·T, C·C and G·G. Middle: The percentage of snapshots registered as kinked versus position. Right: The maximum consecutive kink lifetime in picoseconds versus position. The strong kink criteria requires Roll, Slide and Twist to simultaneously deviate by more than three standard deviations from the sequence specific values associated with free DNA. Such kinks should appear in 2.0E-6% to 0.27% of the snapshots of a simulation of free DNA, as indicated by vertical dotted lines in the Middle plot. Only in AA does a strong kink persist for the entire simulation. However the lifetime of any the individual kink event was only approx 16 ps, i.e. the kinks persisted but were not present continuously.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

3.1. Complete Maps of Kinking

Having evaluated the characteristics of the various criteria for kinking, we present a map of kinking for each of the sixteen step types at all 146 possible positions in Figure 8. Here we utilize expectations for independent helical parameter statistics to weight the data. Thus for K_s and K_{rd} any observation of a kink is scored as kinked, but for K_w more than 516 kinks must be observed during the simulation (1 out of 31 frames) for the location to be considered kinked. Regions where the minor groove faces towards or away from the histone core are marked by dark and light boxes, respectively. Figure 8 also indicates the percentage of the 146 positions that exhibited kinks for each base pair step type. Except for the homopolymers, data for the odd and even positions are obtained from two different simulations. For example A·T data is obtained from even

numbered positions of simulation *AT* and from the odd positions of simulation *TA*. We justify assembling data from two different simulations by pointing out that the neighboring base pairs and hence the sequence context is the same in both simulations.

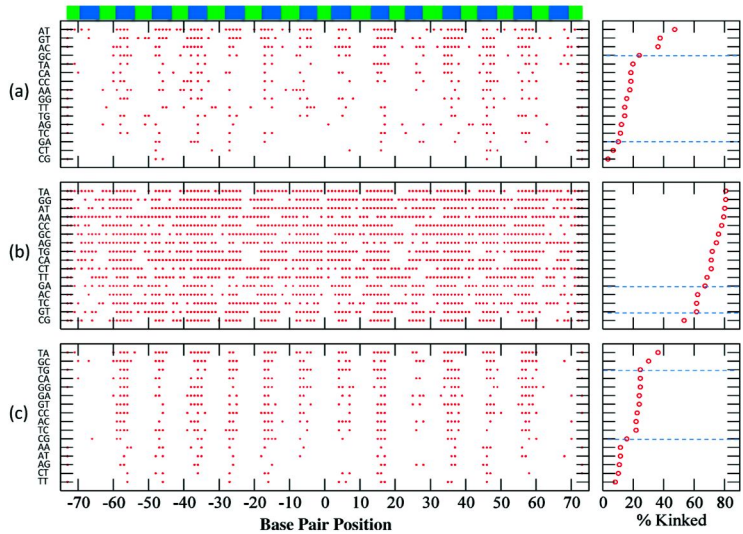


Figure 8. Location of kinks and percentage kinking by (a) K_s (b) K_w (c) K_{rd} for all systems The location of kinks for different base pair steps as obtained from (a) K_s (b) K_w (c) K_{rd} criteria are shown in the left column of plots. The dark and light stripe indicates locations where the minor groove faces in and out, respectively. Indicated in the right column of plot is the percentage of the 146 positions that exhibited kinking for each base pair step type. The horizontal lines are used to identify different categories: kinkable, normal, not kinkable.

Kinking by the strong kink criterion, K_s, is presented in Figure 8(a). The helical parameters are considered to be independent. Strong kinks tend to localize along positions where the minor groove faces towards the histone core, as with the K_{rd} definition, but there also exist a number of exceptions. Based on the percentage of positions kinked, the step types can be separated into three categories, as indicated by the dashed line. Step A·T is decidedly the most kinkable, nearly 50% of all positions in a nucleosome can achieve this type of kink. Steps G·T and A·C are also kinkable. Steps C·T and C·G are the least kinkable. Less than 10% of all positions achieve this type of kink. All other step types exhibit K_s kinks at 10-25% of the positions.

Kinking by the weak kink criterion, K_w, is presented in Figure 8(b). The helical parameters are considered to be independent. Weak kinks, as expected, appear almost everywhere. Surprisingly for each type of step there are locations that do not kink. At these locations the histones support the sequence specific conformation observed for DNA free in solution. Again the step types are divided

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

into three categories based on the percent of locations exhibiting kinks. By this definition, all but the C·G steps exhibit kinking at 60-80% of the positions. C·G steps are the least kinkable, exhibiting kinks at approximately 50% of all positions. The alternative interpretation is that at nearly 50% of all positions C·G steps achieve a conformation that is nearly identical to their conformation in solution. At these locations C·G steps experience no conformational stress from the histones.

Kinking by the Richmond and Davey criterion, K_{rd}, is presented in Figure 8(c). Kinks identified by this criterion require a negative Roll angle and therefore tend to be located where the minor groove faces the histone core. The percentage of positions kinked by this definition most clearly separates the base pair step types into three categories. T·A and G·C are the most kinkable, exhibiting kinks at more than 30% of the positions, whereas, T·T, C·T, A·G, A·T, and A·A are the least kinkable.

Finally, we note that the rank orderings from different criteria do not match. Thus observation of a strong kink does not necessarily imply a weak kink at the same location. The location of kinks by the strong criterion, K_s, is similar to that of Richmond and Davey, but K_s allows identification of additional kinks. We propose K_s as an intuitive and unbiased criterion for defining kinks based on geometry. We have found that K_w is useful for identifying locations that exhibit a free DNA like conformation.

4. Discussion

We have demonstrated that any base pair can be positioned at any location in the nucleosome and remain structurally well-characterized during 16 ns long molecular dynamics simulations. This result is in agreement with the biologic reality that histones, even though favoring some sequences, must be capable of binding to any sequence of DNA. In this regard the simulations and analysis presented here provide the first complete, albeit approximate map of kinking for the approximately $(4)^{147}$ possible nucleosomal DNA sequences.

Our analysis of histone-DNA interaction energies indicates that the interaction between histones and DNA varies more by location than by DNA sequence. The variations are not sinusoidal. We interpret this as an indication that the stresses imposed on DNA by the histone octamer vary considerably from position to position within a nucleosome, regardless of sequence. From a structural engineering point of view, the loads (forces and moments) imposed on the DNA are decidedly non-uniform And since the material properties of DNA vary with sequence the response of a given sequence to such loading is non-trivial. This is further complicated by the fact that as a function of position within the nucleosome the relative contribution from van der Waals versus electrostatics interactions differs significantly. Our results agree with the assessment of electrostatic interactions in (*31*). Thus the effects of ion species and ion concentration, which will alter the electrostatic but not van der Waals interactions, will have a non-uniform effect on histone-DNA interactions. We expect some locations within a nucleosome to be more strongly affected by ionic variations than others.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Our analysis of DNA helical parameters indicates that DNA in the nucleosome tends to be less flexible, as determined by measures of standard deviations, than DNA free in solution. We interpret this as an indication that nucleosome formation results in an effective stiffening of DNA. We first reported a similar observation in 2005 (*18*). The current results extend our initial study of one sequence to suggest this result is generally true. We also performed normality test and found that the six inter base pair helical parameters extracted from our simulations tend to exhibit normal distributions. Explicit tests for normality are not typically included with reports of DNA helical parameters obtained from molecular dynamics simulations. A noted exception is (*27*) where it was demonstrated that for free DNA not all helical parameters exhibit normal behavior (*27*). A normal distribution is a preliminary requirement for the determination of effective force-constants from helical parameter fluctuations as employed for example in (*33*). Normally distributed data is also required to rigorously apply the frequency thresholds used in our characterization of kinking. The absence of a normal distribution does not invalidate our identification of kinks as discussed below.

We tested three definitions of kinking: the definition posed by Richmond and Davey (*11*), K_{rd}; a weak kink criterion, K_w; and a strong kink criterion, K_s. Definition K_{rd} is biased such that kinks appear at sites where the minor groove faces toward the histone core, i.e. a negative Roll is required by K_{rd}. K_w and K_s do not have such a bias. By any of the three definitions employed here, no kink persisted for an entire simulation. The longest lifetime of any kink was approximately 6 ns (6000 consecutive snapshots). This kink was atypical. More typical is the appearance and disappearance of strong kinks, K_s on a 1 to 10 ps time scale and weak kinks, K_w on a 10 to 25 ps time scale. We found that kinks by the strong criterion, K_s, may persist at certain locations, appearing and disappearing. This is the basis of the statement that kinks in nucleosomal DNA are self-healing on the nanosecond time scale.

Ignoring nearest neighbor effects we have categorized the base pair steps as more kinkable, less kinkable or intermediate by each kink criteria. Interestingly categorization by K_s, K_w and K_{rd} leads to different results. By the weak kink criterion T·A, G·G, A·T, and A·A are categorized as the most kinkable and C·G, G·T, T·C, A·C as least kinkable. By the strong kink criterion A·T, G·T and A·C are most kinkable and C·G, C·T are the least. In case of A·T, the tendency to form weak kinks also implies a tendency to form a strong kink. For C·G the tendency to not form weak kinks also implies a lack of strong kinks. However, in case of A·C and G·T, strong kinks do not imply a tendency to form weak kinks. K_w and K_s are not simply correlated in this case. The implication is that kinking behavior itself differs with either location or sequence. In (*19*) we demonstrated that the dynamics of individual base pairs within the nucleosome differs as a function of position. Our present results provide further evidence and provide additional understanding for this observation.

There have been numerous efforts to determine sequence rules governing nucleosome positioning sequences. However, there is no consensus. In one view positioning correlates with A·T steps located at minor groove out and C·G at minor groove in positions (*7*). This motif will kink the least according to our K_s

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

criterion. By our K_w criterion we find this also allows the C·G steps to exhibit nearly the same conformation in the nucleosome as they do in solution. An apparently contrary view maintains that positioning correlates with C·G steps at minor groove out positions and A·T steps at minor groove in positions (*8*). This motif should exhibit maximal kinking according to our K_s criterion. Yet another positioning motif locates A·A, T·T or T·A at the minor groove in and G·C at the minor groove out (*6*). According to our K_s criterion G·C tends to kink more than A·A, T·T or T·A. Thus, this motif places a more kinkable step at minor groove out and less kinkable step at minor groove in locations. This motif will also tend to exhibit fewer kinks at fewer positions.

If as discussed by Crick and Klug, kinking corresponds with a transition from second order to first order variation of the energy as a function of deformation, then kinking can lower the energy barrier associated with superhelix conformation. The energy savings only occur at locations where the transition threshold is exceeded. Nonetheless kinks limit the total energy penalty. Sequences which do not exhibit kinks, or which exhibit fewer kinks, have not reached the transition threshold and should be considered less stressed.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

5. Conclusion

Simplified positioning motifs only aid in our understanding of sequences which exhibit the indicated motif. They provide no insight into sequences which are exceptions to the rule and fall far short of being generally applicable to all possible 147 base pair oligomers or even the subset of sequences which actually occur in natural genomes. In contrast our simulations demonstrate that molecular dynamics simulations can be successfully utilized to investigate nucleosomes in which an arbitrary sequence of DNA has been thread onto the histone core. Combined with our high throughput strategy for conducting and analyzing such simulations, this opens the door for on-demand molecular dynamics simulation studies of the nucleosome in various conformational states and with various chemical modifications. As a first proof of concept we have initiated a genome wide study of nucleosome positioning in *Saccharomyces cerevisiae*. For this study we choose the single most highly occupied and least variable nucleosome positioning sequence for each of the sixteen chromosomes. For each position we consider a sequence window that includes the experimentally determined position as well as 10 positions immediately upstream and 10 positions immediately downstream. Initial results confirm the findings reported here. Namely that any sequence can be thread onto the histone core and that kinks do not simply occur at fixed locations regardless of sequence nor do kinks simply arise in certain sequences of DNA regardless of location. Instead kinking is a complex interaction between the non-uniform stresses imposed on the DNA by the histones and the sequence specific material properties of DNA.

As our current simulations demonstrate, kinks in nucleosomal DNA have at least two time scales. A fast picoseconds time scale that characterizes a given base pair as either kinked or not kinked and a longer nanoseconds time scale that characterizes the tendency of a base pair to kink at a given location. This finding

has important implications for our understanding of both the initial binding and subsequent association of proteins or ligands to nucleosomal DNA.

Acknowledgments

The authors acknowledge support from NIH grant R01-GM76356 "Molecular Dynamics Study of Nucleosome Stability and Receptor Binding", NSF award number EPS-1003897 for LA-SiGMA, the Louisiana Optical Network Initiative (LONI) for computational resources, and NSF award 0619843 for establishment of PetaShare resources.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

References

1. Widom, J. *Q. Rev. Biophys.* **2001**, *34*, 269–324.
2. Pugh, B. F. *Nat. Struct. Mol. Biol.* **2010**, *17*, 923.
3. Kaplan, N.; Moore, I.; Fondufe-Mittendorf, Y.; Gossett, A. J.; Tillo, D.; Field, Y.; Hughes, T. R.; Lieb, J. D.; Widom, J.; Segal, E. *Nat. Struct. Mol. Biol.* **2010**, *17*, 918–920; author reply 920–922.
4. Kaplan, N.; Hughes, T. R.; Lieb, J. D.; Widom, J.; Segal, E. *Genome Biol.* **2010**, *11*.
5. Zhang, Y.; Moqtaderi, Z.; Rattner, B. P.; Euskirchen, G.; Snyder, M.; Kadonaga, J. T.; Liu, X. S.; Struhl, K. *Nat. Struct. Mol. Biol.* **2010**, *17*, 920–923.
6. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thstrm, A.; Field, Y.; Moore, I. K.; Wang, J.-P. Z.; Widom, J. *Nature* **2006**, *442*, 772–778.
7. Trifonov, E. *Phys. Life Rev.* **2011**, *8*, 39–50.
8. Travers, A. *Phys. Life Rev.* **2011**, *8*, 53–55.
9. Sereda, Y.; Bishop, T. C. *J. Biomol. Struct. Dyn.* **2010**, *27*, 867–887.
10. Crick, F. H.; Klug, A. *Nature* **1975**, *255*, 530–533.
11. Richmond, T. J.; Davey, C. A. *Nature* **2003**, *423*, 145–150.
12. Bishop, T. C. *Biophys. J.* **2008**, *95*, 1007–1017.
13. Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
14. Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
15. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
16. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
17. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kal, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
18. Bishop, T. C. *J. Biomol. Struct. Dyn.* **2005**, *22*, 673–686.
19. Ponomarev, S. Y.; Putkaradze, V.; Bishop, T. C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10633–10643.

20. Gower, M.; Cohen, J.; Phillips, J.; Kufrin, R.; Schulten, K. Managing Biomolecular Simulations in a Grid Environment with NAMD-G. Proceedings of the 2006 TeraGrid Conference, 2006.
21. Wang, X.; Kosar, T. Design and Implementation of Metadata System in PetaShare. 21st International Conference on Scientific and Statistical Database Management, 2009.
22. Mukherjee, R.; Fujioka, H.; Thota, A.; Bishop, T. C.; Jha, S. 2011, submitted.
23. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
24. Dickerson, R. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627–634.
25. Dickerson, R. *Nucleic Acids Res.* **1989**, *17*, 1797–1803.
26. Jarque, C. M.; Bera, A. K. *Int. Stat. Rev.* **1987**, *55*, 163–172.
27. Lavery, R.; et al. *Nucleic Acids Res.* **2009**.
28. Tolstorukov, M. Y.; Colasanti, A. V.; McCandlish, D. M.; Olson, W. K.; Zhurkin, V. B. *J. Mol. Biol.* **2007**, *371*, 725–738.
29. Marathe, A.; Bansal, M. *BMC Struct. Biol.* 2011, *11:1*, year.
30. Hall, M. A.; Shundrovsky, A.; Bai, L.; Fulbright, R. M.; Lis, J. T.; Wang, M. D. *Nat. Struct. Mol. Biol.* **2009**, *16*, 124–129.
31. West, S. M.; Rohs, R.; Mann, R. S.; Honig, B. *J. Biomol. Struct. Dyn.* **2010**, *27*, 861–866.
32. Bishop, T. C.; Hearst, J. E. *J Phys. Chem. B* **1998**, *102*, 6433–6439.
33. Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* **2003**, *85*, 2872–2883.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch005

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Chapter 6

The Size of Internal Loops Influences the Unfolding Thermodynamics of DNA Hairpins

Iztok Prislan, Hui-Ting Lee, Cynthia Lee, and Luis A. Marky*

Department of Pharmaceutical Sciences, College of Pharmacy, University of Nebraska Medical Center, 986025 Nebraska Medical Center, Omaha, NE 68198-6025

* **Tel.: (402) 559-4628. Fax: (402) 559-9543. E-mail: lmarky@unmc.edu.**

The main focus is to further our understanding of the physico-chemical properties of unusual DNA structures. We designed a set of stem-loop DNA molecules to mimic a common motif of mRNA i.e., a hairpin with internal loops. We used a combination of UV spectroscopy and differential scanning calorimetry (DSC) techniques to determine the unfolding thermodynamics of a set of hairpins with sequence: d(GCGCT_nGTAACT_5GTTACT_nGCGC), where "T_n" corresponds to internal loops with n = 1, 3 or 5 and "T_5" is an end loop of 5 thymines. UV melts of each hairpin show monophasic transitions with T_Ms that are independent of strand concentration i.e., all hairpins formed intramolecularly. DSC profiles indicate hairpin unfolding results from the typical compensation of an unfavorable enthalpy with a favorable entropy. The increase in the size of the internal loop yielded: a) lower T_Ms and similar enthalpy contributions; b) lower heat capacity values that correlated with the lower releases of structural water; and c) higher ion releases. Therefore, hairpin loops with large internal loops are less stable and amenable for their targeting with complementary oligonucleotides.

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Introduction

The particular structure of a nucleic acid is controlled by its base sequence in a precise and potentially predictable way. However, knowledge of the structure of a particular duplex alone cannot provide an understanding of the forces responsible for maintaining the distinct structures of nucleic acids. To understand how nucleic acids carry out their biological roles, it is essential to have a complete physical description of how their folding takes place. The overall physical properties of a nucleic acid molecule not only depend on its chemical architecture, but on contributions from base pairing, base stacking, ion and water binding.

It has been demonstrated that palindromic sequences embedded in plasmids developed cruciform structures in response to a topological stress (*1*, *2*), which are cleaved by specific endonucleases (*3*, *4*). These findings suggest that the presence of hairpin loops in DNA may play an important role in biological processes. On the other hand, the presence of loops in the secondary structures of RNA is very common. Currently, there is considerable interest in both the structure and overall physical properties for the folding (and unfolding) of nucleic acid hairpin loops. Our current understanding of the structures and stability of both DNA and RNA has been enhanced by thermodynamic investigations of the helix-coil transitions of model oligonucleotide compounds, of known sequence (*5–8*). Our laboratory is primarily interested in understanding the unfolding of single-stranded DNA oligomers that may adopt a variety of intramolecular secondary structures (*9–11*). Their intramolecular unfolding takes place with a lower entropy penalty yielding transition temperatures higher than their bimolecular counterparts, allowing investigations of the physical properties of their 100% helical conformations over a wider temperature range (*12*).

There have been several reports of using nucleic acid oligonucleotides (ODNs) as drugs since ODNs present an exquisite selectivity and are able to discriminate targets that differ by a single base (*13*). There are three main approaches for the use of ODNs as modulators of gene expression: the antisense, antigene, and small interfering RNA (*14–16*). For instance, in the antigene strategy, an ODN binds to the major groove of a DNA duplex, forming a triple helix that inhibits transcription, by competing with the binding of proteins that activate the transcriptional machinery. In the antisense strategy, an ODN binds to messenger RNA, forming a DNA/RNA hybrid duplex that inhibits translation, by sterically blocking the correct assembly of the translation machinery or by inducing an RNase H mediated cleavage of their mRNA target. There are advantages and disadvantages in these two strategies. The main disadvantage is that the ODN needs to cross lipoidal membranes, which hydrophilic ODN duplexes are not able to do. This disadvantage can be circumvented by using stable intramolecular DNA structures containing loops. The presence of the unpaired nucleobases in their loops offsets the hydrophobic-hydrophilic balance of the helical ODN towards slightly more hydrophobic (*17*), allowing them to cross the cellular membranes and/or to interact better with polycationic micelles, which can be used as cellular delivery vectors. In order to target RNA molecules, we must first understand the structure and thermodynamic properties of RNA

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

secondary motifs, since successful control of gene expression depends on the effective binding of a DNA sequence to its target with tight affinity and specificity. The prediction of the secondary structures of RNA molecules has been solved (*18*, *19*); however, current investigations are geared to predict the overall tertiary structure of RNA from its sequence (*20*). Our laboratory has started a program to model the potential secondary structures of RNA using DNA oligonucleotides in the absence of magnesium ions. We have investigated a variety of single-stranded DNA complexes, including hairpin loops, four-way junctions and pseudoknots (*21–25*).

In this work, we present a thermodynamic description of the unfolding of DNA intramolecular stem-loop motifs. Specifically, we used a combination of spectroscopic and calorimetric techniques to investigate a set of hairpins containing internal loops of varied size in their stem, as a model to mimic a common motif of mRNA. The favorable folding of each hairpin results from favorable enthalpy contributions, formation of base-pair stacks, compensated with unfavorable entropy contributions, uptake of ions and water molecules. The overall results should improve our current understanding of how sequence, loops, ion binding, and hydration control the stability and melting behavior of a nucleic acid molecule.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Materials and Methods

Materials

The oligonucleotides and their designations (Scheme 1): d(GCGCTGTAAC T_5GTTACTGCGC), *IL-2*; d(GCGCT_3GTAACT_5GTTACT_3GCGC), *IL-6*; d(GCGCT_5GTAACT_5GTTACT_5GCGC), *IL-10*; were synthesized by the Core Synthetic Facility of the Eppley Research Institute at UNMC, HPLC purified, and desalted by column chromatography using G-10 Sephadex exclusion chromatography. The concentrations of the oligomer solutions were determined spectroscopically at 260 nm and 80 °C using the following molar extinction coefficients for the single strands, in mM^{-1} cm^{-1}: 223.0 (*IL-2*), 255.6 (*IL-6*), 288.1 (*IL-10*). These values were obtained using procedures reported previously (*21*). Inorganic salts from Sigma were reagent grade, and used without further purification. All measurements were made in buffer solutions consistimg of 10 mM sodium phosphate at pH 7, adjusted to the desired salt concentration with NaCl or to the desired osmolyte concentration with ethylene glycol. All oligonucleotide solutions were prepared by dissolving the dry and desalted ODNs in buffer, the solution was then heated to 90 °C for 5 min and cooled to room temperature slowly.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Scheme 1. The 5′ to 3′ sequences and labels of oligonucleotides.

Temperature-Dependent UV Spectroscopy (UV Melting Curves)

Absorbance versus temperature profiles were measured at 260 nm with a thermoelectrically controlled Aviv model 14DS UV-vis spectrophotometer (Lakewood, NJ). The temperature was scanned from 10 °C to 95 °C at a heating rate of 0.6 °C/min. For clarity, some curves were converted to α-curves using a procedure described earlier (*26*), "α" is described as the fraction of strands in the helical state. Shape analysis of the UV melting curves yielded transition temperatures, T_M, and van't Hoff enthalpies ΔH_{vH}, using standard procedures (*26*). The transition molecularity for the unfolding of a particular complex was obtained from the dependence of the T_M on strand concentration.

Circular Dichroism (CD) Spectroscopy

The conformation of each hairpin was determined by inspection of the CD spectra at low temperatures. These CD spectra were obtained from 320 to 200 nm in 1 nm increments, using an Aviv model 202SF CD spectrometer (Lakewood,

NJ) equipped with a peltier temperature control system. CD melting curves were obtained at 250 nm and 280 nm. Analysis of the CD melting curves yielded T_Ms and ΔH_{vH} using procedures reported previously (*26*).

Differential Scanning Calorimetry (DSC)

The total heat required for the unfolding of each hairpin was measured with a VP-DSC differential scanning calorimeter from Microcal (Northampton, MA). The excess heat capacity of a hairpin solution in the sample cell was measured against a reference cell filled with the buffer solution in the temperature range of 1 °C to 95 °C, using a heating rate of 0.75 °C/min. Buffer against buffer scans were also done under similar conditions and subtracted from the sample scans, then normalized for the number of moles. Analysis of the resulting thermograms yielded T_Ms, and standard thermodynamic profiles, ΔH_{cal}, ΔS_{cal} and $\Delta G^o_{(T)}$. These parameters are measured using the following relationships (*26*): $\Delta H_{cal} = \int \Delta C_p(T) dT$ and $\Delta S_{cal} = \int C_p(T)/T dT$, where $\Delta C_p(T)$ represents the anomalous heat capacity during the unfolding process. The Gibbs free energy at any temperature, $\Delta G^o_{(T)}$, is calculated with the Gibbs equation: $\Delta G^o_{(T)} = \Delta H_{cal} - T\Delta S_{cal}$. Alternatively, $\Delta G^o_{(T)}$ can be calculated using the equation $\Delta G^o_{(T)} = \Delta H_{cal}(1 - T/T_M)$ which is rigorously correct for the unfolding of an intramolecular G-quadruplex. Additional DSC experiments were obtained at several salt concentrations to indirectly obtain the associated heat capacity contributions, if any. These heat capacity effects are determined from the slopes of the lines of the ΔH_{cal} vs. T_M plots. The ΔH_{vH} terms were also obtained from the DSC thermograms using the temperatures at the half-width height of the experimental curve (*26*). The $\Delta H_{vH}/\Delta H_{cal}$ ratio provides information about the nature of transition (*26*, *27*).

Determination of the Differential Binding of Counterions and Water

Additional UV melting curves were obtained as a function of salt and osmolyte concentrations to determine the differential binding of counterions, Δn_{Na+}, and water molecules, Δn_w, respectively. These Δn_{Na+} and Δn_w linking numbers are measured experimentally using the following relationships (*28*):

$$\Delta n_{Na^+} = 1.11\ (\Delta H_{cal}/RT_M^2)[\partial T_M/\partial \ln\ [Na^+]] \quad (1)$$

$$\Delta n_w = (\Delta H_{cal}/RT_M^2)[\partial T_M/\partial \ln\ a_W] \quad (2)$$

The $(\Delta H_{cal}/RT_M^2)$ term of eqs 1 and 2 is determined directly from DSC experiments, whereas the terms in brackets are determined from the slopes of the plots of T_M as a function of the concentration of salt and water activity, respectively. The 1.11 value in eq 1 is a constant that is used to convert activity into concentration terms. The activity of water was varied by using different concentrations of a co-solute, ethylene glycol, which does not interact with the hairpin molecule. The osmolality of these solutions were obtained using a UIC vapor-pressure model 830 osmometer, calibrated with standardized NaCl solutions.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Results and Discussion

UV Unfolding of Hairpins

Typical UV melting curves in buffer with 0.1 M NaCl are shown in Figure 1a. All three UV melting curves are sigmoidal and show the characteristic melting behavior of the cooperative unfolding of base pairs and base-pair stacks. The helix-coil transition of each hairpin takes place in monophasic transitions, that show a hyperchromic effect. These hyperchromicities ranged from 13 % (*IL-6*) to 22 % (*IL-2*). The T_Ms of the transitions of each hairpin follow the order: 68.2 °C (*IL-2*) > 57.6°C ($_{IL\text{-}6}$) > 50.4 °C (*IL-10*), and correspond roughly to the percentage of GC base pairs in their helical stems. The T_Ms for the transitions of each oligonucleotide remain constant, despite the ~10 fold increase in strand concentration (Figure 1b) consistent with its intramolecular unfolding (*10–12*). Analysis of the shape of these UV melts yielded averaged ΔH_{vH} values of 64 kcal/mol (*IL-2*), 57 kcal/mol (*IL-6*) and 53 kcal/mol (*IL-10*). The combination of these two observables yielded folding ΔG^o_{20}'s that indicate *IL-2* is the most stable hairpin, while *IL-10* is the least stable one.

CD Spectra and Conformational Analysis

Figure 2a shows representative CD spectra of each hairpin at 5 °C and 95 °C. At 5 °C, all spectra show a large positive band centered at 280 nm and a large negative band at 250 nm. These two bands are equivalent in magnitude, which indicates the hairpins are adopting the B-like conformation. At 95 °C, both bands decreased significantly their magnitude, by ~60% (250 nm) and by ~40% (280 nm), providing two different wavelengths to monitor the CD unfolding of these molecules. This is in contrast to the single wavelength, around 250-255 nm, that is used to follow the base-pair unstacking in the CD unfolding of B-DNA duplexes. Figure 2b shows the α-curves of each hairpin at these two wavelengths in buffer containing 0.1 M NaCl. All CD melts showed monophasic transitions. Analysis of the shape of the CD melts yielded T_Ms of 67.9 °C (*IL-2*), 57.0 °C (*IL-6*), and 49.0 °C (*IL-10*) at both wavelengths, and ΔH_{vH}s of 59 kcal/mol (*IL-2*), 53 kcal/mol (*IL-6*) and 50 kcal/mol (*IL-10*). Overall, the T_M and ΔH_{vH} values are in good agreement with the ones obtained from the UV melts. One observation is that we obtained an additional optical window to follow the CD unfolding of these hairpins at 280 nm, which monitors the changes in the sugar phosphate backbone. The ellipticity at this wavelength remains more or less unchanged in the unfolding of a DNA duplex in the "B" conformation.

DSC Unfolding

DSC melting curves in low salt buffer, and with additional salt, or ethylene glycol, are shown in Figure 3 for each hairpin. The unfolding of each molecule undergoes highly reproducible monophasic transitions. The T_Ms obtained from the calorimetric melts (Table I) are similar to those obtained from UV and CD melts, consistent with their intramolecular formation. The increase in salt concentration shifts all curves to higher temperatures, whereas the addition of ethylene glycol

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

shifts the curves to lower temperatures. The net effect is to increase the T_M, by an average of 10 °C, with the increase in salt; and to decrease the T_M, by an average of 5 °C. Thus, the effect of salt on the helix-coil transition of these hairpins is to stabilize the helical state with the higher charge density parameter, while the effect of osmolyte is to stabilize the random coil state i.e., hydration favors the helical state.

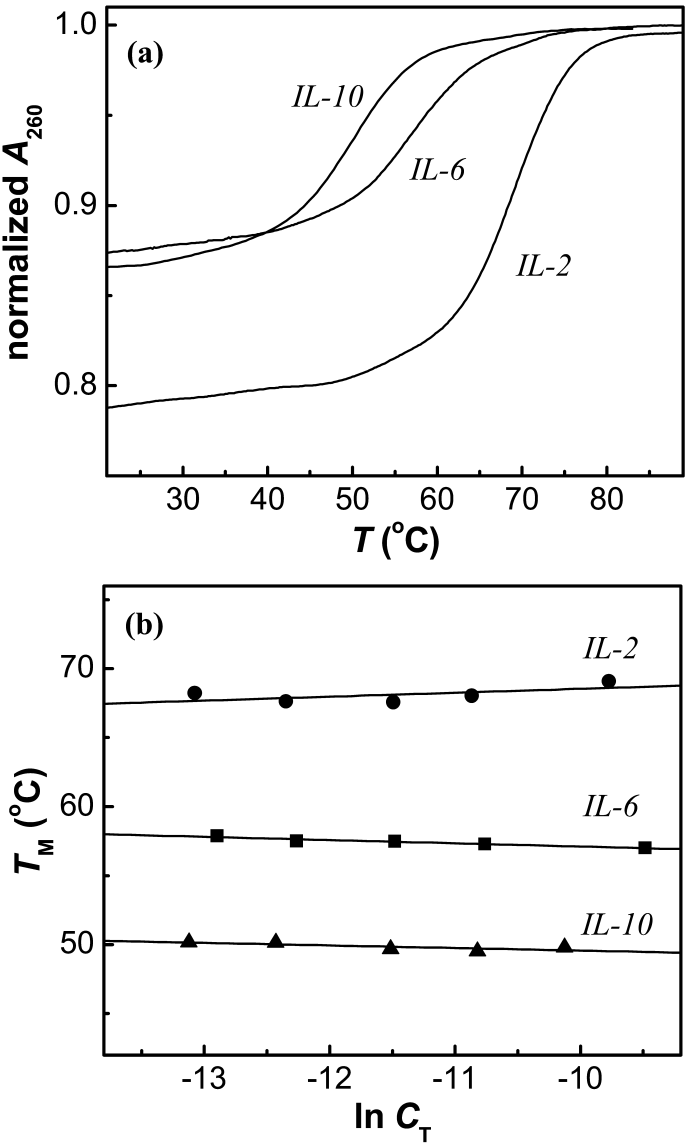


Figure 1. (a) UV melting curves of hairpins at 260 nm in 10 mM sodium phosphate buffer, 0.1 M NaCl, at pH 7. (b) The TM-dependences on total strand concentration, IL-2 (□), IL-6 (■) and IL-10 (▲).

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

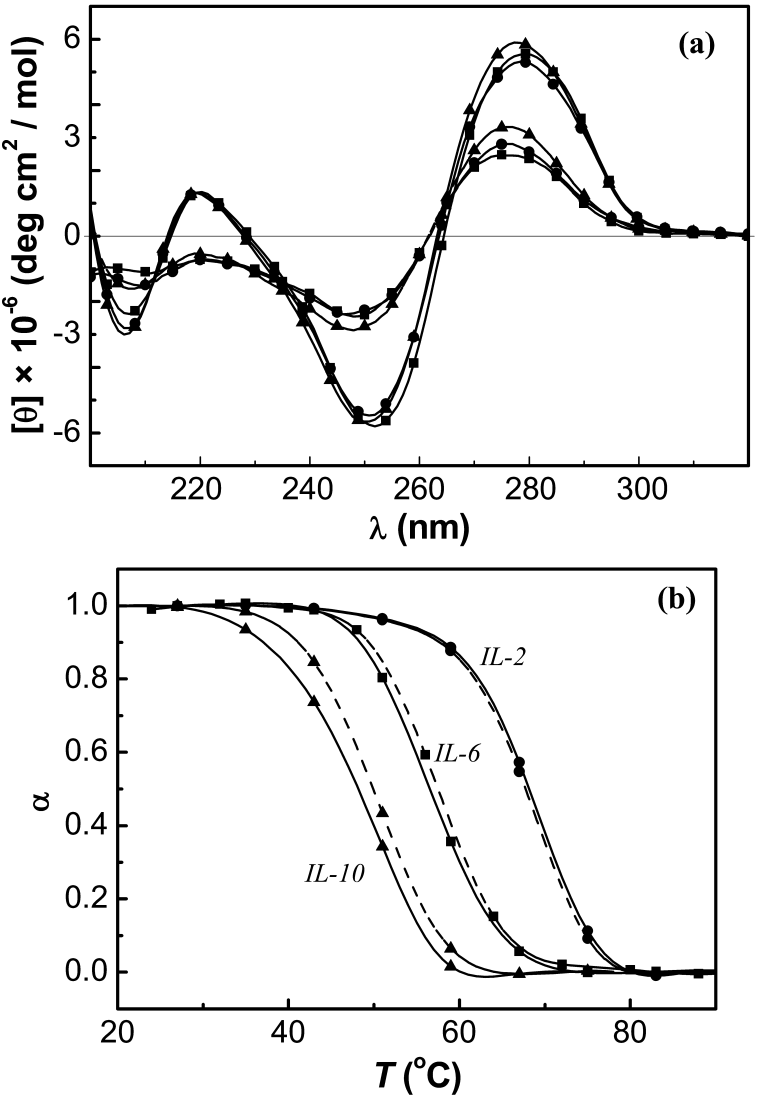


Figure 2. (a) CD spectra of hairpins at 5 °C (solid lines) and at 95 °C (dashed lines), IL-2 (□), IL-6 (■) and IL-10 (▲). (b) CD melting curves at 250 nm (solid lines) and 280 nm (dashed lines), IL-2 (□), IL-6 (■) and IL-10 (▲). All experiments were done in 10 mM sodium phosphate buffer, 0.1 M NaCl, at pH 7.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

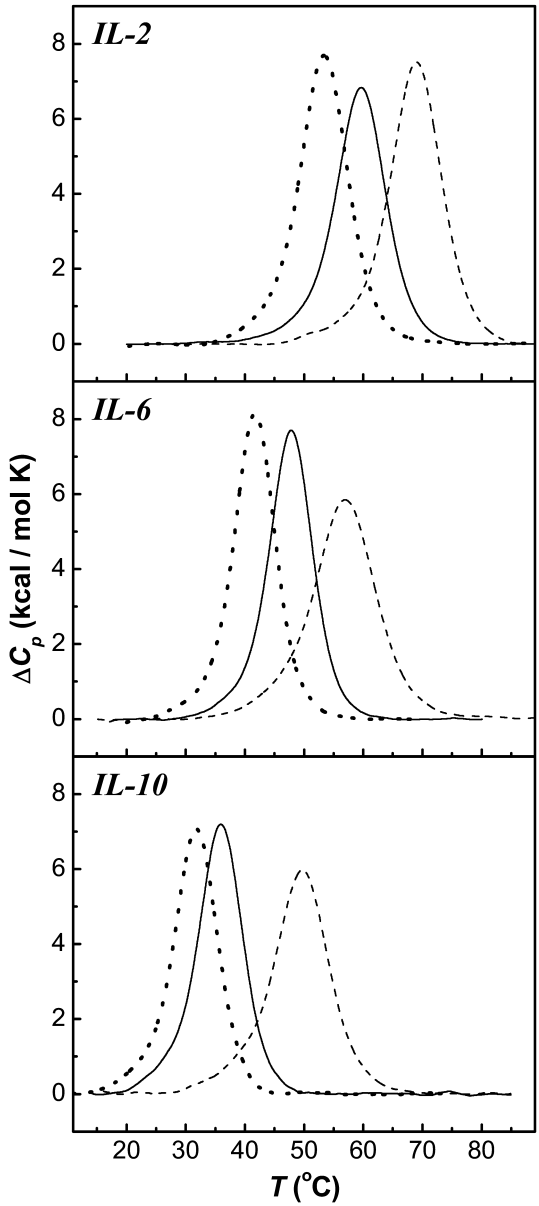


Figure 3. DSC curves of hairpins (~80 μM in total strands) in 10 mM sodium phosphate buffer at pH 7 (solid lines); with the addition of 0.1 M NaCl (dashed lines); or with the addition of 3 m ethylene glycol (dotted lines).

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

The ΔH_{cal}s obtained from the DSC melts are shown in the second column of Table I. In low salt buffer, we measured an average ΔH_{cal} of -75 kcal/mol. The increase in the size of the internal loop slightly drops the favorable ΔH_{cal} term from -77 kcal/mol (*IL-2*), to -76 kcal/mol (*IL-6*), and to -73 kcal/mol (*IL-10*). However, the increase in salt yielded more favorable ΔH_{cal}, by an average of 6 kcal/mol, while the inclusion of osmolyte to the low salt buffer yielded also an increase in favorable enthalpy, by an average of 3 kcal/mol. In both cases slight drops are observed when the size of the internal loop is increased from 2 to 10 thymines (Table I). Thus, the effect of both salt and osmolyte is to improve base-pair stacking interactions within each hairpin. The increase in the size of the internal loop yielded slight enthalpy drops in all solution environments, indicating poor thymine-thymine stacking contributions, as observed earlier (*29*).

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Table I. Thermodynamic Profiles for the Folding of Hairpins at 20 °C[a]

Solution Conditions	T_M *(°C)*	ΔH_{cal} *(kcal/mol)*	ΔC_p *(cal/°C mol)*	ΔH_{vH} *(kcal/mol)*	$T\Delta S$ *(kcal/mol)*	ΔG^o_{20} *(kcal/mol)*
			IL-2			
buffer	59.6	-77.2	-970	-78	-68.0	-9.2
+0.1 M NaCl	69.1	-86.3		-80	-73.9	-12.4
+3 m EG	53.3	-84.1		-77	-75.5	-8.6
			IL-6			
buffer	47.6	-76.1	-630	-83	-69.6	-6.5
+0.1M NaCl	57.1	-82.1		-64	-72.9	-9.2
+3 m EG	41.7	-77.5		-84	-72.2	-5.3
			IL-10			
buffer	36.2	-73.1	-310	-75	-69.3	3.8
+0.1 M NaCl	49.5	-76.4		-65	-69.3	-7.1
+3 m EG	31.7	-73.5		-74	-70.7	-2.8

[a] All experiments were performed in 10 mM sodium phosphate buffer at pH 7.0. Experimental errors are as follows: T_M (± 0.5 °C), ΔH_{cal} (± 5 %), ΔH_{VH} (± 15 %), ΔC_p (± 20 %), $T\Delta S$ (± 5 %), $\Delta G^0_{(20)}$ (± 7%).

We used the DNA nearest neighbor parameters to predict the unfolding enthalpy of just the stem of these hairpins without the internal loops (*30*, *31*), we obtained an unfolding enthalpy of 72.5 kcal/mol; however, the inclusion of a T-T mismatch (*IL-2*) and considering also the two adjacent end loop thymines yielded an unfolding enthalpy of 81 kcal/mol (*31–33*). This value is in excellent agreement with the experimentally determined ΔH_{cal} of 86 kcal/mol of this hairpin. The increase in the size of the internal loop, from 2 to 10 thymines, yielded lower enthalpy values, indicating negligible contributions from thymine-thymine stacking in the internal loops. Most likely the placement of larger internal loops is making the hairpins more flexible, with the internal loop thymines slightly more exposed to the solvent.

The ΔH_{vH} values from DSC measurements (fourth column of Table I) are on the average 15 kcal/mol higher than the average ΔH_{vH} values of all optical melts, and similar to their corresponding ΔH_{cal} values. The ΔH_{vH} terms correspond to the unfolding of a cooperative unit, assuming a two-state transition. The nature of the helix-coil transition of each hairpin is therefore analyzed by simple inspection of the $\Delta H_{vH}/\Delta H_{cal}$ ratio, using the ΔH_{vH}s obtained from DSC experiments. We obtained $\Delta H_{vH}/\Delta H_{cal}$ ratios ranging from 0.9-1.1 for the majority of the transitions, independent of the solution condition used, which indicates that all transitions are two-state (*27*) i.e., hairpin unfolding takes place without the presence of intermediates. The exceptions were the transitions of IL-6 and IL-10 in the higher salt buffer, their $\Delta H_{vH}/\Delta H_{cal}$ ratios were 0.8 and 0.85, respectively. This may be due to the stabilizing effect of salt on improving base-pair stacking, lowering the size of the cooperative unit of these hairpins (*21*).

Heat Capacity Effects

We did not find heat capacity effects between the initial and final states of these curves, i.e., $\Delta C_p = 0$. However, small heat capacity effects may be present but the sensitivity of the VP-DSC calorimeter does not allow their direct determination, as these are within the experimental noise of the DSC baselines (40 cal/°C-mol base pair.). To determine whether heat capacity effects during unfolding we have used an indirect approach, where we measured DSC melting curves at different salt concentrations (Figure 4a). The resulting plots of ΔH_{cal} as a function of T_M are shown in Figure 4b. These plots yielded straight lines with positive slopes, which correspond to positive ΔC_p values (Table I) of 970 cal/°C (*IL-2*), 630 cal/°C (*IL-6*), and 310 (*IL-10*) cal/°C. This indicates that the unfolding of each hairpin is accompanied by an exposure of hydrophobic groups to the solvent; a greater exposure is obtained with *IL-2* while the lowest exposure is with *IL-10* i.e, the thymine bases of the internal loop of *IL-10* have similar exposure to those in the random coil state. If we assume similar interactions of the unfolded state of each hairpin with the solvent water; then the helical state of *IL-2* is the most hydrophilic one, which is consistent with a fully paired stem.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Differential Binding of Counterions

We measured the differential binding of counterions, using eq 1, from the T_M-dependence on salt concentration, and the average $\Delta H_{cal}/RT_M^2$ term obtained in the DSC curves at several concentrations (Table II). The derivation of eq 1 assumes a similar type of sodium binding to the helical and coil states of each hairpin. For each transition, the increase in salt concentration shifts the melting curves to higher temperatures (Figure 5a), due to the shift of the hairpin-coil equilibrium towards the conformation with higher charge density parameter. The T_M-dependence on salt concentration for each hairpin is shown in the inset of Figure 5a, linear dependences are obtained with slope values ranging from 4.5 °C (*IL-2*) to 6.9 °C. The Δn_{Na+} values shown in the third column of Table II correspond to the folding of each hairpin, all values are negative and correspond to an uptake of counterions, while their magnitude indicates the strength of ion binding by the helical state of each molecule. We obtained Δn_{Na+}s ranging from -1.8 mol Na^+/mol hairpin (*IL-2*) to -2.8 mol Na^+/mol hairpin (*IL-10*) and are normalized per phosphate, by taking into account the number of phosphates of the stem plus the two phosphates adjacent to the stem of each hairpin; for instance, *IL-2* has 20 phosphates, while *IL-10* has 28 phosphates. This exercise yields Δn_{Na+} values ranging from 0.09 to 0.1 mol Na^+/mol phosphate. These values are low relative to the 0.17 mol Na^+/mol phosphate value for long DNA duplexes (*34*, *35*), but consistent with the formation of 4-5 base pairs in the hairpin stems (*22*, *36*). However, the similarity of the values indicate that the phosphates of the hairpins with larger internal loops behave electrostatically as helical phosphates, this may be consistent with T-T mismatches having two hydrogen bonds and therefore these thymine bases are intrahelical and somewhat loose.

Differential Binding of Water Molecules

UV melting curves (α-curves) as a function of osmolyte concentration were performed in the 0.5-3 m range (Figure 5b). Increasing the concentration of ethylene glycol (decreasing water activity) had a small effect on the T_M (shifts of 1-4 °C). The T_M dependences on the activity of water are shown in the inset of Figure 5b. The differential binding of water between the hairpin and random coil states were calculated using eq 2. We obtained negative Δn_W values for the formation of each hairpin (Table II), indicating that hairpin folding is accompanied by an uptake of water molecules, i.e., the random coil state is the less hydrated stated in these molecules. Specifically, these Δn_W terms ranged from -63 mol H_2O/mol hairpin (*IL-2*) to -41 mol H_2O/mol hairpin (*IL-10*) and correlate well with the measured heat capacity effects with these hairpins. This emphasizes that the type of water measured with this osmotic stress technique is structural water, which is the one immobilized by polar and non-polar groups.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

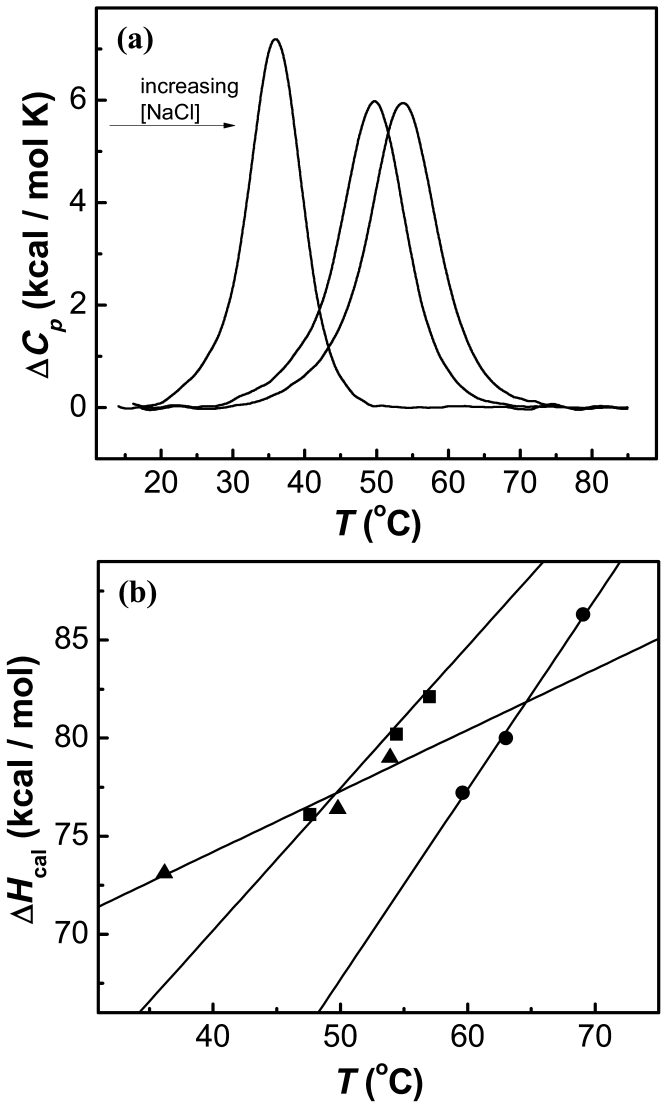


Figure 4. (a) Typical DSC curves of IL-10 in 10 mM sodium phosphate buffer, at pH 7 and at several NaCl concentrations of 0 M, 0.1 M, and 0.2 M. (b) Determination of ΔCp effects from ΔHcal vs. T_M plots, IL-2 (□), IL-6 (■), and IL-10 (▲).

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

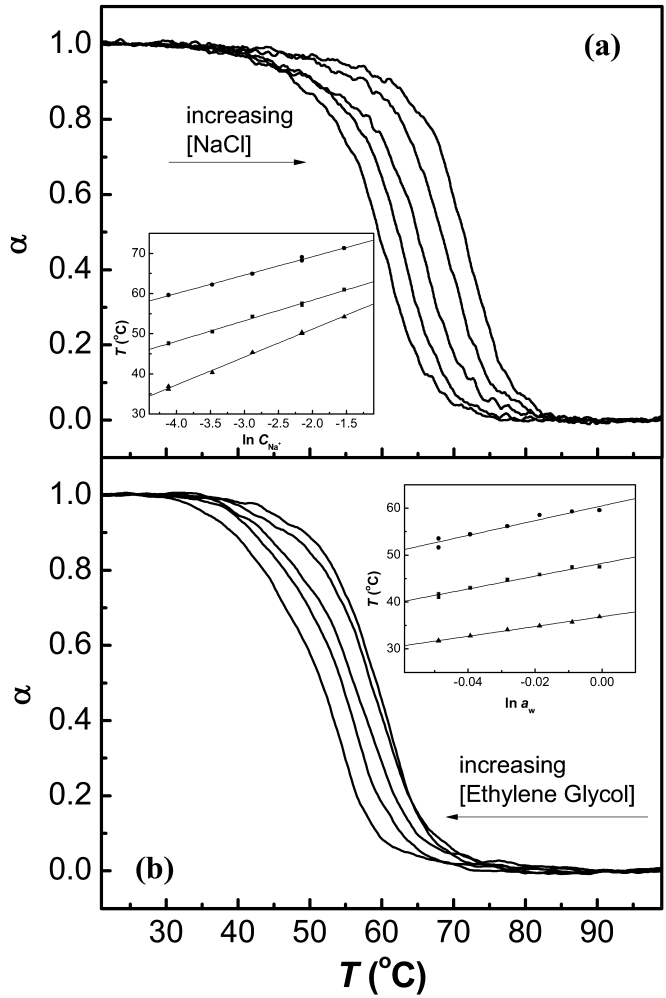


Figure 5. (a) Typical α-curves of IL-2 in 10 mM Na phosphate buffer, at pH 7, adjusted to desired salt concentration with NaCl. Inset: T_M-dependences on [NaCl], IL-2 (□), IL-6 (■) and IL-10 (▲). (b) Typical α-curves of IL-2 in 10 mM Na phosphate buffer, at pH 7, adjusted to the desired ethylene glycol concentration. Inset: T_M-dependences on (H_2O), IL-2 (□), IL-6 (■) and IL-10 (▲).

Table II. Differential Binding of Ions, Δn_{Na+}, and Water, Δn_W[a]

Differential Binding of Counterions			*Differential Binding of Water*		
$\Delta H_{cal}/RT_M^2$ *(kcal/mol)*	$\alpha T_M/\alpha \ln [Na^+]$ *(°C)*	Δn_{Na+} *(per mol hairpin)*	$\Delta H_{cal}/RT_M^2$ *(kcal/mol)*	$\alpha T_M/\alpha \ln a_w$ *(°C)*	Δn_w *(per mol hairpin)*
		IL-2			
0.36	4.5	-1.8	0.38	168	-63
		IL-6			
0.38	5.2	-2.2	0.38	138	-53
		IL-10			
0.37	6.9	-2.8	0.39	104	-41

[a] The salt dependences were done in 10 mM sodium phosphate buffer at pH 7.0 and adjusted to the desired salt concentration with NaCl, while the dependences on the activity of water were determined in the same buffer and adjusted to the desired osmolyte concentration with ethylene glycol. The experimental errors are as follows: Δn_{Na+} (±7%) and Δn_W (±10%).

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Thermodynamic Profiles for the Folding of Hairpin

Thermodynamic profiles for all transitions observed in the folding of each hairpin are listed in Table I and shown as building blocks in Figure 6 (law salt buffer). The thermodynamic parameters of each transition show that the favorable free energy contribution results from the characteristic compensation of a favorable enthalpy contribution with an unfavorable entropy contribution. Favorable heat contributions involve the formation of base pairs and base pair stacks, whereas unfavorable entropy contributions involve the ordering of the strands, ion and water binding. The enthalpy contributions were discussed in the previous sections. The entropy contributions, in energy terms ($T\Delta S_{cal}$), for the folding of each hairpin are shown in the sixth column of Table I. The magnitude of these unfavorable entropy terms follow the same order as those of the enthalpies, and correspond to contributions of the ordering of the oligonucleotide, uptake of counterions, and the immobilization of water molecules by the helical states of each hairpin (Figure 6).

In terms of the overall free energy contribution at 20 °C, ΔG^o_{20}, we obtained favorable ΔG^o_{20} terms (last column of Table I), indicating that all hairpins folded as designed at the three different solution conditions used. In the low salt buffer, the favorable ΔG^o_{20} drops substantially from -9.2 kcal/mol (*IL-2*), to -6.5 kcal/mol (*IL-6*), to -3.8 kcal/mol (*IL-10*) with the increase in the size of the internal loop. The addition of 0.1 M NaCl yielded more favorable ΔG^o_{20} terms, by an average of ~3.1 kcal/mol, whereas the addition of 3 m ethylene glycol yielded less favorable ΔG^o_{20} terms, by ~0.9 kcal/mol. These effects emphasize the overall electrostatic and hydration contributions on the stability of a nucleic acid helix.

The left panel of Figure 6 shows standard thermodynamic profiles for each hairpin in low salt buffer (open buildings) and corrected for heat capacity effects (shaded buildings). The inclusion of heat capacity effects, using standard

thermodynamic relationships, yielded higher enthalpy and entropy corrections for *IL-2* (largest ΔC_p) and lowest terms for *IL-10* (lowest ΔC_p), which in turn yielded less favorable folding ΔG^{o}_{20} terms, -4.6 kcal/mol (*IL-2*), -5.1 kcal/mol (*IL-6*), and -3.6 kcal/mol (*IL-10*). Similar drops were obtained at the higher salt concentration (data not shown). Overall, we obtained an average net drop of 5.4 % of ΔG^{o}_{20} per thymine residue in this 10-100 mM NaCl range. *IL-10* has the lowest favorable free energy of unfolding and it is the less affected with the heat capacity corrections (Figure 6). Therefore, hairpin loops with large internal loops in their stems constitute better targets in reactions with its complementary strands. This is due to the lower stability of the hairpin and the large and favorable free energy contribution from the additional formation of base-pairs and base-pair stacks involving the nucleotide residues of the internal loop.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

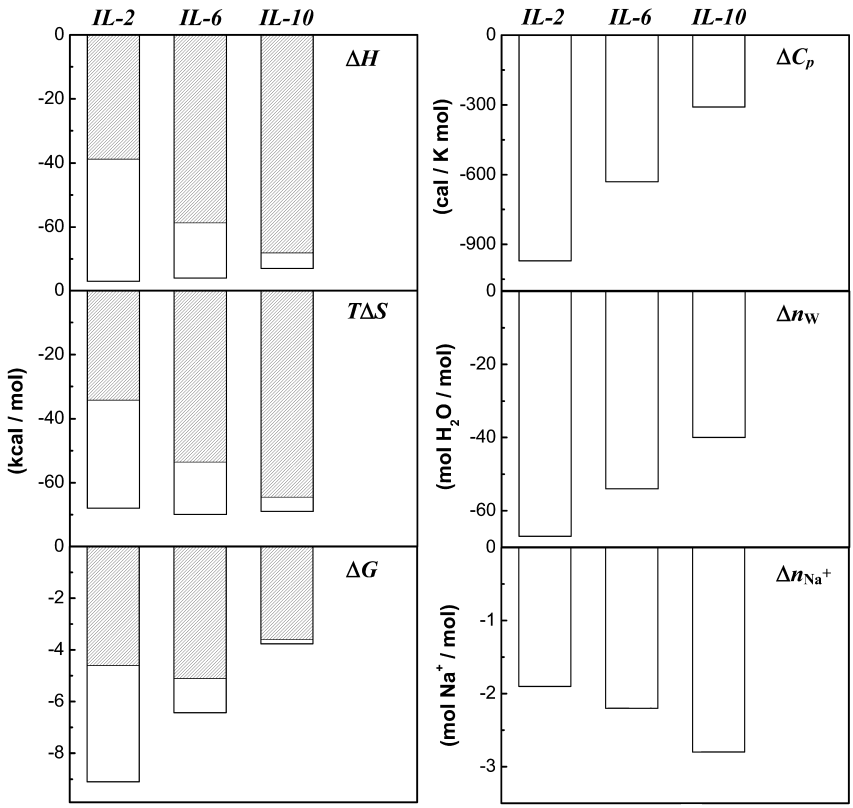


Figure 6. Complete thermodynamic profiles for the folding of hairpins in 10 mM sodium phosphate buffer at pH 7. Uncorrected (open buildings) and corrected for heat capacity effects (shaded buildings).

Conclusions

The unfolding of each hairpin containing internal loops in their stem shows intramolecular monophasic transitions, which result from the typical compensation of an unfavorable enthalpy (breaking of base-pair stacks) and favorable entropy contributions (release of ion and water molecules). The increase in the size of the internal loop from 2 to 10 thymines yielded: a) lower T_Ms and similar enthalpy contributions; b) lower heat capacity values that correlated with the lower releases of structural water molecules; and c) higher ion releases. This investigation has provided new methods, based on physico-chemical principles, to determine how the size of internal loops influences the molecular forces involved in the formation of hairpins. One significant observation is that hairpin loops with large internal loops in their stems are excellent targets in the reactions with its complementary strands. This can be used in the target of RNA molecules with complementary oligonucleotides for the control of gene expression.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Acknowledgments

This work was supported by Grants MCB-0315746 and MCB-0616005 from the National Science Foundation. Partial financial support (to I.P.) from the Slovenian Research Agency (P1 0201) is greatly appreciated.

References

1. Panayotatos, N.; Wells, R. D. *Nature* **1981**, *289*, 466–470.
2. Lilley, D. M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6468–6472.
3. Pollack, Y.; Stein, R.; Razin, A.; Cedar, H. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6463–6467.
4. Lilley, D. M. J. *Nucleic Acids Res.* **1981**, *9*, 1271–1290.
5. Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3746–3750.
6. Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719–14735.
7. SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* **1996**, *35*, 3555–3562.
8. Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211–11216.
9. Rentzeperis, D.; Shikiya, R.; Maiti, S.; Ho, J.; Marky, L. A. *J. Phys. Chem. B* **2002**, *106*, 9945–9950.
10. Marky, L. A.; Maiti, S.; Olsen, C. M.; Shikiya, R.; Johnson, S. E.; Kaushik, M.; Khutsishvili, I. In *Biomedical Applications of Nanotechnology*; Labhasetwar, V., Leslie-Pelecky, D., Ed.; John Wiley & Sons, Inc.: New York, 2007; pp 191−226.
11. Lee, H.-T.; Olsen, C. M.; Waters, L.; Sukup, H.; Marky, L. A. *Biochimie* **2008**, *90*, 1052–1063.

12. Lee, H.-T.; Carr, C.; Siebler, H.; Waters, L.; Khutsishvili, I.; Iseka, F.; Domack, B.; Olsen, C. M.; Marky, L. A. *Methods Enzymol.* **2011**, *492*, 1–26.
13. Crooke, S. T. *Biochim. Biophys. Acta, Gene Struct. Expression* **1999**, *1489*, 31–43.
14. Helene, C. *Eur. J. Cancer* **1991**, *27*, 1466–1471.
15. Helene, C. *Eur. J. Cancer* **1994**, *30A*, 1721–1726.
16. Mahato, R. I.; Cheng, K.; Guntaka, R. V. *Expert Opin. Drug Delivery* **2005**, *2*, 3–28.
17. Maiti, S.; Kankia, B.; Khutsishvili, I.; Marky, L. A. *Biophys. Chem.* **2011**, *159*, 162–171.
18. Abrahams, J. P.; van den Berg, M.; van Batenburg, E.; Pleij, C. *Nucleic Acids Res.* **1990**, *18*, 3055–3044.
19. Knudsen, B.; Hein, J. *Nucleic Acids Res.* **2003**, *31*, 3423–3428.
20. Batey, R. T.; Rambo, R. P.; Doudna, J. A. *Angew. Chem., Int. Ed.* **1999**, *38*, 2326–2343.
21. Marky, L. A; Blumenfeld, K. S.; Kozlowski, S.; Breslauer, K. J. *Biopolymers* **1983**, *22*, 1247–1257.
22. Rentzeperis, D.; Alessi, K.; Marky, L. A. *Nucleic Acids Res.* **1993**, *21*, 2683–2689.
23. Rentzeperis, D.; Medero, M.; Marky, L. A. *Bioorgan. Med. Chem.* **1995**, *3*, 751–759.
24. Marky, L. A.; Kallenbach, N. R.; McDonough, K. A.; Seeman, N. C.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1621–1634.
25. Soto, A. M; Loo, J; Marky, L. A. *J. Am. Chem. Soc.* **2002**, *124*, 14355–14363.
26. Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601–1620.
27. Privalov, P. L.; Potekhin, S. A. *Methods Enzymol.* **1986**, *131*, 4–51.
28. Kaushik, M.; Suehl, N.; Marky, L. A. *Biophys. Chem.* **2006**, *126*, 154–164.
29. Kankia, B. I.; Marky, L. A. *J. Phys. Chem. B* **1999**, *103*, 8759–8767.
30. Marky, L. A.; Breslauer, K. J. *Biopolymers* **1982**, *21*, 2185–2194.
31. SantaLucia, J., Jr.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
32. Peyret, N.; Seneviratne, A.; Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1999**, *38*, 3468–3477.
33. Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406−3415, http://mfold.rna.albany.edu/.
34. Manning, G. S. *Q. Rev. Biophys.* **1978**, *11*, 179–246.
35. Record, M. T., Jr.; Anderson, C. F.; Lohman, T. M. *Q. Rev. Biophys.* **1978**, *11*, 103–178.
36. Rentzeperis, D; Kharakoz, D. P.; Marky, L. A. *Biochemistry* **1991**, *30*, 6276–6283.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch006

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Chapter 7

A Thermodynamic Perspective of sRNA-mRNA Interactions and the Role of Hfq

Kanav Jain, Taylor B. Updegrove, and Roger M. Wartell*

School of Biology and The Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia 30332
***E-mail: roger.wartell@biology.gatech.edu**

The RNA binding protein Hfq facilitates the hybridization of small regulatory RNAs (sRNA) to target mRNAs to regulate gene expression post-transcriptionally. The exact mechanism by which Hfq enhances RNA pairing is not well understood. In this work we examined the interaction of sRNAs and mRNAs and the effect of Hfq on sRNA thermal stability. We employed computational approaches to predict interaction sites of sRNA-mRNA pairs and the effects of single base changes on sRNA-mRNA binding, and compared the predictions to experimental data. The comparisons indicate that thermodynamic calculations predict functional characteristics of sRNA-mRNA binding with good accuracy. Experimental denaturation studies of segments of the sRNA DsrA in the absence and presence of Hfq show that Hfq is a melting protein. This observation and an analysis of DsrA-*rpoS* mRNA binding suggest that Hfq's ability to destabilize DsrA hairpin structures plays a role in enhancing its hybridization to *rpoS* mRNA.

Introduction

A class of short non-coding RNAs, approximately 70 to 150 nt long, have been shown to be involved in post-transcriptional regulation of mRNA expression in *Escherichia coli* and other bacteria. Characterization of a number of sRNAs indicates that many are induced under stress conditions and act in trans by base pairing to target sites on specific mRNAs, often inhibiting and sometimes enhancing translation initiation. A characteristic common to this class of sRNA

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

is their strong binding affinity to Hfq, a RNA binding protein that also binds to mRNA target sites. Many of the sRNAs that have been studied require Hfq to facilitate their regulation of mRNA targets.

Initially identified as a host factor for the replication of the RNA phage Qβ (*1*), the importance of Hfq to cell metabolism in *E. coli* was demonstrated by the widespread pleiotropic effects caused by null mutants of the *hfq* gene (*2*). Phylogenetic analysis showed that Hfq is well conserved in many bacterial species, and is closely related to the Sm family of RNA binding proteins in archaea and eukaryotes (*3–5*). Crystal structures of the protein from different bacterial species (*6–8*) as well as solution studies (*1*, *4*) indicate it forms a hexameric toroid. Over the past ten years *E. coli* Hfq was demonstrated to be important for the regulation of mRNA expression by a number of sRNAs including DsrA (*9*), OxyS (*4*, *10*), RprA (*11*), Spot42 (*5*, *12*), RyhB (*13*) and SgrS (*14*). Hfq homologues in other bacterial species appear to play a similar role in sRNA regulation (*15–18*).

The mechanism by which Hfq binds to the RNAs and enhance sRNA-mRNA pairing is not well understood. Two general models have been proposed that provide conceptual frameworks with which to gain a better understanding of the process (Figure 1). One model is that a Hfq hexamer binds both the sRNA and mRNA target site simultaneously enabling their proximity to enhance the probability of base pair nucleation events between the RNAs. The second model proposes that Hfq alters the structure of one or both RNAs making the complimentary regions accessible for base pairing. In the above models one Hfq is assumed to be involved in binding both RNAs. A variant possibility is that two Hfq are involved, each bound to an RNA. Evidence that Hfq can alter the structure of some sRNA and mRNA lends support to the notion that Hfq acts as a chaperone (*4*, *13*, *19*). Several *in vitro* studies have shown that Hfq enhances the rate of sRNA-mRNA association and the apparent equilibrium constant of sRNA-mRNA formation (*14*, *20–22*).

In this article we have examined the interaction of sRNAs and mRNAs and the role of Hfq from a thermodynamic perspective. Previous investigators have proposed that the requirement of Hfq for sRNA-mRNA pairing may be related to the free energy of formation or binding affinity of sRNA-mRNA complexes (*23*). Several algorithms have been developed to predict sRNA-mRNA interaction based on the empirical nearest neighbor model governing RNA secondary structure and free energy minimization of RNA-RNA binding (*24–27*).

We have employed computational approaches to predict interaction sites of sRNA-mRNA pairs and the effects of single base changes on sRNA- mRNA binding and compared them to experimental data. The comparisons indicate that thermodynamic calculations determine functional characteristics of sRNA-mRNA binding with good accuracy. Although the calculations did not indicate a correlation between the free energy of sRNA-mRNA binding and a requirement for Hfq, analysis of the interaction between DsrA and its target site on *rpoS* mRNA indicate that part of Hfq's mechanism of action is to destabilize hairpin duplex regions of DsrA.

Experimental denaturation studies of segments of DsrA in the absence and presence of Hfq indicate that Hfq is indeed a melting protein. This finding is consistent with previous nuclease footprinting analysis that showed Hfq binding

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

can make specific base pairs of sRNAs or mRNA accessible (*4*, *13*). Our thermodynamic analysis of DsrA-*rpoS* mRNA binding and the melting curve results support the notion that Hfq's ability to destabilize the first and second hairpins of DsrA is part of its mechanism for promoting DsrA-*rpoS* mRNA interaction, particularly at temperatures below 25 °C.

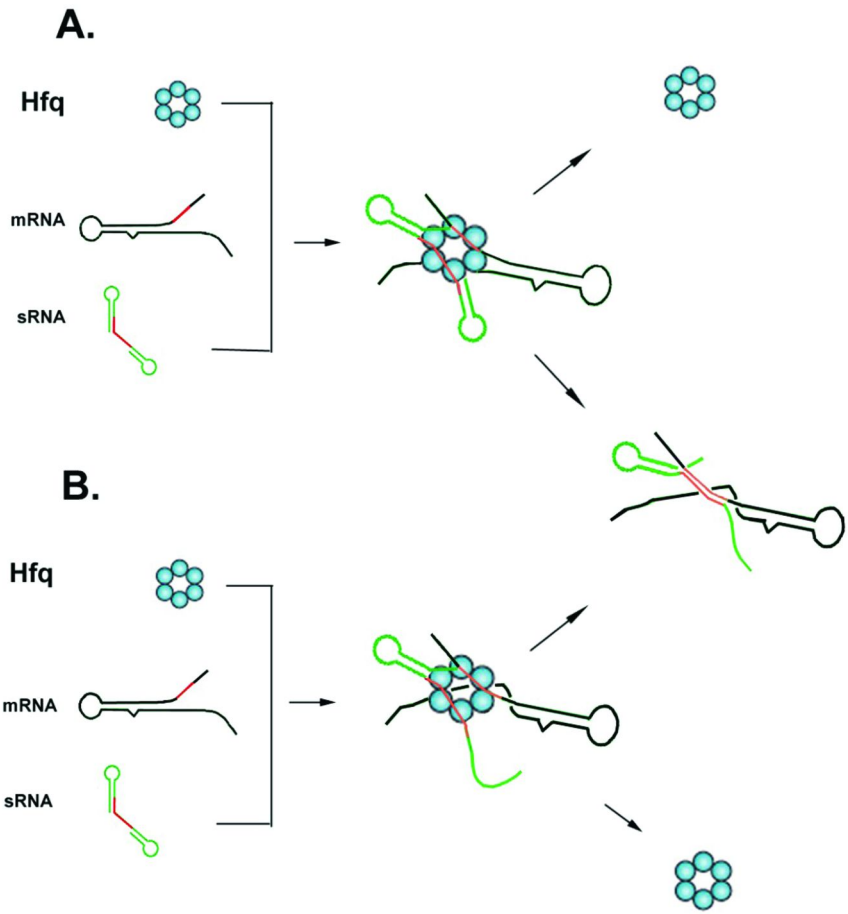


Figure 1. Cartoons illustrating two models of how Hfq may facilitate pairing of an sRNA with its mRNA target site. ***A.*** *Hfq hexamer binds transiently to both the sRNA and mRNA holding them in proximity with an appropriate orientation to enable base pair nucleation.* ***B.*** *Hfq binds to both sRNA and mRNA and induces conformational changes to one or both RNAs that are needed to enable base pair nucleation between the RNAs.*

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Methods

Computational Prediction of sRNA-mRNA Interaction Sites

The RNAup program was used to predict the interaction sites for the sRNA-mRNA pairs listed in Table I (*26*, *28*). RNAup considers the free energy required to open potential binding sites in each RNA (i.e. site accessibility), and the free energy gained from duplex formation. The region of interaction is restricted to one interval for each RNA. The interaction site may include bulges and internal loops but not intrastrand secondary structures with base pairs. The optimal interaction site is determined by calculating all combinations of potential binding sites of both RNAs. The free energy parameter set from Mathews et al 2004 (*29*) was employed and default parameters of the RNAup website were used for most calculations. The maximum interaction length was 25 except for two instances in which the reported interaction site extended to 35 nt and 49 nt. Predictions using the energy parameters from Mathews et al 1999 were also employed (*30*).

The full sRNA sequences and mRNA sequences from the start of transcription to 100 nt downstream of the translation start codon were employed. Locations of transcriptional start points were obtained from EcoCyc for all *E. coli* mRNAs (*31*). Transcription start locations for other species such as S. enterica were found in the literature. Reported sRNA-mRNA pairing regions were found in the literature with the aid of sRNATarBase <http://ccb.bmi.ac.cn/srnatarbase/> (*32*) as well as sRNAMap <http://srnamap.mbc.nctu.edu.tw/> (*33*). Interaction site positions are numbered relative to 'A' (+1) of the mRNA AUG start codon.

RNA-RNA duplex free energies were calculated for segments of sRNA and mRNA using the program RNA Structure (*34*). This approach was employed to ensure free energies were determined at the experimentally reported interaction site. The bimolecular binding free energy of a sRNA segment pairing to a mRNA sequence was denoted as ΔG_{sm}. The intra-molecular binding free energy of a mRNA sequence hybridizing to another sequence on the mRNA was calculated by treating each sequence as a separate strand and evaluating the bimolecular binding energy denoting it as ΔG_{m1m2}.

Preparation of Hfq Protein and $DsrA_{dI}$ and $DsrA_{dII}$

The Impact-CN intein system (New England Biolabs) was used to purify Hfq protein as previously described (*35*). The plasmid used to over express the Hfq protein contained the *E. coli* hfq gene inserted into SapI-SmaI digested pTYB11 plasmid (pEcHfq). Protein purification was carried out according to the recommendation of the manufacturer using strain ER2566. Cell lysis was carried out using a french press, and a chitin column employed to retain the Hfq fused to a chitin binding domain. The column was extensively washed with the lysis/wash buffer of 20 mM Tris (pH 8.3) and 1 M NaCl prior to incubation of the column with this buffer plus 40 mM dithiothreitol. The eluted Hfq was concentrated and buffer-exchanged using centrifugation filtration units. The dominant form of Hfq in solution is a homohexamer (*4*, *36*).Throughout this paper concentrations of Hfq refer to moles hexamer/L.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Table I. Computed sRNA-mRNA Binding Free Energies and Predicted and Reported Interaction Sites. Org is organism with abbreviations EC- *Escherichia coli*, SE- *Salmonella. enterica*, VC- *Vibrio Cholera*, BS- *Bacillus subtilus*, PM- *Pro. MED4*. ΔG_{total} is the total or net free energy calculated by RNAup for each sRNA-mRNA pair. They are determined from ΔG_{duplex}, the free energy gain for sRNA-mRNA pairing, and ΔG_{open}, the free energy cost to open the folded structures of the sRNA and mRNA to allow base pairing. The predicted and reported sites x,y denote the mRNA region hybridized by the sRNA. Positions are numbered relative to the mRNA start codon, +1, with positions upstream numbered -1,-2,-3,.., and positions downstream +2,+3, +4…. Sensitivity and PPV are defined in the text.

sRNA	*mRNA*	*Org.*	*ΔG_{total} (kcal/mol)*	*ΔG_{duplex} (kcal/mol)*	*ΔG_{open} mRNA (kcal/mol)*	*ΔG_{open} sRNA (kcal/mol)*	*Predicted Sites*	*Sensitivity*	*PPV*	*Reported Sites*	*Refer-ence*
RyhB	sodB	EC	-9.34	-17.98	5.54	3.09	-13,+5	1.000	0.600	-4,+5	(*13*)
DsrA	hns	EC	-14.55	-22.87	2.04	6.28	+2,+22	1.000	0.765	+7,+19	(*44*)
MicC	ompC	EC	-17.61	-22.91	1.54	3.76	-33,+15	1.000	0.889	-30,-15	(*45*)
MicF	ompF	EC	-15.11	-18.60	2.96	0.54	-11,+7	0.857	1.000	-11,+10	(*45*)
SgrS	ptsG	EC	-18.30	-26.65	3.72	4.63	-28,-9	1.000	0.471	-22,-15	(*14*)
GcvB	dppA	EC	-16.50	-24.54	0.84	7.21	-31,-11	0.941	0.842	-31,-14	(*46*)
DsrA	rpoS	EC	-10.55	-16.76	4.06	2.15	-126,-109	0.500	0.647	-119,-97	(*47*)
RprA	rpoS	EC	-9.20	-17.90	5.32	3.39	-117,-101	0.722	0.867	-117,-94	(*47*)
RybB	ompC	EC	-10.51	-16.30	4.35	1.44	-49,-42	0.800	1.000	-50,-41	(*37*)
RybB	ompN	EC	-12.65	-17.45	3.13	1.67	+5,+20	1.000	1.000	+5,+20	(*48*)
CyaR	ompX	EC	-10.88	-18.59	3.24	4.47	-9,+15	1.000	0.714	-9,+2	(*49*)

Continued on next page.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Table I. (Continued). Computed sRNA-mRNA Binding Free Energies and Predicted and Reported Interaction Sites. Org is organism with abbreviations EC- *Escherichia coli*, SE- *Salmonella. enterica*, VC- *Vibrio Cholera*, BS- *Bacillus subtilus*, PM- *Pro. MED4*. ΔG_{total} is the total or net free energy calculated by RNAup for each sRNA-mRNA pair. They are determined from ΔG_{duplex}, the free energy gain for sRNA-mRNA pairing, and ΔG_{open}, the free energy cost to open the folded structures of the sRNA and mRNA to allow base pairing. The predicted and reported sites x,y denote the mRNA region hybridized by the sRNA. Positions are numbered relative to the mRNA start codon, +1, with positions upstream numbered -1,-2,-3,.., and positions downstream +2,+3, +4…. Sensitivity and PPV are defined in the text.

sRNA	*mRNA*	*Org.*	*ΔG_{total} (kcal/mol)*	*ΔG_{duplex} (kcal/mol)*	*ΔG_{open} mRNA (kcal/mol)*	*ΔG_{open} sRNA (kcal/mol)*	*Predicted Sites*	*Sensitivity*	*PPV*	*Reported Sites*	*Refer-ence*
CyaR	nadE	EC	-10.97	-17.60	3.75	2.89	-11,-2	0.909	1.000	-11,+3	(*50*)
MicA	ompX	SE	-12.01	-18.70	5.04	1.65	+2,+24	0.950	1.000	+1,+24	(*49*)
GcvB	livJ	SE	-16.27	-27.90	3.89	7.74	-51,-28	1.000	1.000	-51,-28	(*46*)
GcvB	oppA	SE	-15.58	-26.07	5.01	5.48	-8,+16	1.000	1.000	-8,+16	(*46*)
GcvB	gltI	SE	-14.85	-25.81	7.51	3.45	-51,-28	0.923	0.545	-38,-27	(*46*)
FnrS	sodB	EC	-7.15	-14.77	5.54	2.08	-14,+4	0.000	0.000	+13,+20	(*51*)
OmrB	csgD	EC	-16.07	-30.31	7.89	6.35	-79,-61	1.000	1.000	-79,-61	(*52*)
MicA	ompA	EC	-12.55	-18.26	4.87	0.85	-19,-6	0.867	1.000	-21,-6	(*53*)
RyhB	sdhD	EC	-14.19	-26.73	5.90	6.63	-33,-6	0.897	1.000	-33,-3	(*54*)
RyhB	fur	EC	-5.55	-8.40	1.79	1.06	-60,-54	0.214	1.000	-96,-47	(*55*)
Qrr4	luxO	VC	-10.59	-15.50	3.09	1.82	-10,-3	1.000	0.875	-9,-3	(*56*)
Qrr1	hapR	VC	-12.67	-20.70	3.24	4.79	-29,-11	0.727	1.000	-35,-5	(*17*)

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

sRNA	*mRNA*	*Org.*	ΔG_{total} *(kcal/mol)*	ΔG_{duplex} *(kcal/mol)*	ΔG_{open} *mRNA (kcal/mol)*	ΔG_{open} *sRNA (kcal/mol)*	*Predicted Sites*	*Sensitivity*	*PPV*	*Reported Sites*	*Reference*
DicF	ftsZ	EC	-9.73	-24.04	5.52	8.79	-28,-8	0.625	1.000	-28,+2	(*57*)
RybB	ompW	EC	-15.71	-22.54	4.95	1.88	-4,+23	0.720	0.947	-13,+20	(*58*)
FsrA	sdhC	BS	-14.72	-28.44	4.54	9.17	+7,+28	0.563	1.000	-23,+28	(*59*)
IstR-1	tisB	EC	-17.14	-38.11	10.75	10.21	-124,-101	1.000	0.958	-124,-102	(*60*)
VrrA	ompA	VC	-12.51	-20.31	5.10	2.69	-22,+15	0.818	0.692	-14,+21	(*61*)
Yfr1	PM1119	PM	-12.10	-16.90	4.80	0.00	-3,+7	1.000	1.000	-3,+7	(*62*)

To remove residual contaminating nucleic acids, Hfq preparations were subjected to a micrococcal nuclease treatment. 25 µl of 300 units/ml micrococcal nuclease (Worthington Biochemical Corporation) was added to ~1 ml of 0.3 to 0.4 OD of Hfq measured at 274 nm in 0.2 M NaCl, 20 mM Tris (pH 8.5) and 5 mM $CaCl_2$ and incubated at 37 °C for 45 min. This nuclease has a strict dependence on Ca^{2+}. Following this reaction 10 µL of 0.5 M Na_2EDTA was added and the sample was washed and concentrated in 0.5 M or 0.2 M NaCl and 20 mM Tris at pH 8.3 using 30 kD MWCO Millipore centrifugal filter units. The UV spectrum gave an absorbance ratio $A_{274nm}/A_{255nm} \geq 1.8$. The two RNAs employed in the study $DsrA_{dI}$ and $DsrA_{dII}$ were synthesized by IDT Inc and purified by HPLC. Their sequences are illustrated in Figure 6. They were dissolved in the 0.1 M Na+ buffer employed for melting studies (see below) and characterized by UV spectra and polyacrylamide gel electrophoresis.

UV Melting Curve Studies

UV absorbance was employed to monitor the melting transitions of the two RNA molecules in the absence and presence of Hfq. Melting curves were obtained using a 0.1 M Na+ buffer (0.1 M NaCl, 0.02 M Tris, pH 8.0). The concentration of the DsrA fragments employed in the UV melting studies was 1 µM. Samples were placed in 1 cm path length quartz cuvettes and a Cary 100 spectrophotometer (Varian, Inc.) was employed to measure absorbance as a function of temperature at 260 nm with a heating rate of 0.5 °C min^{-1}. Temperature was measured with a platinum resistance probe inserted into a solvent cell adjacent to the sample. Absorbance readings were taken every 0.2 °C over a range of the experiments. For experiments that monitored the influence of Hfq, the protein was added to the RNAs from a stock protein solution to give the concentrations indicated in Figures 6 and 7. A solution of 1.5 µM Hfq alone had an absorbance of 0.025 at 260 nm. It showed no change in absorbance with temperature until ~73 °C when it increased by ~0.012.

Results

Predicting sRNA-mRNA Interaction Sites

To assess the ability of equilibrium thermodynamics to predict sRNA-mRNA interactions, 29 reported sRNA-mRNA interaction sites were compared with their thermodynamically predicted sites. The RNAup algorithm (*28*) was employed using the complete sRNA sequences and their corresponding mRNA targets from their transcription startpoints to 100 nt downstream of the start codons. The lengths of the sRNAs and mRNAs varied from 52 nt to 769 nt. Previous studies using RNAup and IntaRNA with a smaller set of sRNA-mRNA pairs have shown that equilibrium thermodynamics can predict sRNA-mRNA interaction sites with good accuracy (*25*). In the current calculations we employed RNAup with an updated energy parameter set (*29*) to determine if this parameter set improves predictive accuracy compared with previously employed parameters (*30*). We also explored the utility of this empirical thermodynamic approach to provide insight on the

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

effect of RNA sequence changes on gene expression, and the role of Hfq on sRNA-mRNA interactions.

Table I compares the predicted and reported sRNA-mRNA interaction sites. The interaction sites - designated by two positions- x,y – that define the ends of the mRNA region hybridized by a segment of the sRNA. The positions are referenced to the first nucleotide of the mRNA start codon defined as +1. Positions upstream from +1 are numbered -1,-2,-3,… while positions downstream are numbered +2,+3, +4,….. Thus -4,+5 refers to the 9 nt region of sodB mRNA perfectly or imperfectly paired by a segment of RyhB. Evidence supporting the reported sites are given by the references in Table I. Most of the sites were supported by mutational data and/or protection assays such as ribonuclease footprinting. Other sites were based primarily on sequence complementarity between the sRNA and a bounded region of the mRNA. As shown by Kawamoto et al, putative base pairs strongly inferred by sequence complementarity are not always supported by mutation data (*14*). Thus the reported sites may not be precise with regard to their boundaries and the specific base pairs forming the sRNA-mRNA hybrid.

Even with the above caveat and the approximations inherent in the algorithm and parameters, Table I shows excellent overall agreement between the predicted and experimental sites. With one exception, the predicted interaction site overlapped with at least a portion of the reported site. A quantitative assessment of the predictions is given by the sensitivity, the number of correctly predicted base pairs/number of reported base pairs, and the positive predictive value (PPV), the number of correctly predicted base pairs/number of predicted base pairs (*25*). The average PPV value was 0.829 while the sensitivity was 0.856. The 1999 parameter set produced identical predictions for many but not all sRNA-mRNA pairs and gave lower average PPV and sensitivity values (Sensitivity = .753, PPV= 0.767). The updated free energy parameters improve predictions.

The only predicted interaction site that shared no base pairs with the reported site was FnrS and *sodB* mRNA. This exception may result from small errors in the calculated free energies or possibly from the influence of the Hfq protein which is not considered in these calculations. If one changes a single base in FnrS such that a U•G base pair within the predicted site (-14, +4) becomes a G•G mismatch, RNAup now predicts the reported site exactly (+13, +20). This U to G change does not affect the duplex free energy at the correct site and returns a total free energy of -7.05 kcal/mol. This is only 0.1 kcal/mol higher than the value obtained for the correctly predicted site. The experimental interaction site thus appears to be a suboptimal solution slightly higher in total free energy than the predicted site. The location of the predicted site may be sensitive to small errors in parameters or to factors that alter the free energy of accessibility of the sRNA or mRNA.

The Influence of Base Mutations on sRNA-mRNA Binding and Gene Expression

The ability of thermodynamic calculations to explain or help interpret the influence of base mutations on *in vivo* assays of sRNA-mRNA interactions was explored. A comparison of this nature assumes that base changes in the sRNA or mRNA only influence the binding of the sRNA to its mRNA target and does

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

not affect nuclease digestion of the RNAs, Hfq binding, ribosome binding or other factors. One also assumes that expression of the gene regulated by sRNA-mRNA binding is proportional to the probability of sRNA-mRNA formation, P. That is,

$$P \alpha \exp(-\Delta G_{tot})/RT \alpha [\text{protein expressed}] \quad (1a)$$

or,

$$-\Delta G_{tot} \alpha \ln[\text{protein expressed}] \quad (1b)$$

ΔG_{tot} is the total free energy required to form the sRNA-mRNA duplex, R the molar gas constant, T the temperature, and [protein expressed] is the amount of protein expressed in the functional translation assay employed.

A comparison of calculated probabilities of sRNA-mRNA formation as a function of base changes with gene expression data was made using results from three studies. In each work investigators used an *in vivo* translation reporter assay to quantify the effect of base changes to the sRNA and/or mRNA on the expression of the enzyme β-galactosidase (βGz) from the mRNA. A study by Balbontin et al quantified the influence of eight base changes on the interaction of RybB and *ompC* mRNA (*37*). RybB binding to *ompC* mRNA represses translation of this gene. The program RNA Structure was employed to calculate the free energy of duplex formation between the 16 nt sequence at the 5′ end of RybB and the 44 nt sequence at the 5′ end of *ompC* mRNA (Figure 2). We used this approach and the limited sequences in order to focus on the known interaction site and surrounding sequences. Potential effects of the mutations on non-local secondary structure rearrangements that could influence site accessibility were ignored. A plot of the free energy of duplex formation (ΔG_{dup}) vs. ln (βGz) for the data available is shown in Figure 3. The calculated ΔG_{dup} values for the RybB-*ompC* mRNA pair are linearly related to ln(βGz) with a correlation coefficient of $R^2 = 0.71$.

Studies by Majdalani et al evaluated the influence of mutations to DsrA, RprA, and *rpoS* mRNA on the translation regulation of *rpoS* mRNA by these two sRNAs (*11*, *38*). The locations of the single base changes examined in these studies are illustrated in Figure 4. RprA and DsrA bind to the same target region within the 5′ leader sequence of *rpoS* mRNA, which is designated $rpoS_t$, that otherwise forms an intra-molecular duplex with the ribosome binding site or $rpoS_{RBS}$. When DsrA or RprA binds $rpoS_t$ the $rpoS_{RBS}$ sequence is released and translation is enhanced. Unlike the RybB-*ompC* mRNA interaction, accessibility of the *rpoS* mRNA target site was considered since experimental evidence indicates it is a significant factor (*39*). The free energy of the intramolecular $rpoS_t$•$rpoS_{RBS}$ duplex (ΔG_{m1m2}) was calculated using RNA Structure and subtracted from the free energy of forming the sRNA•$rpoS_t$ duplex, ΔG_{sm}. The net or total free energy values ($\Delta G_{sm} - \Delta_{Gm1m2}$) were plotted as a function of ln(βGz) (Figure 5). Linear least squares fit of the RprA and DsrA data gave correlation coefficients of 0.76 and 0.88 respectively. Although it would seem appropriate to employ the RNAup algorithm to calculate the total free energies of RprA or DsrA binding to *rpoS* mRNA, we could not meaningfully employ this approach since the predicted interaction sites did not precisely overlap with the experimental ones (Table I) and did not include the base position of the *rpoS* mRNA mutations.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

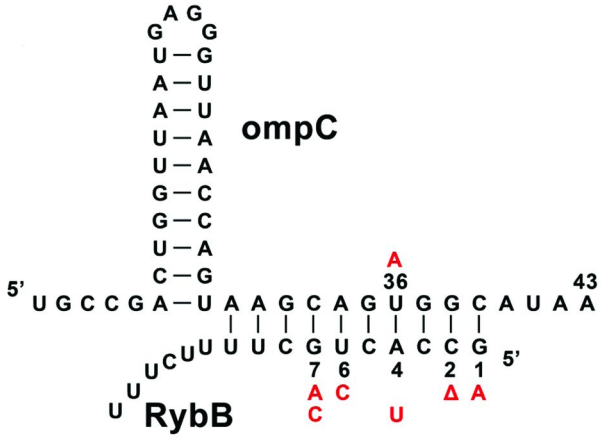


Figure 2. The RybB and ompC mRNA sequences that comprise the interaction site and adjoining bases used in calculating ΔG_{dup}. Single base mutations are shown in red. Table lists sRNA-mRNA pairs for which gene expression data was obtained in ref. (37).

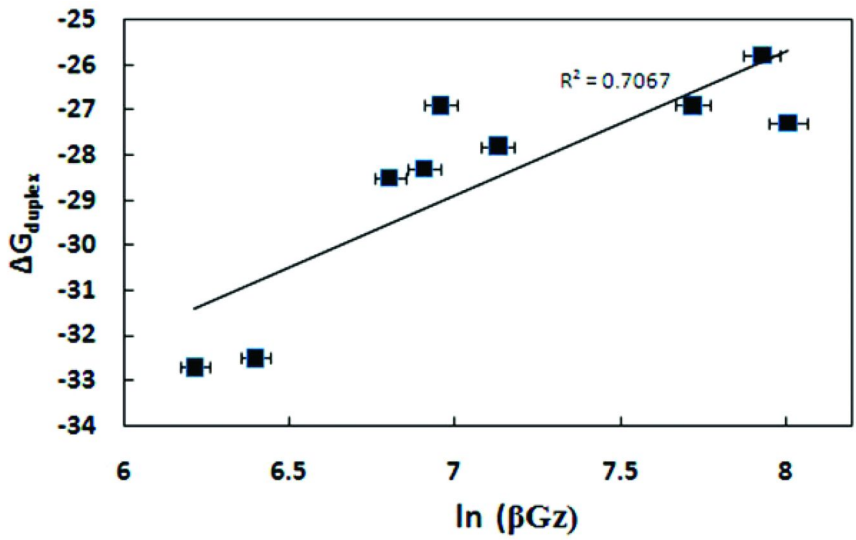


Figure 3. Plot of ΔG_{dup} for RybB-ompC mRNA binding for sequences in Figure 1 vs. ln of β-galactosidase activity measured in ref. (37).

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

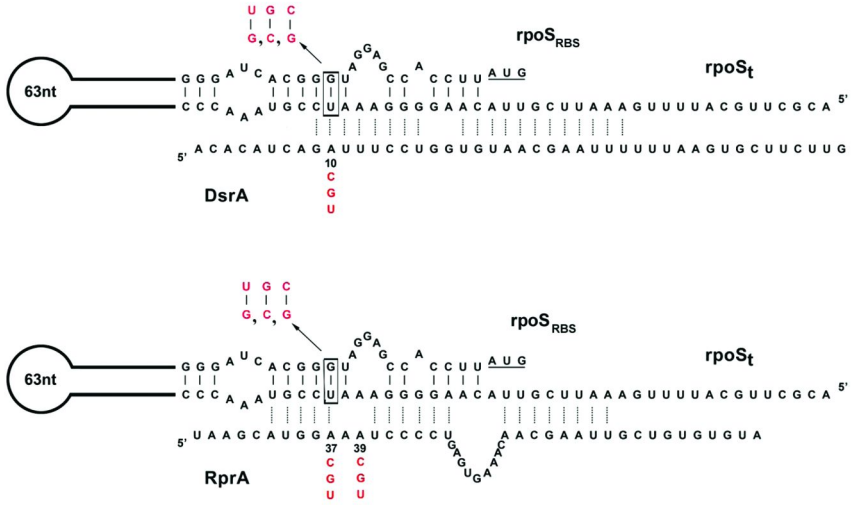


Figure 4. Sequences comprising the DsrA-rpoS mRNA and RprA-rpoS mRNA interaction sites and the intramolecular rpoS$_t$-rpoS$_{RBS}$ duplex. Gene expression data for all combinations of wt and mutant(in red) sequences are in refs. (11, 38).

The correlations observed between the thermodynamic predictions and the binding sites and gene expression data indicate that sRNA-mRNA affinity plays a major role in sRNA function. The calculations do not of course consider the role of Hfq in mediating sRNA gene regulation or other factors such as the impact of mutations on ribosome binding. Many sRNAs require the Hfq protein to function. A previous study proposed that the need for Hfq was linked to the free energy of sRNA-mRNA binding; sRNA-mRNA interactions that were the most stable (lowest in free energy) could dispense with Hfq. Table I does not support this hypothesis. The total free energies of sRNA-mRNA pairs that do not require Hfq are similar to or more positive than the free energies of sRNA-mRNA pairs that require Hfq. However evidence described below indicates that Hfq is a RNA melting protein. This observation supports the notion that Hfq's mechanism for enhancing sRNA-mRNA interaction involves destabilizing one or both RNAs.

The Interaction of DsrA with *rpoS* mRNA

Of the 25 sRNA-mRNA pairs in Table I that require Hfq, three had predicted interaction sites with sensitivity values below 0.6. One of these pairs, DsrA and *rpoS* mRNA, was examined more closely. The predicted target site (-126,-109) was shifted ~9 nt upstream from the experimentally verified mRNA site (-119,-97). The predicted interaction also did not involve bases in the hairpin stem at the 5′ end of DsrA known to base pair with *rpoS* mRNA (Figure 4, Figure 6). A possible explanation for this difference is that the RNAup calculation may have found it energetically too costly to open the 5′ DsrA hairpin to enable the correct interaction. To test this hypothesis, we theoretically examined the effect of single

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

base changes at positions 6 to 9 of DsrA expected to reduce the stability of the 5′ hairpin (Figure 6). Using the four modified DsrA sequences and the full *rpoS* mRNA sequence employed previously, RNAup now predicted the correct target site (-119,-97) and the expected DsrA-*rpoS* mRNA base pairing.

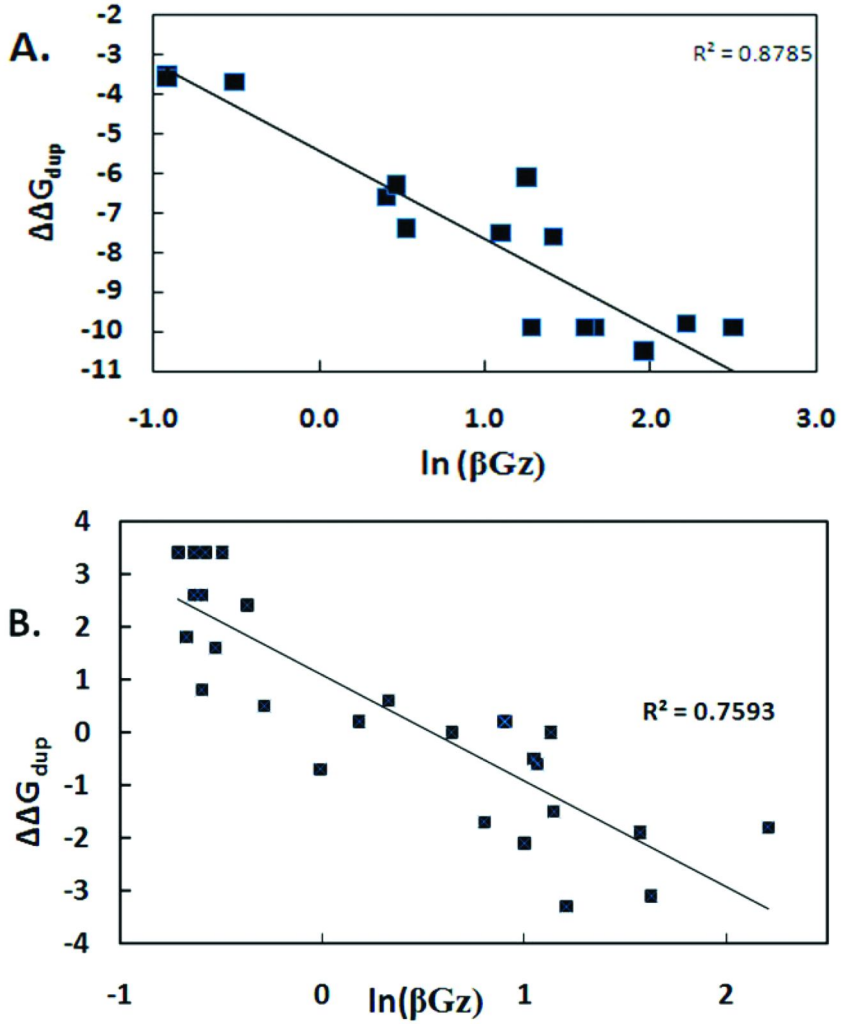


Figure 5. ***A.*** *Plot of ΔG_{dup} for DsrA-rpoS mRNA binding for sequences in Figure 4 vs. ln of β-galactosidase activity measured in ref. (38).* ***B.*** *Plot of ΔG_{dup} for RprA-rpoS mRNA binding for sequences in Figure 4 vs. ln of β-galactosidase activity measured in ref. (11).*

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

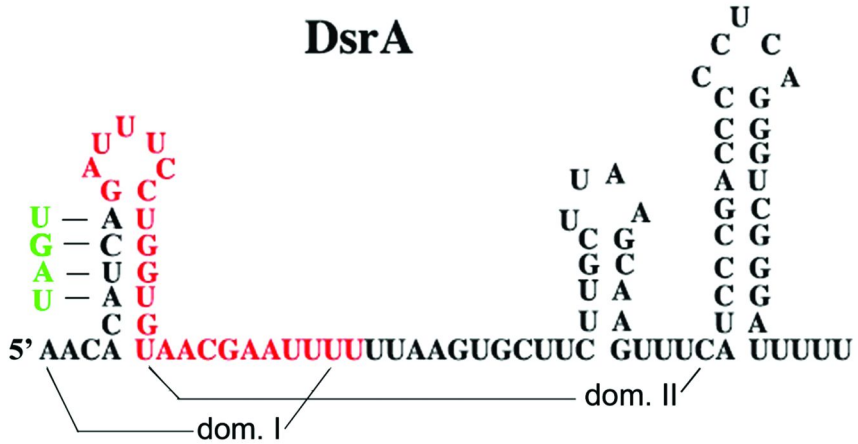


Figure 6. DsrA secondary structure as described in ref. (44). Bases in red bind to rpoS mRNA (Figure 4). The 31 nt domain I sequence and 39 nt domain II sequence are shown. Four single base changes (shown in green) were made in silico to compute the effect of 5′ hairpin stability on the predicted interaction site.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Hfq Is a RNA Melting Protein

The above calculations could simply reflect the approximate knowledge of free energy parameters and their impact on the predicted interaction site for DsrA and *rpoS* mRNA. A more speculative interpretation is that a factor not part of the calculation, such as the Hfq protein, plays a role similar to the simulated destabilization of the 5′ hairpin of DsrA. To explore this experimentally we examined the influence of Hfq on the melting curves of two domains of DsrA. One domain is a 31 nt segment that contains the 5′ hairpin and a portion of the adjoining AU-rich sequence (Figure 6). This is referred to as domain I or $DsrA_{dI}$. Figure 6 also shows a 39 nt sequence of DsrA referred to as domain II or $DsrA_{dII}$. This RNA contains the AU-rich sequence present in $DsrA_{dI}$ and the sequence comprising the middle hairpin. It binds Hfq with high affinity (*40*).

UV absorbance melting curves of $DsrA_{dI}$ in the absence and presence of Hfq are shown in Figure 7. $DsrA_{dI}$ alone exhibited a single sigmoidal transition with a T_m of 62 °C. Addition of Hfq hexamer reduced the melting temperature of the transition. At 0.5:1, 1:1, and 1.5:1 molar ratios of Hfq to $DsrA_{dI}$ the T_m decreased by 1.5 °C, 8 °C, and 11 °C respectively. The addition of Hfq also increased the hyperchromicity of the $DsrA_{dI}$ transition. This implies that Hfq bound to the single stranded form of $DsrA_{dI}$ decreases base stacking relative to the RNA's natural denatured form. At the 1.5:1 ratio of Hfq to $DsrA_{dI}$ a small absorbance increase is observed with a midpoint of T_m ~ 73 °C. A melting profile of 1.5 μM Hfq alone indicates that this transition corresponds to the protein (data not shown). The absorbance of the protein alone (0.025) showed no absorbance change until it increased by ~0.012 between 71 and 75 °C.

Figure 8 shows the melting curve of $DsrA_{dII}$ in the absence and presence of Hfq. The RNA alone exhibits a biphasic transition with T_m's of 26 °C and 63.5 °C. Previous secondary structure predictions and chemical and nuclease protection experiments at 25 °C argued for a secondary structure model of DsrA differing from the one commonly displayed in Figure 6 (*41*). In this alternate model most of the AU rich sequence becomes part of the middle hairpin, extending the duplex stem of this hairpin with A•U and G•U base pairs. The two phase transition of $DsrA_{dII}$ is consistent with the notion that a duplex segment at the end of the middle hairpin melts prior to its more stable upper portion. Addition of Hfq at a 1:1 molar ratio to $DsrA_{dII}$ abolished the first transition completely and reduces the T_m of the second transition by ~9 °C. Thus Hfq can also destabilizes this portion of DsrA.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

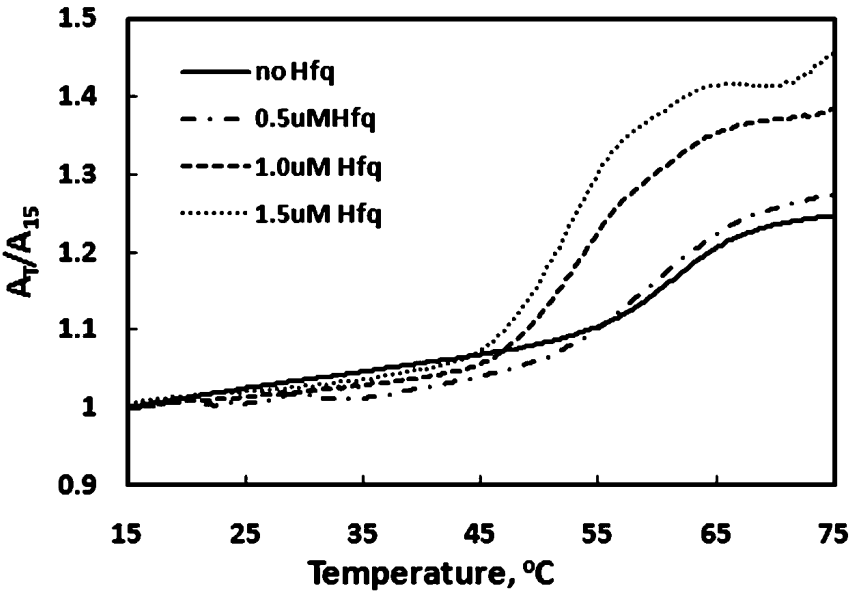


Figure 7. Melting curves of 1 μM $DsrA_{dI}$ in the absence and presence of three Hfq concentrations. Absorbance at 260 nm was normalized to the initial value at 15 °C.

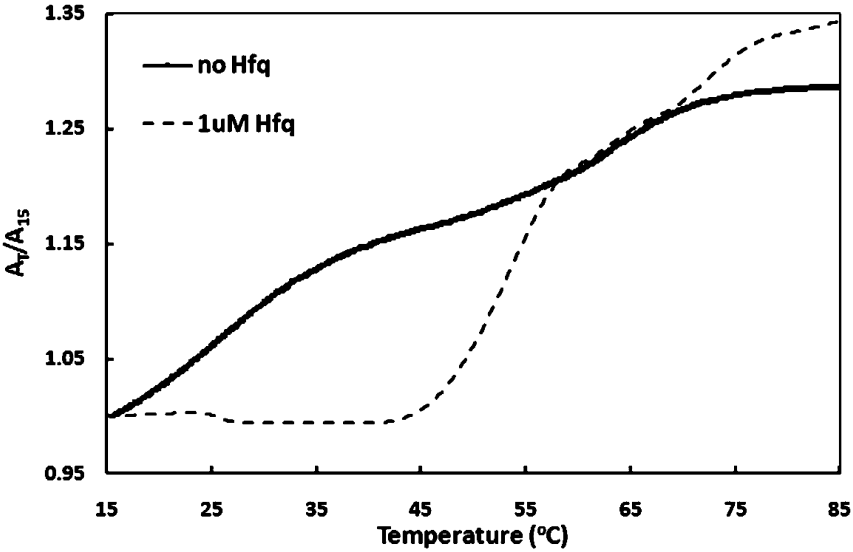


Figure 8. Melting curves of 1 μM $DsrA_{dII}$ alone and in the presence of 1 μM Hfq. Absorbance at 260 nm was normalized to the initial value at 15 °C.

Discussion

The good agreement between the thermodynamically predicted and reported interaction sites of the sRNAs confirms the utility of the RNAup algorithm to help identify mRNA targets of sRNAs. Application of the most current thermodynamic parameter set improved the accuracy of the predicted binding sites compared to previously employed parameters. The results also indicate that thermodynamics of RNA-RNA binding alone is the dominant factor in determining sRNA-mRNA interaction sites. Only three of the 25 sRNA-mRNA pairs that require Hfq gave sensitivity values below 0.6 and two of the inaccurate predictions became highly accurate predictions (precisely aligning with reported sites) with a change to one base in the sRNA. The correlation between the effect of base change on RNA binding and gene expression also supports the importance of sRNA-mRNA binding affinity to biological function.

A correlation between the predicted free energies to form sRNA-mRNA complexes and the requirement or dispensability of Hfq was not observed. This comparison was limited by the relatively few sRNA-mRNA pairs examined that do not require Hfq. We also excluded sRNA-mRNA interactions involving RNAIII from *S. aureus* since some of its pairing interactions involve secondary structures that can not be handled by RNAup, and its length (~515 nt) is much longer than the other sRNAs. A recent algorithm has been developed that modify the partition function approach used by RNAup and allows for RNA-RNA interaction sites that may include intramolecular structures within each RNA (*27*) . The approach improves computational performance and should enable predictions for the class of interactions exemplified by OxyS and *fhlA* or RNaseIII and its

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

target mRNAs which have intramolecular structures with their intermolecular binding site.

Figures 6 and 7 show that Hfq is a RNA melting protein, i.e. it binds with greater affinity to the single stranded conformation than the duplex form of the RNA hairpins. Hfq completely unwinds the segment of $DsrA_{dII}$ that is marginally stable at room temperature. This segment very likely involves the AU-rich sequence that base pairs to *rpoS* mRNA. If this interpretation is correct, DsrA binding to its *rpoS* mRNA target site should be more dependent on Hfq at low temperatures.

For $DsrA_{dI}$ higher temperatures are needed to observe Hfq's ability to denature this more stable hairpin. The absorbance spectrum of $DsrA_{dI}$ does not change significantly when Hfq is added at room temperature (data not shown). This implies that under this condition Hfq binding does not have a major effect on this RNA's average conformation. It does not exclude the possibility that Hfq increases the lifetime of base pair open states. Indeed the observation that Hfq is a 'melting protein' is consistent with the notion that Hfq can trap bases or backbone segments that transiently fluctuate open from a duplex, characteristics proposed for the classical model of a DNA melting protein, Bovine pancreatic ribonuclease (*42*). On the timescale of Hfq forming a ternary complex with two RNAs (*43*) , an increase in base pair opening rate of the 5′ hairpin of DsrA could enhance nucleation events.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Acknowledgments

We gratefully acknowledge the support of a Georgia Tech undergraduate research award to Kanav Jain, a GAANN Fellowship to Taylor Updegrove and funding from the NASA Astrobiology Institute.

References

1. Franze de Fernandez, M. T.; Eoyang, L.; August, J. T. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **1968**, *219*, 588–590.
2. Tsui, H. C.; Leung, H. C.; Winkler, M. E. Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.* **1994**, *13*, 35–49.
3. Sun, X.; Zhulin, I.; Wartell, R. M. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic. Acids Res.* **2002**, *30*, 3662–3671.
4. Zhang, A.; Wassarman, K. M.; Ortega, J.; Steven, A. C.; Storz, G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell* **2002**, *9*, 11–22.
5. Moller, T.; Franch, T.; Hojrup, P.; Keene, D. R.; Bachinger, H. P.; Brennan, R. G.; Valentin-Hansen, P. Hfq: A bacterial Sm-like protein that mediates RNA−RNA interaction. *Mol. Cell* **2002**, *9*, 23–30.
6. Nikulin, A.; Stolboushkina, E.; Perederina, A.; Vassilieva, I.; Blaesi, U.; Moll, I.; Kachalova, G.; Yokoyama, S.; Vassylyev, D.; Garber, M.;

Nikonov, S. Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 141–146.
7. Schumacher, M. A.; Pearson, R. F.; Moller, T.; Valentin-Hansen, P.; Brennan, R. G. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *Embo J* **2002**, *21*, 3546–3556.
8. Sauter, C.; Basquin, J.; Suck, D. Sm-like proteins in Eubacteria: The crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic. Acids Res.* **2003**, *31*, 4091–4098.
9. Sledjeski, D. D.; Whitman, C.; Zhang, A. Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.* **2001**, *183*, 1997–2005.
10. Zhang, A.; Altuvia, S.; Tiwari, A.; Argaman, L.; Hengge-Aronis, R.; Storz, G. The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J.* **1998**, *17*, 6061–6068.
11. Majdalani, N.; Hernandez, D.; Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol. Microbiol.* **2002**, *46*, 813–826.
12. Moller, T.; Franch, T.; Udesen, C.; Gerdes, K.; Valentin-Hansen, P. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.* **2002**, *16*, 1696–1706.
13. Geissmann, T. A.; Touati, D. Hfq, a new chaperoning role: Binding to messenger RNA determines access for small RNA regulator. *EMBO J.* **2004**, *23*, 396–405.
14. Kawamoto, H.; Koide, Y.; Morita, T.; Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **2006**, *61*, 1013–1022.
15. Antal, M.; Bordeau, V.; Douchin, V.; Felden, B. A small bacterial RNA regulates a putative ABC transporter. *J. Biol. Chem.* **2005**, *280*, 7901–7908.
16. Bossi, L.; Figueroa-Bossi, N. A small RNA downregulates LamB maltoporin in Salmonella. *Mol. Microbiol.* **2007**, *65*, 799–810.
17. Lenz, D. H.; Mok, K. C.; Lilley, B. N.; Kulkarni, R. V.; Wingreen, N. S.; Bassler, B. L. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **2004**, *118*, 69–82.
18. Sharma, C. M.; Vogel, J. Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr. Opin. Microbiol.* **2009**, *12*, 536–546.
19. Moll, I.; Leitsch, D.; Steinhauser, T.; Blasi, U. RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep.* **2003**, *4*, 284–289.
20. Lease, R. A.; Woodson, S. A. Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J. Mol. Biol.* **2004**, *344*, 1211–1223.
21. Soper, T. J.; Woodson, S. A. The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **2008**, *14*, 1907–1917.
22. Updegrove, T.; Wilf, N.; Sun, X.; Wartell, R. M. Effect of Hfq on RprA-rpoS mRNA pairing: Hfq-RNA binding and the influence of the 5' rpoS mRNA leader region. *Biochemistry* **2008**, *47*, 11184–11195.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

23. Jousselin, A.; Metzinger, L.; Felden, B. On the facultative requirement of the bacterial RNA chaperone, Hfq. *Trends Microbiol.* **2009**, *17*, 399–405.
24. Muckstein, U.; Tafer, H.; Hackermuller, J.; Bernhart, S. H.; Stadler, P. F.; Hofacker, I. L. Thermodynamics of RNA-RNA binding. *Bioinformatics* **2006**, *22*, 1177–1182.
25. Busch, A.; Richter, A. S.; Backofen, R. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **2008**, *24*, 2849–2856.
26. Muckstein, U. Bioinformatics Research and Development: Second International Conference, BIRD 2008 Vienna, Austria, July 7−9, 2008 proceedings, Vol. 13, 1st ed., Springer, New York, 2008.
27. Huang, F. W.; Qin, J.; Reidys, C. M.; Stadler, P. F. Target prediction and a statistical sampling algorithm for RNA−RNA interaction. *Bioinformatics* **2010**, *26*, 175–181.
28. Gruber, A. R.; Lorenz, R.; Bernhart, S. H.; Neubock, R.; Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **2008**, *36*, W70–74.
29. Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7287–7292.
30. Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940.
31. Keseler, I. M.; Collado-Vides, J.; Santos-Zavaleta, A.; Peralta-Gil, M.; Gama-Castro, S.; Muniz-Rascado, L.; Bonavides-Martinez, C.; Paley, S.; Krummenacker, M.; Altman, T.; Kaipa, P.; Spaulding, A.; Pacheco, J.; Latendresse, M.; Fulcher, C.; Sarker, M.; Shearer, A. G.; Mackie, A.; Paulsen, I.; Gunsalus, R. P.; Karp, P. D. EcoCyc: A comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.* **2011**, *39*, D583–590.
32. Cao, Y.; Wu, J.; Liu, Q.; Zhao, Y.; Ying, X.; Cha, L.; Wang, L.; Li, W. sRNATarBase: A comprehensive database of bacterial sRNA targets verified by experiments. *RNA* **2010**, *16*, 2051–2057.
33. Huang, H. Y.; Chang, H. Y.; Chou, C. H.; Tseng, C. P.; Ho, S. Y.; Yang, C. D.; Ju, Y. W.; Huang, H. D. sRNAMap: Genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Res.* **2009**, *37*, D150–154.
34. Reuter, J. S.; Mathews, D. H. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinf.* **2010**, *11*, 129.
35. Sun, X.; Wartell, R. M. Escherichia coli Hfq binds A18 and DsrA domain II with similar 2:1 Hfq6/RNA stoichiometry using different surface sites. *Biochemistry* **2006**, *45*, 4875–4887.
36. Updegrove, T. B.; Correia, J. J.; Chen, Y.; Terry, C.; Wartell, R. M. The stoichiometry of the *Escherichia coli* Hfq protein bound to RNA. *RNA* **2011**.
37. Balbontin, R.; Fiorini, F.; Figueroa-Bossi, N.; Casadesus, J.; Bossi, L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol Microbiol* **2010**, *78*, 380–394.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

38. Majdalani, N.; Cunning, C.; Sledjeski, D.; Elliott, T.; Gottesman, S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12462–12467.
39. Brown, L.; Elliott, T. Mutations that increase expression of the rpoS gene and decrease its dependence on hfq function in *Salmonella typhimurium*. *J Bacteriol* **1997**, *179*, 656–662.
40. Brescia, C. C.; Mikulecky, P. J.; Feig, A. L.; Sledjeski, D. D. Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA* **2003**, *9*, 33–43.
41. Rolle, K.; Zywicki, M.; Wyszko, E.; Barciszewska, M. Z.; Barciszewski, J. Evaluation of the dynamic structure of DsrA RNA from *E. coli* and its functional consequences. *J Biochem.* **2006**, *139*, 431–438.
42. Jensen, D. E.; von Hippel, P. H. DNA "melting" proteins. I. Effects of bovine pancreatic ribonuclease binding on the conformation and stability of DNA. *J. Biol. Chem.* **1976**, *251*, 7198–7214.
43. Hopkins, J. F.; Panja, S.; Woodson, S. A. Rapid binding and release of Hfq from ternary complexes during RNA annealing. *Nucleic Acids Res.* **2011**.
44. Lease, R. A.; Cusick, M. E.; Belfort, M. Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12456–12461.
45. Chen, S.; Zhang, A.; Blyn, L. B.; Storz, G. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J. Bacteriol.* **2004**, *186*, 6689–6697.
46. Sharma, C. M.; Darfeuille, F.; Plantinga, T. H.; Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* **2007**, *21*, 2804–2817.
47. Majdalani, N.; Vanderpool, C. K.; Gottesman, S. Bacterial small RNA regulators. *Crit. Rev. Biochem. Mol. Biol.* **2005**, *40*, 93–113.
48. Bouvier, M.; Sharma, C. M.; Mika, F.; Nierhaus, K. H.; Vogel, J. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell* **2008**, *32*, 827–837.
49. Johansen, J.; Eriksen, M.; Kallipolitis, B.; Valentin-Hansen, P. Down-regulation of outer membrane proteins by noncoding RNAs: Unraveling the cAMP-CRP- and sigmaE-dependent CyaR-ompX regulatory case. *J. Mol. Biol.* **2008**, *383*, 1–9.
50. De Lay, N.; Gottesman, S. The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol.* **2009**, *191*, 461–476.
51. Durand, S.; Storz, G. Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol. Microbiol.* **2010**, *75*, 1215–1231.
52. Holmqvist, E.; Reimegard, J.; Sterk, M.; Grantcharova, N.; Romling, U.; Wagner, E. G. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.* **2010**, *29*, 1840–1850.
53. Rasmussen, A. A.; Eriksen, M.; Gilany, K.; Udesen, C.; Franch, T.; Petersen, C.; Valentin-Hansen, P. Regulation of ompA mRNA stability: The

role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.* **2005**, *58*, 1421–1429.
54. Masse, E.; Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 4620–4625.
55. Vecerek, B.; Moll, I.; Blasi, U. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. *EMBO J.* **2007**, *26*, 965–975.
56. Svenningsen, S. L.; Tu, K. C.; Bassler, B. L. Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *EMBO J.* **2009**, *28*, 429–439.
57. Tetart, F.; Bouche, J. P. Regulation of the expression of the cell-cycle gene ftsZ by DicF antisense RNA. Division does not require a fixed number of FtsZ molecules. *Mol. Microbiol.* **1992**, *6*, 615–620.
58. Johansen, J.; Rasmussen, A. A.; Overgaard, M.; Valentin-Hansen, P. Conserved small non-coding RNAs that belong to the sigmaE regulon: Role in down-regulation of outer membrane proteins. *J. Mol. Biol.* **2006**, *364*, 1–8.
59. Gaballa, A.; Antelmann, H.; Aguilar, C.; Khakh, S. K.; Song, K. B.; Smaldone, G. T.; Helmann, J. D. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11927–11932.
60. Darfeuille, F.; Unoson, C.; Vogel, J.; Wagner, E. G. An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell* **2007**, *26*, 381–392.
61. Song, T.; Mika, F.; Lindmark, B.; Liu, Z.; Schild, S.; Bishop, A.; Zhu, J.; Camilli, A.; Johansson, J.; Vogel, J.; Wai, S. N. A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Mol. Microbiol.* **2008**, *70*, 100–111.
62. Richter, A. S.; Schleberger, C.; Backofen, R.; Steglich, C. Seed-based INTARNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. *Bioinformatics* **2010**, *26*, 1–5.

Downloaded by PENNSYLVANIA STATE UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch007

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

Chapter 8

Protein: Feeling the Groove of DNA

Derrick Watkins,[1] Lydia-Ann Harris,[2] Gerald B. Koudelka,[2] and Loren Dean Williams*,[1]

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia, 30332-0400
[2]Department of Biological Sciences, 607 Cooke Hall, State University of New York at Buffalo, Buffalo, NY 14260
***E-mail: loren.williams@chemistry.gatech.edu**

B′-form DNA differs from standard B-form DNA by a more narrow minor groove, which is filled in the B′-form by a highly ordered and geometric array of water molecules. A-tracts (sequences of contiguous 5′ApA3′ and/or ApT steps without TpA steps or C·G base pairs) have the greatest propensity to adopt the B′-form. We show that binding of a protein (P22 c2 repressor N-terminal domain) to DNA induces the B to B′ transition of the central non-contacted region of the operator, regardless of the DNA sequence. P22R NTD provides an experimental system for studying B′-DNA of any sequence, even those with C·G base pairs. One can determine how hydration and distributions of ions of B′-DNA change with sequence, while holding DNA conformation fixed. Here we describe three-dimensional structures of P22R NTD-DNA complexes containing 5′ApT3′, TpA, GpC or CpG at the center of the non-contacted region of the DNA operator. All these dinucleotide steps are in the B′-form, with a narrow minor grove and a spine of hydration. Many features of the water interactions within the minor groove are conserved for the sequences ApT, TpA and GpC. The primary water molecules, at the base of the hydration motif, are highly restrained in translation and rotation. These water molecules are in discrete sites, interacting exclusively with hydrogen bond acceptors in optimally spaced binding pockets formed by DNA functional groups on the floor and walls of the minor groove. The water

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

interactions at the CpG step differ fundamentally from those of ApT, TpA or GpC. At the CpG step, the primary water molecule interacts with a mixture of hydrogen bond donors and acceptors at closely spaced and overlapping binding sites. The translational and rotational restraints on the primary water molecules are released by the 2-amino groups of CpG step, breaking the entire hydration motif. The work presented in this chapter suggests that stability of DNA complexes with minor groove ligands is modulated in a structurally comprehensible way by the entropy of release of water molecules from the minor groove.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

Introduction

Double-stranded DNA can adopt a variety of conformations (*1*) including linear and bent, and A, B and B′ [reviewed by Hud (*2*)] and Z [reviewed by Rich (*3*)]. The propensity to adopt a given conformation is influenced by sequence and environment.

B-DNA

Rosalind Franklin collected two distinct diffraction patterns from DNA fibers, which she reversibly inter-converted by changing the level of hydration. Without knowledge of the 3D structures, she called the highly hydrated form the "B structure" and the less hydrated form the "A Structure" (*4*). We now know that the vast majority of genomic DNA, with sequences containing a mix of A·T and C·G base pairs, is found in the B-form. In B-DNA, the bases stack like pennies, with their normals in line with the helical axis (Figure 1A). The phosphodiester backbones are on the outside of the duplex where they form the walls of two grooves. The edges of the base pairs provide the floors of the grooves. In B-form DNA, the major groove is deep and wide. The minor groove is deep and narrow, but there is sufficient space there to allow two strings of water molecules to combine to form a "ribbon of hydration" along the floor of the groove (*5*). B-form DNA is characterized by significant frequency of cations in the major groove, adjacent to the N-7 and the O-6 atoms of guanines (*6*).

B′-DNA

The B′-form of DNA (*7*), shown in Figure 1B, is a variant of B-form DNA. It was first described by Arnott (*8*), and has been shown by Tullius and coworkers (*9–11*) to form spontaneously in A-tracts, which contain contiguous $^{5'}$ApA$^{3'}$ and/or ApT steps. TpA steps provide a barrier to formation of the B′-form. That barrier can be overcome by environment: the B′-form can be induced by proteins (*12*, *13*), crystal lattice effects (*14*) and some minor groove binders (*15*). C·G base pairs appear to provide the largest barrier to the transition from B- to B′-form. Only in complexes of certain proteins (*12*, *13*) are C·G base pairs observed in B′-form.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

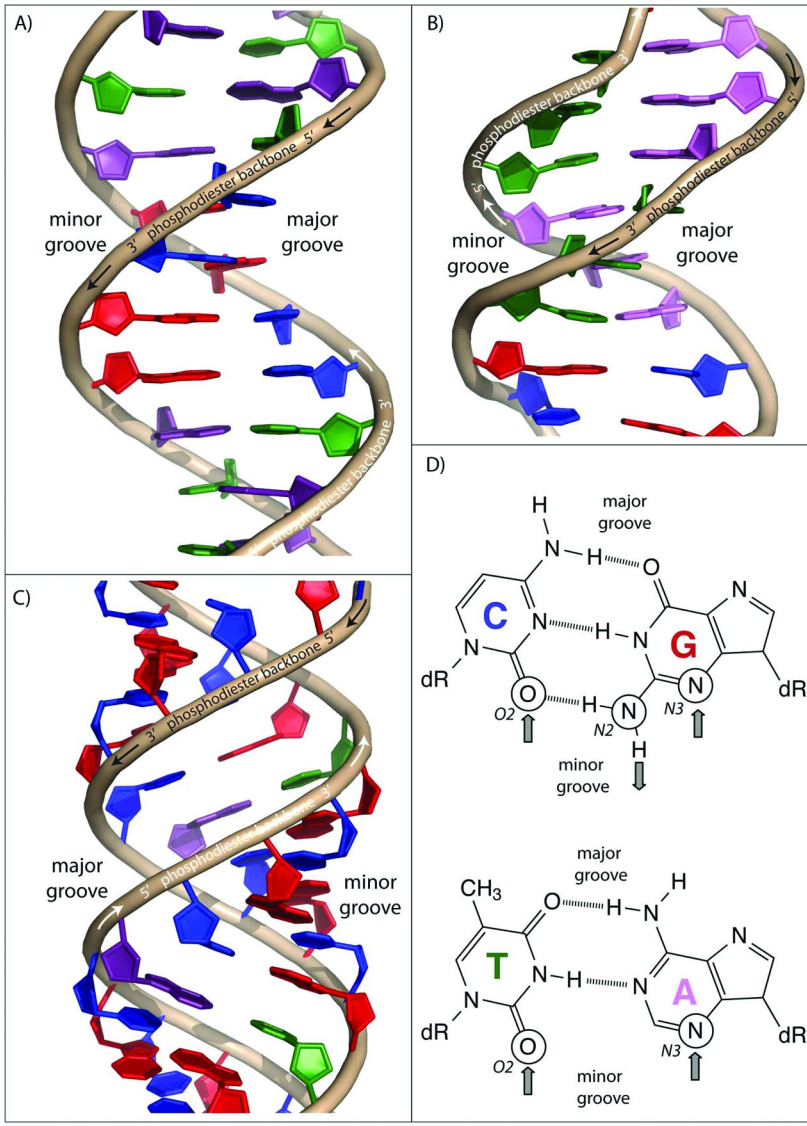


Figure 1. A) B-form DNA from idealized fiber coordinates. B) B' DNA from the central region of the single crystal structure of [d(CGCAAATTTGCG)]$_2$. (PDB entry 1S2R) C) A-form DNA from idealized fiber coordinates. D) Watson-Crick base pairs with the major and minor grooves indicated, along with the hydrogen bonding functionalities on the floor of the minor groove. The up arrows indicate hydrogen bond acceptors, and the down arrows in indicated hydrogen bond donors.

The B′-form differs from standard B-form DNA by (i) a narrow minor groove (*9*, *16*), (ii) a one-layer thick sheet of water molecules containing a zig-zag spine of hydration (*17*, *18*) as a foundation for higher-level minor groove assemblies (*19–21*), (iii) a propensity to bend DNA when appropriately phased (*22–24*), (iv) monovalent cations within the minor groove and a depletion of cations from the major groove (*20*, *25–28*), (v) negative propeller twisting of base pairs (*29*, *30*), (vi) unusual rigidity (*31*, *32*), and (vii) a 'pre-melting' thermal transition that is distinct from duplex melting (*33–35*). The B′-form helical axis is thought to be linear (*36*, *37*). Axial bends arise at the junctions between B′-form and flanking B-form DNA (*2*). The hydration within the B-form minor groove appears to be monolithic and cooperative, with an extended and interdependent hydration assembly (*13*).

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

A-DNA

DNA is converted from B-form to A-form by partial dehydration, or by addition of a 2′-hydroxyl group. Duplex RNA is found exclusively in A-form. DNA base composition and sequence is an important influence of the B/A balance. GC rich DNA most readily converts from B- to A-form. In the A-form, the bases are pushed away from the helical axis and are inclined relative to the penny-like bases of B-form DNA (Figure 1C). The DNA is compressed along the helical axis. The phosphodiester backbones are pulled together across the major groove, which is very deep and very narrow. In comparison with B-DNA, the major grove of A-DNA contains greater populations of cations, which are more extensively localized (*38*).

Protein Recognition and Deformation of DNA

The binding of proteins to specific sequences of DNA is a basic requirement for biological transactions ranging from gene expression, to DNA restriction and modification, to regulating chromatin structure. Proteins use both direct and indirect readout to read DNA sequence. In direct readout, DNA bases make contacts with protein sidechains and backbone atoms. Direct readout can employ hydrogen bonding (*39*) and/or shape complementarity (*12*).

In indirect readout, DNA sequence is sensed at a distance by its propensity to assume or to resist various altered conformations. The propensity for DNA to alter its state is modulated by sequence. Proteins exploit sequence-dependent DNA conformational polymorphism and deformability in recognizing sequence. Previously the role of DNA conformation and deformability in protein recognition was put in a quantitative framework by Olson and Zhurkin (*40*). Williams has described how DNA electrostatic potential and charged species can alter DNA conformation, especially minor groove width (*41*). Koudelka and Carlson (*42*) and Crothers and Zhakked (*43*) experimentally determined relationships between DNA flexibility and protein affinity. Honig (*44*) has proposed that electrostatic potential and minor groove width are recognized by proteins.

We have determined a series of high-resolution crystal structures to help understand direct and indirect readout of DNA sequence by helix-turn-helix proteins. The results indicate a mechanism of sequence recognition in which the affinity of a given protein for DNA is attenuated by the sequence-dependence of the free energy required to induce a specific type of DNA deformation (*13*, *14*). DNA sequence can be 'read' by the ease with which it can be converted from the B- to the B′-form. Structures of complexes of the P22 c2 repressor N-terminal domain (P22R NTD) with DNA operators (*13*, *14*) confirm that direct readout of helix-turn-helix proteins involves interactions of recognition helices with the floor of the major groove (Figure 2). P22R NTD directly reads DNA sequence primarily by shape complementarity rather than hydrogen bonding interactions. Shape complementarity between the DNA and the P22R NTD recognition helix is consistent with effects of mutations on P22R NTD - DNA complex stability (*40*, *41*).

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

A Molecular-Level Model of Indirect Readout

Solution experiments demonstrated P22R NTD indirectly reads the sequence of the non-contacted bases at the center of its operator (*45–48*). These non-contacted bases are indicated in Figure 2. The sequence of the central non-contacted bases modulates the affinity of P22R NTD for DNA. The 3D structures suggest that P22R NTD induces a transition of the non-contacted region from the B-form to the B′-form (*12*, *13*). In the P22R NTD – DNA complex, regardless of the sequence, the minor groove of the non-contacted region is narrow, and is hydrated by an zig-zag spine of hydration. P22R NTD clearly induces a B- to B′-form transition of the non-contacted region of the DNA, no matter the sequence. The differential effects of DNA sequence on the barrier to this transition allow the protein to sense the non-contacted DNA sequence.

Exploring the Sequence Landscape of B′-Form DNA

P22R NTD provides an experiment system for studying the B′ form of DNA for any sequence. Within the central region of the DNA operator (Figure 2), binding by P22R NTD will force any sequence, even sequences containing C·G base pairs, into B′-form. The DNA conformation is held fixed by the protein, allowing one to characterize the hydration and cation interactions of a variety of sequences in a conformationally well-controlled system. We have determined x-ray structures of the P22R NTD in complexes with a series of synthetic operators. These complexes are the same except that they contain ApT, TpA, GpC or CpG at the center of the non-contacted region of the DNA operator (indicated by the highlighted base pairs in Figure 2). DNAs containing these sequences vary in their affinity for P22R NTD.

Nonetheless, they all contain B′-form DNA, with a narrow minor grove and a spine of hydration. The similarities in the groove widths and minor groove hydration in this series are observable in Figures 3 and 4. Although the sequences

differ, the water interactions within the minor groove are essentially conserved for ApT, TpA and GpC, which are the three highest affinity sequences.

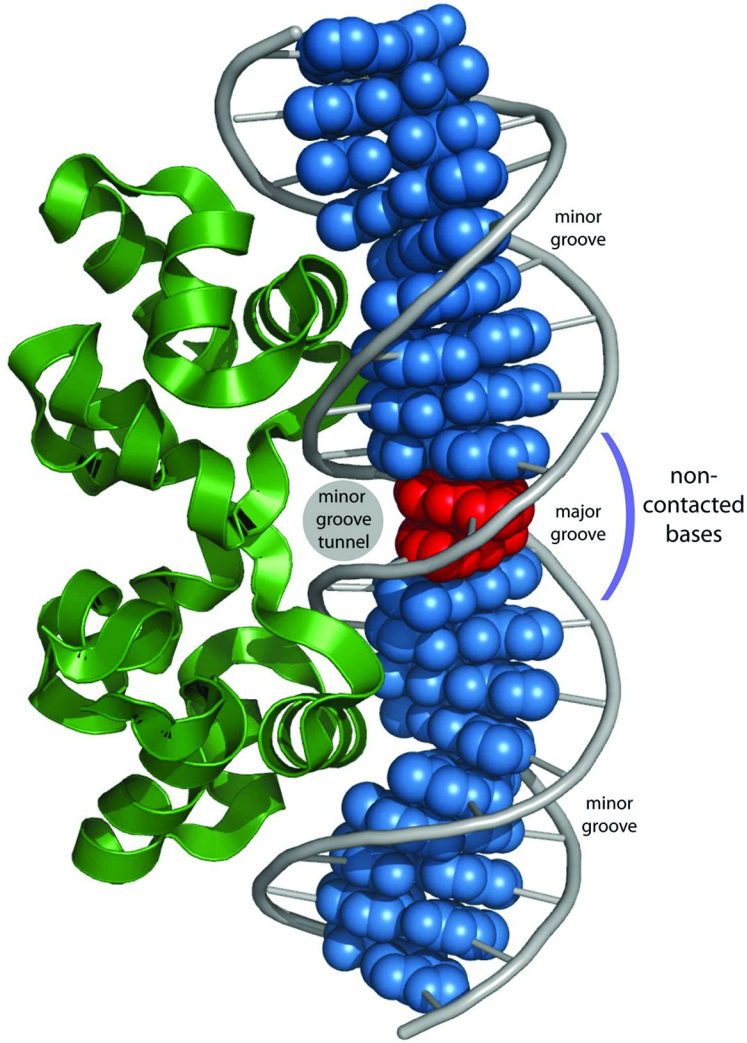


Figure 2. P22R NTD-DNA Complexes. The P22R NTD (ribbon) binds as a dimer to a 20 base pair consensus sequence, making direct contacts with the major groove of the outer bases of the DNA operator. The protein does not contact the central four base pairs (indicated), but forces the DNA into the B'-form, reading the sequence indirectly. The central sequence has a narrow minor groove and a spine of hydration. The dimer interface is located above the minor groove creating a tunnel with the non-contacted region (indicated by the grey circle). The two central DNA bases (shaded) have been varied in the series of structures described here as shown in Figures 3 and 4. The DNA bases are rendered as spheres, and the DNA backbone is rendered as a tube.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

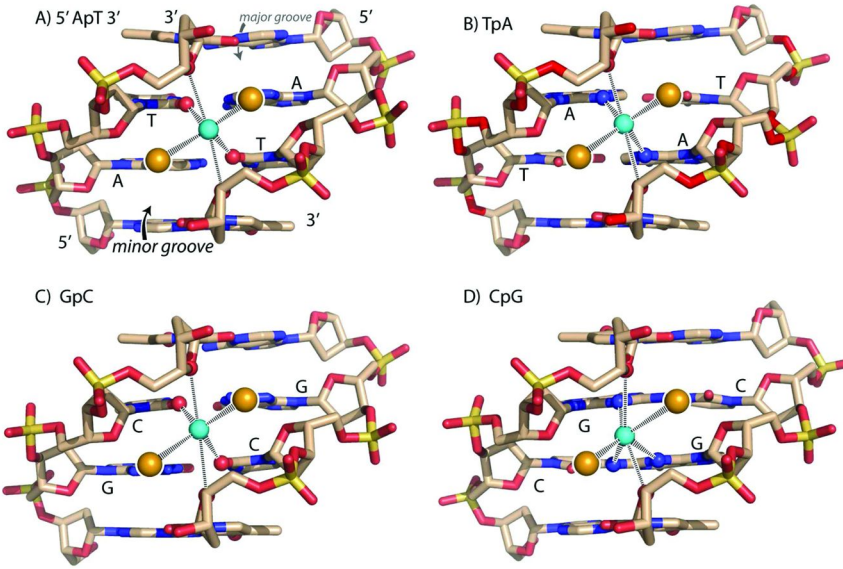


Figure 3. Water Coordination in the B′-form Minor Groove. The primary water of the spine of hydration, interacts with the floor of the minor groove, is hexa-coordinated (indicated by dashed lines) by two base functional groups, by two O4′ atoms of deoxyriboses and by two additional water molecules in the secondary layer. The coordination by the O4′ atoms and the secondary layer water molecules are independent of sequence. The base atoms coordinatingthe waters and their relative dispositions vary with the DNA sequence. Shown are the coordination at an A) ApT step; B) TpA step; C) GpC step; D) CpG step. The coordination geometries are conserved at ApT, TpA and GpC but not CpG.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

ApT steps stabilize B′-form DNA. At an ApT step within B′-DNA, the primary water molecule, at the base of the zig-zag spine of hydration, is highly constrained to a discrete site, forming four hydrogen bonds with the DNA (Figures 3A and 4A) in a well-defined pocket. The DNA uses the O2 atoms of adjacent, cross-strand T's, and two cross-strand O4′ ribose atoms to form the binding pocket for this water molecule (the numbering scheme for the atoms on the floor of the minor groove is indicated in Figure 1D). All four atoms of the DNA are hydrogen bond donors, restraining not only the translation but the rotation of this water molecule. This water molecule cannot freely rotate because it contributes two hydrogens simultaneously to four hydrogen bond acceptors of the DNA (*13*). ApA steps, which also stabilize B′-form DNA, forms the same four-HB acceptor pocket as ApT (*19*), except that one of the O2 atoms is replaced by the N3 of an A. The combined data suggest that hydrogen bonds between O2(T) and water are stronger than hydrogen bonds between N3(A) and water.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

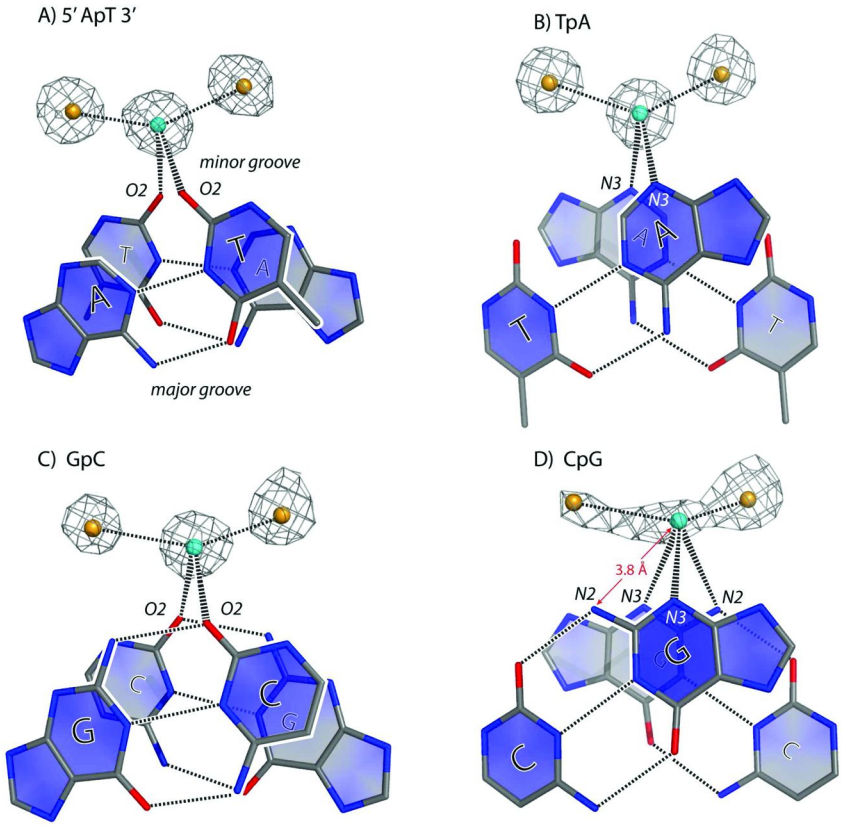


Figure 4. Water Localization in the Minor Groove Tunnel. Electron density (2Fo-Fc, gray mesh) surrounding the primary and secondary water molecules within the B′-form minor groove. Water molecules are highly localized, and are contained within spherical electron density in the A) the ApT step; (B) the TpA step; and the (C) the GpC step. However, the electron density at the D) CpG step shows elongated tube of electron density consistent with delocalization - multiple overlapping positions - for these water molecules. The primary water molecule at the CpG step occupies alternate positions by shifting from one 2-amino group to another. The maps were constructed from single crystal x-ray diffraction data of the P22R NTD-DNA complexes. All structures refined at a resolutions better than 1.8 Å and the 2Fo-Fc maps are contoured at 1 σ. Refinement statistics are R-cryst/R-free; ApT, 18.4/21.4; TpA, 20.4/22.5; GpC, 18.9/22.8; CpG, 18.7/23.6

When DNA is forced into the B′ form, this elaborate hydrogen-bonding scheme can be accomplished by other sequences. An TpA step is considered a B′-breaker in free DNA. However when forced by the protein into the B′-form, the TpA step conserves the hydrogen bonds with the primary water molecule, replacing both O2's of the T's with N3's of A's (Figures 3B and 4B).

Similarly at the GpC step, the hydrogen bonds of the primary water molecule are conserved in B′-form DNA, by replacing the O2's of T's with the O2's of C's

(Figures 3C and 4C). In each case the water molecule is fixed in position and rotation, by simultaneous interactions with four hydrogen bond with acceptors of the DNA (including two O4′ atoms of the deoxyribose). Thus the hydrogen bonding pattern in the minor groove is highly degenerate. The degree of localization of these water molecules can be inferred as high by the well-formed spheres of electron electron density, as shown in the maps in Figures 4A-C.

The hydration at CpG, the weakest P22R NTD binder of the four sites examined, differs fundamentally from in ApT, TpA or GpC. CpG contains two adjacent amino groups on the floor of the minor groove (Figure 3D and 4D). These amino groups extend the binding region, delocalizing the primary water molecule along the minor groove. The delocalizing effects of the 2-amino groups are clearly observable in the electron density map (Figure 4D) and in Koudelka's chemical footprinting results (*13*). The primary water molecule teeter-totters between the two 2-amino groups, forming hydrogen bonds with one 2-amino group, or with the other, but never with both. This water molecule is not fixed at a specific site as in the other three dinucleotides.

At the CpG step, the primary water molecule interacts with a mixture of hydrogen bond acceptors (N3 and O4′) and acceptors (2-amino groups). Therefore the rotational restraints on this water molecule are released at the CpG step. The delocalization of the primary water molecule disturbs the entire hydration assembly, with water molecules in the secondary layer shifting and rotating in concert with those in the primary layer. Since the hydration structure of B′-DNA is globally interdependent, these 2-amino-groups alter patterns of hydration at proximal sites.

Feeling the Minor Groove

The minor groove is a natural binding cleft that has been exploited by nature and in biotechnology. Natural antibiotics such as distamycin (*49*, *50*), and proteins such as TATA box binding protein (*51*) target the minor groove. Synthetic polyamide hairpins can be programmed to bind in the minor groove of DNA to any sequence (*52*). Guanine destabilizes DNA complexes of many minor groove ligands. It was clear from inspection of DNA-ligand models (*49*) and crystal structures (*53*) that guanine, in the form of the 2-amino group, can afford a steric impediment to binding. The 2-amino group of guanine extends further out from the floor of minor groove than any other functional group of DNA. Our recent work, as described here, suggests another effect of the 2-amino group that could be of equal or even greater thermodynamic importance. The free energy gain from release of water molecules from the minor groove can be tuned by their constraints within the groove. Release of translationally and rotationally constrained water molecules of an A-tract would afford significantly greater entropic gain than release of the translationally and rotationally unconstrained waters associated with CpG steps. Therefor the entropy of association of water molecules within the minor groove, and the entropy of release of those waters to bulk upon ligand or protein binding, appears to be an important element of DNA recognition.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

References

1. Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
2. Hud, N. V.; Plavec, J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **2003**, *69*, 144–158.
3. Rich, A. DNA comes in many forms. *Gene* **1993**, *135*, 99–109.
4. Franklin, R. E.; Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **1953**, *171*, 740–741.
5. Prive, G. G.; Heinemann, U.; Chandrasegaran, S. K.; S., L.; Kopka, M. L.; Dickerson, R. E. Helix geometry, hydration, and GA mismatch in a B-DNA decamer. *Science* **1987**, *238*, 498–504.
6. Howerton, S. B.; Sines, C. C.; VanDerveer, D.; Williams, L. D. Locating Monovalent Cations in the Grooves of B-DNA. *Biochemistry* **2001**, *40*, 10023–10031.
7. Segal, E.; Widom, J. Poly(dA:dT) tracts: Major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **2009**, *19*, 65–71.
8. Arnott, S.; Hukins, D. W. L. Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504–1509.
9. Burkhoff, A. M.; Tullius, T. D. The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell* **1987**, *48*, 935–943.
10. Burkhoff, A. M.; Tullius, T. D. Structural details of an adenine tract that does not cause DNA to bend. *Nature* **1988**, *331*, 455–457.
11. Price, M. A.; Tullius, T. D. How the structure of an adenine tract depends on sequence context: A new model for the structure of T_nA_n DNA Sequences. *Biochemistry* **1993**, *32*, 127–136.
12. Watkins, D.; Hsiao, C.; Woods, K. K.; Koudelka, G. B.; Williams, L. D. P22 C2 repressor-operator complex: Mechanisms of direct and indirect readout. *Biochemistry* **2008**, *47*, 2325–2338.
13. Watkins, D.; Mohan, S.; Koudelka, G. B.; Williams, L. D. Sequence recognition of DNA by protein-induced conformational transitions. *J. Mol. Biol.* **2010**, *396*, 1145–1164.
14. Mack, D. R.; Chiu, T. K.; Dickerson, R. E. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: A comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.* **2001**, *312*, 1037–1049.
15. Tevis, D. S.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D. Large, sequence-dependent effects on DNA conformation by minor groove binding compounds. *Nucleic Acids Res.* **2009**, *37*, 5550–5558.
16. Alexeev, D. G.; Lipanov, A. A.; Skuratovskii, I. Y. Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove. *Nature* **1987**, *325*, 821–823.
17. Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **1983**, *163*, 129–146.
18. Drew, H. R.; Dickerson, R. E. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **1981**, *151*, 535–556.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

19. Shui, X.; Sines, C.; McFail-Isom, L.; VanDerveer, D.; Williams, L. D. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **1998**, *37*, 16877–16887.
20. Tereshko, V.; Minasov, G.; Egli, M. A "hydrat-ion" spine in a B-DNA minor groove. *J. Am. Chem. Soc.* **1999**, *121*, 3590–3595.
21. Woods, K. K.; Maehigashi, T.; Howerton, S. B.; Sines, C. C.; Tannenbaum, S.; Williams, L. D. High-resolution structure of an extended A-tract: $[d(CGCAAATTTGCG)]_2$. *J. Am. Chem. Soc.* **2004**, *126*, 15330–15331.
22. Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci., U.S.A.* **1982**, *79*, 7664–7668.
23. Wu, H.-M.; Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **1984**, *308*, 509–513.
24. Zinkel, S. S.; Crothers, D. M. DNA bend direction by phase sensitive detection. *Nature* **1987**, *328*, 178–181.
25. Woods, K.; McFail-Isom, L.; Sines, C. C.; Howerton, S. B.; Stephens, R. K.; Williams, L. D. Monovalent cations sequester within the A-tract minor groove of $[d(CGCGAATTCGCG)]_2$. *J. Am. Chem. Soc.* **2000**, *122*, 1546–1547.
26. Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **1998**, *37*, 8341–8355.
27. Hud, N. V.; Sklenar, V.; Feigon, J. Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. Mol. Biol.* **1999**, *286*, 651–660.
28. Cesare Marincola, F.; Denisov, V. P.; Halle, B. Competitive Na(+) and Rb(+) binding in the minor groove of DNA. *J. Am. Chem. Soc.* **2004**, *126*, 6739–6750.
29. Wing, R.; Drew, H.; Takano, T.; Broka, C.; Takana, S.; Itakura, K.; Dickerson, R. E. Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980**, *287*, 755–758.
30. Dickerson, R. E.; Drew, H. R. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* **1981**, *149*, 761–786.
31. Rhodes, D. Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucleic Acids Res.* **1979**, *6*, 1805–1816.
32. Simpson, R. T.; Kunzler, P. Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. *Nucleic Acids Res.* **1979**, *6*, 1387–1415.
33. Herrera, J. E.; Chaires, J. B. A premelting conformational transition in poly(dA)-poly(dT) coupled to daunomycin binding. *Biochemistry* **1989**, *28*, 1993–2000.
34. Chan, S. S.; Breslauer, K. J.; Austin, R. H.; Hogan, M. E. Thermodynamics and premelting conformational changes of phased $(dA)_5$ Tracts. *Biochemistry* **1993**, *32*, 11776–11784.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

35. Augustyn, K. E.; Wojtuszewski, K.; Hawkins, M. E.; Knutson, J. R.; Mukerji, I. Examination of the premelting transition of DNA A-tracts using a fluorescent adenosine analogue. *Biochemistry* **2006**, *45*, 5039–5047.
36. Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. The structure of an oligo(dA)•oligo(dT) tract and its biological implications. *Nature* **1987**, *330*, 221–226.
37. Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. A bifurcated hydrogen-bonded conformation in the d(A-T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 8385–8389.
38. Hud, N. V.; Polak, M. DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.* **2001**, *11*, 293–301.
39. Seeman, N. C.; Rosenberg, J. M.; Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804–808.
40. Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11163–11168.
41. McFail-Isom, L.; Sines, C. C.; Williams, L. D. DNA structure: Cations in charge? *Curr. Opin. Struct. Biol.* **1999**, *9*, 298–304.
42. Koudelka, G. B.; Carlson, P. DNA twisting and the effects of non-contacted bases on affinity of 434operator for 434 repressor. *Nature* **1992**, *355*, 89–91.
43. Zhang, Y.; Xi, Z.; Hegde, R. S.; Shakked, Z.; Crothers, D. M. Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8337–8341.
44. Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248–1253.
45. Hilchey, S. P.; Wu, L.; Koudelka, G. B. Recognition of nonconserved bases in the P22 operator by P22 repressor requires specific interactions between repressor and conserved bases. *J. Biol. Chem.* **1997**, *272*, 19898–19905.
46. Hilchey, S. P.; Koudelka, G. B. DNA-based loss of specificity mutations. Effects of DNA sequence on the contacted and non-contacted base preferences of bacteriophage P22 repressor. *J. Biol. Chem.* **1997**, *272*, 1646–1653.
47. Wu, L.; Vertino, A.; Koudelka, G. B. Non-contacted bases affect the affinity of synthetic P22 operators for P22 repressor. *J. Biol. Chem.* **1992**, *267*, 9134–9139.
48. Wu, L.; Koudelka, G. B. Sequence-dependent differences in DNA structure influence the affinity of P22 operator for P22 repressor. *J. Biol. Chem.* **1993**, *268*, 18975–18981.
49. Wartell, R. M.; Larson, J. E.; Wells, R. D. Netropsin: A specific probe for A/T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.* **1974**, *249*, 6719–6731.
50. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 1376–1380.

51. Burley, S. K. The TATA box binding protein. *Curr. Opin. Struct. Biol.* **1996**, *6*, 69–75.
52. Doss, R. M.; Marques, M. A.; Foister, S.; Dervan, P. B. DNA minor-groove recognition by 3-methylthiophene/pyrrole pair. *Chem. Biodivers* **2004**, *1*, 886–899.
53. Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6332–6336.

Downloaded by OHIO STATE UNIV LIBRARIES on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch008

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Chapter 9

Chemistry and Biology of Aflatoxin-DNA Adducts

Michael P. Stone,* Surajit Banerjee, Kyle L. Brown, and Martin Egli

Department of Chemistry, Center in Molecular Toxicology, Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee 37235
***E-mail: michael.p.stone@vanderbilt.edu**

Aspergillus flavus is a fungal contaminant of stored rice, wheat, corn, and other grainstuffs, and peanuts. This is of concern to human health because it produces the mycotoxin aflatoxin B_1 (AFB_1), which is genotoxic and is implicated in the etiology of liver cancer. AFB_1 is oxidized *in vivo* by cytochrome P450 to form aflatoxin B_1 epoxide, which forms an N7-dG adduct (AFB_1-N7-dG) in DNA. The latter rearranges to a formamidopyrimidine (AFB_1-FAPY) derivative that equilibrates between α and β anomers of the deoxyribose. In DNA, both the AFB_1-N7-dG and AFB_1-β-FAPY adducts intercalate above the 5′-face of the damaged guanine. Each produces G→T transversions in *Escherichia coli*, but the AFB_1-β-FAPY adduct is more mutagenic. The *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) provides a model for understanding error-prone bypass of the AFB_1-N7-dG and AFB_1-β-FAPY adducts. It bypasses the AFB_1-N7-dG adduct, but it conducts error-prone replication past the AFB_1-FAPY adduct, including mis-insertion of dATP, consistent with the G→T mutations characteristic of AFB_1 mutagenesis in *E. coli*. Crystallographic analyses of a series of binary and ternary complexes with the Dpo4 polymerase revealed differing orientations of the N7-C8 bond of the AFB_1-N7-dG adduct as compared to the N^5-C8 bond in the AFB_1-β-FAPY adduct, and differential accommodation of the intercalated AFB_1 moieties within the active site. These may modulate AFB_1 lesion bypass by this polymerase.

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Introduction

The fungi *Aspergillus flavus* contaminates stored rice, wheat, corn, and other grainstuffs, and peanuts. This is of concern to human health because it produces the mycotoxin aflatoxin B_1 (AFB_1) (*1–4*). This mycotoxin is among the most genotoxic natural products. It is mutagenic in bacteria (*2*, *5–7*), tumorogenic in fish (*8*, *9*), carcinogenic in rodents (*10*, *11*), and is implicated in the etiology of liver cancer (*4*, *12*, *13*). Aflatoxin exposures are implicated in mutations to the p53 tumor suppressor gene (*14–20*).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Chemistry of AFB_1-Induced Alkylation of DNA

The genotoxicity of AFB_1 is associated with its oxidation to the electrophile AFB_1-*exo*-8,9-epoxide *in vivo* by cytochromes P450 (Chart 1) (*21–25*). The synthesis of AFB_1-*exo*-8,9-epoxide was reported by Baertschi *et al.* (*26*) and involved oxidation of AFB_1 with dimethyldioxirane (*27*, *28*). The availability of this epoxide facilitated investigations with respect to both the chemical and biological consequences of AFB_1-induced DNA alkylation. The short-lived epoxide (*25*) reacted efficiently with DNA to yield the AFB_1 N7-dG adduct *trans*-8,9-dihydro-8-(*N*7-guanyl)-9-hydroxyaflatoxin B_1 (*2*, *29*). This was attributed to its intercalation above the 5′ face of dG in DNA (*30*), facilitating nucleophilic attack by the N7 nitrogen at the C8 carbon of the epoxide (*31*). At acidic pH, the AFB_1-N7-dG adduct depurinated to yield a potentially genotoxic abasic site. In contrast, hydrolysis of the guanine imidazole ring formed the AFB_1 formamidopyrimidine adduct (AFB_1-FAPY) (*32*). In DNA, the AFB_1-FAPY adduct was longer-lived (*33–35*). Smela *et al.* (*36*) demonstrated that an oligodeoxynucleotide containing an AFB_1-FAPY adduct equilibrated between two separable species, one of which was mutagenic whereas the other blocked DNA replication. These corresponded to the α and β deoxyribose anomers of the AFB_1-FAPY adduct; the mutagenic species was the AFB_1-β-FAPY adduct (Chart 2) (*37*). In duplex DNA the AFB_1-β-FAPY anomer was favored, but in single strand DNA, a 2:1 α:β equilibrium mixture of anomers was observed (*37*). Additionally, the AFB_1-FAPY adduct underwent conformational rearrangements involving atropisomers about the C5-N^5 bond and geometrical isomers of the formyl moiety (*37*). Note the change in numbering in going from the AFB_1-N7-dG adduct to the AFB_1-FAPY adduct; the C5-N^5 bond in the AFB_1-FAPY adduct corresponds to the C5-N7 bond in the AFB_1-N7-dG adduct.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Site-Specific Mutagenesis of the AFB_1-N7-dG Adduct in *E. coli*

Essigmann and co-workers developed methodology allowing individual DNA adducts, such as those induced by exposures to AFB_1-*exo*-8,9-epoxide, to be situated at defined sites in the *E. coli* virus M13mp8 (*38*). These modified phage genomes could be introduced into *E. coli*, where replication occurred, subject to the actions of endogenous replication and repair enzymes. When the levels and types of DNA sequence alteration(s) induced in progeny phage at the site originally occupied by the adducts were examined, details of the biological processing of these adducts by *E. coli* emerged (*38*). Such site-specific mutagenesis assays demonstrated that the AFB_1-N7-dG adduct was mutagenic in *E. coli*, yielding primarily G→T transversions at levels of about 5% (*39*). This was significant because earlier studies had indicated that G→T transversions were the primary mutations associated with random AFB_1 mutagenesis in *E. coli* (*6*). Consequently, these site-specific mutagenesis data (*39*) implicated the AFB_1-N7-dG adduct as a pre-mutagenic lesion induced by AFB_1 in *E. coli*.

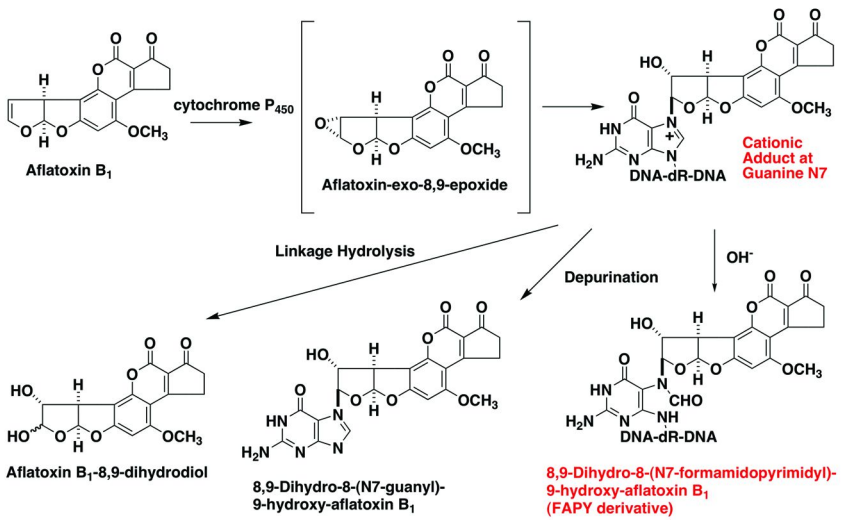


Chart 1. Chemistry of AFB_1-induced DNA alkylation. Cytochrome P450-mediated oxidation of AFB_1 forms aflatoxin B_1-exo-8,9-epoxide, a reactive electrophile. This alkylates DNA regioselectively at the N7 position of guanine, forming the AFB_1-N7-dG adduct. At neutral and acidic pH, this adduct is prone to depurination, but at basic pH, it converts to the AFB_1-FAPY adduct.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

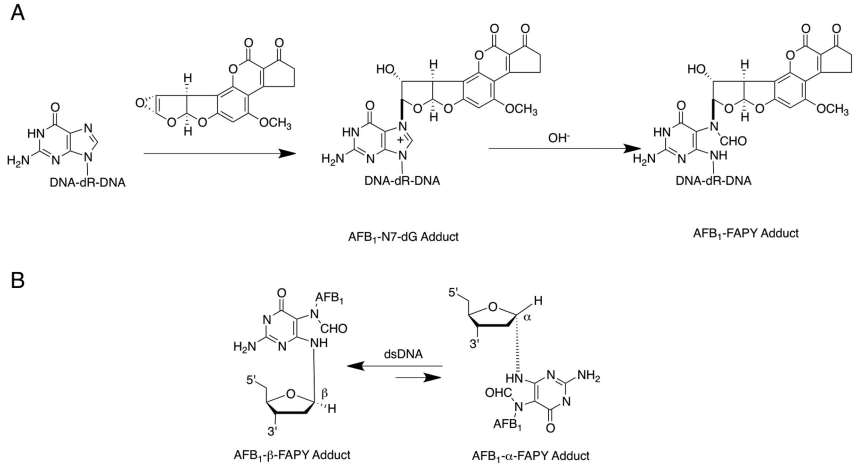


Chart 2. Chemistry of the AFB_1-FAPY adduct. ***A.*** *Formation of the AFB_1-FAPY adduct via base-catalyzed ring opening of the AFB_1-N7-dG adduct.* ***B.*** *The AFB_1-FAPY adduct interconverts between α and β anomers; the equilibrium is dependent on single strand vs. duplex DNA environments. Note the change in atomic numbering of the AFB_1-FAPY adducts, where the N7 nitrogen of guanine becomes the N5 nitrogen of the AFB_1-FAPY adducts. In nucleosides and nucleotides, two atropomers, R_a and S_a, are possible about the AFB_1-FAPY C5-N^5 bond. Furthermore, Z and E geometrical isomers are possible about the formamide bond. In duplex DNA, rotation about the C5-N^5 bond is restricted because of the intercalated AFB_1 moiety, and the R_a atropisomer predominates.*

Structural Studies of the AFB_1-N7-dG Adduct in Oligodeoxynucleotides

Using NMR (*40*, *41*), structures of oligodeoxynucleotide duplexes containing AFB_1 N7-dG adducts were obtained by our laboratory. These revealed that the AFB_1-N7-dG adduct intercalated above the 5′ face of the modified dG in two oligodeoxynucleotides: d(ATC<u>X</u>AT)•d(ATCGAT), as shown in Figure 1, and d(AT<u>X</u>CAT)$_2$, X = AFB_1-N7-dG adduct (*42–44*). A structure of the AFB_1 N7-dG adduct in a duplex containing an extra dA opposite the AFB_1 moiety was also obtained (*45*). A structurally-related adduct of sterigmatocystin formed a similar intercalated structure (*46*). Unlike many DNA adducts, the AFB_1 N7-dG adducts thermodynamically stabilized the duplex, as evidenced by increases in duplex melting temperatures (T_m studies), as monitored either by UV spectroscopy or by NMR (*47*, *48*). These observations provided insight into the structural alterations of the DNA duplex that may modulate the repair of AFB_1 lesions, and have led to the suggestion that the AFB_1 N7-dG adduct may be refractory to DNA repair (*48*). This remains to be established. It is established that the AFB_1-N7-dG adduct is a substrate for nucleotide excision repair (*49*). DNA damage frequently involves the "flipping" of damaged bases into an active site pocket of the repair enzyme, particularly in the case of base excision repair (*50*), but possibly also in

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

the case of nucleotide excision repair (*51*). One can speculate that an intercalated and thermodynamically stabilizing adduct might be more difficult to recognize and repair.

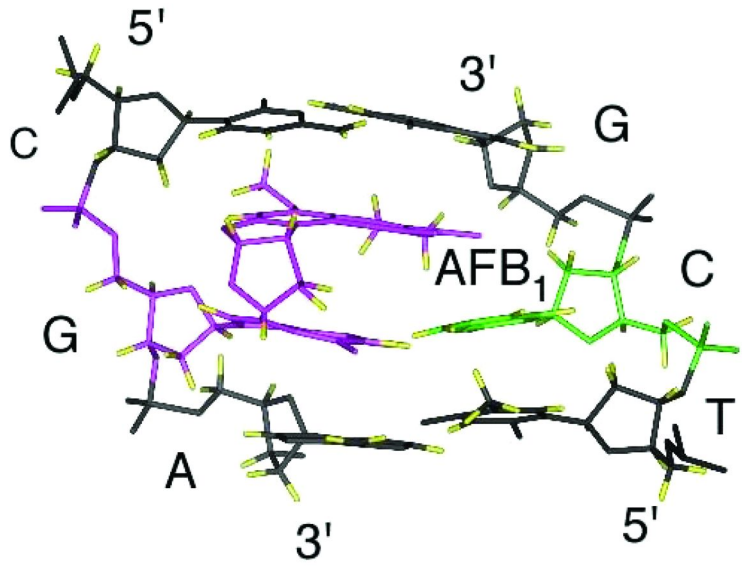


Figure 1. Structure of the AFB_1-N7-dG adduct in the 5'-d(CXA)-3':5'-d(TCG)-3' sequence as determined by NMR; X = AFB_1-N7-dG adduct. The adduct intercalates above the 5'-face of the modified guanine base and increases the thermal melting temperature (T_m) of the duplex. Reprinted with permission from Giri and Stone (48). Copyright 2002 John Wiley and Sons.

Site-Specific Mutagenesis of the AFB_1-FAPY Adduct in *E. coli*

Opening of the imidazole ring of the initially formed AFB_1-N7-dG adduct (*2*, *29*) yields the AFB_1-β-FAPY adduct (*32*, *37*). Site-specific mutagenesis experiments carried out by Essigmann and co-workers revealed that the AFB_1-β-FAPY adduct yielded G→T transversions at levels as high as 36% in *E. coli* (*36*), significantly higher than those observed for the AFB_1-N7-dG adduct (*39*). This, combined with observations that the AFB_1-β-FAPY adduct was persistent *in vivo* (*33–35*), suggested that it may be the most genotoxic lesion formed by AFB_1. In single strand DNA, the lack of the complementary strand facilitates the epimerization of the deoxyribose and the formation of an equilibrium mixture of α and β FAPY anomers, with the α anomer predominating in this sequence (*36*). In *E. coli*, the AFB_1-α-FAPY adduct blocked replication (*36*).

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Structural Studies of the AFB_1-FAPY Adducts in Oligodeoxynucleotides

As shown in Figure 2, in DNA the AFB_1-β-FAPY adduct intercalated with the AFB_1 moiety on the 5′ face of the pyrimidine moiety of the adducted nucleotide (*48*, *52*), similar to the AFB_1-N7-dG adduct (*42–45*). The stability of the AFB_1-β-FAPY adduct in DNA was attributed to interstrand stacking interactions (*48*, *52*). The AFB_1-α-FAPY adduct also intercalated above the 5′ face of the damaged base (*53*). The lower stability of the AFB_1-α-FAPY as compared to the AFB_1-β-FAPY adduct (*48*, *52*) was attributed to structural perturbations in the DNA and reduced interstrand stacking (*53*). The similar 5′-intercalation of the AFB_1-N7-dG (*42–45*) and AFB_1-β-FAPY (*52*) adducts was consistent with each producing G→T transversions in *E. coli*, but did not readily explain the increased mutagenicity of the AFB_1-β-FAPY adduct (*36*).

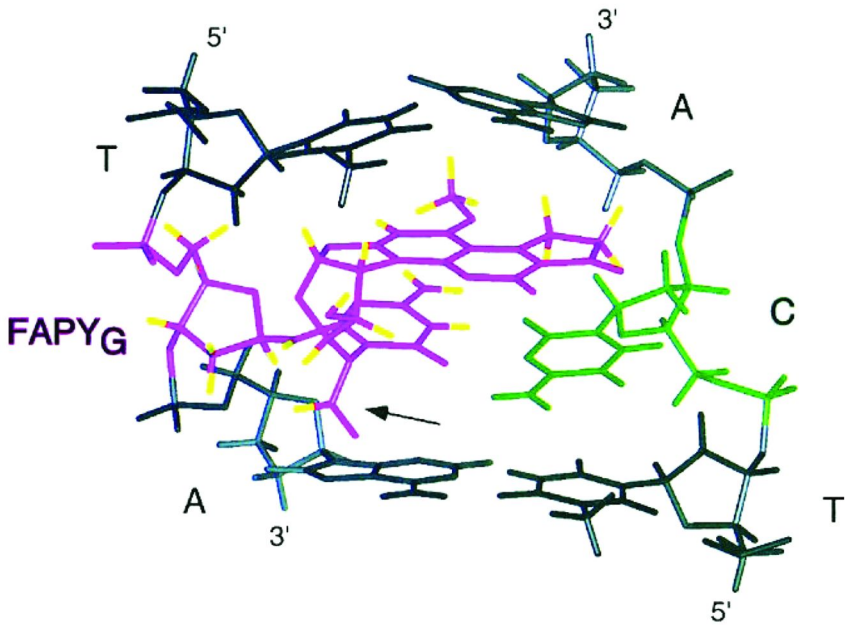


Figure 2. Structure of the AFB_1-β-FAPY adduct in the modified 5′-d(TXA)-3′:5′-d(TCA)-3′ sequence as determined by NMR; X = AFB_1-β-FAPY adduct. The adduct intercalates above the 5′-face of the modified guanine and increases the thermal melting temperature (T_m) of the duplex. The arrow indicates the positioning of the formamide moiety in the major groove. Reprinted with permission from Giri and Stone. (48). Copyright 2002 John Wiley and Sons.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Bypass of AFB_1 Adducts by the *Sulfolobus solfataricus* P2 DNA Polymerase IV (Dpo4)

In an effort to understand why the AFB_1-β-FAPY adduct is more mutagenic than is the AFB_1-N7-dG adduct (*36*), we examined the bypass of the AFB_1-N7-dG (*42–45*) and AFB_1-β-FAPY (*52*) adducts by the *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) (*54*), a DinB homologue (*55*). In one series of experiments, the 18-mer template 5′-d(TCATTXAATCCTTCCCCC)-3′ (X = AFB_1-N7-dG or AFB_1-β-FAPY adduct) was annealed with the Sequence I 12-mer primer 5′-d(GGGGGAAGGATT)-3′, leading to a template•primer primed for insertion of dNTPs opposite the adduct (Chart 3) (*54*). In a second series of experiments, the same 18-mer template was annealed with the Sequence II 13-mer primer 5′-d(GGGGGAAGGATTC)-3′, leading to a template•primer primed for extension of dNTPs beyond the dX:dC primer terminus (Chart 3) (*54*).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

```
5'-GGGGGAAGGATT-3'
3'-CCCCCTTCCTAAXTTACT-5'

5'-GGGGGAAGGATTC-3'
3'-CCCCCTTCCTAAXTTACT-5'
```

Chart 3. Sequences used for gel extension assays and crystallography. ***Top:*** *The 12-mer Sequence I primer used for examining insertion of dNTPs opposite the AFB_1 adducts.* ***Bottom:*** *The 13-mer Sequence II primer used for examining extension of dNTPs opposite either the AFB_1-N7-dG:dC or AFB_1-FAPY:dC base pairs. X denotes either the AFB_1-N7-dG or AFB_1-β-FAPY adducts.*

Figure 3 (top panel) shows the results of single nucleotide insertion assays using the Sequence I template:primer (Chart 3) (*54*). The Dpo4 polymerase correctly inserted dCTP opposite the AFB_1-N7-dG lesion. In the presence of all four dNTPs, the polymerase extended the primer to the full-length product. Figure 3 (bottom panel) shows the results of single nucleotide insertion assays using the Sequence II template:primer (Chart 3) with X the AFB_1-N7-dG adduct (*54*). This monitored single nucleotide extension beyond the correctly inserted dC at the 3′-terminus of the primer. The polymerase inserted dATP twice, corresponding to the positioning of the two thymines 5′ to the AFB_1-N7-dG lesion in the template. In the presence of all four dNTPs, the polymerase extended the primer to the full-length product.

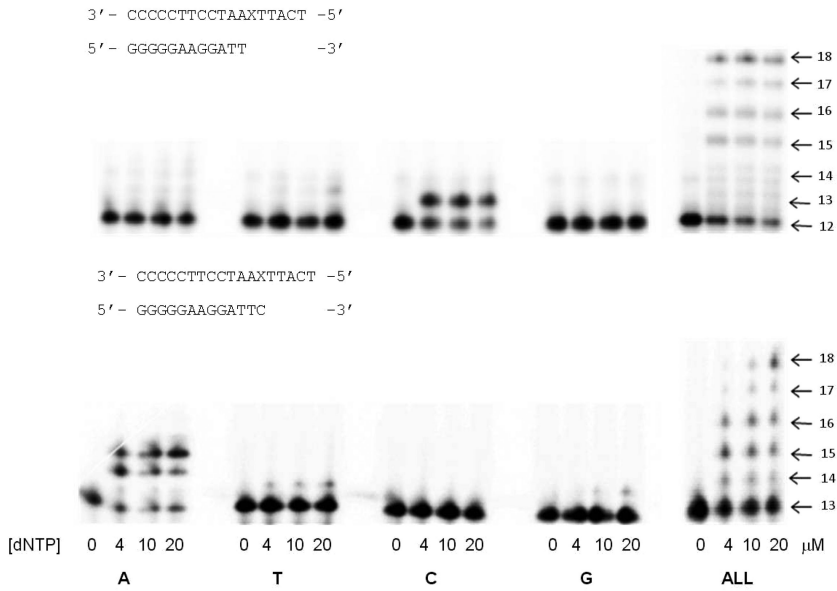


Figure 3. Replication bypass of the AFB$_1$-N7-dG modified Sequence I and Sequence II template:primers with S. solfataricus P2 DNA polymerase Dpo4. The sequences I and II are displayed with the gels. The concentrations of the dNTPs are provided below the gels. The designations A, T, C, G represent single nucleotide insertion experiments; the designation ALL represents the full-length extension assay incorporating all four dNTPs. Each assay was incubated for 1 hr at 37 ºC. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.

Figure 4 (top panel) shows the results of single nucleotide insertion assays using the Sequence I template:primer (Chart 3) with X the AFB$_1$-β-FAPY adduct (*54*). The Dpo4 polymerase was less efficient at inserting nucleotides opposite the AFB$_1$-β-FAPY lesion. The correct nucleotide dCTP was inserted but the incorrect nucleotide dATP was also inserted. In the dATP lanes, weak bands were observed corresponding to 14-mer and 15-mer products, suggesting that multiple insertions of dATP, presumably involving the two 5′-neighboring template dT nucleotides in the template. Thus the AFB$_1$-β-FAPY adduct exhibited both correct insertion of dCTP and mis-insertion of dATP for this template:primer. In the presence of all four dNTPs, the polymerase extended the primer to the full-length product. Figure 4 (bottom panel) shows the results of single nucleotide insertion assays using the Sequence II template:primer (Chart 3). This monitored single nucleotide extension beyond the correctly inserted dC at the 3′-terminus of the primer. The polymerase was less efficient at extending nucleotides past the AFB$_1$-β-FAPY adduct, as compared to the AFB$_1$-N7-dG adduct. It did not efficiently insert any dNTP in the single-nucleotide extension reactions, although again in the dATP lanes, weak bands probably corresponded to multiple insertions of dATP opposite

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

the 5′-neighboring dT nucleotides in the template. When all four dNTPs were included, the primer was extended to the full-length product.

These data recapitulated aspects of the site-specific mutagenesis for the AFB_1-N7-dG and AFB_1-FAPY adducts (*36*, *39*). Predominantly error-free bypass was observed for the AFB_1-N7-dG adduct (Figure 3), which correlated with its lower mutagenicity (*39*). For the AFB_1-N7-dG adduct the polymerase both inserted the correct dCTP opposite the damaged base and in the presence of all four dNTPs, extended the primer to a full-length product (Figure 3). With the AFB_1-β-FAPY adduct, bypass was less efficient and error-prone, correlating with its greater mutagenicity (Figure 4) (*36*). Also, the polymerase mis-inserted dATP, albeit inefficiently, when challenged by the AFB_1-β-FAPY adduct (Figure 4). This would be anticipated to lead to a G→T transversion, in agreement with site-specific mutagenesis studies (*36*).

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

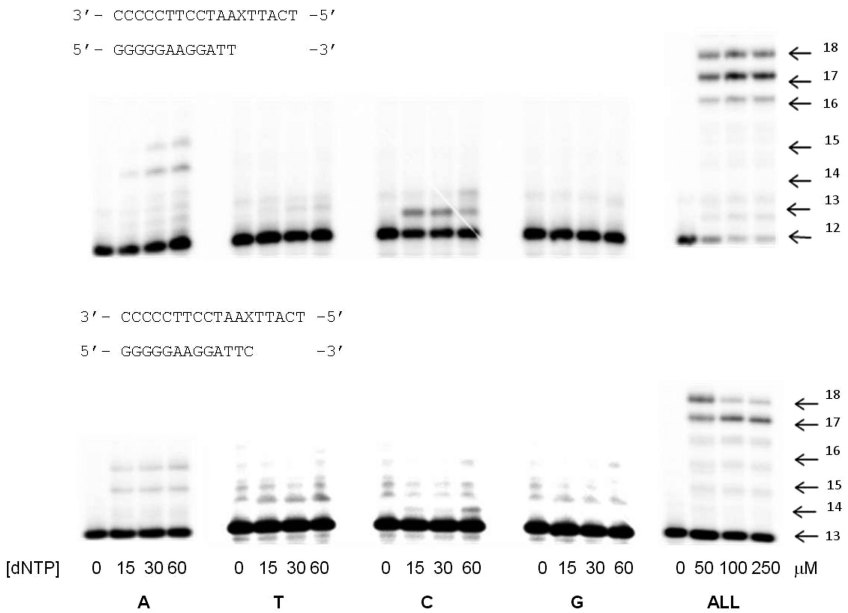


Figure 4. Replication bypass of the AFB_1-β-FAPY modified Sequence I and Sequence II template:primers with S. solfataricus P2 DNA polymerase Dpo4. The sequences I and II are displayed with the gels. The concentrations of the dNTPs are provided below the gels. The designations A, T, C, G represent single nucleotide insertion experiments; the designation ALL represents the full-length extension assay incorporating all four dNTPs. Each assay was incubated for 1 hr at 37 ºC. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.

Crystallography of AFB_1-Adducted Template:Primers Complexed with the Dpo4 Polymerase

Structures of the Dpo4 polymerase in complex with DNA and incoming dNTPs model the structural features that determine lesion bypass efficiency and fidelity with Y-family polymerases (*56*). The active sites of Y-family polymerases are solvent accessible and in some cases can accommodate two template bases (*56–61*). The nascent base pair is less constrained than for replicative polymerases. The spacious active site relaxes geometric selection for incoming dNTPs (*58*), compromising the efficiency and fidelity of replication. The bypass ability, accuracy, and efficiencies of these polymerases vary (*55*, *62–66*) and depend on the types of DNA adducts (*67–77*).

The ternary complex of the AFB_1-N7-dG adduct with the Sequence I 12-mer primer, showing correct insertion of dCTP opposite the adducted guanine (Figure 5) provided the first glimpse of the AFB_1-N7-dG adduct during replication bypass (*54*). The intercalation of the AFB_1 moiety above the 5′-face of the adducted guanine, with the methoxy group facing the minor groove of the template, and the keto oxygens facing into the nascent duplex, was similar to that observed in DNA (*53*). The damaged nucleotide was accommodated within the polymerase active site without changes to the conformation of the AFB_1 adduct in DNA. The adducted guanine base and AFB_1 moiety were 16º out of plane. This was a consequence of maintaining the bond between N7-dG and the C8 carbon of the AFB_1 moiety in plane with the damaged guanine base and had been inferred from NMR data (*43*, *44*). This allowed for stacking of the AFB_1 moiety above the 5′-face of the adducted guanine base, but created a wedge in the DNA. During bypass, this perhaps facilitates the insertion of the incoming dCTP between the AFB_1 moiety and the adducted guanine. The potential for Watson-Crick hydrogen bonding between the incoming dCTP and the guanine base remained intact (Figure 5). The distance between the 3′-OH group of the primer to the α-phosphate suggested that phosphodiester bond formation could be facilitated. This may account for the ability of the polymerase to correctly insert dCTP opposite the AFB_1-N7-dG lesion (Figure 3), consistent with the lower mutagenicity of this lesion in *E. coli* (*2*, *39*).

Figure 6 shows the polymerase active site in the Sequence II 13-mer primer complex, with the correct incoming dATP placed opposite to the template 5′-neighbor dT (*54*). The incoming dATP stacked above the AFB_1 moiety and was positioned to form Watson-Crick hydrogen bonds with the template dT (Figure 6). Watson-Crick base pairing was maintained between the guanine base of the AFB_1-N7-dG adduct and the primer 3′-terminus dC. Both the guanine base and dC of the primer tilted out of plane toward the 3′-direction of the template. The AFB_1 moiety remained intercalated above the 5′-face of the modified guanine. The distance between the α-phosphate of the incoming dATP and the 3′-OH group of the primer was 6.7 Å.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

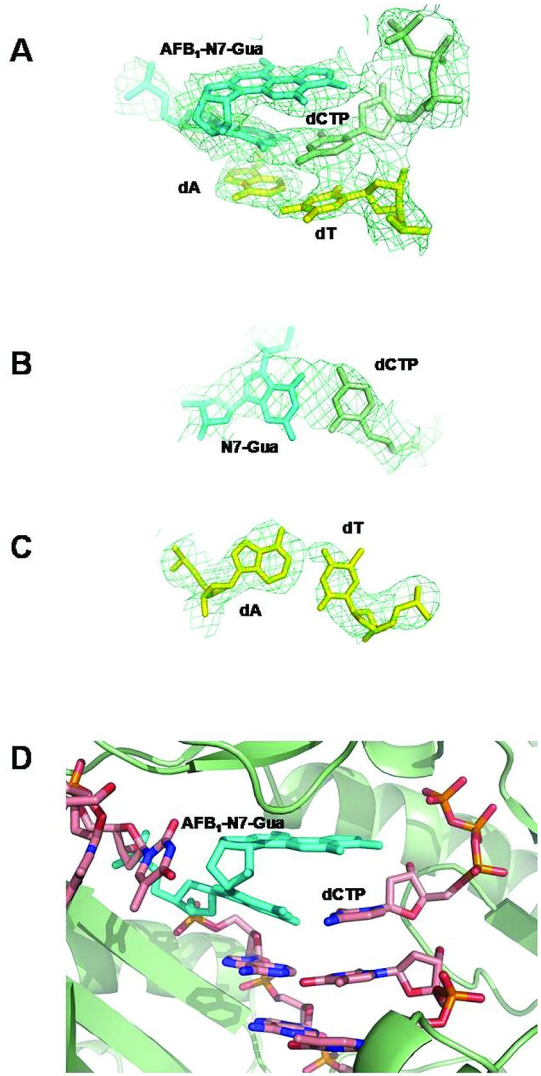


Figure 5. Structure of the ternary AFB$_1$-N7-dG modified Sequence I template:primer complex with the S. solfataricus P2 DNA polymerase Dpo4 and incoming dCTP. ***A.*** *Electron density at the active site.* ***B.*** *Watson-Crick base pair between AFB$_1$-N7-dG and 3'-primer terminus dC.* ***C.*** *Watson-Crick dA:dT base pair involving the template 3'-neighbor dA.* ***D.*** *Active site with the modified template:primer and the dCTP along with the polymerase. The Dpo4 is colored green and the AFB$_1$-N7-dG adduct is colored cyan. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.*

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

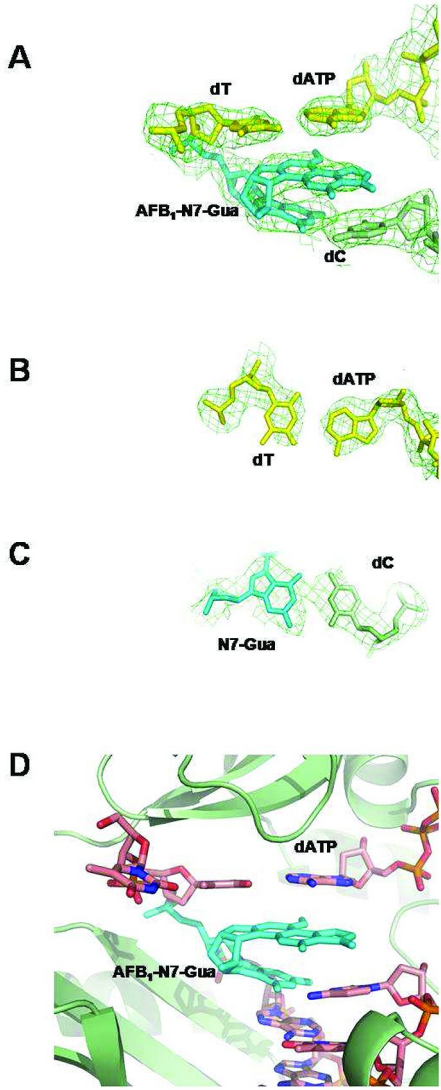


Figure 6. Structure of the ternary AFB$_1$-N7-dG modified Sequence II template:primer complex with the S. solfataricus P2 DNA polymerase Dpo4 and incoming dATP. ***A.*** *Electron density at the active site.* ***B.*** *Watson-Crick base pair between the 5'-template neighbor T and incoming dATP.* ***C.*** *Watson-Crick base pair between AFB$_1$-N7-dG and 3'-primer terminus dC.* ***D.*** *Active site with the modified template:primer and the dATP along with the polymerase. The Dpo4 is colored green and the AFB$_1$-N7-dG is colored cyan. The Ca^{2+} ions are shown as blue spheres. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.*

Figure 7 shows the binary complex formed between the template containing the AFB_1-FAPY adduct and Sequence II, the 13-mer primer (Chart 3) (*54*). The AFB_1 moiety intercalated above the 5′-face of the FAPY base. Similar to the AFB_1-N7-dG lesion, this placed the methoxy group facing the minor groove of the template, with the keto oxygens of AFB_1 facing the major groove of the template. The deoxyribose was in the β-anomeric configuration. The C5-N^5 bond of the pyrimidinyl moiety was in the R_a configuration as observed for the AFB_1-β-FAPY adduct in DNA (*48*, *52*), and the nucleoside and the FAPY base (*37*). Rotation about the C5-N^5 bond allowed the bond between the alkylated N^5 formamido nitrogen and the C8 of the AFB_1 moiety to orient out of plane with respect to the FAPY base and toward the 5′-direction, as had been inferred from NMR (*48*, *52*, *53*). This allowed the AFB_1 moiety to stack with the FAPY base. The 3′-primer terminus dC formed a Watson-Crick bonding interaction with the FAPY base. Unlike the binary complex involving native DNA (*78*) in which the primer terminus reached to the end of the active site, here the site was occupied by the AFB_1-β-FAPY adduct and the template 5′-neighbor dT (Figure 7). This was attributed to the intercalation of the AFB_1 moiety above the 5′-face of the FAPY base and the stacking of the AFB_1 moiety with the template 5′-neighbor dT. The distance between the primer terminus AFB_1-FAPY:dC pair and AFB_1 was ~3.7 Å, similar to the helicoidal rise in B-DNA. The conformations of the protein side chains for Y10, Y48 and R51 were similar to those for the unmodified binary complex (*78*).

For the ternary complex showing correct insertion of dATP from the AFB_1-β-FAPY:dC Sequence II primer (Chart 3) the AFB_1 moiety was parallel with the DNA base pairs and stacked between the FAPY base and 5′-neighbor dT (Figure 8) (*54*). The deoxyribose was in the β-anomeric configuration. The helicoidal rise between the AFB_1 moiety and the FAPY base was ~ 3.7 Å. The incoming dATP paired with the 5′-neighbor template base dT, with conservation of Watson-Crick hydrogen bonding (Figure 8). At the 3′-terminus of the primer, Watson-Crick hydrogen bonding was maintained between the FAPY base and the primer dC (Figure 8). This resulted in a gap of 6.9 Å between the 3′-hydroxyl of the primer dC and the α-phosphate of the dATP (Figure 8). The formamide oxygen participated in a water-mediated hydrogen bond with Arg332 and made van der Waals contacts with Ile 295. Three bound Ca^{2+} ions were identified (Figure 8). The first two were in the active site. The third ion was 3.0 Å from the peptide carbonyl oxygen of Ala181 in the thumb domain of the polymerase. One Ca^{2+} ion at the active site was 3.5 Å distant from the primer 3′-terminus hydroxyl, suggesting that it was positioned to faciliate the reaction.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

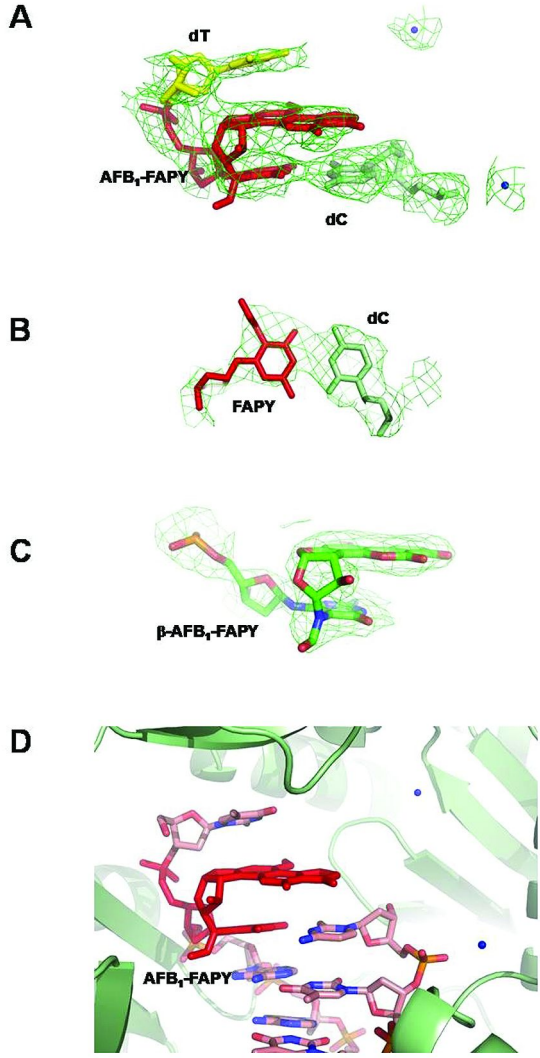


Figure 7. Structure of the binary AFB$_1$-β-FAPY modified Sequence II template:primer complex with the S. solfataricus P2 DNA polymerase Dpo4. ***A.*** *Electron density at the active site.* ***B.*** *Watson-Crick base pair between the FAPY base of the AFB$_1$-β-FAPY adduct and 3'-primer terminus dC.* ***C.*** *The electron density of the AFB$_1$-b-FAPY nucleoside.* ***D.*** *Active site with the modified template:primer along with the polymerase. The Dpo4 is colored green and the AFB$_1$-N7-dG is colored cyan. The Ca^{2+} ions are shown as blue spheres. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.*

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

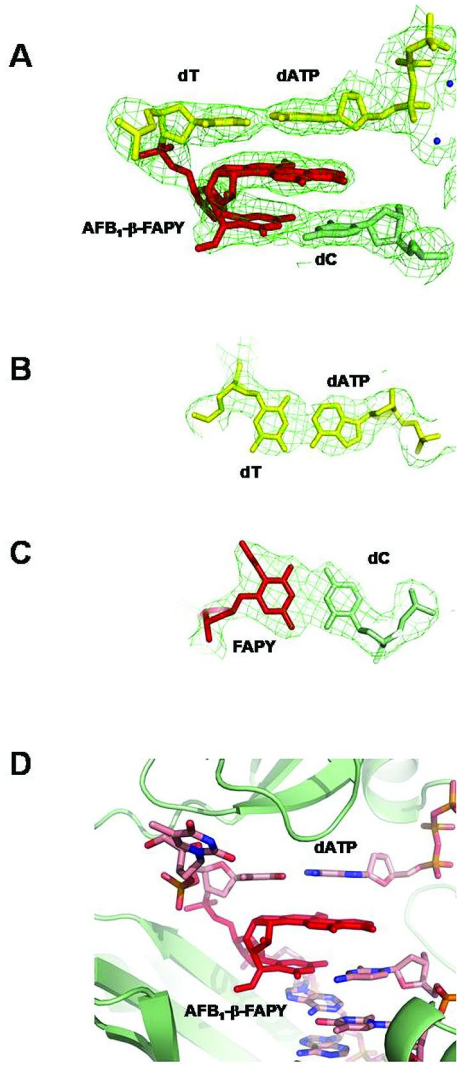


Figure 8. Structure of the ternary AFB$_1$-β-FAPY modified Sequence II template:primer complex with the S. solfataricus P2 DNA polymerase Dpo4 and incoming dATP. ***A.*** *Electron density at the active site.* ***B.*** *Watson-Crick base pair between the 5'-template T and the incoming dATP.* ***C.*** *Watson-Crick base pair between the FAPY base of the AFB$_1$-β-FAPY adduct and 3'-primer terminus dC.* ***D.*** *Active site with the modified template:primer along with the polymerase and the dATP. The Dpo4 is colored green and the AFB$_1$-β-FAPY is colored cyan. The Ca^{2+} ions are shown as blue spheres. Reprinted with permission from Banerjee et al. (54). Copyright 2011 American Chemical Society.*

Structure-Activity Relationships

The conformational differences of the AFB_1 moiety within the active site of the Dpo4 polymerase, which result from the differing orientations of the N7-C8 bond of the AFB_1-N7-dG *vs.* the N^5-C8 bond of the AFB_1-β-FAPY adduct may, in part, modulate AFB_1 lesion bypass by this polymerase. For the AFB_1-β-FAPY adduct, parallel stacking of the AFB_1 moiety with the FAPY base may hinder access to the incoming dNTP (Figure 7). For the AFB_1-N7-dG adduct, the adducted guanine base and AFB_1 moieties are 16º out of plane. The resulting wedge between the adducted guanine base and the AFB_1 moiety might faciliate access for the incoming dCTP (Figure 5), consistent with the error-free bypass of the AFB_1-N7-dG adduct (Figure 3) (*54*).

The data suggest that following correct insertion of dCTP opposite the AFB_1-β-FAPY adduct, the polymerase can insert dATP opposite the template 5′-neighbor dT. In the ternary complex with the Sequence II 13-mer primer and incoming dATP the FAPY base conserves Watson-Crick hydrogen bonds with the 3′-primer terminus dC (Figure 8). The incoming dATP forms Watson-Crick hydrogen bonds with the template 5′-neighbor dT. The structure seems unlikely to be catalytically efficient since the 3′-OH of the primer and the α–phosphate of the incoming dATP are more than 6.9 Å apart. It is unclear what rearrangements might facilitate phosphodiester bond formation although one could envision transient molecular dynamics bringing these atoms sufficiently close. The insertion of dATP opposite the template 5′-neighbor dT is reminiscent of the "Type II" structure observed for ternary complexes of the polymerase with undamaged DNA (*56*). The polymerase accommodates two template nucleotides, and it inserts the dATP opposite the template 5′-neighbor nucleotide, rather than the damaged nucleotide. Remarkably, the active site accommodates the FAPY base, the AFB_1 moiety, and the template 5′-neighbor dT. Thus, this could be considered to be a "pseudo" Type II structure (*54*).

Summary

Progress has been made in understanding the DNA chemistry associated with human exposures to the mycotoxin AFB_1. Both the initially formed AFB_1-N7-dG adduct and its rearrangement product, the AFB_1-FAPY adduct, exhibit intercalation of the AFB_1 moiety above the 5′-face of the damaged base. Site-specific mutagenesis conducted in *E. coli* shows that both the AFB_1-N7-dG and the AFB_1-FAPY adducts are mutagenic and induce G→T transversions. However, the AFB_1-FAPY adduct is significantly more mutagenic. Crystallographic studies using site-specifically modified template:primers with the Y-family Dpo4 polymerase indicate that conformational differences of the AFB_1 moiety within the active site of the polymerase, which result from the differing orientations of the N7-C8 bond of the AFB_1-N7-dG *vs.* the N^5-C8 bond of the AFB_1-β-FAPY adduct may, in part, modulate AFB_1 adduct bypass. Future studies will examine the site-specific mutagenesis of these adducts in mammalian cells and the replication bypass of these adducts by human Y-family polymerases.

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Laboratory Safety Statement

AFB_1 is a potent liver toxin and is genotoxic, and it should be presumed that AFB_1-*exo*-8,9-epoxide is toxic and genotoxic. Crystalline aflatoxins are hazardous due to their electrostatic nature. AFB_1 can be destroyed by oxidation with NaOCl. Manipulations should be carried out in a well-ventilated hood with suitable containment procedures.

Data Deposition

Complete structure factor and final coordinates were deposited in the Protein Data Bank (www.rcsb.org): PDB ID codes for the ternary complexes of the AFB_1-N7-dG adduct with dCTP, 3PW7; with dATP, 3PW4; for the binary complex of the AFB_1-β-FAPY, 3PVX; for the ternary complex of the AFB_1-β-FAPY adduct with dATP, 3PW0.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Acknowledgments

This work was presented, in part, at the 66th Southwestern Regional Meeting and 62nd Southeastern Regional Meeting of the American Chemical Society, New Orleans, Louisiana, December 2010. This work was funded by NIH grants R01 CA-55678 (M.P.S.), the Vanderbilt University Center in Molecular Toxicology, P30 ES-00267, and the Vanderbilt-Ingram Cancer Center, P30 CA-68485. Vanderbilt University and the Vanderbilt Center for Structural Biology assisted with the purchase of in-house crystallographic instrumentation. Crystallographic data were collected on the 21-ID-F beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Supporting institutions may be found athttp://ls-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract W-31109-Eng-38.

References

1. Busby, W. F. Jr.; Wogan, G. N. Aflatoxins. In *Chemical Carcinogens*, 2nd ed.; Searle, C. E., Ed.; American Chemical Society: Washington, DC, 1984; pp 945−1136;
2. Smela, M. E.; Currier, S. S.; Bailey, E. A.; Essigmann, J. M. *Carcinogenesis* **2001**, *22*, 535–545.
3. Bennett, J. W.; Klich, M. *Clin. Microbiol. Rev.* **2003**, *16*, 497–516.
4. Kensler, T. W.; Roebuck, B. D.; Wogan, G. N.; Groopman, J. D. *Toxicol. Sci.* **2011**, *120* (Suppl 1), S28–48.
5. McCann, J.; Spingarn, N. E.; Koburi, J.; Ames, B. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 979–983.
6. Foster, P. L.; Eisenstadt, E.; Miller, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 2695–2698.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

7. Foster, P. L.; Groopman, J. D.; Eisenstadt, E. *J. Bacteriol.* **1988**, *170*, 3415–3420.
8. Bailey, G. S.; Williams, D. E.; Wilcox, J. S.; Loveland, P. M.; Coulombe, R. A.; Hendricks, J. D. *Carcinogenesis* **1988**, *9*, 1919–1926.
9. Bailey, G. S.; Loveland, P. M.; Pereira, C.; Pierce, D.; Hendricks, J. D.; Groopman, J. D. *Mutat. Res.* **1994**, *313*, 25–38.
10. McMahon, G.; Davis, E. F.; Huber, L. J.; Kim, Y.; Wogan, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1104–1108.
11. Soman, N. R.; Wogan, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2045–2049.
12. Yang, M.; Zhou, H.; Kong, R. Y.; Fong, W. F.; Ren, L. Q.; Liao, X. H.; Wang, Y.; Zhuang, W.; Yang, S. *Mutat. Res.* **1997**, *381*, 25–29.
13. Groopman, J. D.; Kensler, T. W. *Toxicol. Appl. Pharmacol.* **2005**, *206*, 131–137.
14. Bressac, B.; Kew, M.; Wands, J.; Ozturk, M. *Nature* **1991**, *350*, 429–431.
15. Hsu, I. C.; Metcalf, R. A.; Sun, T.; Welsh, J. A.; Wang, N. J.; Harris, C. C. *Nature* **1991**, *350*, 427–428.
16. Greenblatt, M. S.; Bennett, W. P.; Hollstein, M.; Harris, C. C. *Cancer Res.* **1994**, *54*, 4855–4878.
17. Shen, H. M.; Ong, C. N. *Mutat. Res.* **1996**, *366*, 23–44.
18. Soini, Y.; Chia, S. C.; Bennett, W. P.; Groopman, J. D.; Wang, J. S.; DeBenedetti, V. M.; Cawley, H.; Welsh, J. A.; Hansen, C.; Bergasa, N. V.; Jones, E. A.; DiBisceglie, A. M.; Trivers, G. E.; Sandoval, C. A.; Calderon, I. E.; Munoz Espinosa, L. E.; Harris, C. C. *Carcinogenesis* **1996**, *17*, 1007–1012.
19. Lunn, R. M.; Zhang, Y. J.; Wang, L. Y.; Chen, C. J.; Lee, P. H.; Lee, C. S.; Tsai, W. Y.; Santella, R. M. *Cancer Res.* **1997**, *57*, 3471–3477.
20. Mace, K.; Aguilar, F.; Wang, J. S.; Vautravers, P.; Gomez-Lechon, M.; Gonzalez, F. J.; Groopman, J.; Harris, C. C.; Pfeifer, A. M. *Carcinogenesis* **1997**, *18*, 1291–1297.
21. Shimada, T.; Guengerich, F. P. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 462–465.
22. Raney, K. D.; Shimada, T.; Kim, D. H.; Groopman, J. D.; Harris, T. M.; Guengerich, F. P. *Chem. Res. Toxicol.* **1992**, *5*, 202–210.
23. Ueng, Y. F.; Shimada, T.; Yamazaki, H.; Guengerich, F. P. *Chem. Res. Toxicol.* **1995**, *8*, 218–225.
24. Gallagher, E. P.; Kunze, K. L.; Stapleton, P. L.; Eaton, D. L. *Toxicol. Appl. Pharmacol.* **1996**, *141*, 595–606.
25. Johnson, W. W.; Harris, T. M.; Guengerich, F. P. *J. Am. Chem. Soc.* **1996**, *118*, 8213–8220.
26. Baertschi, S. W.; Raney, K. D.; Stone, M. P.; Harris, T. M. *J. Am. Chem. Soc.* **1988**, *110*, 7929–7931.
27. Murray, R. W.; Jeyaraman, R. *J. Org. Chem.* **1985**, *50*, 2847–2853.
28. Adam, W.; Bialas, J.; Hadjiarapoglou, L. *Chem. Ber.* **1991**, *124*, 2377–2377.
29. Essigmann, J. M.; Croy, R. G.; Nadzan, A. M.; Busby, W. F., Jr.; Reinhold, V. N.; Buchi, G.; Wogan, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1870–1874.

30. Gopalakrishnan, S.; Byrd, S.; Stone, M. P.; Harris, T. M. *Biochemistry* **1989**, *28*, 726–734.
31. Iyer, R. S.; Coles, B. F.; Raney, K. D.; Thier, R.; Guengerich, F. P.; Harris, T. M. *J. Am. Chem. Soc.* **1994**, *116*, 1603–1609.
32. Hertzog, P. J.; Smith, J. R. L.; Garner, R. C. *Carcinogenesis* **1982**, *3*, 723–725.
33. Hertzog, P. J.; Lindsay Smith, J. R.; Garner, R. C. *Carcinogenesis* **1980**, *1*, 787–793.
34. Croy, R. G.; Wogan, G. N. *Cancer Res.* **1981**, *41*, 197–203.
35. Groopman, J. D.; Croy, R. G.; Wogan, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 5445–5449.
36. Smela, M. E.; Hamm, M. L.; Henderson, P. T.; Harris, C. M.; Harris, T. M.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6655–6660.
37. Brown, K. L.; Deng, J. Z.; Iyer, R. S.; Iyer, L. G.; Voehler, M. W.; Stone, M. P.; Harris, C. M.; Harris, T. M. *J. Am. Chem. Soc.* **2006**, *128*, 15188–15199.
38. Essigmann, J. M.; Fowler, K. W.; Green, C. L.; Loechler, E. L. *Environ. Health Perspect.* **1985**, *62*, 171–176.
39. Bailey, E. A.; Iyer, R. S.; Stone, M. P.; Harris, T. M.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1535–1539.
40. Reid, B. R. *Q. Rev. Biophys.* **1987**, *20*, 2–28.
41. Patel, D. J.; Shapiro, L.; Hare, D. *Q. Rev. Biophys.* **1987**, *20*, 35–112.
42. Gopalakrishnan, S.; Harris, T. M.; Stone, M. P. *Biochemistry* **1990**, *29*, 10438–10448.
43. Jones, W. R.; Johnston, D. S.; Stone, M. P. *Chem. Res. Toxicol.* **1998**, *11*, 873–881.
44. Giri, I.; Jenkins, M. D.; Schnetz-Boutaud, N. C.; Stone, M. P. *Chem. Res. Toxicol.* **2002**, *15*, 638–647.
45. Johnston, D. S.; Stone, M. P. *Biochemistry* **1995**, *34*, 14037–14050.
46. Gopalakrishnan, S.; Liu, X.; Patel, D. J. *Biochemistry* **1992**, *31*, 10790–10801.
47. Gopalakrishnan, S.; Stone, M. P.; Harris, T. M. *J. Am. Chem. Soc.* **1989**, *111*, 7232–7239.
48. Giri, I.; Stone, M. P. *Biopolymers* **2002**, *65*, 190–201.
49. Alekseyev, Y. O.; Hamm, M. L.; Essigmann, J. M. *Carcinogenesis* **2004**, *25*, 1045–1051.
50. Tainer, J. A. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *68*, 299–304.
51. Min, J. H.; Pavletich, N. P. *Nature* **2007**, *449*, 570–575.
52. Mao, H.; Deng, Z.; Wang, F.; Harris, T. M.; Stone, M. P. *Biochemistry* **1998**, *37*, 4374–4387.
53. Brown, K. L.; Voehler, M. W.; Magee, S. M.; Harris, C. M.; Harris, T. M.; Stone, M. P. *J. Am. Chem. Soc.* **2009**, *131*, 16096–16107.
54. Banerjee, S.; Brown, K. L.; Egli, M.; Stone, M. P. *J. Am. Chem. Soc.* **2011**, *133*in press.
55. Boudsocq, F.; Iwai, S.; Hanaoka, F.; Woodgate, R. *Nucleic Acids Res.* **2001**, *29*, 4607–4616.
56. Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. *Cell* **2001**, *107*, 91–102.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

57. Silvian, L. F.; Toth, E. A.; Pham, P.; Goodman, M. F.; Ellenberger, T. *Nat. Struct. Biol.* **2001**, *8*, 984–989.
58. Goodman, M. F. *Annu. Rev. Biochem.* **2002**, *71*, 17–50.
59. Alt, A.; Lammens, K.; Chiocchini, C.; Lammens, A.; Pieck, J. C.; Kuch, D.; Hopfner, K. P.; Carell, T. *Science* **2007**, *318*, 967–70.
60. Reissner, T.; Schneider, S.; Schorr, S.; Carell, T. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 3077–3080.
61. Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K. *Nature* **2010**, *465*, 1039–1043.
62. Johnson, R. E.; Prakash, S.; Prakash, L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3838–3843.
63. Zhang, Y.; Yuan, F.; Wu, X.; Rechkoblit, O.; Taylor, J. S.; Geacintov, N. E.; Wang, Z. *Nucleic Acids Res.* **2000**, *28*, 4717–4724.
64. Ohashi, E.; Bebenek, K.; Matsuda, T.; Feaver, W. J.; Gerlach, V. L.; Friedberg, E. C.; Ohmori, H.; Kunkel, T. A. *J. Biol. Chem.* **2000**, *275*, 39678–39684.
65. Gerlach, V. L.; Feaver, W. J.; Fischhaber, P. L.; Friedberg, E. C. *J. Biol. Chem.* **2001**, *276*, 92–98.
66. Rechkoblit, O.; Malinina, L.; Cheng, Y.; Kuryavyi, V.; Broyde, S.; Geacintov, N. E.; Patel, D. J. *PLoS Biol.* **2006**, *4*, e11.
67. Johnson, R. E.; Prakash, S.; Prakash, L. *Science* **1999**, *283*, 1001–1004.
68. Ohashi, E.; Ogi, T.; Kusumoto, R.; Iwai, S.; Masutani, C.; Hanaoka, F.; Ohmori, H. *Genes Dev.* **2000**, *14*, 1589–1594.
69. Haracska, L.; Yu, S. L.; Johnson, R. E.; Prakash, L.; Prakash, S. *Nat. Genet.* **2000**, *25*, 458–461.
70. Johnson, R. E.; Washington, M. T.; Haracska, L.; Prakash, S.; Prakash, L. *Nature* **2000**, *406*, 1015–1019.
71. Frank, E. G.; Sayer, J. M.; Kroth, H.; Ohashi, E.; Ohmori, H.; Jerina, D. M.; Woodgate, R. *Nucleic Acids Res.* **2002**, *30*, 5284–5292.
72. Rechkoblit, O.; Zhang, Y.; Guo, D.; Wang, Z.; Amin, S.; Krzeminsky, J.; Louneva, N.; Geacintov, N. E. *J. Biol. Chem.* **2002**, *277*, 30488–30494.
73. Huang, X.; Kolbanovskiy, A.; Wu, X.; Zhang, Y.; Wang, Z.; Zhuang, P.; Amin, S.; Geacintov, N. E. *Biochemistry* **2003**, *42*, 2456–266.
74. Washington, M. T.; Minko, I. G.; Johnson, R. E.; Haracska, L.; Harris, T. M.; Lloyd, R. S.; Prakash, S.; Prakash, L. *Mol. Cell Biol.* **2004**, *24*, 6900–6906.
75. Washington, M. T.; Minko, I. G.; Johnson, R. E.; Wolfle, W. T.; Harris, T. M.; Lloyd, R. S.; Prakash, S.; Prakash, L. *Mol. Cell Biol.* **2004**, *24*, 5687–5693.
76. Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2010**, *17*, 379–388.
77. Wong, J. H.; Brown, J. A.; Suo, Z.; Blum, P.; Nohmi, T.; Ling, H. *EMBO J.* **2010**, *29*, 2059–2069.
78. Wong, J. H.; Fiala, K. A.; Suo, Z.; Ling, H. *J. Mol. Biol.* **2008**, *379*, 317–330.

Downloaded by STANFORD UNIV GREEN LIBR on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch009

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Chapter 10

Examining Flanking Sequence Specificity and Topological Specificity in the Binding of Various Molecular Types to DNAs Using Restriction Endonuclease Activity Assays

S. A. Winkle,*,[1] E. Duran,[1] J. Pulido,[1] G. Santil,[1] M. Talavera,[1] C. Winkle,[1] R. D. Sheardy,[2] and V. Ramsauer[3]

[1]Department of Chemistry and Biochemistry, Florida International University, Miami, Florida 33199
[2]Department of Chemistry and Physics, Texas Woman's University, Denton, Texas 76205
[3]College of Pharmacy, East Tennessee State University, Johnson City, Tennessee 37614
***Email: winkles@fiu.edu**

Observing alteration of restriction enzyme activity has been employed frequently to determine the sequence specificity of the binding of many types of molecules to DNAs. Generally, these studies employed restriction enzymes which cut the target DNA several times. The effects of binding to sequences flanking the restriction enzyme cleavage sites may have been obscured. In this study, we report on restriction enzyme activity assays of the binding of the intercalators actinomycin D, ametantrone and ethidium, the groove binder netropsin, and the covalent binding cisplatin to a mixture of supercoiled and relaxed phiX 174 RF DNA using restriction enzymes which cleave this DNA once or twice. Sequence selectivities and topological selectivities were observed for these ligands. In some cases restriction enzymes not containing the reported preferred binding sites had altered activities, suggesting binding to flanking sequences affects activity in neighboring DNA sequences.

© 2011 American Chemical Society
In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Introduction

The sequence selectivity of DNA binding has been examined for a wide variety of ligands including those used in this study- actinomycin D, Ametantrone, cisplatin, ethidium and netropsin. Methods employed have included optical tirations, nmr studies, calorimetric titrations, circular dichroism studies and enzymatic digestion studies. Varying degrees of binding and sequence selectivities have been observed.

Actinomycin D was initially shown by several groups (*1–4*) to bind intercalatively with a strong preference for GpC sequences. This specificity arises in part from the interaction of the pentapeptide moieties of actinomycin with DNA bases above and below the intercalated actinomycin ring system. Subsequent work by Graves and coworkers showed that the bases flanking the GpC are important in the sequence specificity (*5*). They further showed that $T(G_n)$ sequences may also serve as high affinity binding sites (*6*).

Ametantrone has also been shown by Wilson, et.al., to display a GpC specificity (*7*). Ametantrone possesses an anthraquinone ring system which intercalates and two side chains which can interact with the groove of the DNA. In contrast, netropsin, which is primarily a groove binder, prefers AT-rich sequences (*8*). Ethidium, a "simple" intercalator is fairly promiscuous in its binding and showed no marked sequence selectivity (*9*). Cisplatin binds covalently to guanines with the formation of intrastrand crosslinks and interstrand crosslinks (*10*) though binding to single guanines has been observed (*11*). Intrastrand GpG binding appears preferential.

Actinomycin (*12*, *13*), mitoxantrone (which is structurally related to ametantrone) (*14*), ethidium (*15*) and netropsin (*16*) have been shown to bind cooperatively to various DNAs. We have also shown ametantrone to bind cooperatively. With all of these molecules, the cooperativity can be described using an allosteric model in which initial ligand binding alters the structure of the DNA lattice to enhance the binding of subsequent ligand molecules. For ligands displaying such cooperativity, the malleability of the DNA lattice should affect binding behavior.

One means of detecting sequence selectivity in ligand binding to DNAs is to use restriction enzyme activity assays. This method has been employed to examine the binding of species as diverse as ethidium (*17*), cisplatin (*18*), and the carcinogen N-acetoxy-N-acetyl-2-aminofluorene (*19*). The initial assumption of this method was that the presence of a bound ligand at the cleavage site for a particular restiction enzyme would block the restriction enzyme from the site – thus inhibiting cleavage. However, it must be remembered that, like most DNA binding enzymes, restriction enzymes initially bind "randomly" to the DNA, then move along the DNA until the target sequence is located. Binding of a ligand in a region flanking the target sequence could affect the movement of the enzyme and thus affect the subsequent activity of the enzyme. For ligands which bind cooperatively in an allosteric manner, binding of a ligand to a sequence flanking an enzyme site could alter the structure of that site – also affecting the cleavage. Thus observation of a change in restriction enzyme activity at a particular location

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

may suggest that a ligand is binding in the vicinity of that reaction site, rather than specifically at that site.

In this chapter we report on the results of restriction enzyme activity assays of the binding of actinomycin D, ametantrone, cisplatin, ethidium and netropsin (structures of these molecules are given in Figure 1) to the restriction enzyme cleavage sites on phiX 174 RF DNA given in Figure 2. These molecules were chosen because they represent a variety of binding types – intercalation, groove binding, or covalent binding. These restriction ezymes were selected because they cleave this DNA once or twice and because they cleave in the vicinity of a variety of types of sequences. Use of enzymes that only cut the DNA once or twice removes the possibility of confusion which could arise with inhibition-caused partial digestion of enzymes which cleave multiple times. The amount of restriction enzyme used and the digestion time were limited so that partial digestion was observed in the absence of the binding molecules. This allows for the detection of enhanced cleavage as well as inhibition. A mixture of supercoiled and relaxed circular DNA molecules was used so that we could evaluate the competitive binding of each molecule to these two types of topoisomers. The results presented here indicate that using this methodology, topological preferences for the binding of these molecules can be detected. Additionally, inhibition of restriction enzymes not possessing the reported binding sites for the molecules listed indicates that binding at flanking sequences may alter enzyme activity. Depending upon enzyme and the sequence in the vicinity of the enzyme site, enhancement of cleavage as well as inhibition was observed.

We also compared effects of actinomycin binding on the cleavage, by the restriction enzyme EcoRV, of a linear DNA fragment possessing a $(CG)_4$ segment to the cleavage effects for binding of the same sequence missing the $(CG)_4$. Partial sequences of these fragments are given in Figure 3. As can be seen, the Eco RV site is several base pairs from the $(CG)_4$. EcoRV was chosen because it does not possess the GpC sequence preferred by actinomycin. With the fragment possessing the $(CG)_4$, enhanced cleavage is observed at low actinomycin/base pair ratios and inhibition is observed at higher actinomycin/base pair ratios.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Materials and Methods

PhiX 174 RF DNA (a mixture of the supercoiled and relaxed forms) and all restriction enzymes were obtained from Promega/Fisher Scientific. DNA concentrations were determined by uv spectroscopy.

Actinomycin D, cisplatin, ethidium bromide and netropsin were obtained from Sigma/Aldrich. Ametantrone was synthesized in our laboratory from 1,4, 9, 10-tetrahydroxyanthracene and 2-[N-(2′-hydroxyethyl)amino}ethylamine following the procedure of Murdoch, et.al. (*19*). Product verification was accomplished using ^{1}H-NMR and uv spectroscopy. All ligand molecules, except cisplatin, were dissolved in water for use and concentrations of stock solutions were determined by uv spectroscopy. Cisplatin concentrations were determined by weight.

To prepare the DNA fragment containing the $(CG)_4$ segment, a twenty base pair oligomer (sequence is underlined in Figure 3) containing sticky ends compatible with the Sal I restriction enzyme site was synthesized on an Applied Biosystems 380B DNA synthesizer and was inserted into the Sal I site of the plasmid pGEM-5zf(+). The insert-containing plasmid, pCW001, was cloned into E.coli strain JM 110 and the DNA isolated using standard protocol. A 367-base pair fragment containing the insert was isolated by digesting the plasmid with the restriction enzyme HinfI, followed electrophoresis on 6% polyacrylamide gels. A 347-base pair control fragment was isolated from the parent pGEM plasmid. Partial sequences of the fragments are given in Figure 3. Both fragments were ^{32}P end-labeled using alpha-^{32}P-dATP and Klenow fragment prior to use. Fragment concentrations were determined by uv spectroscopy.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

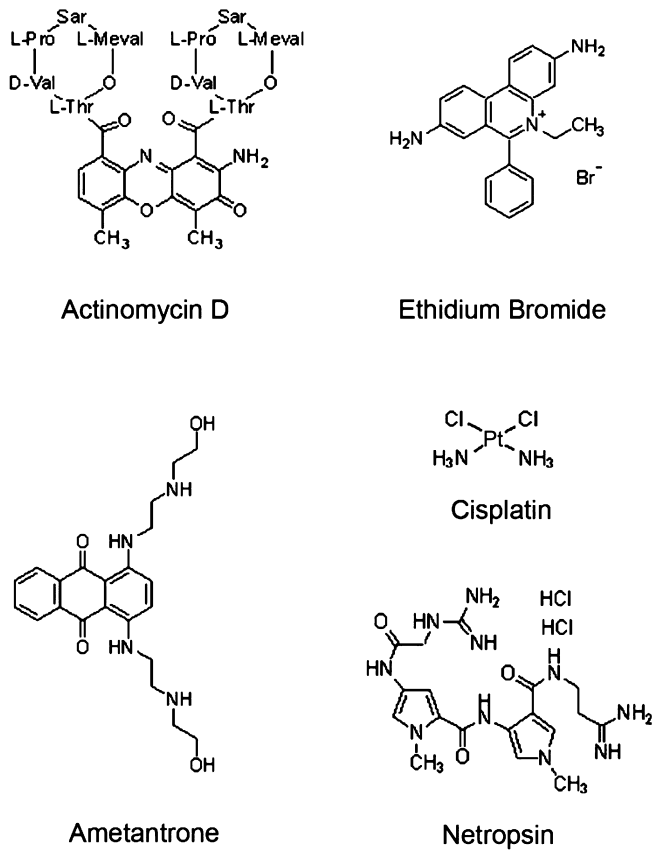


Figure 1. Structures of the DNA-binding ligands used in this study.

RESTRICTION ENZYME CLEAVAGE SITES
W/ FLANKING SEQUENCES

XhoI (162) TTCGACCTATCCTTGCGCAGCTCGAGAAGCTCTTACTTTGCGACCT

Mlu I (221) ACGATTCTGTCAAAAACTGACGCGTTGGATGAGGAGATGGCTTA

Dra I (221) AGAGATTCTCTTGTTGACATTTTAAAAGAGCGTGGATTACTATCT

Dra I (1406) ATTTGATATCTATAGCGTATTTTAAAGCGCCGTGG---ATGCCTGA

Mlu I (2146) CGTTCTGGTGATTTGCAAGAACGCGTACT-TCGCAACCATGATT

Nru I (2260) TATCGCAATCTGCCGACCACTCGCGATTCAATCATGACTTCGTG

Nci I (2800) TTGACGTCCTTCCCCGTACGCCCGGGAATAACGTCTACGTTGGT

Nar I (2970) TTGCTTCTGCTCTTGCTGGTGGCGCCATGTCTAAATTGTTTGGAG

Nru I (4424) GAACGCCCTCTTAAGGATATTCGCGATGAGTATAATTACCCCAA

Stu I (4486) TGAGTGTTCAAGATTGCTGGAGGCCTCCACTAAGATATCGCGTA

Alw44(4779) GCCGTCTTCATTTCCATGCGGTGCACTTTATGCGGACACTTCCTA

Ava II (5042) GTTAACAAAAAGTCAGATATGGACCTTCGCTAAAGGTCTAGGAG

BsshII (5348) ACAAATCTGCTCAAATTTATGCGCGCTTCGATAAAAATGATTGG

Pst I (5382) AAATGATTGGCGTATCCAACCTGCAGGAGTTTTATCGCTTCCAT

Figure 2. Reaction sites with flanking sequences for the restriction enzymes employed in this study.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

pCW 367

GGCCGCGG***GATATC***ACTAGTGCGGCCGCCTGCAGGTCGA
CGCGCGCGATCAGTCATCGAC

pGEM 347

GGCCGCGG***GATATC***ACTAGTGCGGCCGCCTGCAGGTCGA
C

Figure 3. Partial sequences for the pCW001 367 base pair fragment and the pGEM 347 fragment. Cleavage site for Eco RV is bolded and the $(CG)_4$ – containing insert on pCW001 367 fragment is underlined.

Restriction enzyme assays were conducted with each ligand on phiX174RF DNA using the restriction ezymes given in Figure 2. For each enzyme, the buffer supplied by Promega/Fisher was used. 0.2 ug DNA samples were incubated with the indicated ligand for 15 min at 37 C. For actinomycin, ametantrone, ethidium and netropsin, ligand/base pair ratios ranged from 0 – 20 ligand molecules/base pair, while for cisplatin the ratio range was 0 – 0.023 cisplatin molecules/base pair. Samples were then digested for 30 min at 37 C with 3 units of restriction enzyme and the reactions were stopped by the addition of sodium dodecyl sulfate to a concentration of 0.3% and heating at 65 C for 5 min. Digestion products were separated on 1% agarose gels. DNA bands were visualized by ethidium staining and the relative band intensities determined optically. Ethidium stained band intensities were determined to be DNA sequence independent.

Restriction enzyme assays with the DNA fragments were conducted in a similar manner – with the following modifications. Digestion fragments were isolated on 6% acrylamide gels and were visualized using x-ray film. The films were scanned with a densitometer and the relative amounts in the densitometer bands determined by weight.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Results and Discussion

The effects of the binding on restriction enzyme activity can be readily and rapidly observed using gel electrophoresis. Examples of gels for the non-covalent binding ligands are shown in Figure 4. Figure 4a gives the results for the effects of the presence of actinomycin D on the cleavage of phiX 174RF DNA by the restriction enzyme Mlu I which cuts this DNA twice. The first lane, containing DNA which was not treated with either actinomycin or Mlu I, gives the positions of the supercoiled and relaxed DNAs. In the second lane, the DNA was reacted with Mlu I only. A limited amount of Mlu I was employed so that residual uncut supercoiled and relaxed DNA bands are observed as well as the band for partial digestion linear DNA and the two bands for complete digestion. Lanes 3-6 show the results for the addition of increasing concentrations of actinomycin. The complete digestion bands fade out as [actinomycin] increases. Initially, the partial digestion linear band remains, suggesting that one Mlu I site is preferentially inhibited. At the highest actinomycin concentration shown complete inhibition of MluI is observed.

Figure 4b displays a gel for the effects of ametantrone on the digestion of phiX174RF DNA by Pst I which has one cleavage site. Again the first lane shows untreated DNA; lane two has DNA reacted with Pst I only and lanes 3-6 show the effect of increasing [ametantrone]. Inhibition of cleavage on both relaxed and supercoiled substrates is observed with some preference for inhibition with the relaxed DNA.

In Figure 5, gel results for the effects of covalently bound cisplatin on the activity of BssH II, with one cleavage site (Figure 5a) and Dra I, with two cleavage sites (Figure 5b) are given. One gel feature observed with the covalently binding cisplatin that is not observed with the noncovalently binding ligands is the increasing mobililty shift of the supercoiled DNA with increasing [cisplatin].

Previously, Dabrowiak and co-workers have reported such cisplatin-induced shifts and attributed the shift to unwinding of the supercoiled DNA (*17*). Band shifts for the noncovalent ligands are not observed because the ligand/DNA complexes were disrupted by the addition of sodium dodecyl sulfate to quench the enzyme reactions. With both BsshII and Dra I, the gels reveal the inhibition of cleavage by the addition of cisplatin.

Intensities of the electrophoresis bands were quantitated to obtain digestion percentages. To compare the digestion in the presence of a ligand with the digestion in its absence, ratios of % digestion in the presence to % digestion of DNA alone were determined. These ratios are then plotted versus [ligand]/base pairs. A decreasing ratio indicates inhibition of enzyme activity and an increasing ratio indicates enhanced cleavage activity.

In Figure 6a, the results for the effects of actinomycin binding on BssH II activity are shown. The plots illustrate that actinomycin affects BssH II activity with a relaxed DNA substrate but not with a supercoiled substrate – thus suggesting a topological preference for actinomycin binding. The Bssh II reaction sequence is GCGCGC which would initially appear to be a preferred actinomycin binding site though it has been indicated by Graves and coworkers (*5*, *6*) that flanking bases, forming sequences such as TpGpC, may affect the specificity. Restriction enzyme probing of the Bssh II site suggests that this sequence may be structurally distinctive in a supercoiled DNA (*20*). Since actinomycin has been indicated to bind to poly(dGdC) in an allosterically cooperative manner, perhaps actinomycin cannot alter the structure of a superhelically strained Bssh II site to provide a favored binding site.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

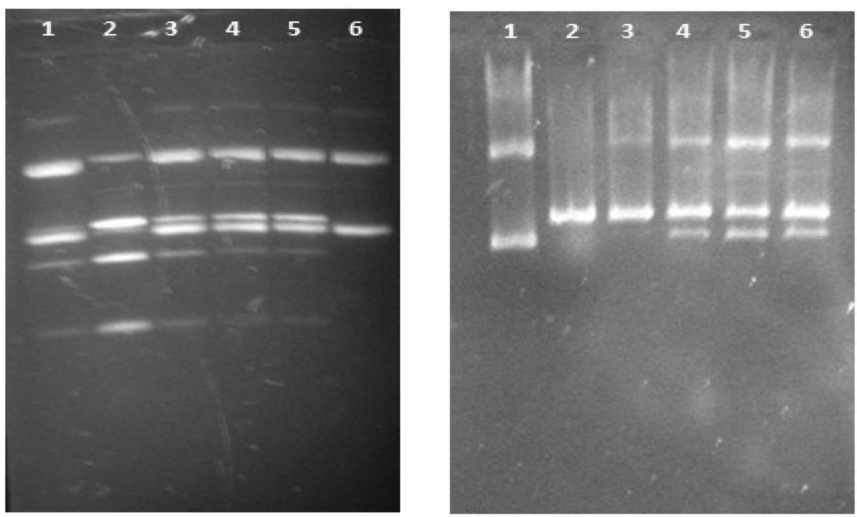


Figure 4. Photographs of 1% agarose gels showing the effects of ligands on restriction enzyme activity on phiX 174 RF DNA. 4a: effects of actinomycin D on Mlu I activity. 4B: effects of ametantrone on Pst I activity. For both, lane I – DNA alone; lane 2 –DNA cut by the enzyme; lanes 3-6, cleavage by the enzyme in the presence of increasing [ligand].

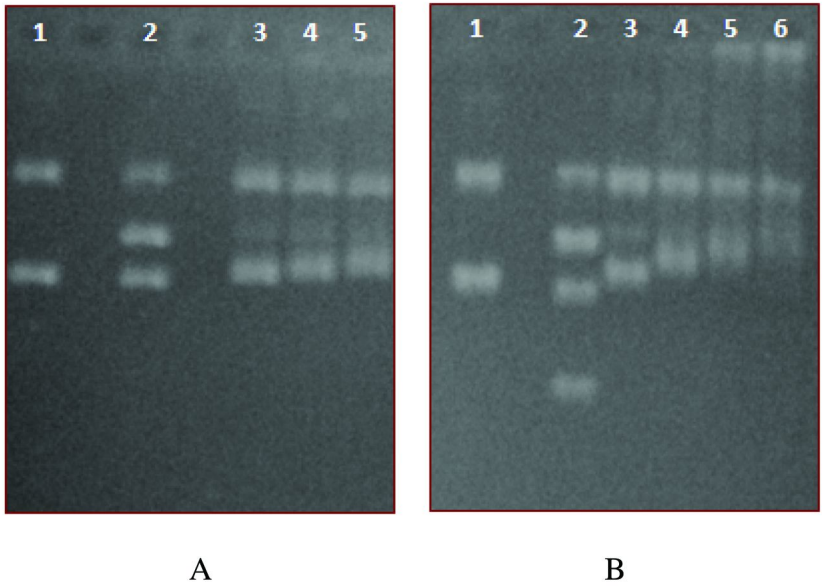


Figure 5. Photographs of 1% agarose gels showing the effects of cisplatin on restriction enzyme activity on phiX 174 RF DNA. 5A: effects of cisplatin on BssH II activity; 5B: effects of cisplatin on Dra I activity. For both, lane 1 – DNA alone; lane 2 – DNA cut by the enzyme; lanes 3-5 (Figure 5a) and lanes 3-6 (Figure 5B), cleavage in the presence of increasing [cisplatin].

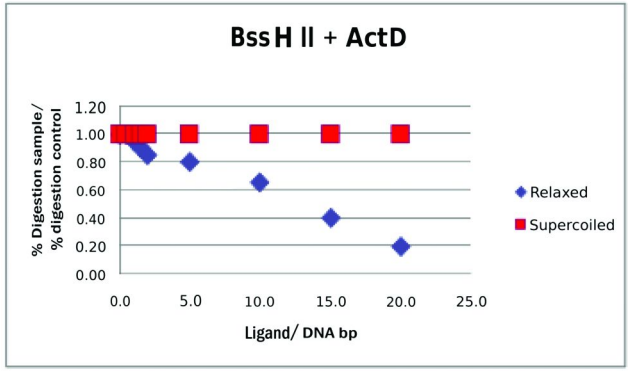
Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Figure 6b shows the results for the effects of ametantrone on Ava II activity. The reaction sequence for Ava II is GGACC. Ametantrone inhibits cleavage of the supercoiled DNA but not cleavage of the relaxed DNA. This topological distinction in the effects of ametantrone binding may arise from the observations that ametantrone has also been shown to bind cooperatively (*13*) and that ametantrone possesses side chains capable of interactions with the groove or backbone of the DNA. The preferred sequence for ametantrone binding has been reported to be a GpC, which is not found in the Ava II site. There is a GpC near the site, perhaps suggesting that cleavage at flanking locations alters restriction enzyme activity.

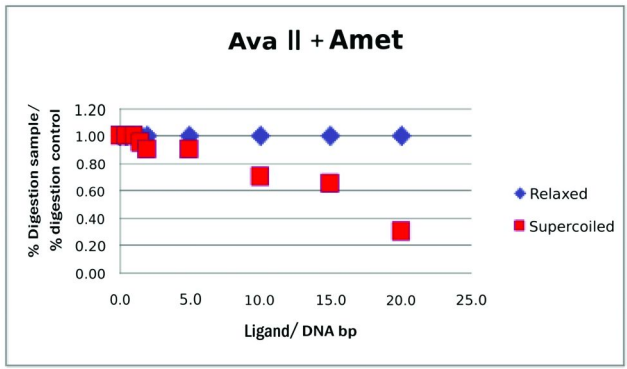
In Figure 6c, the results for how ethidium binding affects Ava II activity are given. In contrast to the effects of ametantrone, ethidium shows no topological specificity.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

A



B



C

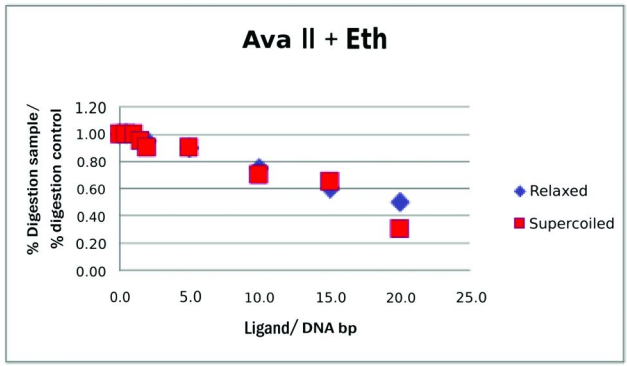
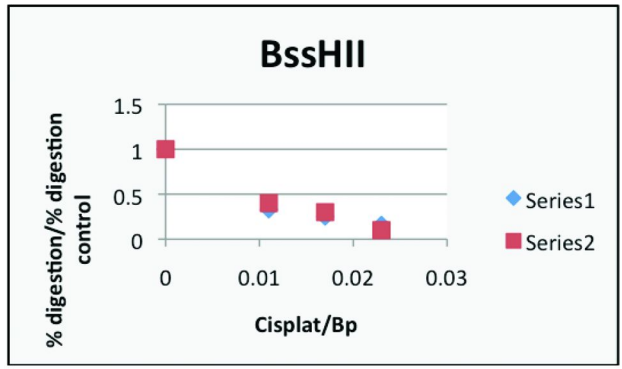


Figure 6. Plots of % digestion sample/% digestion DNA alone versus [ligand]/[base pairs]. 6A: actinomycin with BssH II; 6B: ametantrone with Ava II; 6C: ethidium with Ava II.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

A



B

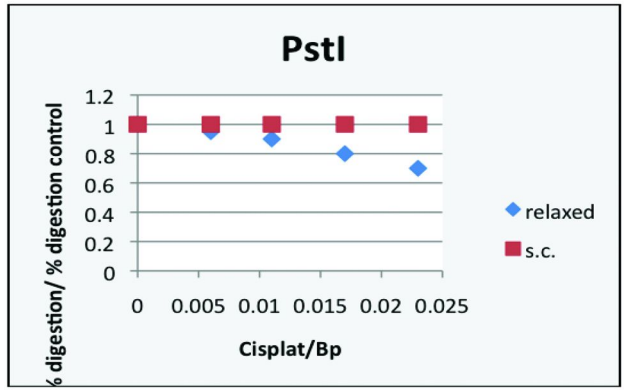


Figure 7. Plots of % digestion sample/% digestion DNA alone versus [cisplatin]/[base pairs]. 7A: cisplatin with BssH II; 7B: cisplatin with Pst I.

In Figure 7a, results with cisplatin and Bssh II are presented. The bound ligand/base pair ratios are much lower than those for the noncovalent ligand. However, the presumption is that all cisplatin molecules are covalently bound, while with the noncovalent ligands the [ligand] are for the total amount of ligand present, not the bound species. The cisplatin/ base pair ratios are also much lower than those employed in previous studies (*17*), indicating that cisplatin can affect the activity of enzymes on DNAs at very low ratios. Bssh II activity is inhibited on both supercoiled and relaxed DNA. The Bssh II reaction site does not contain a GpG and the closest GpG locations are 30 and 20 base pairs to either side. Either the cisplatin is forming an interstrand crosslink at the Bssh II site or the presence of a bound cisplatin can have a long range effect on activity. Noting that restrction enzymes bind at random on a DNA and subsequently move to the reaction site, noting that cisplatin does unwind the DNA, and noting that, in this study, cisplatin effectively alters restrction enzyme activity at low ratios, the second explanation could be plausible.

The data in Figure 7b suggests that cisplatin may show topological specificity in its effects at certain locations or with certain restriction enzymes. Reactivity of Pst I with relaxed DNA is inhibited by the presence of cisplatin while reactivity

with supercoiled DNA is unaffected. Although the Pst I site does not contain a preferred cisplatin target, there are cisplatin reaction sites in the vicinity, explaining why the cleavage of linear DNA is inhibited. The lack of an observed effect on the supercoiled DNA could suggest that either supercoiling alters the DNA in this location to prevent cisplatin binding or that cisplatin, on binding, does not twist the DNA in a manner that prevents Pst I activity.

The results for effects of all ligands in the study, actinomycin, ametantrone, ethidium, netropsin and cisplatin are given in Table 1. Certain results are not surprising. For example, ethidium inhibited activity of virtually all of the restriction enzymes employed. This is not surprising, given that ethidium exhibits no great sequence specificity in its binding to DNA (*9*). Actinomycin and ametantrone showed inhibition of restriction enzymes possessing a GpC in the reaction sites – when the substrate DNA is relaxed. Again this is expected, given the reported sequence specificities (*1–7*). Netropsin inhibits cleavage by Dra I which recognizes TTTAAA when the DNA is relaxed. This follows the observed specificity (*8*). Cisplatin inhibits restriction enzymes which have a GpG in their cleavage sites – AvaII, Nci I, Stu I – when the DNA is relaxed.

Other results are more interesting. For example, actinomycin was found to inhibit Dra I with linear DNA. Obviously, there is no actinomycin binding site in the Dra I reaction site. There are GpC's in the vicinity and thus the results provide a clear example of the effect of binding to a flanking sequence on restriction enzyme activity. Actinomycin also generally did not inhibit restriction enzymes when the DNA was supercoiled. As noted above, this may be the result of the need by actinomycin to twist the DNA on binding. Generally, inhibition of cleavage was observed for the various ligands. Actinomycin, at lower [actinomycin]/base pair ratios, produced enhanced cleavage by the restriction enzyme Pst I. Enhanced activity by a restriction enzyme would result from the DNA being twisted into a more reactive form. This could arise from actinomycin altering the DNA structure on binding.

Netropsin also demonstrated inhibition of several enzymes that do not possess the traditional AT –type sequence in the reaction site. Again, there are AT sequences in the vicinity and it was observed that longer AT sequences gave more inhibition. Netropsin has been observed to bind in an allosterically cooperative manner. Cisplatin demonstrates effects on restriction enzyme activity when the only GpG's are in flanking sequences. One notable case presented is that of Dra I. Obviously, the Dra I site is GpG-less. The closest GpG's are at 25 and 11 base pairs from the first Dra I site and at 25 and 6 base pairs from the second. Clearly, binding of cisplatin can affect activity over a distance. These experiments were performed with a mixture of supercoiled and relaxed DNAs, providing for a comparision of the competitive binding to these two substrates. As shown in the table, in certain locations on the substrate phiX174RF DNA, certain ligands showed inhibition of restriction enzyme activity on relaxed DNA but not inhibition with supercoiled DNA – or visa versa. This suggests that the ligands actinomycin, ametantrone, netropsin and cisplatin do show topological specificity when binding to certain sequences. Previous work by Ren and Chaires using a competitive equilibrium dialysis method also showed topological specificity for various ligands (*21*).

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Table 1. Results of Restriction Enzyme Activity Assays

	Alw	*Ava-II*	*BssH*	*Dra-I*	*Mlu*	*Nci*	*Nru*	*Pst*	*Stu*	*Xho-I*
Act	IR	IR	IR	PIR	IR	PIR	IR	ER	PIR	ER
	NSC	ISC	NSC	NSC	ISC	NSC	NSC	ESC	NSC	NSC
Amt	NR	IR	NR	IR	NR	IR	NR	PIR	IR	NR
	ISC	NSC	NSC	PISC	ISC	ISC	ISC	ISC	NSC	ISC
Eth	ER	PIR	PIR	IR	IR	IR	IR	PIR	ER	IR
	ESC	PISC	PISC	ISC	PISC	NSC	PISC	ISC	ESC	PISC
Net	IR	IR	ER	IR	IR	IR	PIR	IR	PIR	PIR
	PISC	PISC	ESC	ISC	ISC	ISC	ISC	PISC	NSC	PISC
Cis		IR	IR	IR	IR	IR		PIR	IR	IR
		PISC	ISC	ISC	ISC	PISC		NSC	PISC	ISC

Table Legend: IR = Inhibition Relaxed DNA; ISC = Inhibition Supercoiled DNA; PIR = Partial Inhibition Relaxed; PSC = Partial Inhibition Supercoiled; ER = Enhancement Relaxed; ESC = Enhancement Supercoiled; NR = No Effect Relaxed; NSC = No Effect.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

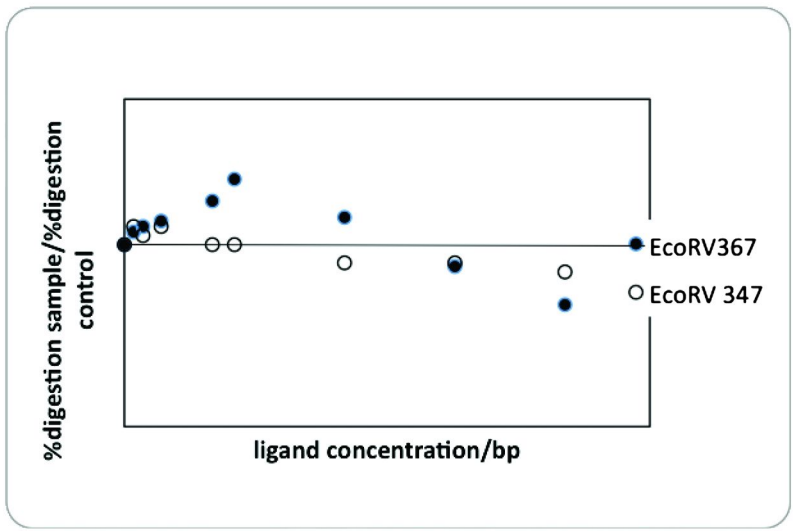


Figure 8. Plot of % digestion sample/% digestion DNA alone versus [actinomycin]/[base pairs] for showing the effects of actinomycin on Eco RV activity on the pCW 367 and pGEM 347 fragments.

In Figure 8, data showing the effects of actinomycin D binding on the activity of Eco RV on two linear DNA fragments are presented. One of these fragments, pCW 367 possesses a $(CG)_4$ insert twenty five bases pairs from the Eco RV site (Figure 3). Comparison of the reactivity of this fragment with Eco RV in the presence of actinomycin with that of a fragment (pGEM 347) which is missing the insert indicates that alteration of the DNA sequence at some distance from a restriction enzyme site also alters how a binding ligand affects enzyme reactivity. As Figure 8 shows, actinomycin binding initially significantly enhances the activity of Eco RV on the $(CG)_4$ containing fragment. Enhancement of cleavage could arise from the actinomycin, in its cooperative binding, changing the structure of the DNA. This structural alteration has previously been observed by Krugh and Walker (*12*). The same actinomycin concentrations affect reactivity on the control fragment to a much smaller degree. While there are other GPC's in the area, the $(CpG)_4$ may be a major target for actinomycin binding. If the actinomycin is binding there, this suggests that actinomycin binding can affect DNA structure and functioning over a distance. Even if the actinomycin is binding between the insert and the Eco RV site, the presence of the $(CG)_4$ clearly alters how the actinomycin affects the DNA structure.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Conclusions

This study demonstrates the efficacy of using restriction enzyme activity assays to explore various facets of the binding of small molecules to DNAs. Relatively small amounts of the DNA substrate are required; the binding at several types of sequences and multiple topologies can be examined fairly rapidly; the technique works with a simple intercalator (ethidium), a groove binder (netropsin), groove-interacting intercalators (actinomycin and ametantrone) and a covalent binder (cisplatin); and the analysis can be done with gel electrophoresis.

This study specifically showed that the binding of various small molecules can enhance restriction enzyme activity as well as inhibit it. Ligand binding to sequences flanking restriction enzyme sites was inferred from the data. Further, the topology of the DNA does affect the binding of ligands possessing moieties capable of "grabbing" the DNA.

Acknowledgments

This work was supported in part by PHS MBRS GM 08205-07. The authors would also like to thank Dr. Nicholas Combates for assistance in preparing the pCW001 clone.

References

1. Muller, W.; Crothers, D. M. *Eur. J. Biochem.* **1975**, *54*, 267–277.
2. Krugh, T. R.; Neely, J. W. *Biochemistry* **1973**, *12*, 1775–1782.
3. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271–9285.

4. Lane, M. J.; Dabrowiak, J. C.; Vournalis, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3260–3264.
5. Bailey, S. A.; Graves, D. E.; Rill, R.; Marsch, G. *Biochemistry* **1993**, *32*, 5881–5887.
6. Bailey, S. A; Graves, D. E.; Rill, R. *Biochemistry* **1994**, *33*, 11493–11500.
7. Lown, J. W.; Morgan, A. R.; Yen, S.-F.; Wang, Y.-H.; Wilson, W. D *Biochemistry* **1985**, *24*, 4028–4035.
8. Marky, L. A.; Breslauer, K. J *Biochemistry* **1987**, *84*, 4359–4363.
9. Waring, M. J. *J. Mol. Biol.* **1965**, *13*, 269–282.
10. Sherman, S. E.; Lippard, S. J. *Chem. Rev.* **1987**, *87*, 1153.
11. Winkle, S. A.; Krugh, T. R. *Nucleic Acids Res.* **1982**, *9*, 3175–3181.
12. Walker, G. T.; Stone, M. P.; Krugh, T. R. *Biochemistry* **1991**, *24*, 7471–7479.
13. Rosenberg, L. S.; Balakrishnan, M. S.; Graves, D. E.; Lee, K. R.; Winkle, S. A.; Krugh, T. R. In *Biological Activity of Polymers*; Carraher, C. E., Gebelein, C. G., Eds.; ACS Symposium Series 186; American Chemical Society: Washington, DC, 1982; pp 269−281.
14. Winkle, S. A.; Rosenberg, L. S.; Krugh, T. R. *Nucleic Acids Res.* **1982**, 8211–8223.
15. Dattagupta, N.; Hogan, M.; Crothers, D. M. *Biochemistry* **1980**, *19*, 5998–6005.
16. Hardwick, J. M.; Von Sprecken, R. S.; Yielding, K. L.; Yielding, L. W. *J. Biol. Chem.* **1984**, *259*, 11090–11097.
17. Zhang, B.; Seki, S.; Akiyama, K.; Tsutsui, K.; Li, T.; Nagao, K. *Acta Med. Okayama* **1992**, *46*, 427–434.
18. Mallamaci, M. A.; Bascoy, M. L.; Brown, J; Combates, N. J.; Winkle, S. A. *J. Biomol. Struct. Dyn.* **1992**, *10*, 83–96.
19. Murdock, K. C. *J. Med. Chem.* **1979**, *22*, 1024–1030.
20. Thomson, H. R; Winkle, S. A. Florida International University, unpublished results.
21. Ren, J.; Chaires, J. B.; *Drug−Nucleic Acids Interactions (Methods in Enzymology)*; Chaires, J. B., Waring, M. J., Eds.; Academic Press, Inc.: San Diego, 2001; Vol. 340, pp 99−108.

Downloaded by CORNELL UNIV on June 2, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ch010

Subject Index

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ix002

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ix002

E

F

G

H

I

In Frontiers in Nucleic Acids; Sheardy, R., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ix002

T

U

W

X

Downloaded by 89.163.35.42 on June 1, 2012 | http://pubs.acs.org
Publication Date (Web): October 25, 2011 | doi: 10.1021/bk-2011-1082.ix002