Specific Binding Affinity of the Non-Catalytic Domain of Eukaryotic Like Type IB Topoisomerase of vaccinia Virus

Benjamin R. Reed

The Graduate Center, City University of New York
SPECIFIC DNA BINDING AFFINITY OF THE NON-CATALYTIC DOMAIN OF EUCARYOTIC LIKE TYPE IB TOPOISOMERASE OF vaccinia VIRUS

BY

BENJAMIN REED

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2016
SPECIFIC DNA BINDING AFFINITY OF THE NON-CATALYTIC DOMAIN OF EUCARYOTIC LIKE TYPE IB TOPOISOMERASE OF vaccinia VIRUS

by

Benjamin Reed

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Ranajeet Ghose

Chair of Examining Committee

Date

Brian R. Gibney

Executive Officer

Supervisory Committee:

Ranajeet Ghose

Ruth Stark

Nancy Greenbaum

Stewart Shuman

David Cowburn

THE CITY UNIVERSITY OF NEW YORK
Abstract

SPECIFIC DNA BINDING AFFINITY OF THE NON-CATALYTIC DOMAIN OF EUCARYOTIC LIKE TYPE IB TOPOISOMERASE OF vaccinia VIRUS By

by
Benjamin Reed

Advisor: Ranajeet Ghose

Topoisomerases are ubiquitous proteins that alter supercoiling in double stranded DNA (dsDNA) during transcription and replication and vaccinia and the closely related poxvirus variola virus, at 314 amino acids in length, encode the smallest of the type I topoisomerases(TopIB). TopIB is a two domain protein that recognizes the sequence 5’-T/CCCTT, cleaves at the 3’-end and relaxes supercoiling through rotation. The C-terminal domain (CTD) alone contains the catalytic activity and specificity. Deletion of the N-terminal domain results in a greatly reduced rate of relaxation and rapid dissociation. Biochemical data suggests that the N-terminal domain (NTD) is important for pre-cleavage binding and affinity for the target site. A combination of NMR-based interaction studies, the measurement of backbone dynamics using $^{15}$N relaxation measurements, and isothermal calorimetry (ITC) is used in this work to show that the NTD is capable of independently binding to DNA. Additionally, it is shown that the nature of the engagement of dsDNA by the NTD, in terms of affinity and characteristics of the binding modes, differs between sequences containing the 5’-CCCTT segment from those that do not. An attempt is made to extend these observations to the full length protein.
Acknowledgements

I would like to thank my wife who has been my primary cheerleader. Without her love and support I would not be where I am today. I would also like to thank my parents and step-parents for encouraging me and their unwavering faith in my ability to succeed.

I would like to thank Dr. Ranajeet Ghose. His expertise and insight has been an inspiration. He has always been my strongest advocate at CCNY and throughout the CUNY system. I can think of no other professor I’d rather have as a mentor.

I would like to thank the current and former members of Ronnie’s research group. Especially Dr. Deniz Temel, Dr. Zhen Ren, Dr. Venkatesh Ramakrishnan, Dr. Andrea Piserchio and Dr. Sebastien Alphonse for teaching me the basics in biochemistry lab techniques. I would especially like to thank Dr. Piserchio and Dr. Alphonse for the many interesting conversations we’ve had as I propose my ideas to them. I would also like to thank current lab members Imane Djemil, Jonathan Lee, Nathan Will and Alicia Sponholtz for their help and support.

I thank my committee members Dr. David Cowburn, Dr. Nancy Greenbaum, Dr. Stewart Shuman and Dr. Ruth Stark for their valuable feedback at our every meeting.

I would like to extend a special thank you to the current and former staff of the New York Structural Biology Center for their invaluable help in collecting and interpreting NMR experiments.
**Table of Contents**

**ABSTRACT**

**acknowledgements**

**TABLE OF CONTENTS**

**LIST OF FIGURES**

**LIST OF TABLES**

**ABBREVIATIONS**

1 **INTRODUCTION**

1.1 DNA structure and topology

1.2 Topoisomerases

1.2.1 Classification of topoisomerases

1.2.2 Type I topoisomerases

1.3 Protein-DNA Interactions

1.4 vaccinia / variola Type IB topoisomerase function

1.4.1 Functional roles of the TopIB N-terminal and C-terminal domains

1.4.2 Reaction kinetics of TopIB catalytic cycle

1.4.3 Overall structure of TopIB in the free and bound forms

1.4.4 Role of the N-terminal domain in DNA recognition

1.5 NMR spectroscopy

1.5.1 Backbone and sidechain assignment experiments

1.5.2 Study of protein-ligand interactions by NMR

1.5.3 Study of protein dynamics by spin relaxation measurements

1.6 Isothermal titration calorimetry
1.7 Questions to be addressed 27

2 MATERIALS AND METHODS 28

2.1 Cloning, expression and purification of TopN 28

2.2 Cloning, expression, purification and refolding of TopC 31

2.3 Cloning, expression, purification and refolding of TopIB and TopIB<sub>Y274F</sub> 32

2.4 NMR spectroscopy 33

2.4.1 Backbone assignment of TopN 33

2.4.2 Backbone assignment of TopC 35

2.4.3 Transfer of backbone assignments to TopIB and TopIB<sub>Y274F</sub> 36

2.4.4 NMR-based titrations with TopN in isolation and in the context of TopIB 37

2.4.5 NMR spin-relaxation measurements 40

2.5 Isothermal Titration Calorimetry measurements of TopN/DNA interactions 42

3 RESULTS 42

3.1 Measurement TopN/DNA interactions using ITC 42

3.2 Resonance assignment of TopN 45

3.3 Evidence for TopN binding to dsDNA 45

3.4 Comparison of TopN binding to nsDNA and spDNA 49

3.5 The effect of salt on CSPs of nsDNA and spDNA bound to TopN 61

3.6 nsDNA and spDNA interactions of the NTD in the full length protein 65

3.7 Hydrodynamic and dynamic features of TopN in the absence or presence of dsDNA 67

4 CONCLUSION 75

REFERENCES 83
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DNA Supercoiling</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Catalytic mechanisms of topoisomerase</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Structures of major classes of topoisomerase</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Structure of variola topoisomerase IB</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Specific contacts of NTD with DNA</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>NMR experiments for protein backbone and sidechain assignment</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>TopN FPLC shows monomeric peak</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>spDNA and nsDNA sequences</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>Measurement of the affinity of TopN to DNA duplexes</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>Possible loss of interactions with TopN in the nsDNA construct</td>
<td>44</td>
</tr>
<tr>
<td>11</td>
<td>Probability plot of TopN secondary structure</td>
<td>46</td>
</tr>
<tr>
<td>12</td>
<td>Assigned 15N, 1H HSQC spectrum of TopN (600 MHz)</td>
<td>48</td>
</tr>
<tr>
<td>13</td>
<td>Salt titration CSPs</td>
<td>51</td>
</tr>
<tr>
<td>14</td>
<td>CSPs induced on TopN in the presence of dsDNA</td>
<td>52</td>
</tr>
<tr>
<td>15</td>
<td>Nature of perturbations for specific TopN resonances in the presence of dsDNA molar ratios beyond 1:1</td>
<td>53</td>
</tr>
<tr>
<td>16</td>
<td>Multiple binding nsDNA Glu69 side chain</td>
<td>53</td>
</tr>
<tr>
<td>17</td>
<td>Differences in CSPs for nsDNA referenced to spDNA</td>
<td>60</td>
</tr>
<tr>
<td>18</td>
<td>Effects of ionic strength on TopN/DNA interactions</td>
<td>64</td>
</tr>
<tr>
<td>19</td>
<td>Normalized salt effect on TopN in the presence of nsDNA and spDNA</td>
<td>66</td>
</tr>
<tr>
<td>20</td>
<td>CSPs of NTD of full length TopIBY274F in presence of DNA</td>
<td>66</td>
</tr>
<tr>
<td>21</td>
<td>15N relaxation measurements for TopN</td>
<td>73</td>
</tr>
</tbody>
</table>
Figure 22 Differences in high frequency motions of apo TopN vs. spDNA bound TopN 73

List of Tables
Table 1 Sequence identity between vaccinia and variola Type IB topoisomerases 10
Table 2 Contacts between vaccinia NTD and spDNA substrate. 14
Table 3 Residues most perturbed by the presence of dsDNA 51
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTD</td>
<td>Carboxyl terminal domain of TopIB</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded DNA</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>NTD</td>
<td>Amino terminal domain of TopIB</td>
</tr>
<tr>
<td>TopIB</td>
<td>vaccinia/variola type 1B topoisomerase</td>
</tr>
<tr>
<td>TopC</td>
<td>carboxyl terminal truncation mutant of TopIB (aa 81-313)</td>
</tr>
<tr>
<td>TopN</td>
<td>amino terminal truncation mutant of TopIB (aa 1-80)</td>
</tr>
<tr>
<td>TROSY</td>
<td>Transverse Relaxation Optimized Spectroscopy</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 DNA structure and topology

The double helical structure of DNA with its inward facing bases has been recognized as imparting chemical stability to the bases.\(^1\), \(^2\) In speculating on DNA being the source of genetic information Watson and Crick also correctly identified the necessity of separating the bases to access the information stored therein. Additionally they anticipated that a double helix, unlike a single strand, is susceptible to the formation of tangled structures.\(^3\) Fully relaxed DNA has 10.5 base pairs per turn. In circular DNA and compacted cellular DNA the DNA can be treated as if it is fixed at both ends. Most cellular DNA is less tightly wound (under-wound). This facilitates compaction and lowers the energetic cost of strand separation for proteins that interact with DNA. DNA can also be more tightly wound (over-wound). The degree of over or under-winding can be described analytically by counting the total number of turns in a DNA structure. This number is called the linking number (Lk). Fully relaxed circular double stranded DNA with 2,100 base pairs will have 200 \((2,100/10.5)\) total turns. This is referred to as \(Lk_0\) and represents the relaxed state linking number. If the DNA were to be under-wound by two turns the new linking number would be \(Lk = 198\). The difference in linking number is \(\Delta Lk = Lk – Lk_0 = -2\). The state of winding in DNA is thus measured as the specific linking difference defined as \(\Delta Lk = (Lk – Lk_0) / Lk_0\). In the previous example \(\Delta Lk = -0.02\). In the process of transcription for example the DNA becomes overwound preceding the transcription bubble and the following strands become under-wound.\(^4\) With sufficient under-winding the DNA double helix responds by wrapping around itself forming a super helical structure. Cells rely on this in order to compact the genome and promote unwinding to aid in protein interactions of replication and promoter regions.\(^5\) If the
tension of over winding preceding transcription and replication is not released it can lead to strand breakage and apoptosis.\textsuperscript{6}

In order to be more quantitative in describing this winding behavior of DNA there has been defined two terms. Twist describes the number of times a strand of DNA is wrapped around another. By convention, a right-handed helix has a positive twist and a left-handed helix has negative twist. Writhe describes the number of times that two strands wrap around two other strands. Excessive twist will lead to writhe and form supercoils as previously described. (\textbf{Figure 1} below)

1.2 Topoisomerases

Topoisomerases are ubiquitous enzymes that are found in all domains of life. Their function is required for DNA processes such as transcription, replication\textsuperscript{7-9}, recombination\textsuperscript{10}, and repair\textsuperscript{11, 12}. Their control of DNA topology is essential for maintaining genome integrity and stability.\textsuperscript{5, 13} They are typified by their ability to break and re-seal the DNA backbone and pass DNA strands through the gap.\textsuperscript{14, 15} With this function they are able to unwind, unknot and untangle DNA and resolve complex structures.\textsuperscript{16} Inhibition of topoisomerases is the basis of several classes of antiviral and anticancer chemotherapeutics.\textsuperscript{17} Included in the category of type I topoisomerase inhibitors are camptothecins such as irinotecan and topotecan. In human type IB topoisomerase (hTopIB) these compounds act by locking the enzyme in the covalent complex with DNA thus preventing dissociation.\textsuperscript{18} Pox virus TopIB is insensitive to this class of drugs due to its minimal structure which lacks the regions that are contacted by the drugs.\textsuperscript{19} The fluoroquinolone enrofloxacin inhibits DNA relaxation by Vaccinia topoisomerase I by causing the enzyme to supercoil relaxed DNA.\textsuperscript{20}
Figure 1 DNA Supercoiling

DNA topology showing the effects of positive and negative turns of DNA, called twist (top), and positive and negative supercoiling, also called writhe (bottom). Topoisomerases alter the twist and writhe of DNA. In doing so they alter the linking number of the DNA strand. Image provided for the public domain by Ascendinglotus2 (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons
1.2.1 Classification of topoisomerases

There are two major classes of topoisomerases. Type I topoisomerases act by cutting one strand of DNA while simultaneously making a covalent bond with the backbone on one side of the cut. They allow the cut strand (scissile strand) to relax by rotating around the uncut strand (non-scissile strand) for a certain number of turns and thus alter the twist of DNA. (Figure 1 above) This is followed by a reversal of the ligation reaction rejoining the DNA phosphate bonds accompanied by the breaking of the protein/DNA covalent bond. They change the linking number by single integer values. Type I topoisomerases relax positive supercoils for example by reducing the twist to the point that the writhe decreases.

Type II topoisomerases remove super-helical tension by covalently bonding to one double helix of a super-helix, cutting both strands of the DNA backbone and pulling another double helical strand through the break, re-annealing the DNA backbone and breaking the protein/DNA covalent bond and thus they alter the writhe of DNA directly. (Figure 2 below) All type II topoisomerases require ATP as an energy source to drive the reaction. They alter the linking number by values of 2. Type II topoisomerases are not the topic of this dissertation and will not be discussed further. Reviews of their structure and function have been covered by Berger⁷, Sharma⁸ and Schoffler⁹
Type IA topoisomerases bind a single-stranded region of dsDNA. They cut one strand and transfer the other single strand through the break. Consequently, they increase the twist by $Tw = +1$. Type IB topoisomerases create a break in one of the strands (scissile strand) and relax DNA by permitting rotation about the un-cut strand (non-scissile strand). Therefore, the change in twist is a somewhat random integer $n$: $Tw = \pm n$. Type II topoisomerases bind to a dsDNA segment (the gate, G-segment), cut both strands. They then pull a second dsDNA (T-)segment through the break in the first. In that process, they consume two ATP molecules and change the writhe of supercoiled DNA by two units: $Wr = \pm 2$. Figure modified from Annu. Rev. Biophys. Biomol. Struct. 34, 201-19 (2005).
Figure 3 Structures of major classes of topoisomerases
Examples of the classes of topoisomerases showing the diversity of structures. **Type IA** (PDB: 1I7D, e. coli Topoisomerase III), **Type IB** (PDB: 3IGC, vaccinia virus Topoisomerase I), **Type IC** (PDB: methanopyrus kandleri Topoisomerase V), **Type IIB** (PDB: 1D3Y, methanococcus jannaschii Type IIB Topoisomerase DNA binding core, full structure not available)
1.2.2 Type I topoisomerases

Type I topoisomerases are further divided into three groups. (see Figure 2 above) Type IA topoisomerases are found in eukaryotes, bacteria and archaea. They are structurally different from the other two classes of type I topoisomerases. (see Figure 3 above) Their activity is dependent upon divalent metal ions (Mg$^{2+}$, Zn$^{2+}$), they can only relax positive supercoils and they change the linking number of a circular strand of DNA by only one. In their catalytic cycle, type IA topoisomerases form a covalent bond with the 5’ end of the DNA backbone.

In contrast type IB topoisomerases can relax positive and negative supercoils, are metal independent though low levels of divalent metals and some nucleosides enhance or inhibit their activity. Originally thought to only exist in eukaryotes and poxviruses they have since been found in all domains of life. Unlike Type IA topoisomerases, members of the type IB family make a covalent bond with the 3’ end of the phosphate backbone where they nick the DNA. (Figure 2 above) Type IC topoisomerase is comprised exclusively of archaeal topo V. It is structurally distinct from type IB topoisomerases in spite of making a 3’ link to DNA, relaxing positive and negative supercoils and being ATP independent. (Figure 3 above)

1.3 Protein-DNA Interactions

Early studies of protein-DNA binding generally assumed that there was a simple code to explain sequence specificity. The first attempts at developing this code therefore analyzed the specific hydrogen bonding patterns of amino acid side chains and the DNA bases. Since that time as more DNA bound protein structures have been published it has become apparent that these simple rules are insufficient to fully explain binding specificity. Comprehensive studies of side chain hydrogen bonding have shown that these specific interactions do exist however the majority of hydrogen bonding is to the DNA backbone and is largely not base specific. Some
helix-turn-helix proteins, including the 434 repressor, have shown binding that is dependent on linker region sequences between the binding sites even though linker bases are not contacted directly.\textsuperscript{32}

Such results have led to the recognition of the importance of direct, (hydrogen bonding, steric fit, etc.) and indirect (DNA structural adaptation) components. Rohs et al. have argued that direct and indirect readouts describe idealized extremes that rarely exist in real protein-DNA complexes and propose that a better understanding can be developed using the concepts of base readout and shape readout.\textsuperscript{33} Base readout is accomplished by specific base contacts including hydrogen bonding and hydrophobic contact. Shape readout involves the DNA and protein having structural complementarity of binding surfaces in order to maximize relevant interactions. It has generally been recognized that binding specificity involves a combination of these recognition components.\textsuperscript{34}

1.4 vaccinia / variola Type IB topoisomerase function

Vaccinia and the closely related variola virus type IB topoisomerases were chosen for this study because of their similarity to eukaryotic type IB topoisomerases. These poxviral topoisomerases are biochemically identical, share >99\% sequence identity\textsuperscript{35} and are similar to their eukaryotic counterparts in that they relax positive and negative supercoils in DNA and they are stimulated by but not dependent on divalent cations.\textsuperscript{36} (Additionally, these viruses encode the smallest of the eukaryotic-like type IB topoisomerases. This is valuable as its small size makes it amenable to NMR spectroscopy allowing an analysis of weak interactions that may not be seen in crystallo and dynamics that may be related to function. Vaccinia topoisomerase (TopIB) is a two-domain protein (314 residues; 36.6 kDa) consisting of an N-terminal domain (NTD; 1-80) and a C-terminal catalytic domain (CTD; 81-314) connected by a flexible linker that is sensitive
to proteolysis.\textsuperscript{37} The active site of apo TopIB is not pre-assembled prior to binding with DNA.\textsuperscript{38} TopIB recognizes the sequence 5’-(T/C)CCTT-3’ in dsDNA, leading to the assembly of the active site, cleavage and simultaneous formation of a covalent bond between the 3’-end of the target backbone phosphate and a conserved tyrosine (Y274).\textsuperscript{39, 40} Relaxation of both positive and negative supercoils occurs by frictionally constrained rotation of the scissile strand around the non-scissile strand.\textsuperscript{41} Following relaxation, the catalytic cycle reverses with the re-sealing of the DNA backbone phosphodiester bond and the protein moves to a new site to repeat the process.

\section*{1.4.1 Functional roles of the TopIB N-terminal and C-terminal domains}

TopIB has been shown to bind dsDNA containing the consensus 5’-CCCTT-3’ sequence with a 7 to 10-fold higher affinity than a sequence missing the consensus pentamer.\textsuperscript{42} Cheng and Shuman showed that the CTD alone is capable of catalytic activity while retaining specificity for the target sequence but truncation of the NTD slows the single turnover rate of cleavage of DNA by greater than 10,000-fold compared to the full-length enzyme. In addition, they showed that TopIB relaxes DNA completely while the CTD alone results in partial relaxation indicating premature dissociation with minimal effect on the rate of single turnover re-ligation. The CTD alone was shown to bind dsDNA with one-fifth the affinity of the full-length protein. Additionally, the rate of cleavage by full-length TopIB is largely insensitive to NaCl and is stimulated by Mg\textsuperscript{2+}. In contrast the CTD alone is highly sensitive to salt and Mg\textsuperscript{2+}. All of these signs indicate a direct or indirect role for the N-terminal domain in facilitating DNA binding.\textsuperscript{43}
Table 1 Sequence identity between vaccinia and variola Type IB topoisomerases
Sequence identity of vaccinia and variola virus type IB topoisomerases from amino acids 1-180 are shown. The full sequences (1-314 aa) show 99% identity and 99.7% sequence similarity. Amino acids 181-314 have complete identity and are not shown. Differences are indicated by * symbol.

| Vaccinia | 1 MRALFYKDGLFTDNNFLNPVSDDNPAYEVLQHVKIPTHLTDVVYYEQTWEELALRIFV |
| Variola | 1 MRALFYKDGLFTDNNFLNPVSDDNPAYEVLQHVKIPTHLTDVVYYGQTWEELALRIFV |

vaccinia
61 GSNSKGRRQYFYGKMHVQRNRAKDRIFRVRVYNNVMKRINCIFNKNIKKSSTDNSNYQLAVF
variola
61 GSNSKGRRQYFYGKMHVQRNRAKDRIFRVRVYNNVMKRINCIFNKNIKKSSTDNSNYQLAVF

vaccinia
121 MLMETMFFIRFGKMKLYKETVGLLTLKNKIEISPDIEVIKFGKDVSHEFVHSKSN
variola
121 MLMETMFFIRFGKMKLYKETVGLLTLKNKIEISPDIEVIKFGKDVSHEFVHSKSN

**Figure 4 Structure of variola topoisomerase IB**
A) variola TopIB bound to dsDNA (PDB: 3IGC). The N terminal domain (NTD) and C terminal domains (CTD) are colored in green and light brown respectively. The catalytic tyrosine 274 is colored red and the scissile thymine is colored orange. The scissile and non-scissile strands are colored yellow and white respectively. B) Interaction of the NTD with duplex DNA. The secondary structure elements of TopN are labeled and the DNA duplex is shown in surface representation.
1.4.2 Reaction kinetics of TopIB catalytic cycle

Crystal structures of TopIB complexed with DNA containing the CCCTT sequence show that it binds DNA circumferentially.\textsuperscript{44} This involves a clamping down of the structure concurrent with assembly of the active site.\textsuperscript{45} Under low salt conditions \(\text{Mg}^{2+}\) increases the activity of wild type enzyme \(\sim 9\) fold with excess DNA by enhancing the product off rate.\textsuperscript{46,47} Kinetic studies by Kwon and Stivers have suggested a two-state reaction pathway, attributable to binding/clamping leading up to catalytic cleavage/relaxation.\textsuperscript{48} Under the reaction conditions that the used (200 mM NaCl) the cleavage was shown to be rate limiting with binding/clamp closing being at least 14-fold faster and opening/dissociation occurring at least 5-8 fold more quickly. In contrast, reactions performed under stringently low salt conditions or with 0.1 M NaCl in the absence of \(\text{Mg}^{2+}\), dissociation is rate limiting.\textsuperscript{47} Experiments performed under high salt had no effect on the maximal rate of cleavage but decreased binding affinity by approximately 50-fold showing the importance of non-covalent binding, specifically electrostatic interactions, to reaction efficiency.

1.4.3 Overall structure of TopIB in the free and bound forms

The X-ray crystal structures of the NTD both as a 1-77 truncation mutant\textsuperscript{49} (PDB:1VCC) and in the amino terminal domain of dsDNA bound full-length TopIB\textsuperscript{45,50} (PDB: 2H7F, 2H7G, 3IGC) show a fairly orderly structure composed of two alpha helices and a five stranded antiparallel beta sheet with connecting loops Figure 4 (above). In PDB structure 1VCC the carboxyl terminal residues (73-77) are largely unstructured. In the dsDNA bound full-length x-ray structures residues 73-106 form a bent helix (\(\alpha_3\)) bridging the two domains and curving around the DNA.

The catalytic domain of TopIB in the absence of DNA is composed of ten \(\alpha\) helices and a three stranded \(\beta\) sheet that is highly conserved among all eukaryotic Type IB topoisomerases.\textsuperscript{38}
Catalytically important residues which form the active site are distributed on one face of the protein. Arg130 is located at the loop between $\alpha_4$ and $\alpha_5$, Lys167 is on a turn between $\beta_7$ and $\beta_8$, Arg223 is on $\alpha_9$, His265 is in a loop between $\alpha_{10}$ and $\alpha_{11}$ and Tyr274 which makes the covalent bond with the DNA backbone is also on $\alpha_{11}$. Upon binding to DNA multiple secondary structural rearrangements occur. $\alpha_3$ extends from Gly73 to Ile106 as described above. $\alpha_4$, which is composed of the residues from Tyr115 to Thr125 in the free protein, extends spanning Ser113 to Phe127. $\alpha_5$ Asn140 - Leu146 in the free protein shifts to Met134-Asn140 upon binding. A short helix forms ($\alpha_7$, Lys167-Lys169) from the loop between $\beta_7$ and $\beta_8$. The helix that contains the key residue Tyr274 (a12, Ile269-Met282) breaks in to two $\alpha$ helices at Tyr274 ($\alpha_{12}$, Pro267-Ala273 and $\alpha_{13}$, Thr277-Met282). These shifts in secondary structure are accompanied by a 23° rotation of the segment spanning $\alpha_4$-$\alpha_7$ and the beta sheet composed of $\beta_6$-$\beta_8$ relative to helices $\alpha_8$-$\alpha_{12}$ of free TopIB. This rearrangement is essential to the formation of the active site as it assembles the active site comprised of Arg130, Lys167, Arg223 and His265 and Tyr274 and brings the catalytic Tyr in range of the minor groove to allow the formation of the covalent phosphodiester bond.

1.4.4 Role of the N-terminal domain in DNA recognition

The crystal structures of TopIB by Perry et. al revealed the close contacts of the NTD with the dsDNA oligo Figure 4 (above). The fifth $\beta$-strand of the NTD is inserted deep in to the major groove of the DNA and makes several base specific contacts with the CCCTT sequence. Gln69 makes a double hydrogen bond with the adenine base in the +2 position; Tyr70 and Tyr72 lie along the major groove and make van der Walls contact with the CCCTT sequence of bases and sugars and electrostatic contact with the DNA backbone phosphates. Additional interactions
are seen between the side-chains of several basic residues (Lys35, His39, Lys65, Arg67, His76, Arg80) and the sugar-phosphate backbone of the DNA duplex (Figure 5).

Of these residues Arg67, Gln69, Tyr70 and Tyr 72 have been identified as being important for specific binding site affinity without altering catalytic specificity.\textsuperscript{45, 51-55} Alanine mutations of Arg67 and Arg80 in TopIB stimulated relaxation of supercoiled pUC19 plasmid however the addition of 5 mM Mg\textsuperscript{2+} caused a reduction in relaxation rate in contrast to the wild type enzyme that is stimulated by similar amounts of Mg\textsuperscript{2+}. The pUC19 plasmid also showed signs of intermediate topoisomers in the R80A mutant in the presence of Mg\textsuperscript{2+} indicating premature dissociation.\textsuperscript{51} The addition of 0.1 M NaCl and Mg\textsuperscript{2+} compounded the rate reduction.\textsuperscript{52} An alanine mutation of Tyr70 did not have a detrimental effect on relaxation rates in the presence of salt but was inhibited by magnesium. The Y72A mutation had a very detrimental effect on DNA relaxation. In the presence of salt, the rate was decreased by about 8 times in relation to the wild type enzyme and intermediate topoisomers were apparent. The addition of Mg\textsuperscript{2+} and salt slowed the rate to < 1% of the wild type. None of these mutations had a significant effect on the rate of single turnover relaxation. This indicates that they are not essential for catalysis but rather are important for binding affinity. A further discussion of mutations of Tyr 70 and Tyr72 is presented in the RESULTS section.
Figure 5 Specific contacts of NTD with DNA
Two views of selected sidechain contacts made by TopN with the specific sequence element.

<table>
<thead>
<tr>
<th>Residue</th>
<th>DNA position</th>
<th>PDB Structure</th>
<th>DNA Side Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3IGC Bond Distance</td>
<td>2H7G Bond Distance</td>
</tr>
<tr>
<td>K35</td>
<td>Between +2/+3</td>
<td>3.0</td>
<td>n/a</td>
</tr>
<tr>
<td>H39</td>
<td>-1(sugar), Between -1/-2</td>
<td>3.3</td>
<td>4.4</td>
</tr>
<tr>
<td>K65</td>
<td>Between +1/+2</td>
<td>3.1</td>
<td>4.7</td>
</tr>
<tr>
<td>K67</td>
<td>Between +1/+2</td>
<td>2.6, 3.1</td>
<td>2.5, 3.0</td>
</tr>
<tr>
<td>Q69</td>
<td>+2</td>
<td>2.8, 3.1</td>
<td>2.9, 3.1</td>
</tr>
<tr>
<td>Y70</td>
<td>Between +4/+5, +2:+3 (base), +3 (sugar)</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Y72</td>
<td>Between +3/+4, +2:+3 (base), +3 (sugar)</td>
<td>2.8</td>
<td>2.7</td>
</tr>
<tr>
<td>H76</td>
<td>Between +2/+3</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>R80</td>
<td>+1 (base), Between +2/+3</td>
<td>3.0, 3.1</td>
<td>2.8, 2.9</td>
</tr>
</tbody>
</table>

Table 2 Contacts between vaccinia NTD and spDNA substrate.
Residues above the dashed line make contact with the non-scissile strand of DNA. Those below make contact with the scissile strand. Distance measurements are from side chain nitrogen to DNA backbone oxygen except in the case of Q69. No backbone amide contacts are seen in the interaction of NTD and spDNA. H: hydrogen bond, V: van der Waals contact. K35 Bond distances in 2H7G, 2H7F are not shown because DNA is not present in a location to allow contact. Q69 distances are between glutamine side chain nitrogens and NH$_2$(6), N(7) of adenine base. Ne is 3.0 Å from O6 of D63 in 3IGC, 3.3 Å in 2H7G, 2.8 Å in 2H7F and 2.9 Å in 1VCC structures suggesting a salt bridge between D63 and Q69 in addition to the bonds between Q69 and the +2 DNA base. R80 hydrogen bonds involve primary and secondary amines.
1.5 NMR spectroscopy

Nuclear magnetic resonance is a powerful tool for the study of protein-ligand interactions with atomic level resolution. NMR has been used to characterize protein-peptide / protein–protein, and protein-nucleic acid interactions.\textsuperscript{56-58} Experiments carried out in the solution state are capable of testing under near physiological conditions. One of the major considerations and limitations of NMR is the size of the protein/complex under study. Larger molecules have a slower tumbling rate (rotational correlation time; $\tau_c$). This is detrimental as it allows for a greater relaxation rate resulting in the rapid loss of signal and consequently poor sensitivity. Modern NMR experimental techniques such as TROSY\textsuperscript{59} and CRINEPT\textsuperscript{60} along with deuteration have greatly increased the size of complexes available for analysis by NMR.\textsuperscript{61} NMR signals are very sensitive to the magnetic environment around the nuclei. Changes in dihedral angle in the protein backbone and electron density around the nuclei being probed will alter the resonance frequencies and therefore the positions of the signals in NMR spectra. Perturbations in NMR signals can be evidence of van der Waals binding, hydrogen bonding, electrostatic interaction and conformational changes in proteins. For single state binding on a timescale that is fast compared to the chemical shift timescale one can track the chemical shift changes as the protein goes from the majority unbound to the majority bound state. (vide infra)

NMR is a phenomenon in which nuclei, when placed in a magnetic field, absorb electromagnetic energy and emit it at specific frequencies in the radio frequency (RF) range. When placed in a magnetic field the magnetic moments of nuclei and electrons rotate about the axis of the magnetic field in a process called Larmor precession. The precession frequency is proportional to the strength of the magnetic field at the nucleus and property of nuclei called the gyromagnetic ratio. When irradiated at the Larmor frequency, also called the resonant frequency,
the nuclei emit a radio frequency signal due to a property called spin. The isotopes of biologically important nuclei $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$ and $^{31}\text{P}$ have two distinct energy levels associated with their nuclear spin ($S = \pm \frac{1}{2}$) called $\alpha$ and $\beta$ states. The RF pulses cause a transition between the two levels, which generates magnetization that can be relayed to other nearby nuclei and/or detected.

Specific interactions between nuclei lead to changes in their energy levels. These interactions may occur through the bonding electrons via scalar coupling also called J-coupling. This type of coupling is generally not effective beyond 4 bonds. Nuclei may also interact through space by dipolar coupling. The magnitude of this interaction is dependent on the angle between the axis of the bulk magnetic field and a vector connecting the interacting nuclei. Isotropic tumbling in solution averages this coupling to zero due to this orientational dependence.

Electrons around the nucleus will circulate in the magnetic field. This in turn generates a magnetic field that interacts vectorially with the applied magnetic field. This alters the overall magnetic field experienced by a nucleus in a phenomenon that is referred to as shielding. Differences in the electron density around the nuclei due to different chemical environments lead to differences in the local magnetic field and therefore alter the resonant frequency of particular nuclei. Molecular bonds are anisotropic and therefore produce a spatially variant magnetic field. This alters the resonant frequencies of nearby nuclei in a manner that is dependent on their electronic and therefore chemical environment. Therefore, unique chemical environments produce unique resonant frequencies that are reflective of molecular structure. These unique signatures relative to bare nuclei are called chemical shifts.

Fourier transform NMR spectroscopy involves an excitation pulse followed by conversion of the time domain signal (FID) into frequency domain information (the spectrum).
For many pulse sequences the FID depends on a second time variable. If one terms the first time interval as $t_1$, and the interval in which the FID is recorded as $t_2$, then the FID may be considered as a function of both times. In a 2D experiment acquisition is separated from the excitation stage by intermediate stages in which spins can be allowed to precess (evolution) for a variable period of time and often after additional pulses a second period of time (mixing). This is followed by the observe pulse and signal acquisition. Successive experiments are acquired in which time $t_1$ is varied and a complex FID signal is acquired. With sufficient increments in $t_1$ values a 2D Fourier transform is performed. The spectrum is represented as a topographical map of frequency in dimension one, frequency in dimension two and intensity at each point. In experiments of this type magnetization transfer is measured. This can be through bonds as in the COSY and TOCSY experiments, to another type of nucleus such as in a $^1$H,$^{15}$N-HSQC or through space as in a NOESY experiment. Thus it is possible to measure the changes in the chemical shifts of the $^1$H and $^{15}$N nuclei of the protein backbone amides upon interaction with ligand.

### 1.5.1 Backbone and sidechain assignment experiments

The ultimate goal of backbone assignments in this work is to correlate each backbone amide $^1$H and $^{15}$N chemical shift pair to a particular amino acid in the protein sequence. This allows the possibility of residue level analysis of protein behavior under various test conditions such as changes in salinity, pH or the response to the presence of a ligand.

Consider an amino acid in a polypeptide chain. We’ll call it the ‘i’th residue. The amino acid attached to the amino terminal end of the ‘i’th is referred to as the ‘i-1’th residue to the preceding residue. The amino acid connected to the carboxyl end of residue ‘i’ is the ‘i+1’ or following residue. Triple resonance experiments have been developed to correlate the backbone and side-chain carbon resonances to the backbone amides.$^{62}$ A standard procedure for backbone
assignment entails collecting an experiment that all other experiments are referenced to. A
common experiment, which was chosen in this work, is the HNCO experiment. This correlates
the ‘i’th residue amide $^1$H-$^{15}$N resonances to the ‘i-1’ carboxyl carbon resonance. The
HN(CA)CO experiment correlates the amide resonances to the ‘i’th and ‘i-1’th carboxyl carbon
resonances. To ascertain the identity of the amino acid associated with each amide resonance
HN(CO)CACB and HN(CA)CO experiments are frequently used. The former correlates the
‘i’th residue amide signals with the ‘i-1’ Cα and Cβ carbon resonances. The latter produces
resonances of Cα and Cβ for both the ‘i’th and ‘i-1’ residues. By comparing these two it is
possible to determine which of the resonances in the HNCACB experiment are from the ‘i’
residue and which are from ‘i-1’. The Cα and Cβ chemical shifts of the 20 amino acids each
have a specific pattern though for some amino acids there is some overlap, e.g. phenylalanine
(Cα δ = 58.13 ± 2.59 ppm, Cβ δ = 39.94 ± 2.07 ppm) and tyrosine (Cα δ = 58.17 ± 2.51 ppm,
Cβ δ = 39.28 ± 2.15 ppm), leading to ambiguities. The HNCA and HN(CO)CA experiments are
similar to the HNCACB and HN(CO)CACB experiments however they only correlate the amide
signals to the Cα carbons. The smaller spectral window required for the Cα chemical shift range
allow for the acquisition of higher resolution data in a reasonable time. This facilitates the
assignment of Cα resonances that are highly degenerate. To help address any ambiguities of
amino acid type for each amide C(CO)NH-TOCSY can be performed. This correlates side
chain protons and carbons of the ‘i-1’ residue to the ‘i’th residue amide signal. Each amino acid
type has a distinctive pattern. If the 3D structure of the protein is known the use of the
HNNNOESY-TROSY experiment can be performed to correlate amide protons and nitrogens
that are within 6 Å via polarization transfer due to the nuclear Overhauser effect.
Figure 6 NMR experiments for protein backbone and sidechain assignment

HNCO correlates the $^1$H, $^{15}$N amide signals to the i-1 $^{13}$C carbonyl. HN(CA)CO correlates the amide signals of the ith residue to the ith and i-1 $^{13}$C carbonyls. HNCACB correlates the amide signals to the $^{13}$C signals of the ith and i-1 C$\alpha$ and C$\beta$. HN(CO)CACB correlates the amide signals to the i-1 C$\alpha$, C$\beta$ signals only. HCCONH-TOCSY correlates the amide signals of the ith residue to all of the non-carbonyl carbons of the i-1 backbone and sidechain. $^{15}$N NOESY TROSY correlates the $^{15}$N of amides to all protons near in space (~6Å). Images from Protein NMR a Practical guide (http://www.protein-nmr.org.uk/).
1.5.2 Study of protein-ligand interactions by NMR

As described previously the chemical shift of a nucleus is extremely sensitive to changes in the local environment, be it changes in bond angle, hydrogen bonding or electrostatic interaction with charged ligands such as DNA. All of these interactions will alter the density and geometry of the electronic environment around the nucleus. In general, the largest changes in chemical shift occur at the interaction site. If a ligand does not bind then there will be no appreciable chemical shift changes seen.\textsuperscript{57} Due to this it is possible to track changes on the per residue basis.

The timescale of the interaction has a large influence on the chemical shift changes seen in a protein during the course of titration with ligand.\textsuperscript{69} When the rate of the chemical exchange between states (k\textsubscript{ex}) is much greater than the frequency difference of the resonances for each state (|ω\textsubscript{1} - ω\textsubscript{2}| = Δω) it is said to be in fast exchange. The observed chemical shift is located at the population-weighted average of the chemical shifts of the states (p\textsubscript{1}δ\textsubscript{1} + p\textsubscript{2}δ\textsubscript{2}). Fast exchange is generally suggestive of weak binding. If single site binding on the fast timescale occurs the visible resonances will progress linearly from the position of the free state to that of the bound state with increasing concentration of ligand. Ultimately the protein will reach saturation and the positions of the resonances will cease to change appreciably. If multiple site binding occurs resonances following non-linear paths will be observed throughout the titration. At the opposite extreme is slow exchange in which k\textsubscript{ex} \ll Δω. During a titration one will see a gradual decrease of the resonances of state one and the appearance and increase in intensity of the resonances due to state two. Slow exchange is generally suggestive of strong binding. Additionally, intermediate exchange, where k\textsubscript{ex} \approx Δω results in broadening effects. In any of these time regimes, changes mapped on to structure indicate the location the binding surface.
1.5.3 Study of protein dynamics by spin relaxation measurements

At room temperature proteins are inherently flexible. These motions are essential for protein function. Proteins are not locked in to a single structure but sample some set conformations in a time dependent fashion. A wide range of functions are affected by dynamic motion including catalytic turnover of enzymes, and binding via induced fit or conformational selection.

NMR allows for the study of local structural dynamics at atomic level resolution due to the influence of dynamics on the relaxation of nuclei that may be quantitatively estimated using specific measurements. A common probe for protein dynamics is the backbone amide \(^{15}\)N nucleus. The advantages of this type of analysis include the fact that the \(^{1}\)H-\(^{15}\)N group form an isolated spin pair to a good approximation and is often well resolved in \(^{1}\)H-\(^{15}\)N correlation spectra.

Nuclear spin relaxation in solution is facilitated by temporal fluctuations in the magnetic field at each observed nucleus due to molecular tumbling and internal motions. These fluctuations are due to alterations in dipolar and scalar interactions with surrounding nuclei and changes in the electronic environment that affect chemical shielding. For backbone \(^{15}\)N nuclei the major sources of spin relaxation due to dynamics are the dipolar interaction between \(^{15}\)N and its attached proton and the anisotropy of the \(^{15}\)N chemical shift tensor.

Three relaxation parameters that are most often measured in the study of protein dynamics through \(^{15}\)N relaxation are the longitudinal or spin-lattice relaxation rate (\(R_1\)), the transverse or spin-spin relaxation rate (\(R_2\)), and the steady state Nuclear Overhauser Effect (ssNOE). \(R_1\) relaxation corresponds to the process of establishing or re-establishing the normal
population distribution of $\alpha$ and $\beta$ spin states in the magnetic field. $R_1$ relaxation is caused by transient magnetic fields, usually due to molecular motion, at the Larmor precession frequency.

$R_2$ relaxation is caused by transient magnetic fields, usually due to molecular motion, at all frequencies and chemical shift swapping or changing coupling constants called chemical exchange, $R_{ex}$. These types of interactions lead to phase decoherence of signals in the X-Y plane. Slower tumbling, $R_{ex}$ in the $\mu$s-ms range and inhomogeneity in the static magnetic field increases $R_2$ relaxation rates. Motions in the ps-ns range and anisotropic rotation cause a decrease in $R_2$.

The NOE is an induced population change in the $\alpha$ and $\beta$ states of nuclei transmitted through dipolar interactions. The $\{^1H\}^{15}N$ NOE is reduced in the presence of internal flexibility and can be below zero due to the fact that the gyromagnetic ratio of $^{15}N$ is negative.

For a protein that is internally rigid and undergoes isotropic tumbling then the $R_1$, $R_2$ and $\{^1H\}^{15}N$ NOE values will be approximately the same for each residue. Any values far from the average are indicative of residue specific internal motions and/or anisotropic molecular tumbling.

The Fourier transforms of the correlation functions describing the reorientations of internuclear bond vectors are the spectral density functions, $J(\omega)$. They are directly related to the longitudinal, $R_1$, transverse, $R_2$, and the ssNOE, relaxation parameters. The value of $J(\omega)$ represents the magnitude of motions at frequency $\omega$. Equations (1) - (3) describe the relationship between the spectral density functions and relaxation parameters for the backbone $^1H$-$^{15}N$ pair.73,
\[ R_1 = 3\left( d^2 + c^2 \right) J(\omega_N) + d^2 \left[ J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N) \right] \]  

\[ R_2 = \frac{1}{2} \left( d^2 + c^2 \right) \left[ 4J(0) + 3J(\omega_N) \right] + \frac{1}{2} d^2 \left[ J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) \right] + R_{ex} \]  

\[ NOE = 1 - \left[ \frac{\gamma_H}{\gamma_N} \right] \frac{d^2 \left[ 6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) \right]}{R_1} \]  

Where \( d = \mu_0 \gamma_H \gamma_N/(4\pi r^3) \) is the strength of H-N dipolar coupling, \( c = -\omega_N \text{CSA}/3 \) represents the influence of chemical shift anisotropy, \( \omega_H \) and \( \omega_N \) are the resonant frequencies of \(^1H\) and \(^{15}N\) nuclei respectively, \( \mu_0 \) is the vacuum permeability, \( h \) is the reduced Planck’s constant, \( r \) is the length of the H-N amide bond, \( R_{ex} \) is the conformational exchange contribution to \( R_2 \), and \( \gamma_{H,N} \) are the proton and nitrogen gyromagnetic ratios respectively.

Of the frequencies to be considered, those near the \(^1H\) frequency i.e. \( J(\omega_H) \) are considered to be the most diagnostic for motions in the ps-ns range.\(^74\) Farrow et. al. observed that the high frequency terms of the spectral density function are approximately equal to the that of proton spectral density function.\(^75\) i.e.;

\[ J(\omega_H) \approx J(\omega_H \pm \omega_N) \]  

Leading to the supposition that these terms may be substituted with a single term \( J(\omega_h) \). This approximation was supported by experimental data presented by Peng and Wagner\(^77,78\).

If one assumes that the spectral density function is of the form \( J(\omega) = \lambda_c/\omega^2 + \lambda_i \), where \( \lambda_c \) is the contribution to \( J(\omega) \) due to molecular tumbling and \( \lambda_i \) is the contribution from internal motions, and the requirement that \( 6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) = cJ(\omega_c) \), where \( c \) is some constant and \( J(\omega_c) \) is some equivalent frequency, it is possible to represent \( \omega_c \) as;
\[ \omega_c = \frac{5}{\sqrt{\left[ 1 + \frac{\gamma_H}{\gamma_N} \right]^2 - \left[ 1 - \frac{\gamma_H}{\gamma_N} \right]^2}} = 0.87 \omega_H \quad (5) \]

Through a similar process it is possible to recast the frequency relationships in equations (1) - (3) as:

\[ J(0.870\omega_H) = \frac{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)}{5} \quad (6) \]
\[ J(0.921\omega_H) = \frac{J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)}{7} \quad (7) \]
\[ J(0.955\omega_H) = \frac{J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)}{13} \quad (8) \]

If the correlation function describing dynamics is a sum of exponentials and if \(\tau_c \gg 1\), where \(\tau_c\) is the molecular rotational correlation time, and \(\tau_i \ll 1\) where \(\tau_i\) are the internal motional correlation times then \(J(\omega) = \lambda_c/\omega^2 + \lambda_i\) provides exact solutions. Based on numerical simulations by Farrow et. al. the above is also accurate for all values of \(\tau_c\) if the spectral density function is a sum of Lorentzians.

If the form of the spectral density function is as described by Lipari and Szabo\textsuperscript{79, 80}, equation (9) below, with \(\tau_c, S^2\) and \(\tau_i\) having the usual range of values for proteins, \(2 \text{ ns} \leq \tau_c, 0.1 \leq S^2 \geq 1, 0.001 \text{ ns} \leq \tau_i\) then the differences between the two sides of equations (6) - (8) are less than 0.8%.

\[ J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1-S^2) \tau}{1 + \omega^2 \tau^2} \right] \quad (9) \]

\((1/\tau = 1/\tau_c + 1/\tau_i)\)
Thus it is possible to represent equations (1) - (3) as;

\[ R_1 = \frac{d^2\left[3J(\omega_N) + 7J(0.921\omega_H)\right]}{4} + c^2J(\omega_N) \quad (10) \]

\[ R_2 = \frac{d^2\left[4J(0) + 3J(\omega_N) + 13J(0.955\omega_H)\right]}{8} + c^2\left[3J(\omega_N) + 4J(0)\right] \quad (11) \]

\[ NOE = 1 + \frac{d^2\gamma_H 5J(0.870\omega_H)}{4R_i\gamma_N} \quad (12) \]

Where d, c, and the \( \gamma \)s have the previously defined values. The value of \( J(0.870\omega_H) \) may therefore be calculated from NOE and \( R_1 \) data by way of equation (12). It is important to note that the above methods do not require explicit knowledge of the spectral density function and molecular tumbling does not need to be isotropic. They do require the assumption that the spectral density function is a slowly varying function between \( \omega_H + \omega_N \geq \omega \geq \omega_H - \omega_N \). It also does not accurately account for residues with a high \( R_{ex} \) contribution to relaxation unless data is collected at multiple fields.

### 1.6 Isothermal titration calorimetry

In the study of protein ligand interactions an important question is how tightly does a protein bind to its ligand? An important analytical technique that can answer this is isothermal titration calorimetry (ITC). Nearly all chemical reactions and physical changes are accompanied by a change in heat (enthalpy). The first instance of ITC used to study binding of a biological system was performed by Beaudette and Langerman in 1978 to measure the free energy of binding of ADP to bovine liver glutamate dehydrogenase.\(^8\) It is now routinely used to characterize the thermodynamics of bimolecular binding interactions.\(^8\)\(^2\)\(^,\)\(^3\)

ITC is a two-cell calorimetric method in which the sample is placed in the test cell which is constantly stirred to ensure solution homogeneity. A reference cell is connected to the sample
cell by a thermally conducting material. Surrounding both cells is an insulating jacket. Constant voltage is applied to a heater for the reference cell. A sensitive thermocouple circuit detects temperature differences between the two cells. This controls a heater located in the sample cell. During the experiment ligand is titrated in to the sample cell causing either heating or cooling of the cell. The time-dependent change in power needed to maintain a constant temperature between the two cells is recorded. This reading can be directly correlated to the amount of heat removed or evolved from the system. As the protein approaches saturation of the ligand less and less heat is evolved. The resultant titration curve can be used to calculate the binding constant $K_B$ or dissociation constant $1/K_B = K_D$. The relevant thermodynamic relationships for binding are listed below:

\[
K_{eq} = \left(\frac{[\text{Complex}]}{[\text{Protein}]\times[\text{Ligand}]\right)}
\]

\[
\Delta G^\circ = -RT\ln K_{eq}
\]

\[
\Delta G = \Delta G^\circ + RT\ln \left(\frac{[\text{Complex}]}{[\text{Protein}]\times[\text{Ligand}]\right)}
\]

\[
\Delta G = \Delta H - T\Delta S
\]

Where $K_{eq}$ is the equilibrium constant. In the case of binding $K_{eq} = K_B$, $[X]$ is the equilibrium concentration of species $X$, $\Delta G^\circ$ is the standard Gibbs fee energy change, $R$ is the molar gas constant, $T$ is the temperature in Kelvin $\Delta G$ is the Gibbs free energy change, $\Delta H$ is the enthalpy change and $\Delta S$ is the entropy change for complex formation. If done under optimum conditions one is able to calculate $K_B$, $\Delta G$, $\Delta H$, $\Delta S$ and $n$, the stoichiometry of the interaction. Optimal conditions require that the heat change for each injection be measurable and the heat varies sufficiently with each injection to produce a curved thermogram. Even under less than optimal
conditions it is possible to fit the curve and calculate $K_B$ if the curve includes the inflection point and ends sufficiently close to saturation.

1.7 Questions to be addressed

The biochemical and structural data discussed above in section 1.4.4 suggest that TopIB NTD may be involved in scanning the DNA for the specific sequence. If this is the case the following conditions must be true. (1) The NTD must have an inherent DNA-binding ability. For this result to be biologically relevant it must also bind DNA in the same mode as in full-length TopIB. The detrimental effect of removing the NTD seen in the CTD truncation protein and the salt and magnesium sensitivity imparted by select alanine mutations on the NTD as discussed in section 1.4.4 suggest that reasonable to expect this. Hydrogen bonding is exothermic and due to the high degree of hydrogen bonding present in the crystal structures that include the DNA bound NTD it is reasonable to expect binding by the NTD to DNA be exothermic. ITC is therefore an ideal experimental method for determining if TopN binds to DNA in a similar manner to the NTD of the full length protein. If topN binds to DNA in a similar mode to the full length protein, then the ITC curve is expected to be strongly exothermic. If there is a poorly defined or non-existent binding curve, then one can conclude that TopN is binding to DNA in a different manner than in the full length protein or not binding at all. By assigning the $^1H$-$^{15}N$ NMR signals of the backbone and amide side-chains of TopN and tracking the chemical shift perturbations as described in 1.5.2 upon addition of DNA it will be possible to identify the binding surface and thereby verify that it is binding DNA with a similar mode to the full length protein. (2) The NTD must possess a binding preference for DNA duplexes that contain the consensus CCCTT sequence over those that do not. The extreme level of specificity exhibited by TopIB and specific contacts made by sidechains within the NTD to the target sequence suggest
that this may be so. ITC will determine if there is a higher binding affinity for specific DNA than for non-specific DNA. If TopN is able to detect a difference in the DNA sequence, there should be a detectable difference in binding behavior between DNA that contains the target sequence and DNA that does not. This will be apparent in the $^1$H-$^1$H NMR experiments. If chemical shift perturbations are identical in the presence of nsDNA or spDNA then this suggests that there is no specific response to the two DNA sequences. If, however, chemical shift perturbations show divergent behavior then this will suggest that there is a binding specificity. An example of divergent behavior would be the chemical shift perturbations following different trajectories in the presence of nsDNA than is observed in the presence of spDNA. Due to the specific contacts made by the NTD with the target sequence one would also expect the chemical shift perturbations to be reflective of single site binding in the presence of spDNA. No such requirement would be necessary with nsDNA.

2 Materials and Methods

2.1 Cloning, expression and purification of TopN

The region encoding the N-terminal 80 residues of vaccinia virus topoisomerase (1-80; TopN) was excised from a pET-21b plasmid coding for the full-length protein (1-314) by NdeI/BamH1 digestion and inserted between the NdeI and BamH1 restriction sites of a pET-15b expression vector. The resultant plasmid encoding TopN fused to an N-terminal thrombin cleavable (His)$_6$-tag was transformed into E. coli BL21 CodonPlus (DE3) RIL competent cells (Agilent Technologies). Transformed cells from a single colony were introduced into M9 media containing ampicillin (100 mg/mL) and grown at 37 °C until the $A_{600}$ reached a value of 0.6. The M9 medium contained 1 g/l of $^{15}$NH$_4$Cl (for uniform $^{15}$N-labeling) and 1.5 g/l $^{13}$C$_6$-D-glucose (for uniform $^{15}$N, $^{13}$C-labeling) as the only sources of nitrogen and carbon, respectively.
Expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 400 µM. The temperature was reduced to 16 °C following induction and cells were grown under agitation for 24 hours. Subsequently cells were harvested by centrifugation at 6750 g for 20 minutes at 4 °C and re-suspended in lysis buffer A: 50mM Tris, 1 mM NaCl, 10 % glycerol, pH 7.5 supplemented with 1 tablet of protease inhibitor (Complete mini EDTAfree, Roche). Cells were lysed by sonication on ice (0.3s on, 1s off) and debris removed by centrifugation at 24,400 g for 60 minutes. The soluble fraction was loaded onto a Talon cobalt metal affinity-column (Clontech Laboratories Inc.) pre-equilibrated with buffer B: 50 mM Tris, 150 mM NaCl, 10 % glycerol, pH 8.0, and incubated with shaking at 4 °C for 30 minutes.

The (His)$_6$-tag protein was eluted using buffer B containing 400 mM imidazole and the eluate was dialyzed into buffer C: 50 mM Tris, 150 mM NaCl, 2.5 mM CaCl$_2$, pH 8.5 and cleaved with 1 unit of human thrombin (Enzyme Research) per 10 mg of fusion protein. The completion of the cleavage reaction was verified by SDS-PAGE. TopN was further purified using size exclusion chromatography with a Superdex™ 75 10/300 GL gel-filtration column (GE Biosciences) pre-equilibrated with either Buffer D: 50 mM sodium phosphate, 150 mM NaCl, 1 mM DTT or Buffer E: 25 mM 2-(N-morpholino)ethanesulfonic acid (MES) at pH 6.4. Assignment of backbone resonances of TopN was carried out in Buffer D and DNA titrations and relaxation measurements were carried out in Buffer E. All NMR samples contained 10% D$_2$O to facilitate locking. When partial deuteration was required, the H$_2$O-based M9 media (described above) were replaced D$_2$O-based M9 media and the expression and purification of TopN was carried out as before.
Figure 7 TopN FPLC shows monomeric peak
FPLC chromatogram of TopN (9,597 Da) showing single peak at 14.8 ml elution volume on Superdex 735 10/300 column. The calibration curve for this system shows Ribonuclease A (13,700 Da) eluting at 13.1 ml. A dimer of TopN would have a molecular weight of 19,194 and would be expected to elute with a smaller retention volume than the calibration standard.
2.2 Cloning, expression, purification and refolding of TopC

pET-21b vector coding for the C terminal domain of vaccinia virus topoisomerase (81-314; TopC) with a C terminal SLE linker and non-cleavable (His)$_6$-tag was obtained from the laboratory of Dr. Stewart Shuman. The plasmid was transformed into E. coli BL21 CodonPlus (DE3) RIL competent cells (Agilent Technologies). Transformed cells from a single colony were introduced into LB media and grown overnight. An aliquot of the overnight growth was diluted to an OD$_{600}$ of 0.3 in M9 medium containing 10% D$_2$O. The culture was allowed to grow at 37 °C to OD$_{600}$ of 0.6. An aliquot of this solution was diluted to OD$_{600}$ of 0.3 with M9 prepared 25% D$_2$O and 75% H$_2$O. The process was repeated at 37 °C with progressively higher percentages of D$_2$O to adapt cells for growth in 100% D$_2$O-based M9 media. The culture in 100% D$_2$O was grown overnight at 37 °C, following which cells were harvested by centrifugation at 4,000 G for 30 minutes and added to M9 growth solution containing $^{15}$NH$_4$Cl and $^{13}$C$_6$H$_6$-D-glucose. The culture was allowed to grow to an OD$_{600}$ of 0.6. Expression was induced by addition of IPTG to a final concentration of 400 μM. The temperature was reduced to 16 °C following induction and cells were grown under agitation for 24-36 hours. Subsequently cells were harvested by centrifugation at 6750 g for 20 minutes at 4 °C.

Cells were re-suspended in lysis buffer F: Tris HCl 50 mM, 1 M NaCl, imidazole 10 mM, glycerol 10% (v/v), triton X-100 0.1% (v/v), pH 8.0 supplemented with 1 tablet of protease inhibitor (Complete mini EDTAfree, Roche). Cells were lysed by sonication on ice (0.3 s on, 0.7 s off) and debris removed by centrifugation at 24,400 g for 60 minutes. The soluble fraction was loaded onto a Ni-NTA Agarose metal affinity-column (Qiagen) pre-equilibrated with buffer G: Tris HCl 50 mM, NaCl 1 mM, Imidazole 20 mM, glycerol 10% (v/v), pH 8.0, and incubated with shaking at 4 °C for 60 minutes. The (His)$_6$-tagged protein was eluted using buffer G
containing 500 mM Imidazole. The eluate was combined with sufficient guanidinium HCl to make a 6 M solution. The resultant solution was dialyzed overnight at 4 °C in to buffer H: Bis-Tris propane 25 mM, NaCl 150 mM, dithiothreitol (DTT) 5 mM and, glycerol 1% (v/v) pH 6.8. Some precipitation was visible in the dialysis product. The precipitate was separated from the soluble fraction by centrifugation at 4,000 g for 30 minutes. The soluble fraction was further purified by size exclusion chromatography using a Superdex 75 10/300 GL gel-filtration column (GE Biosciences) pre-equilibrated with buffer H. A doublet of peaks was often seen in the chromatogram around the expected retention time for TopIB. The properly folded protein is expected to have the smallest hydrodynamic radius therefore fractions of the later eluting peak (~37 kDa) were collected. Assignment of backbone resonances of TopC was carried out in buffer H.

### 2.3 Cloning, expression, purification and refolding of TopIB and TopIBY274F

pET-21b vector coding for the vaccinia virus topoisomerase (1-314; TopIB) with a C terminal SLE linker and non-cleavable (His)$_6$-tag was obtained from the lab of Dr. Stewart Shuman. A pET-21b vector coding for the catalytically inactive Y274F mutant of TopIB (TopIBY274F) was also obtained from Dr. Shuman. The expression, purification and, refolding protocols for both products are the same. The plasmids were transformed into E. coli BL21 CodonPlus (DE3) RIL competent cells (Agilent Technologies). Transformed cells were adapted to D$_2$O as described above with the following change. The growth medium was LB with the appropriate % compositions of D$_2$O. The final overnight growth was in LB with 100% D$_2$O. Cells from overnight growth were collected by centrifugation at 4,000 g for 30 minutes. Resultant cell pellet was re-suspended in M9 media supplemented with $^{15}$NH$_4$Cl in 100% D$_2$O.
Cells were grown to an OD$_{600}$ of ~0.8 and induced with IPTG to a final concentration of 400 μM. The culture was allowed to grow at 37 °C for 6-8 hours before harvesting as described for TopC.

The resultant cell pellets were re-suspended in lysis buffer I: MOPS 50 mM, NaCl 300 mM, pH 7.0 supplemented with glycerol 10% (v/v), Triton X-100 0.1% (v/v) and, protease inhibitor (Complete mini EDTAfree, Roche). Cells were lysed and debris removed as described for TopC. Soluble fraction was loaded on to a TALON nickel metal affinity column pre-equilibrated with buffer I and agitated at 4 °C for one hour. Protein was unfolded on column by addition of 2x5 bed volumes of buffer I supplemented with 6 M guanadinium HCl. Beads were agitated in this solution for 10 minutes, allowed to settle and, the protein was refolded by addition of 2x5 bed volumes of buffer I supplemented with 5 mM imidazole. TopIB or TopIB$_{Y274F}$ was eluted with 2x4 bed volumes of buffer I supplemented with 150 mM Imidazole. The eluate was further purified by size exclusion chromatography with a Superdex 75 10/300 GL gel-filtration column (GE Biosciences) pre-equilibrated with buffer E supplemented with 5 mM DTT or buffer H.

2.4 NMR spectroscopy

2.4.1 Backbone assignment of TopN

NMR spectra were measured using $^{15}$N-labeled, $^{15}$N, $^{13}$C-labeled, $^{15}$N, $^2$H-labeled or $^{15}$N, $^{13}$C, $^2$H-labeled TopN or TopC on Bruker Avance spectrometers operating at 500 MHz, 600 MHz, 700 MHz, 800 MHz or 900 MHz or a Varian Inova spectrometer operating at 600 MHz. All spectrometers were equipped with cryogenic probes capable of applying pulsed field gradients along the z-axis. NMR experiments were performed at 25° C unless otherwise stated. Raw spectral data were processed using NMRPipe$^{84}$ and analyzed using the NMRViewJ$^{85}$ software suite.
Backbone assignments were performed on TopN (~300 μM) using a standard backbone-directed triple resonance strategy. The following experiments were performed: HNCO (600 MHz; 512, 32 and 32 complex points with sweep-widths of 13.3, 33 and 13.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), CBCANH, CBCA(CO)NH (600 MHz; 512, 32 and 52 complex points with sweep-widths of 13.3, 33 and 70 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), and CC(CO)NH-TOCSY (600 MHz; 512, 32 and 64 complex points with sweep-widths of 13.3, 33 and 80 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively. A mixing time of 18.2 ms was used). While a significant number of resonance assignments were transferable to the DNA-bound states of TopN from those in the apo-state, additional experiments were necessary to complete the resonance assignments in the presence of DNA. For TopN in the presence of specific DNA (spDNA, see below) assignments were for a sample containing a protein : DNA ratio of 1:2 in buffer E. The following experiments were acquired: HNCO (700 MHz; 512, 32, 50 complex points and sweep-widths of 12, 35 and 13 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCA (700 MHz; 512, 32, and 60 complex points and sweep-widths of 12, 35 and 36 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), and CBCA(CO)NH (700 MHz; 512, 32, 50 complex points and sweep-widths of 12, 35 and 65 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively). These spectra were acquired using non-uniform sampling and a 30% sampling schedule. FIDs were reconstructed using the MDDGui software. The secondary structure was assessed from the chemical shifts of apo TopN and its complex with spDNA using the PINE server. In order to study the full length TopIB is was necessary to assign the resonances of the catalytic domain. $^2$H, $^{13}$C, $^{15}$N labeled samples were prepared from minimal media with $^{15}$NH$_3$Cl and $^{13}$C$_6$ D-glucose as the sole sources of N and C.
Prior to addition of labeled nutrients cells were adapted to D$_2$O as described above. Growth, induction and expression of adapted cells were carried out in D$_2$O.

2.4.2 Backbone assignment of TopC

Spectral resolution in $^1$H, $^{15}$N TROSY spectra of TopC (300-400 µM) showed a marked improvement at 40 ºC compared to 25 ºC and 30 ºC therefore all TopC experiments were performed at this temperature unless stated otherwise. The following experiments were performed: HNCO (600 MHz; 1024, 32 and 32 complex points with sweep-widths of 15.0, 30.0 and 13.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCA (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 32.0 and 32.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCA (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 28.0 and 30.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCCAC (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 36.0 and 20.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOB (600 MHz; 1024, 32 and 58 complex points with sweep widths of 15.0, 28.0 and 70.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCCACOB (600 MHz; 1024, 28 and 54 complex points with sweep widths of 15.0, 28.0 and 70 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively). HNCO (600 MHz; 1024, 32 and 32 complex points with sweep-widths of 15.0, 30.0 and 13.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCA (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 32.0 and 32.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCA (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 28.0 and 30.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCCAC (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 36.0 and 20.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCA (600 MHz; 1024, 32 and 32 complex points with sweep widths of 15.0, 28.0 and 30.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCCACOB (600 MHz; 1024, 28 and 54 complex points with sweep widths of 15.0, 28.0 and 70 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively).
widths of 15.0, 28.0 and 70.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNOCACB (600 MHz; 1024, 28 and 54 complex points with sweep widths of 15.0, 28.0 and 70 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively). A large number of resonances apparent in the HNCO experiment appeared to be unresolved or at the limit of detection in the complementary 3D experiments. Acquisitions at higher magnetic field increase resolution and sensitivity therefore the following experiments were repeated in order to increase the number of assigned resonances; HNCO (700 MHz; 1024, 32 and 37 complex points with sweep widths of 13.0, 26.5 and 12.0 ppm, 800 MHz; 1024, 32, 30 complex points with sweep widths of 13.9, 26.5 and 12.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCA (900 MHz; 1024, 32 and 40 complex points with sweep widths of 12.0, 24.5 and 29.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCA (900 MHz; 1024, 32 and 32 complex points with sweep widths of 12.0, 27 and 28.6 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCACO (900 MHz; 1024, 30 and 24 complex points with sweep widths of 13.0, 27.0 and 12 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCACB (700 MHz; 1024, 32 and 45 complex points with sweep widths of 13.0, 26.5 and 66.0 ppm, 800 MHz 1024, 32 and 47 complex points with sweep widths of 13.9, 26.5 and 60.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively).

As a result of these experiments 101 residues were able to be unambiguously assigned.

### 2.4.3 Transfer of backbone assignments to TopIB and TopIBY274F

Many amide resonance assignments were directly transferrable from the TopN and TopC spectra to TopIB. To address ambiguous assignments the TROSY versions of the following experiments were performed on TopIB in buffer H at 40 °C; HNCO (800 MHz 1024, 35 and 35 complex points with sweep widths of 12.0, 27.0 and 15.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCA (800 MHz; 1024, 32 and 40 complex points with sweep widths of 12.0, 24.5 and 29.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCA (800 MHz; 1024, 32 and 32 complex points with sweep widths of 12.0, 27 and 28.6 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCACO (800 MHz; 1024, 30 and 24 complex points with sweep widths of 13.0, 27.0 and 12 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCACB (800 MHz; 1024, 32 and 45 complex points with sweep widths of 13.0, 26.5 and 66.0 ppm, 800 MHz 1024, 32 and 47 complex points with sweep widths of 13.9, 26.5 and 60.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively).
widths of 12.0, 27.0 and 66.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNCOCACB (800 MHz; 1024, 32 and 43 complex points with sweep widths of 12.0, 27.0 and 66.0 ppm for the $^1$H, $^{15}$N and $^{13}$C dimensions, respectively), HNN-NOESY-TROSY (800 MHz; 1024, 45 and 48 complex points with spectral widths of 12.0, 27.0 and 27.0 ppm for the $^1$H, $^{15}$Ny and $^{15}$Nz dimensions, respectively). A mixing time of 150 ms was used for buildup of NOEs.

Of the expected 313 peaks (10 prolines of 314 residues of TopIB plus SLE linker and 6H histag), 222 resonances were visible in the HNCO. However, a large number of these amide peaks were at the limit of detection in the 3D experiments and only 147 could be unambiguously transferred with 71 assignments in the NTD and 76 in the CTD.

2.4.4 NMR-based titrations with TopN in isolation and in the context of TopIB

All NMR-based titrations were carried out using 200 μM samples of TopN ($^{15}$N-labeled for salt and DNA titrations at lower fields; $^{15}$N, $^2$H-labeled for DNA titrations at 800 MHz) in buffer E. The influence of salt on the chemical shifts of TopN was determined by acquiring $^1$H, $^{15}$N HSQC spectra (600 MHz; 512, 64 complex points and spectral-widths of 11.7 and 34 ppm for the $^1$H and $^{15}$N dimensions respectively) for salt concentrations of 0, 50, 100 and 150 mM. Two additional spectra were acquired for the sample containing 150 mM salt by addition of phosphate to final concentrations of 25 and 50 mM.

The influence of DNA on TopN was also determined by utilizing $^1$H, $^{15}$N HSQC spectra at various concentrations of specific (spDNA) and non-specific (nsDNA) duplexes (24-mer). The spDNA sequence consisted of a duplex previously used by Morham and Shuman containing a CCCTT↓ sequence targeted by vaccinia topoisomerase. The nsDNA oligo was generated from spDNA by replacing the CCCTT sequence (and its corresponding complementary bases in the
duplex) by ACGTG, while keeping the rest of the bases identical between the two constructs.

**Figure 8** (below). DNA was purchased (IDT) as single stranded oligomers and annealed in a PCR oven overnight. For spDNA $^1$H, $^{15}$N HSQC spectra (500 MHz; 512, 64 complex points and sweep-widths of 12.0 and 35 ppm for the $^1$H and $^{15}$N dimensions, respectively) were acquired for samples with final DNA concentrations of 0, 10, 21.6, 25, 50, 59.6, 78.6, 100, 150, 200, 250, 300 and 400 μM. Additional $^1$H, $^{15}$N HSQC spectra (800 MHz; 512, 128 complex points and spectral-widths of 13 and 26 ppm for $^1$H and $^{15}$N dimensions, respectively) were acquired for samples with final DNA concentrations of 400, 600 and 800 μM.

![spDNA and nsDNA sequences](image)

DNA constructs used in this study. The 5'-CCCTT-3' target sequence in the 24-mer spDNA duplex is shown in red lettering (the equivalent sequence in nsDNA is blue). The strand of spDNA containing the CCCTT sequence is the scissile stand, the arrow indicates the site of cleavage by TopIB.

Corresponding titrations for nsDNA were performed using $^1$H, $^{15}$N-HSQC spectra acquired at 600 MHz (512, 128 complex points with spectral widths of 13.3 and 33 for the $^1$H and $^{15}$N dimensions, respectively) or 800 MHz (512, 128 complex points and spectral widths of 13 and 26 ppm for the $^1$H and $^{15}$N dimensions, respectively) for samples with final DNA concentrations of 20, 40, 60, 80, 100, 120, 140, 160, 200, 250, 300 and 400 μM at 600 MHz or 400, 600 and 800 μM at 800 MHz. Salt titrations of TopN DNA complex were performed as $^1$H, $^{15}$N-HSQC spectra at 800 MHz (512, 128 complex points and spectral widths of 13 and 26 ppm for the $^1$H and $^{15}$N dimensions, respectively) with 200 μM TopN and 400 μM of spDNA or...
nsDNA. Samples were prepared in buffer E to which 50 mM, 100 mM, or 150 mM NaCl was added.

In order to understand the TopN results in the context of the full-length TopIB, DNA titrations were carried out using the catalytically dead mutant TopIB\textsubscript{Y274F}. \textsuperscript{2}H, \textsuperscript{15}N labeled samples of TopIB\textsubscript{Y274F} were prepared as described above. All TopIB\textsubscript{Y274F} titrations involving salt and/or DNA were carried out in buffer E supplemented with 5 mM DTT at 25 °C. \textsuperscript{1}H, \textsuperscript{15}N-HSQC spectra were acquired at 800MHz (512, 80 complex points and 14, 28 ppm for the \textsuperscript{1}H and \textsuperscript{15}N dimensions, respectively) of apo TopIB\textsubscript{Y274F} with 0, 50, 100, 150 mM NaCl added. DNA titrations were performed with 200 \textmu M TopIB\textsubscript{Y274F} and DNA concentrations of 20, 60, 100, 140, 200, 400, 600 or 800 \textmu M of spDNA or nsDNA in buffer E supplemented with 5 mM DTT. Salt titrations were performed at 800 MHz (512, 80 complex points and 14, 28 ppm for the \textsuperscript{1}H and \textsuperscript{15}N dimensions, respectively) with 200 \textmu M TopIB\textsubscript{Y274F} and 400 \textmu M nsDNA or spDNA in buffer E supplemented with 5 mM DTT and either 50 mM, 100 mM or 150 mM NaCl.

Induced chemical shift perturbations (CSP, \(\Delta\delta\)) were calculated per residue using the following equation;

\[
\Delta\delta = \sqrt{\left(\Delta\delta_H\right)^2 + \left(0.154\Delta\delta_N\right)^2}
\]  

(17)

Where \(\Delta\delta_H\) and \(\Delta\delta_N\) represent the chemical shift differences in the \textsuperscript{1}H and \textsuperscript{15}N dimensions, respectively in the presence of a titrant using the corresponding chemical shifts of the apo state as reference. Substantially perturbed resonances were determined using the iterative procedure described by Piserchio et. al.\textsuperscript{90} In summary, the average and standard deviation of CSPs of all assigned resonances was calculated and CSPs > 3 SD were excluded. The process repeated until the average and standard deviation of all remaining resonances converged.
2.4.5 NMR spin-relaxation measurements

Amide $^{15}$N, R$_1$, R$_2$ and $\{^1$H$\}^{15}$N NOE relaxation measurements were performed on a 400 µM apo sample of sample of $^2$H, $^{15}$N-labeled TopN in buffer E. Additionally 200 µM samples of TopN were prepared in the presence of unlabeled 800 µM spDNA or 800 µM nsDNA. Spin-lattice relaxation rates (R$_1$)$^{91}$ were measured using relaxation delays of 0.016, 0.064, 0.144, 0.256 (x2), 0.384, 0.544, 0.800 and 1.200 s for the apo state and the spDNA complex (800 MHz, 512, 50 complex points and sweep widths of 13.9, 26 ppm for the $^1$H and $^{15}$N dimensions, respectively). Relaxation delays used for the measurement of spin-spin relaxation rates (R$_2$)$^{91}$. Relaxation delay times for apo and nsDNA bound TopN were 0.00, 0.016, 0.033, 0.049 (x2), 0.065, 0.082, 0.98 and 0.114 s (800 MHz, 512, 50 complex points and sweep widths of 13.9, 26 ppm for the $^1$H and $^{15}$N dimensions, respectively) and 0.00, 0.016, 0.033, 0.049 (x2), 0.065, 0.115, 0.131 and 0.163 s in the presence of spDNA (800 MHz, 512, 50 complex points and sweep widths of 13.9, 26 ppm for the $^1$H and $^{15}$N dimensions, respectively). $\{^1$H$\}$, $^{15}$N-steady-state NOEs were acquired using the procedures described by Ferrage et. al.$^{92, 93}$ A recycle delay of 6.5 s (800 MHz, 512, 128 complex points and sweep widths of 13.9, 26 ppm for the $^1$H and $^{15}$N dimensions, respectively) was used. The corresponding saturation period was 3 s. Relaxation data were analyzed using the built-in functions available in the nmrViewJ suite.

The rotational diffusion tensor of TopN in the apo state and complexed with 4 molar equivalents of spDNA or nsDNA were determined with a combination of exact and approximate methods as described by Ghose et. al.$^{94}$ Residues that showed ($^1$H), $^{15}$N NOE values less than 0.6 and those with R$_2$/R$_1$ ratios larger than 1.5, (indicating significant exchange contributions to R$_2$ values), were excluded from the calculations. Additionally, Phe71, which shows multiple orientations in the crystal structure, and Tyr72 that is adjacent to it were also excluded.
Backbone amide unit vectors were calculated by addition of hydrogens to a DNA bound full-length structure of TopIB (PDB ID: 3IGC) using UCSF Chimera. Reduced spectral density functions of the apo state and with spDNA were obtained at a frequency of 0.87ωH as described earlier. d and γN,H have the previously described meanings.

\[ J(0.87\omega_H) = -\frac{\gamma_N}{5d\gamma_H}(1 - \text{NOE})R_1 \]  

(18)

Higher values of J(0.87ωH) indicate more high frequency/dynamic motions. In order to compare the apo and bound states their difference is defined as ΔJ(0.87ωH) = J(0.87ωH,apo) - J(0.87ωH,spDNA). The values of ΔJ(0.87ωH) were considered statistically significant if they met the following condition;

\[ \Delta J(0.87\omega_H) \geq \sigma J(0.87\omega_{H,\text{apo}}) + \sigma J(0.87\omega_{H,\text{spDNA}}) \]  

(19)

The values σJ(0.87ωH, apo) and σJ(0.87ωH, spDNA) are the errors in the high-frequency spectral density function of the apo state and the spDNA : TopN complex respectively. σJ(0.87ωH) is calculated for each residue as follows;

\[ \sigma J(0.87\omega_H) = J(0.87\omega_H)\sqrt{\left(\frac{T_{error}}{T_1}\right)^2 + \left(\frac{\text{NOE error}}{1 - \text{NOE}}\right)^2} \]  

(20)

The values deemed statistically significant were binned in to three groups, (I, II, III), by the following classification scheme;
Isothermal Titration Calorimetry (ITC) measurements were performed with an iTC200 microcalorimeter (GE Healthcare Biosciences) at 25 °C. Previously purified TopN solutions were extensively buffer exchanged and concentrated to 794 μM by centrifugal filter concentration (Milipore) into degassed buffer E. 4 mM dsDNA stock solutions were prepared as described above and used to prepare test solutions of ~25 μM in degassed buffer E. Reverse mode titrations were performed with DNA solutions or buffer E in the sample cell and TopN solution in the syringe. Titrations consisted of a preliminary 0.4 μL injection followed by 38 injections of 0.9 μL. The results were fitted to a simple single-site binding model using the Origin 7.0 ITC data analysis module (OriginLab, Northampton, MA). All measurements were performed in duplicate.

3 Results
3.1 Measurement TopN/DNA interactions using ITC

ITC experiments were performed with TopN in the presence of spDNA or nsDNA. They yielded a $K_D$ value of 17.0 ± 1.4 μM for spDNA and 42.1 ± 6.6 μM for nsDNA indicating an ~2.5 fold higher apparent affinity for spDNA containing the CCCTT sequence. In comparison, using a filter-binding assay, an apparent $K_D$ value of 4 nM was obtained for full-length TopIB towards duplex DNA containing the specific sequence under similar conditions$^{39}$ (50 mM Tris-HCl, pH 7.5). The term apparent affinity is used due to the differing natures of the titration
curves. In the case of the spDNA curve a smooth decrease in the intensities of the enthalpy curve is seen. The sigmoidal shape of this curve is in accordance with what is expected for a single site binding curve. In the case of the nsDNA titration a more erratic shape is seen in the enthalpy curve, with the latter half of the curve (Molar Ratio 3+ mol) levelling off abruptly. This deviation from a simple sigmoidal shape may be evidence of complex binding behavior.

A two site model was applied to the nsDNA I attempted to fit the nsDNA curve to a two site binding model. The Chi Square value, which is a measure of goodness of fit of the calculated trend line, is lower for the two site model (1 site:7318, 2 site: 1740) indicating that the two site model is a better fit. An additional important parameter to consider is the N-value which is an indication of the stoichiometry of binding. A value of one indicates 1:1 stoichiometric binding with values below one indicating sub stoichiometric binding. The single site model applied to the nsDNA curve results in an N value of 0.78 mol. The two site model results in N values of 5.6×10⁻³ mol for the first binding constant and 0.80 mol for the second binding constant. Looking at the first and second binding constants they are very similar in value (K₁: 633.3×10³ ± 780.5×10³ mol⁻¹, K₂: 146.5×10³ ± 54.7×10³ mol⁻¹). This great difference in N values should be accompanied by binding constants that differ by several orders of magnitude. This suggests that the single site binding model is a more reasonable fit of the data. The lower chi squared value calculated for the two site model is likely caused by the greater number of fitting parameters and is not reflective of the suitability of the row site model.
**Figure 9 Measurement of the affinity of TopN to DNA duplexes**

Measurement of the affinity of nsDNA (A) and spDNA (B) duplexes to TopN using isothermal calorimetry (ITC) measurements. The data shown are representative of duplicate measurements. Corresponding DNA duplexes are shown on the tops of the ITC traces. Bases of the specific CCCTT (underlined) sequence on the spDNA (red lettering) that have been altered to generate the nsDNA (blue lettering) are indicated. Arrow indicates the site of cleavage by TopIB.

**Figure 10 Possible loss of interactions with TopN in the nsDNA construct**

Schematic representation of contacts between TopN and a portion of the 24-mer DNA duplex containing the CCCTT consensus sequence. Blue circles represent bases/sugars; red circles represent backbone phosphate groups. Bases that differ between the spDNA and nsDNA duplexes are outlined in red. Specific contacts that are present in spDNA but would not be possible with nsDNA if the overall binding mode were maintained in the latter case, are indicated by red arrows. Hydrophobic contacts and hydrogen bonds are represented with solid and dashed arrows, respectively. The scissile phosphate, represented by the black circle, lies between the T(+1) and A(-1) positions on the top strand.
3.2 Resonance assignment of TopN

The $^1$H $^{15}$N HSQC NMR spectrum of the TopN showed good dispersion as can be seen in the $^1$H, $^{15}$N HSQC in Figure 12 (below). This allowed the assignment of 78 of the 79 main chain amide $^1$H, $^{15}$N resonances (1 to 80 plus GS histag remnant, 99 %), 78 of 79 $^1$H$\alpha$ (99 %), 77 of 79 C$\square$ (97 %), 72 of 72 C$\beta$ (100%) and 122 of 186 C’ (66 %) resonances for non-proline residues. For the three prolines 18 of the 18 C$\alpha$, C$\beta$ and C’ (100 %) resonances could be assigned. Secondary structure inference by chemical shift of TopN agrees well with the x-ray structure in Figure 4. As can be seen in Figure 11 below, secondary structure inferred from chemical shifts of residues 74-80 shows evidence of alpha helicity in the apo protein. In addition, $^1$H, $^{15}$N resonance assignments were also obtained for sidechain amino positions (and the indole NH of W50) of N15, N16, N19, N25, Q32, Q69, Q78 and Q79. Gel permeation chromatography of the apo TopN samples showed a single monomeric peak at MW ~ 9kD prior to addition of ligand.

3.3 Evidence for TopN binding to dsDNA

$^{15}$N, $^1$H HSQC spectra of TopN in the presence of salt (NaCl with or without phosphate) display small chemical shift perturbations (CSPs) (Figure 13 below) with largest perturbations apparent at the extreme N-terminus. Ser2 shows CSPs of 0.16 ppm and 0.23 ppm in the presence of 150 mM NaCl (average $\Delta\delta = 0.06 \pm 0.02$) or 150 mM NaCl and 50 mM phosphate (average $\Delta\delta=0.06\pm0.03$). The corresponding values for Ala3 are 0.11 ppm and 0.14 ppm. Some significant CSPs are also seen between $\alpha_1$ and $\beta_3$, most notably at Lys35 with $\Delta\delta = 0.12$, 0.15 ppm respectively. These data indicate that these regions of TopN have a significant response to the ionic strength of the buffer however the overall response is minimal.
Figure 11 Probability plot of TopN secondary structure
Secondary structure inferred from $^1$H, $^{13}$C and $^{15}$N NMR chemical shift positions shows good correlation with secondary structure of full length NTD. Probability of inferred secondary structural elements are represented as follows, Green bars alpha helical, Blue bars beta strand, red bars boundary regions. Secondary structure from X-ray crystal structures are shaded and labeled. The C-terminal segment (shaded cyan, labeled $\alpha_3$) is helical only in the full-length structure in complex with DNA. NMR data was collected in buffer E at 25 °C and a field strength of 600 MHz.
In the presence of two-fold excess dsDNA, (~93% bound for spDNA based on $K_D$ values mentioned above, average $\Delta \delta = 0.10 \pm 0.10$ ppm; ~87% bound for nsDNA, average $\Delta \delta = 0.11 \pm 0.11$ ppm; Figure 14) the largest blocks with significant CSP values were centered on residues 1-3, 34-40, 70-71 and 76-79 the majority of which are at the C-terminus and in the $\alpha_1$-$\beta_3$ and $\beta_4$-$\beta_5$ loops. With spDNA the largest blocks are on residues 34-39, 69-72 and 76-79. The magnitudes of the CSPs of residues 1-3 in the presence of DNA are comparable to those in the presence of salt as previously mentioned suggesting that they are primarily due to changes in ionic strength and will not be discussed further. It can be seen in Figure 14 that block 34-40 has several significantly perturbed residues including Lys35, which is shown in crystal structures to make hydrogen bond contact through its side-chain amide with the phosphate backbone of DNA and His39 which also makes a hydrogen bond, through its side chain $N_{\delta}$, with the DNA backbone and hydrophobic contact with the sugar of the thymine in the -1 position. Block 69-72 contains vital residues Tyr70 and Tyr72. These tyrosines have been shown when mutated to alanine in the full-length protein to induce DNA binding and relaxation behavior that is similar to the C domain truncation mutant. Additionally, it contains Gln69, which makes the most specific contact with the DNA bases. Residues 76-79 are located in $\alpha_3$ described above and include His76, which makes a hydrogen bond, though its side chain $N_{\varepsilon}$, with the DNA backbone in x-ray crystal structures of DNA bound TopIB. As can be seen in Figure 14 the most significant perturbations in the presence of two fold nsDNA or spDNA are positioned on or adjacent to the binding surface shown in the crystal structures of TopIB bound to DNA in suggesting that binding between the TopN and dsDNA in solution involve the same surface as the NTD.
Figure 12 Assigned $^{15}$N, $^{1}$H HSQC spectrum of TopN (600 MHz)

The backbone assignments are labeled in black and the side-chain assignments have been labeled in red. $^{15}$N, $^{1}$H HSQC spectrum collected at 600 MHz. Protein concentration: 300 mM in buffer E at 25 °C.
3.4 Comparison of TopN binding to nsDNA and spDNA

Residues that are perturbed by 1, 2 and 3 SD, using the method described by Piserchio et al. in the MATERIALS AND METHODS, in the presence of spDNA vs. apo, nsDNA vs. apo and nsDNA vs. spDNA at a 2:1 ratio (dsDNA : TopN) are listed in Table 3 below. And the CSPs are mapped on the surface of TopN in Figure 14. Below is a discussion of how these residues support the argument that the binding surface for the NTD in the context of full-length TopIB is maintained in the context of TopN. A comparison of the engagement characteristics of nsDNA and spDNA by TopN is also presented.

Significant CSPs in the α1-β3 loop in the presence of spDNA or nsDNA are noted for Val34 (Δδ = 0.24 ppm for spDNA, Δδ = 0.30 ppm for nsDNA), Lys35 (0.25, 0.25), Thr38 (0.28, 0.38) and His39 (0.22, 0.44). Lys35 in the crystal structure of the transition state mimic of TopIB makes a hydrogen bond with the phosphate between the -2 and -3 positions on the non-scissile strand. The degree of perturbation is nearly identical for Lys35 however Glu34 is more perturbed by 0.06 ppm in the presence of nsDNA. Higher DNA : protein ratios of nsDNA continue to show perturbation of both residues. With higher ratios of spDNA there is essentially no change in position. His39 makes van der Waals contact with the sugar of the -1 thymine. In the structure of the transition state mimic the δ amide of its side chain is 3.3 Å from the phosphate between the +1 and -1 position of the DNA suggesting the possibility of backbone hydrogen bonding as well. In the presence of spDNA large perturbations are apparent in Thr38 and to a lesser extent in His39. With nsDNA Thr38 is essentially unperturbed. Perturbations of His39 are more extreme and extend to Leu40, which is the most perturbed residue in the presence of nsDNA. Additionally, Leu40 is perturbed along a different trajectory than is apparent with spDNA.
Figure 15. The side-chain of Leu40 is 3.5 Å away from that of His39 in DNA bound structures of TopIB whereas in the structure of the non-DNA bound NTD (PDB: 1VCC) they are 3.9 Å apart suggesting an increase in van der Walls contact between the two side chains in the presence of nsDNA. At higher concentrations of spDNA Thr38 and His39 are slightly perturbed (4:1 vs 2:1, Thr38: 0.02, His39: 0.01 ppm respectively) and the position of Leu40 is static. With higher concentration of nsDNA His39 and Leu40 show further perturbations and follow a non-linear trajectory suggesting deviations from a simple 1-site binding behavior (Figure 16, below).

Thr49 shows a high degree of perturbation in the presence of nsDNA (~1.5 SD: 0.27 ppm) but is nearly unperturbed by spDNA (0.08 ppm). In the crystal structures, Thr49 is located on the end of $\alpha_2$ facing away from the dsDNA. This helix is adjacent to $\alpha_3$, which is described above. At ratios above 2:1 nsDNA : TopN this resonance continues along the same trajectory away from the apo location. Ala53 is perturbed by approximately 1SD in the titration with spDNA but less than 1SD with nsDNA (0.15, 0.13 ppm). This residue is in the center of $\alpha_2$ with its hydrophobic side chain oriented toward the core of the protein. The perturbation trajectories for these resonances are different between the spDNA and nsDNA constructs being largely unchanged in the presence of higher molar ratios of spDNA protein resonances are essentially unchanged (other than a slight up-field shift in the $^{15}$N dimension), while higher concentrations of nsDNA induce an up field shift in both the $^1$H and $^{15}$N dimensions in the direction of the apo resonance. A comparison of 2:1 to 4:1 ratio of nsDNA shows the following shift differences up-field in each dimension; (0.05 ppm $^1$H, 0.23 ppm $^{15}$N) indicating that Ala53 shows additional response in the presence of nsDNA in molar equivalents beyond 1.
Figure 13 Salt titration CSPs
Chemical shift perturbations induced in apo TopN by the presence of 150 mM NaCl (black solid bars) in the buffer and by 150 mM NaCl + 50 mM phosphate (red open bars).

<table>
<thead>
<tr>
<th></th>
<th>1 SD</th>
<th>2 SD</th>
<th>3 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsDNA</td>
<td>3, 34, 35, 40, 49, 63, 70, 74, 76, 78, 79</td>
<td>1, 2, 38, 39, 77</td>
<td>None</td>
</tr>
<tr>
<td>spDNA</td>
<td>34, 35, 39, 53, 63, 64, 69, 70, 71, 74, 78</td>
<td>1, 2, 3, 38, 76, 79</td>
<td>77</td>
</tr>
<tr>
<td>nsDNA-spDNA</td>
<td>1, 34, 41, 43, 44, 51, 53, 60, 61, 62, 63, 65, 67, 74, 77</td>
<td>66, 69, 71, 80</td>
<td>38, 39, 40, 49, 64, 68</td>
</tr>
</tbody>
</table>

Table 3 Residues most perturbed by the presence of dsDNA
The first and second rows list residues of TopN (200 μM) significantly perturbed in the presence of nsDNA (400 μM) or spDNA (400 μM) respectively in buffer E. “nsDNA – spDNA” is calculated by way of the nsDNA chemical shifts referenced to the spDNA chemical shifts as described in the MATERIALS AND METHODS. Large differences in CSPs indicate nsDNA resonances that are in significantly different positions from those in the spDNA HSQC spectrum.
Figure 14 CSPs induced on TopN in the presence of dsDNA
A) CSPs of TopN with two molar equivalents of spDNA (black) or nsDNA (red). Titrations were carried out in low salt buffer E. Regions of secondary structure are shaded and labeled. The C-terminal segment (shaded cyan, labeled $\alpha_3$) is helical only in the full-length structure in complex with DNA.
B) CSPs in the presence of spDNA (top) and nsDNA (bottom) mapped onto the TopN surface using blue-to-red gradient. Key residues displaying significant perturbations are labeled. The nsDNA gradient is superimposed on to the crystal structure of the transition state mimic of TopI\B bound to the consensus sequence for comparison. Atomic resolution details of TopN bound to nsDNA is not known. The CTD is omitted for clarity.
Figure 15 Nature of perturbations for specific TopN resonances in the presence of dsDNA molar ratios beyond 1:1

Expanded view of the His39 and Leu40 resonances in $^{15}$N, $^1$H HSQC (800 MHz) of TopN in the presence of 2, 3 and 4 molar equivalents of nsDNA (top) or spDNA (bottom). The spectra in the presence of spDNA indicate saturation at approximately 2 molar equivalents (as expected from the measured $K_D$ value), while the spectra in the presence of nsDNA indicate additional events.

Figure 16 Multiple binding nsDNA Glu69 side chain

Glu 69 side chain amide resonances of TopN show a decrease in signal intensity with increasing concentrations of spDNA. With nsDNA a decreased intensity of signals near the positions of the apo state is seen. Accompanying this is the appearance and increase in intensity of a pair of resonances. This is evidence of slow exchange between two distinct binding states. Spectra were collected in buffer E at 25 °C, (spDNA 600 MHz, nsDNA 800 MHz).
Asp63, which is located in the β4-β5 loop, is highly perturbed (2 SD: 0.26 ppm) in the presence of spDNA and nsDNA (1 SD: 0.19 ppm). It is in position to form bridging hydrogen bonds between Arg67 and Gln69 in the apo crystal structure. A possible deeper insertion into the minor groove of spDNA, could be implied by the higher CSP values; this would allow Gln69 to more effectively hydrogen bond with the +2 adenine base thus weakening the hydrogen bond with Asp63. Ser64 shows a significant CSP value only for spDNA (1SD: 0.23 ppm) with minimal CSP in the presence of nsDNA (0.09 ppm). Arg67 hydrogen bonds with the phosphate backbone between the +1 and +2 positions on the non-scissile strand and in doing so could weaken the hydrogen bond between Asp63 and Arg67 side chains. Ser64 and Lys65 side chains are only 2.7 Å apart in the structure of the NTD suggesting that they are likely hydrogen bonded to each other. In the DNA bound structures Lys65 is tuned away from Ser64 and hydrogen bonded to the phosphate between the +1 Cytosine and the +2 adenine on the non-scissile strand. This puts Ser64 and Lys65 ~6.8 Å apart forcing Ser64 to find alternate hydrogen bonding partners. In the DNA bound structures the Lys65 side chain makes hydrogen bond contact with the phosphate between the +1 and +2 position. With a slight rotation of the side chain it is able to make a bond with the phosphate between the +1 and the -1 position.

Ser64 is significantly perturbed in the presence of spDNA (1 SD: 0.23 ppm) and much less perturbed with nsDNA (0.09 ppm). The crystal structures show that Ser64 makes a water mediated hydrogen bond with the backbone phosphate between the +1 and -1 bases. Additionally, the resonance for Ser64 follows a different trajectory in the nsDNA titration relative to the spDNA titration. At 2:1 ratio and above this resonance shows an arcing rather than linear trajectory in the presence of nsDNA. Gln69, which is on β5 makes bi-dentate hydrogen bond contact with the adenine in the +2 position in DNA bound crystal structures. It is the most
specific base contact present between TopN and spDNA. This resonance is significantly perturbed in the 2:1 titration with spDNA (0.16 ppm) as would be expected. Interestingly it is much less perturbed in the titration with nsDNA (0.06 ppm) even though adenine is still present in the +2 position (Figure 10, above). Apparently the necessarily different contacts formed by Tyr70 and Tyr72, likely prevent effective binding of Gln69 with the adenine base. Titration points at 2:1 and above show an arcing trajectory in the presence of nsDNA whereas the position of this resonance in the spDNA titration is unchanged at the same ratios (Figure 16, above). A conservative Q69N mutation by Tian and Shuman showed a decrease in steady state relaxation of supercoiled DNA in the presence of salt and Mg2+ in contrast to enhancement of relaxation rate seen with TopIB. Increasing the distance between the side chain and +2 adenine resulted in behavior similar to the loss of the Gln69 interaction in that the Q69N mutant behaved in a manner similar to Q69A. This led them to surmise that the hydrogen bonding potential of the sidechain amide is the main contributor to Gln69 site affinity and that this interaction is highly distance dependent.

Tyr70 makes hydrophobic contact with the bases of the +4 and +3 cytosines in DNA containing the pentameric sequence. As can be seen in Figure 10 equivalent binding is not possible with nsDNA as the +3 cytosine has been replaced with a guanine. The position of this resonance is effectively unchanged at ratios of 2:1 and greater with spDNA indicating saturation. For nsDNA in the same concentration range there is continued progressive perturbation of the resonance. Interestingly the progression is in the direction of the resonance position of the apo state. It is likely that the local interactions involving Gln69 and Tyr70 are less robust in the presence of nsDNA compared to spDNA in spite of the fact that these residues contact bases that are unchanged between the two duplexes. This is in contrast to the behavior of Tyr72 that does
contact a base that differs between the two constructs. In this case the magnitude of perturbation of Tyr72 is nearly equal in both nsDNA and spDNA titrations at 2:1 ratio (0.13, 0.14 ppm). Higher ratios show no further perturbations. This is interesting because the +3 cytosine is replaced with guanine in nsDNA therefore this residue is not able to make equivalent base-specific contact. In the crystal structure of covalently bound TopIB Lys74 is 3.4 Å from the phosphate backbone of the adenine in the +2 position. In the other structures it is turned away from the DNA and in towards the core of the protein. A deeper insertion into the major groove could put the sidechain close enough to facilitate a long distance interaction with the DNA backbone and could explain its greater degree of perturbation in the spDNA titration (spDNA 1SD: 0.26, nsDNA 1SD: 0.21 ppm). His76 is less than 3 Å from the phosphate between the +2 and +3 position on the scissile strand and makes hydrogen bond contact. This resonance is highly perturbed in the presence of both types of DNA (spDNA 1SD: 0.27, nsDNA 1SD: 0.29 ppm). At higher DNA concentrations this residue behaves in a similar manner to Tyr70. Namely that the resonance position is essentially unchanged in the presence of spDNA and shows continued chemical shift changes with the resonance moving in the direction of the apo state (a reversal of direction) in the presence of nsDNA. Val77, Gln78 and Asn79 are similarly perturbed in both spDNA and nsDNA. These residues are located in the region that forms α3 in full length TopIB bound to DNA. The sidechain of Val77 faces inward the DNA binding surface but is too short to make direct hydrophobic contact. The side-chains of Gln78 and Asn79 are pointed away from the DNA binding surface. Their perturbation likely indicates a shift in the structure possibly including the formation of α3 upon binding to DNA. In the 2:1 to 4:1 titration points the resonances for all three show arcing trajectories with nsDNA and their positions are unchanged with spDNA. In addition, Asn79 and Arg80 show multiple adjacent peaks trailing back toward
the position of the apo resonance. This suggests slow exchange between the bound state and at least partial return to the apo state.

Side chain amide resonances were assigned for Asn15, Asn16, Asn19, Asn25, Gln32, Gln48, Trp50, Gln69, Gln78 and, Asn79. The addition of DNA at the lowest DNA : TopN ratio (0.1 : 1.0) causes the Asn25 and Gln48 signals to disappear likely due to exchange broadening. This occurs both with nsDNA and spDNA. Asn25 lies on the loop connecting $\beta_2$ and $\alpha_1$. The sidechain carbonyl of Asn25 makes a hydrogen bond with the backbone amide of Ala27. Ala27 forms part of a cluster of hydrophobic residues that connect $\alpha_1$, $\beta_1$ and $\beta_2$ through van der Waals contacts. Gln48 is located on the loop between $\alpha_2$ and $\beta_3$. One of the $\text{H}_\text{E}$ protons on the side chain of Gln48 makes a hydrogen bond with the carbonyl of Leu57. In the presence of nsDNA the backbone resonance of Leu57 follows a different trajectory than is seen in the spDNA titration. The Gln69 side chain resonances of TopN in the nsDNA titration gradually diminish with increasing concentration of DNA. Concurrent with this decrease in intensity another set of resonances appear that are slightly downfield in both the $^1\text{H}$ and $^{15}\text{N}$ dimensions. They increase in intensity with higher nsDNA concentrations and their rate of increase and placement suggests that they also belong to the Gln69 side chain (Figure 16, above). Assignment of these resonances was confirmed by an HNCACB experiment performed on TopN in the presence of two molar excess nsDNA. This is interpreted to mean that Gln69 is in slow exchange between the apo state and a bound state. In the spDNA titration the $\text{H}_\text{E}_{21}$ resonance (arbitrarily labeled) side chain amide signal becomes less intense with increased DNA concentration but is still detectable at a 2:1 ratio. The $\text{H}_\text{E}_{22}$ resonance becomes undetectable with as little as 0.1:1 spDNA : TopN in the titration data collected at 600 MHz. This resonance is visible in the 800
MHz titration data but is significantly less intense than the Hε21 resonances. The reemergence of this resonance at 800 MHz is likely due to the greater signal to noise at higher field.

Calculating the difference in resonance positions of 1H 15N HSQC NMR spectra of nsDNA: TopN referenced to spDNA: TopN (2:1) the majority of the CSP patterns appear to be similar however there are small but significant differences with an average Δδ of 0.04 ± 0.05 ppm. Nineteen resonances in the nsDNA titration exhibit significant differences in resonance position compared to spDNA. The majority of the differences were in blocks from amino acids 60 - 62 and 64 - 67 however the most divergent residues as seen in Figure 17 are His39 and Leu40 (Δδ = 0.22, 0.25 respectively). The side chain of Leu40 is within 3 Å of one of the rotamers of the side chain of Phe71 in the structure of the transition state mimic of TopIB (PDB: 3IGC). Additionally, the backbone amide proton of Leu40 is 3.3 Å from the carboxyl oxygen of Thr38 in the apo structure and 2.9 Å distant in the transition state mimic structure. As shown in Error! Reference source not found. if there is specificity encoded in the NTD then Tyr70 and Tyr72 are unable to make the same contacts with DNA. The large difference in position of resonances for His39 and Leu40 in the two test cases is possible evidence for the different binding modes of Tyr70, Tyr72 being telegraphed through the side chains from Phe71 to Leu40. This may explain why His39 which makes van der Waals contact with the backbone sugar of the thymine in the -1 position in both DNA constructs also shows highly divergent binding behavior.

At concentrations beyond the 1:2 TopN: spDNA, no significant changes in the CSP values are seen for either His39 or Leu40 in contrast with increasing perturbations seen for these residues beyond 1:2 ratio of nsDNA as previously discussed. Thr49 shows the third highest level of divergent binding behavior (Δδ = 0.19 ppm). It is located at the contact point between α1 and α2. The γ methyl of Thr49 on α2 is within 3 Å of the β position of Pro26 on α1 in the transition
state structure. In the apo structure the same methyl is turned slightly away from Pro26 showing a reorientation of $\alpha_2$ in response to the presence of DNA. The large difference in position of these resonances suggests that $\alpha_1$ and $\alpha_2$ in the presence of spDNA are in a different orientation than in the presence of nsDNA. Another possibility is that this is due to a slight difference in the average orientation of the extreme N-terminus which is close to Thr49. The N-terminus itself is flexible on the chemical shift timescale and does not itself show very large differences in perturbations of nsDNA referenced to spDNA. This is partially supported by the increase in fast dynamics (Figure 21, below) for Met1 and an above average difference in CSPs (0.08 ppm) in the presence of nsDNA compared to spDNA.

The largest block of residues showing differences of 1SD or greater includes residues 60 - 69. This region starts in the middle of $\beta_4$, includes the loop connecting $\beta_4$ and $\beta_5$ and ends mid-$\beta_5$. The two most highly divergent residues in this block are Ser64 and Arg68. Ser64 is one amino acid before the phosphate backbone-contacting residue Lys65. Since Lys65 makes a non-specific contact with the backbone the fact that Ser64 and Lys65 show differences in the presence of the two DNA sequences suggests a different orientation of these residues in the presence of the two constructs. Arg68 is between two residues that make contact with the DNA. Arg67 makes a hydrogen bond with the backbone phosphate between the +1 and +2 bases. Gln69 makes specific contact with the +2 adenine base in the structures that contain the canonical sequence as previously described. The side chain of Arg68 bridges $\beta_5$ and $\beta_4$ and fits in to a pocket on $\beta_4$ created by Gly61 in all crystal structures. This interaction could possibly explain the deviations between CSPs of Val60 and Gly61 in the presence of spDNA or nsDNA. This divergence propagates to Val43 and Val44, which are on $\beta_3$ and directly adjacent Val60, and Gly61.
Figure 17 Differences in CSPs for nsDNA referenced to spDNA

(A) Differences in chemical shifts for TopN in the presence of two molar equivalents of spDNA or nsDNA. Resonances in the presence of spDNA were used as reference to calculate the perturbations using Equation (17). Residues with the largest differences are shown in red. (B) Residues with significant differences in resonance positions in the presence of spDNA or nsDNA are mapped onto the TopN surface and colored red. His39 and Leu40 that exhibit the most significant differences are labeled in larger font. The backbone of the scissile and non-scissile strands are colored grey and yellow, respectively.
At higher DNA concentrations the positions of resonances in the spDNA were essentially unchanged with a mean CSP of 4 : 1 relative to 2 : 1 ratio being 0.006 ppm and the most perturbed residue (Val77) having a CSP of 0.015 ppm. In the case of nsDNA, higher perturbations are apparent when comparing the CSPs of the 4 : 1 ratio relative to the position of the 2 : 1 ratio resonances. The average CSP is 0.030 and the most perturbed residue (Val77) is perturbed by 0.13 ppm. All of the significant differences between nsDNA and spDNA lie along the surface making contact with the non-scissile strand except for Arg80 which makes base-specific contact with the scissile strand at the +1 thymine (red bars Figure 17A and red surface in Figure 17B). Only 18 of the 77 total of 77 resonances in the nsDNA titration do not change. The rest are perturbed in the general direction of the apo state, follow arcing paths, or continue to move away from the apo resonance. This behavior is not observed in the NMR spectra of the spDNA titration where saturation is reached at a 1:2 TopN : spDNA molar ratio.

3.5 The effect of salt on CSPs of nsDNA and spDNA bound to TopN

In order to explore electrostatic contribution to interaction between TopN and the DNA oligomers a titration of NaCl with DNA bound complex was performed to gauge the effect on the stability of the complex as exhibited by changes in the CSPs. In Figure 18 it can be seen that spDNA-bound TopN backbone resonances are much more sensitive to salt content. At 50 mM NaCl spDNA perturbations are highly diminished. At salt levels between 50 to 150 mM NaCl Phe71 could not be assigned for the spDNA : TopN complex. Residues that retained chemical shift perturbations of 2 SD or above include Tyr70, His76, Val77, Gln78 and Asn79. With the exception of Tyr70 all of these residues are in the section of TopN that becomes α3 in the full length DNA bound structures.
At higher salt concentrations the chemical shift perturbations continue to diminish with Tyr70, His76, Val77, Gln78 and Asn79 remaining the most highly perturbed residues. At 150 mM NaCl most resonances are very close to the positions of the apo residues at the same level of salt. Interestingly, the side chain resonances of Gln48 do not re-appear with 50 mM NaCl nor does the weaker of the two Gln69 side chain resonances. The more intense Gln69 side chain resonance decreases in intensity relative to its intensity without the addition of salt. At 100 mM NaCl all side chain resonances are detectable and their intensities increase with increasing NaCl concentrations. nsDNA perturbations with 50 mM NaCl are somewhat comparable in magnitude to their positions in buffer E. The side chain resonances of Asn25, Gln48 and, Gln69 remain undetectable at this level of salt. Perturbations in the nsDNA complex continued to be reduced at 100 and 150 mM NaCl with the exception of Thr38 and Leu40, which retained essentially the same level of perturbation throughout the titration. At 100 mM NaCl the side chain resonances of Asn25, the stronger of the two resonances of Gln48 side chain resonances and both Gln69 side chain resonances are detectable. By 150 mM NaCl all of the side chain resonances that were suppressed in the presence of nsDNA in buffer E are again detectable.

To parse out the residues that are highly sensitive to salt in the presence of DNA, the following parameter was calculated as defined in equation (22);

\[
\Delta \delta_{\text{norm,salt}} = \left( 1 - \frac{\Delta \delta_{\text{high,low}}}{\Delta \delta_{\text{high,low}}^{\max}} \right) \left( \frac{\Delta \delta_{\text{apo,NSDNA}}}{\Delta \delta_{\text{apo,NSDNA}}^{\max}} \right)
\]

Equation (22)

\( \Delta \delta_{\text{apo,NSDNA}} \) is the CSPs calculated for dsDNA TopN (nsDNA or spDNA) versus the apo state in buffer E using equation (21). \( \Delta \delta_{\text{high,low}} \) is calculated using the resonance positions two molar excess of either nsDNA or spDNA in buffer E with 150 mM NaCl vs. their positions in buffer E.
with no added salt also using equation (17). $|\Delta \delta_{\text{apo,DNA}}|_{\text{max}}$ and $|\Delta \delta_{\text{high,low}}|_{\text{max}}$ are the largest CSPs of each type and are used as normalization factors. The $\Delta_{\text{norm,salt}}$ values vary from 0 to 1, with larger numbers representing residues that are significantly perturbed by the presence of DNA but show the least sensitivity to salt.

The largest differences between nsDNA and spDNA as can be seen in Figure 19 are at His39 and Leu40. Based on the crystal structures of DNA bound TopIB the interaction between His39 and DNA were proposed to be involved in a van der Waals interaction given the fact that the distance of the side-chain $\delta$ amide of His40 from the DNA phosphate backbone is greater than 3 Å but less than 4 Å (2H7F 3.8 Å, 2H7G 4.4 Å, 3IGC 3.3 Å). It seems that in the presence of spDNA electrostatics contributions to this interaction are dominant. The transition state mimic (3IGC) with a distance of 3.3 Å between the $\delta$ amide nitrogen and the DNA backbone is in position to form a weak hydrogen bond. It appears that this structure has the most accurate portrayal of the interaction of His39 with spDNA. This is reflected by the CSPs of Thr38, His39 and Leu40 are significantly reduced with increasing salt suggesting a weakened interaction. In contrast in the presence of nsDNA, His39 and Leu40 and to a lesser extent Thr38 show a reduced dependence on salt suggesting that van der Waals interactions play a more significant role in the interactions involving at least part of the $\alpha_1$–$\beta_3$ loop and nsDNA. For the other regions involved in binding DNA, some small differences in $\Delta_{\text{norm,salt}}$ values are also seen in the C-terminus (Lys74, His76, Asn79, Arg80) suggesting some differences in electrostatic interactions between spDNA and nsDNA involving the C-terminus of TopN likely due to a difference in conformation of the $\alpha_3$ helices.
Figure 18 Effects of ionic strength on TopN/DNA interactions
Chemical shift perturbations induced on TopN by two molar equivalents of spDNA (A) or nsDNA (B) in NMR buffer containing 0 (black), 50 mM (red), 100 mM (green), or 150 mM (blue) NaCl. The chemical shifts are referenced to TopN alone containing an equivalent amount of salt.
3.6 nsDNA and spDNA interactions of the NTD in the full length protein

\(^2\text{H},^{15}\text{N}\) labeled TopIBY274F was prepared in Buffer E as described in the MATERIALS AND METHODS. Separate samples were prepared for each molar ratio of DNA to TopN. Upon addition of either of the dsDNA oligos precipitation was apparent at DNA ratios from 0.1:1 to 0.7:1 (mole DNA to mole protein). At equimolar DNA : protein and above it was possible to break up any precipitates with mild agitation. Samples were centrifuged at 16,100 g for five minutes and concentrations were estimated by absorbance at 280 nm. At molar ratios below 1:1 concentrations were extremely low and consequently could not be measured by NMR. This behavior was not observed in the TopN truncation mutant titrations. The lack of spectra at intermediate DNA : protein ratios made it difficult to transfer assignments if there was ambiguity due to spectral crowding. This made assignment of the most highly perturbed resonances particularly difficult.

Lys35 (nsDNA, spDNA), Thr38 (nsDNA) and Lys40 (nsDNA, spDNA) are significantly perturbed in the NTD at 1:2 TopF\textsubscript{Y274F} : DNA though the nsDNA perturbations are smaller than seen in TopN. The majority of the missing assignments of the NTD were on the carboxyl terminal end probably because this segment, starting with Gly73 is largely disordered in the structure of the isolated N-domain, folds into a helix (\(\alpha_3\)) as seen in the crystal structures of TopIB bound to DNA as previously mentioned. This combined with fewer overall observed resonances as described below greatly reduced the number of assigned resonances in the titration spectra of the full length protein.
Figure 19 Normalized salt effect on TopN in the presence of nsDNA and spDNA

$\Delta_{\text{norm, salt}}$ values using equation (22) for TopN : nsDNA 1 : 2 (black bars) and TopN : spDNA 1 : 2 (open red bars) plotted against residue number.

Figure 20 CSPs of NTD of full length TopIBY274F in presence of DNA

(A) CSPs in NTD of TopIBY274F in the presence of 2x molar equivalents of either spDNA (black) or nsDNA (red).

(B) CSPs for NTD in TopFY274F in the presence of two molar equivalents of spDNA or nsDNA. Resonances in the presence of spDNA were used as reference to calculate the perturbations using equation (17). Titrations were carried out under low salt conditions (buffer E). Regions of secondary structure are shaded and labeled. The C-terminal segment (shaded cyan, labeled $\alpha_3$) is helical only in the full-length structure in complex with DNA.
The peak intensities of full length TopIB\textsubscript{Y274F} : dsDNA 1 : 2 complex were compared with that of the 1 : 4 complex. Chemical exchange between the free and bound state will cause peak broadening and lower peak intensity. In the case of single site binding peak intensity should increase with greater concentration of ligand as equilibrium is driven toward the bound state. If multi-site binding is predominant the peak intensity will vary in a much more complicated manner with increasing concentration. All TopIB\textsubscript{Y274F} : spDNA 1:4 resonances on the binding surface are more intense than those of the 1:2 complex. When looking at the totality of resonances in the NTD, with the exception of non-essential resonances Leu18 and Asn19, all resonances are more intense.

Taking the ratio of the intensities of only the NTD in the spectrum of TopIB\textsubscript{Y274F} : spDNA (1 : 4/1 : 2) yields an average of 1.30. The total number of visible resonances in the full spectrum of TopIB\textsubscript{Y274F} increases from 126 (1 : 2) to 145 (1 : 4). Comparing the intensities of TopIB\textsubscript{Y274F} : nsDNA 1 : 4 resonances to those of TopIB\textsubscript{Y274F}:nsDNA 1 : 2 reveals a much less clear trend. Approximately 1/3 of the assigned resonances of the NTD have a lower intensity in the 1 : 4 spectrum. Focusing on the binding surface 1/3 of the resonances are greater than in the 1 : 2 complex and 2/3 are less intense. The ratio of intensities of the resonances of the NTD in the spectra of TopIB\textsubscript{Y274F} : nsDNA (1:4/1:2) has an average of 1.07. The total number of resonances in the spectrum of TopIB\textsubscript{Y274F} remains unchanged (83) with increasing concentration of nsDNA. These results suggest that the NTD in the context of the full-length protein participates in single site binding in the presence of spDNA. More complicated binding behavior appears to predominate in the presence of nsDNA.

3.7 Hydrodynamic and dynamic features of TopN in the absence or presence of dsDNA

Stoke’s law for an spherical object can be used to calculate an approximate value of $\vec{c}^{96}$;
\[ t_c = \frac{4\pi \eta_w r_H^3}{3k_B T}, \quad r_H = \sqrt[3]{\frac{3V}{4\pi N_A}} \]

Where \( \eta_w \) is the viscosity of the buffer, \( r_H \) is the radius of hydration, \( V \) is the approximate specific volume of a protein (0.73 cm\(^3\)/g), \( M_r \) is the molecular weight of the protein, \( k_B \) is the Boltzmann constant, \( T \) is the absolute temperature, \( N_A \) is Avogadro’s number. Assuming a hydration radius of 3.2 Å (1 hydration shell), and using the molecular weight of TopN (9.6 kDa) the approximate \( t_c \) is 4.8 ns in a 10% D\(_2\)O solution at 25 °C. Due to the simplistic nature of the approximation including the use of a spherical model, one can expect the correlation time to be underestimated however this sets a reasonable lower bound for \( t_c \).

For a more comprehensive analysis we used the longitudinal (\( R_1 \)) and transverse (\( R_2 \)) relaxation rates and the \( ^1\text{H} \)-\( ^{15}\text{N} \) steady state NOE values collected at 800 MHz were used to calculate the diffusion tensors and consequently the rotational correlation times of apo TopN and TopN : DNA (1 : 4) complex as summarized in the MATERIALS AND METHODS section and fully described by Ghose et. al.\(^{94} \). The F value quantifies the relationship between the relative increase in sum-of-squares and the relative increase in degrees of freedom. If the simpler model is correct, the relative increase in the sum of squares (going from more complicated to simpler model) is expected to equal the relative increase in degrees of freedom. If the more complicated model is correct, then you expect the relative increase in sum-of-squares (going from complicated to simple model) to be greater than the relative increase in degrees of freedom. If the simpler model is correct you expect to get an F ratio near 1.0. If the ratio is much greater than 1.0, there are two possibilities, the more complicated model is correct or, the simpler model is correct, but random scatter led the more complicated model to fit better. The P value tells you how rare this coincidence would be. The P value answers the question, “If the simpler model is
really correct, what is the chance that you would randomly obtain data that fits the more complex model better?” If the P value is low, you can conclude that the complex model is significantly better than the simpler model. Otherwise you can conclude that there is no compelling evidence supporting the more complex model. So the simpler model should be accepted.

The resulting calculations on the apo protein produce an F value of 9.3 with the probability P that the improvement in the fit when applying the more complex (axially symmetric) model being 0.0003. The F value when comparing the fully asymmetric to the axially symmetric model is 1.0418 with a P (the probability that the more complex model was obtained by coincidence; values < 0.05 are considered significant) value of 0.31136. This suggests the axially symmetric diffusion tensor is the most likely model. The tensor shows a small departure from isotropic behavior \( \eta = \frac{D_{\text{para}}}{D_{\text{perp}}} = 2.9981 \times 10^7 \text{ s}^{-1}/2.6097 \times 10^7 \text{ s}^{-1} = 1.15 \pm 0.06 \). The rotational correlation time is \( \tau_c = \frac{1}{2D_{\text{para}} + 4D_{\text{perp}}} = 6.1 \pm 0.1 \text{ ns} \).

The rotational correlation time of the ribonuclease inhibiting protein barstar with a molecular weight of 10.2 kDa and a structure encompassing three beta strands and five alpha helices was measured to be 5.7 ± 0.1 ns and has an axially symmetric diffusion tensor at 32 °C. This is equivalent to a correlation time of 6.8 at 25 °C. Calsensin a 9 kDa invertebrate neuronal calcium binding protein with four alpha helices and two fairly long loops has a measured \( \tau_c \) of 6.7 ± 0.1 at 25 °C. The rotational correlation time of 10 kDa AF-6 PDZ domain comprised of 6 beta strands and one alpha helix is 7.25 ns with an axially symmetric diffusion tensor at 25 °C. The correlation time of TopN therefore appears to be within the expected range for similarly sized proteins. In the presence of spDNA the axially symmetric model is still the most viable (F, P[axially symmetric vs. isotropic] = 5.92, 0.005; F, P[fully asymmetric vs. axially symmetric] = 1.5504, 0.22) and \( \tau_c \) increases to 16.03±0.9 ns as one would expect for the formation of a larger overall complex. An
accompanying increase in the $\eta$ value to $1.42 \pm 0.22$ exhibits the expected behavior of the elongated shape produced by a small protein bound to significantly larger duplex DNA. This is additional and more direct evidence that TopN indeed binds to the spDNA oligo. In the presence of nsDNA the F statistic for the comparison of the fully asymmetric model vs. the axially symmetric model is 3.36 with a probability of 0.08 and for axially symmetric vs. isotropic is 2.3 with a probability of 0.12. This analysis suggests that the isotropic model is the viable choice. For the isotropic model $\tau_c = 1/6D_{iso} = 18.5 \pm 0.1$ ns, a value that is significantly larger than the value for spDNA. If one applies the axially symmetric model the $\tau_c$ calculation produces a higher value of $19.1 \pm 1.1$ ns. The higher value of $\tau_c$ may be indicative of multiple complexes involving two or more units of TopN binding to nsDNA. The anisotropy decreases to $\eta=1.15 \pm 0.20$ which implies that TopN in the presence of nsDNA has rotational behavior closer to the apo state. Multiple small populations of TopN bound in different ways to nsDNA could increase the effective spin-spin relaxation rates with the increased influence of chemical exchange term to the $R_2$ relaxation rates\textsuperscript{101} and may produce an average orientation that is more homogenous.

The calculations of hydrodynamic properties described above depend on excluding residues with significant exchange contributions. If there is significant exchange in a majority of the rates, then the hydrodynamic calculations will not be accurate. The continued and asymmetric changes in TopN : nsDNA CSPs between the 1:2 and 1:4 complexes indicate that there are multiple binding events occurring. In contrast the binding curves of TopN : spDNA for most residues appear to be stoichiometric and the CSPs of the 1:2 and 1:4 complexes are essentially the same suggesting a single binding mode. This could suggest that both the measured affinity values by ITC and indeed the hydrodynamic calculations discussed above are more accurate in the complex with spDNA and less so in the presence of nsDNA.


\(^{15}\text{N}\) relaxation rates were used to calculate the high frequency dynamics as described above. This analysis was chosen due to the fact that direct comparison of fast ps-ns dynamics between the apo state and spDNA bound TopN was possible since the effects of R2 (and therefore the overall correlation time and indeed exchange contributions) do not enter into the calculation as can be seen in equation (18) above.

An analysis of the difference between the J(0.87\(\omega_H\)) values of apo TopN and spDNA bound, reveals a general reduction in J(0.87\(\omega_H\)) values is seen in the spDNA state. This suggests a restriction of the fast motions involving the TopN backbone upon binding to spDNA restricts its motions (average \(\Delta J(0.87\omega_H) = 0.7 \pm 2.5 \text{ ps/\(\omega_H\)}\) (Figure 22 below). In the apo crystal structure of TopN (AA 1-77, PDB: 1VCC), the side chain carboxyl of Glu29, on \(\alpha_1\), is 2.7 Å from the backbone amide of Trp50, on \(\alpha_2\). The helix \(\alpha_2\) extends from Trp50 to Thr55, in the DNA bound structure it shifts by one amino acid ranging from Thr49 to Leu54, this turns the backbone amide of Tyr50 away from the side chain carboxyl of Glu29 increasing the distance between them to 3.5 Å. Breaking this hydrogen bond results in the increased flexibility of Glu29. Tyr50 is no longer the end residue of \(\alpha_2\), which results in a slight decrease in high frequency oscillations. The shift in \(\alpha_2\) has implications for the dynamic behavior of Ala53 and Thr55. In the apo structure the backbone carbonyl of Ala53 is 2.9 Å from the backbone amide of Thr55 allowing hydrogen bonding. In the DNA bound structure due to the shift in secondary structure they are 3.1 Å apart thus weakening hydrogen bonding. This slightly increases the dynamic motion of Ala53 (\(\Delta J(0.87\omega_H) = -1.2 \text{ ps} \pm 1.2 \text{ ps}\)) and as Thr55 (\(\Delta J(0.87\omega_H) = -4.9 \text{ ps} \pm 3.2 \text{ ps}\)) is no longer in alpha helical secondary structure it shows a much greater increase in flexibility. The increase in flexibility exhibited by Ser64 (\(\Delta J(0.87\omega_H) = -4.0 \text{ ps} \pm 3.0 \text{ ps}\)) can be explained by the
breaking of side chain hydrogen bonding with Lys65 previously described in the CSP analysis section. Interestingly, Tyr70 becomes the most dynamic in the presence of spDNA ($\Delta J(0.87\omega_H) = -7.2 \pm 1.1 \text{ ps/rad}$). This is surprising since it lies across the sugar of the cytosine in the +4 position, makes contact with the bases of the +3 and +4 cytosines and forms a hydrogen bond with the backbone phosphate between +4 and +5 on the scissile strand. In the structure of the transition state mimic of TopIB, Phe71 is shown in two orientations each with 50% occupancy. In one orientation the Phe71 sidechain is able to participate in $\pi$ stacking with Phe59 on $\beta_4$ however it also has a minor clash with the methyl of the T(-1) base. In the other orientation, the Phe71 side chain is unable to $\pi$ stack with Phe51 however it is not sterically hindered by -1T though it may clash slightly with Lys74 side-chain which is located on $\alpha_3$. It should be noted that Phe51 shows a slight decrease in dynamics in the presence of spDNA ($\Delta J(0.85\omega_H) = 2.4 \text{ ps} \pm 1.2 \text{ ps}$). This duality of orientations is not seen in the covalent and non-covalent complexes because the DNA in this region of these structures are distorted due to the use of a “suicide” strand in which the downstream scissile strand DNA is not present. Mutational studies by Tian and Shuman$^{52}$ show that a F59A mutant of TopIB showed an aberrant decrease in relaxation rate in the presence of salt and Mg$^{2+}$ compared to the wild type suggesting that it is important contributor to the NTD affinity for the CCCTT target site. They speculate that this is due to $\pi$ stacking with Phe71 and may in fact be a part of the binding surface as the crystal structures show Phe59 stacked over the T(+1) base producing a three ring stack (Phe59, Phe71, +1 thymine).
Figure 21 $^{15}$N relaxation measurements for TopN
$R_1$, $R_2$ and $\{^1H\}^{15}$N ssNOE data for TopN in the absence (black) and presence of nsDNA (blue) or spDNA (red). Data was collected at 800 MHz in Buffer E at 25 °C. Concentrations of TopN were; apo (400 μM) or TopN (200 μM) : DNA (400 μM).

Figure 22 Differences in high frequency motions of apo TopN vs. spDNA bound TopN
Differences in the values of high-frequency spectral density functions between the apo state and bound to spDNA, $\Delta J(0.87\omega_0)$, plotted against residue number. Only those residues where the differences are statistically significant have been plotted. Group I, II and III residues (Equation (21)) are represented by black, red and blue bars respectively. Group III residues with large $\Delta J(0.87\omega_0)$ values have been labeled.
The region of Tyr70-Tyr72 is bounded by Gln69, on the N terminal extreme of β5, and Gly73 at the transition between β5 and α3. Gln69 is held rigid by hydrogen bonding as previously described. A certain degree of flexibility in the C-terminal end of β5 would be expected in order to allow Phe71 to move out of the way of the scissile strand and to accommodate any reorientation of α3 during relaxation. The mutation of Phe71 to the smaller amino acid alanine showed DNA relaxation rate and single turnover covalent adduct formation equivalent to the wild type enzyme in the absence of salt.

Detailed mutational analysis of Tyr70 was carried out by Sekiguchi and Shuman. Mutations to alanine, phenylalanine, histidine, arginine, and glutamine showed wild type DNA relaxation rates under low salt conditions. The phenylalanine and histidine mutants, which are planar rings and have smaller side chains than tyrosine, showed a slight increase in relaxation rate in the presence of 100 mM NaCl. The other mutations showed rates equivalent to the wild type. In the presence of 5 mM MgCl₂, which stimulates the wild type enzyme by increasing dissociation, the Y70F and Y70H mutations showed no stimulation, the Y70A mutation showed a slight decrease in DNA super-helical relaxation rate and the rates for the other two mutants was dramatically lower. Under conditions ideal for WT relaxation (100 mM NaCl, 5 mM MgCl₂) the rates of relaxation of the Phe and His mutants were 1/4 - 1/2 that of WT while the relaxation rates of the Ala, Arg and Gln mutants are 1 - 5% that of the WT. Optimal function of the enzyme appears to depend on there being an amino acid side chain with a planar ring and aromaticity in the position of Tyr70, something that the other mutants do not provide. The increased relaxation rate in the presence of salt of the Phe and His mutants suggest that an increase in flexibility imparted by their smaller size is beneficial to enzyme function. The main benefit of tyrosine in this position appears to be the stability in the presence of Mg²⁺ imparted by the well-positioned
hydrogen bonding with the DNA backbone. There appears to be a tight balancing act between binding with DNA and flexibility in this position. The $J(0.87 \omega_H)$ of Phe71 was not calculated due to overlap with the resonance of Asn25. The increase in flexibility of Val77 is puzzling however it should be noted that the amide of this residue shows an unusually large CSP as described previously. The decrease in high frequency motions of residues 78-80 is consistent with their mobility being restricted by Arg80 binding with DNA.

4 Conclusion

Topoisomerases are ubiquitous enzymes that are essential for maintaining DNA topology. Many DNA binding proteins rely on specific DNA topologies for target recognition. Topoisomerase interference is the basis of the function of several classes of chemotherapeutics and is the topic of ongoing research.

The eukaryotic-like type IB topoisomerase of the poxviruses vaccinia and variola is the smallest type IB topoisomerases (TopIB). This two domain protein recognizes the 5'-'CCCTT sequence, cleaves and makes a covalent bond at the 3' end and relaxes DNA supercoiling through rotation of the scissile strand around the non-scissile strand. Flanking residues to the consensus sequence have been shown to have an impact on binding site affinity. The catalytic domain alone is capable of catalytic specificity but shows deficient affinity for DNA including a marked decrease in DNA relaxation rate in the presence of salt and Mg$^{2+}$ when compared to the full length protein. Structures of TopIB bound to DNA have shown that eight residues located on the carboxyl terminal end of the non-catalytic domain make direct contact with the DNA backbone and bases of the consensus sequence. Biochemical studies have shown these residues to be important to DNA binding affinity and salt sensitivity. The purpose of this dissertation has
been to elucidate the nature of the non-catalytic domain affinity for DNA lacking and containing the consensus sequence of bases by attempting to answer the questions does the NTD have an inherent DNA-binding ability and does it bind DNA in a similar manner to full-length TopIB and does the NTD possess a binding preference for DNA duplexes that contain the consensus CCCTT sequence over those lacking this sequence?

The results obtained using ITC, analysis of NMR chemical shift changes and spin relaxation measurements described in Section 9 above, suggest that the isolated NTD displays subtly divergent chemical shift perturbations in the presence of DNA containing the target sequence (spDNA : [C/T]CCCTT) compared to DNA lacking the target sequence (nsDNA). These perturbations are largely localized in the region of the NTD known to make contact with DNA in the context of full-length protein. ITC measurements reveal a ~2-fold higher affinity for spDNA compared to nsDNA. This suggests that not only does the NTD possess an inherent affinity for duplex DNA, it also possess a slightly higher affinity for duplexes encompassing the consensus sequence over those that do not.

If TopN were incapable of interacting with dsDNA there would be minimal if any CSPs in response to increasing concentrations of ligand. Certainly the pattern of CSPs would be expected to be similar in magnitude and location to those in response to addition of salt and phosphate. The fact that there are pronounced CSPs in the presence of dsDNA shows that there is definitely a binding interaction. The majority of the perturbations are visible on the C terminus of TopN at the loops between $\alpha_1$ and $\beta_3$, The loop between $\beta_4$ and $\beta_5$, residues on $\beta_5$ and on $\alpha_3$ (Figure 14). These regions have previously been identified as containing DNA contacting in complex with full-length TopIB and mutational data have revealed several of these residues as being important for binding site affinity. Hydrodynamic calculations using the $^{15}$N relaxation
data provides additional evidence for binding between TopN and dsDNA. The decreased R$_1$ and increased R$_2$ values (Figure 21) are consistent with the formation of a larger complex. ITC measurements indicate that TopN has a slightly (~2.5-fold) higher affinity for spDNA than it does for nsDNA.

The association of spDNA with TopN appears to be stoichiometric and the binding appears to occur through a single-site on the protein. The trajectories of chemical shift changes noted in the CSPs of TopN in the presence of increasing levels of spDNA show linearity implying a transition from unbound to a single bound state. At ratios of 2:1 (spDNA : TopN) and above there is minimal change in the positions of resonances implying saturation. In the case of nsDNA titration points above 2:1 show additional perturbations. Importantly, the trajectories of these titration points are not linear. Non-linear and especially arcing trajectories is evidence for multi-site binding or other such complex phenomena that could involve binding coupled to conformational transitions. Hydrodynamic features of the complex estimated from $^1$H,$^{15}$N backbone amide NMR relaxation measurements show an increase in axial rotational symmetry in the presence of spDNA compared to TopN alone and an increase in rotational correlation time as one would expect for binding of a globular protein to a large linear DNA molecule. Again the presence of nsDNA produces divergent behavior. The rotational correlation time is higher than with spDNA indicating the possibility of multiple copies of TopN associating with a single nsDNA molecule. Rotational symmetry is more closely represented by a fully isotropic model even though there is apparent binding which may be due to the averaging of multiple orientations. This suggests a lower binding site specificity for nsDNA than for spDNA.

The dramatic reduction in backbone amide CSPs in the TopN : spDNA complex with the addition of NaCl and the relatively smaller $\Lambda_{\text{norm,abc}}$ values suggest that binding to spDNA is
highly dependent on electrostatic interactions while the relatively smaller effects of salt on chemical shifts and $\Delta_{\text{norm, salt}}$ values in certain regions of TopN in the presence of nsDNA suggests that electrostatic interactions play a less significant role in these interactions. If one assumes that binding to nsDNA is not centered on the target sequence region but involves multiple binding sites throughout the length of the DNA then one would expect that dissociation from the CCCTT sequence by TopN with the addition of salt would induce chemical shift changes similar to those visible in the nsDNA : TopN complex due to TopN moving to the flanking DNA sequences. This behavior is not observed and in fact the resonances appear to be returning to the location of the apo state in a linear fashion in response to the addition of salt. Another possible explanation is that TopN has a higher affinity for the nsDNA sequence. This is not supported by the ITC data, which shows that TopN has a slightly higher affinity for the spDNA oligomer. Therefore, a picture emerges in which TopN binds duplex DNA that contains the consensus sequence in a mode that is similar to that in full length TopIB. For duplex DNA that lacks this sequence, though similar regions of TopN are engaged, the interaction is non-stoichiometric and the data suggests complex behavior including additional binding events that may signify sampling of additional parts of DNA by TopN in contrast to a locked in state when the consensus sequence is located. This suggests that TopN is preformed to bind with the CCCTT sequence.

The data presented here cannot answer the question of whether TopN can bind to dsDNA with different orientations in the presence of nsDNA and a single one with spDNA. Other methods are required to provide such structural information. One such method that could be used to attempt to answer this question is Förster Resonance Energy Transfer (FRET), also called Fluorescence Resonance Energy Transfer. FRET may be used as spectroscopic ruler in various
areas such as structural elucidation of biological molecules and their interactions in vitro assays, in vivo monitoring in cellular research and nucleic acid analysis.\textsuperscript{102} It is a microscopy technique used to measure the distance between two fluorophores. Resonance energy transfer occurs only over very short distances, generally less than 10 nm. It involves a direct transfer of excited state energy from a donor fluorophore to an acceptor fluorophore as an alternative to fluorescence emission from the donor. With energy transfer, the acceptor molecule enters an excited state from which it decays through emission (always a longer wavelength than that of the acceptor emission). Thus, by exciting the donor and then monitoring the relative donor and acceptor emissions, it is possible to determine when FRET has occurred and at what efficiency. Since fluorophores can be employed to specifically label biomolecules and the distance condition for FRET is of the order of the diameter of most biomolecules, FRET is often used to determine when and where two or more biomolecules, interact within their surroundings. Aside from spatial proximity, for efficient FRET to take place the FRET dye pair must also exhibit significant overlap of the donor's excitation spectrum with the acceptor's absorption spectrum. One of the experimental complications of FRET is that the spectral profiles of the FRET pair cannot be so separated that we have poor overlap, yet "cross-talk" between the two imaging channels should be avoided. Ideally the donor emission detection frequency should collect only the light from the donor and none from the acceptor, and vice versa. In general, this can be somewhat realized by choosing detection frequencies that detect light from only the shorter wavelength side of the donor emission and the longer wavelength side of the acceptor emission. This can limit somewhat the sensitivity for both donor and acceptor during a typical measurement, especially since these measurements are best performed under reduced excitation power. High light intensity in the presence of oxygen can cause damage to the fluorophore which
reduces the intensity of emissions. This is referred to as photobleaching. Reduced excitation power minimizes the level of photobleaching. Based on the mechanism of FRET a variety of novel chemical sensors and biosensors have been developed.\textsuperscript{103}

The efficiency of energy transfer from the donor to acceptor fluorophore is dependent on the energy transfer rate, their distance apart and, the orientation of their dipoles. Energy transfer rate \((k_d(r))\) between the donor and acceptor fluorophores can be defined by the following equation;

\[
k_d(r) = \frac{1}{\tau_D} \left( \frac{R_0}{r} \right)^6
\]

\text{(24)}

Where \(\tau_D\) is the rate of decay of the fluorescence signal. \(r\) is the donor and acceptor and \(R_0\) is the Förster radius given by the spectral overlap between the fluorescence spectrum of the donor and the absorption spectrum of the acceptor. The distance at which resonance energy transfer is 50\% efficient, is called the Förster distance. At \(r = R_0\), the transfer efficiency is 50\% and at this distance the donor emission would be decreased to half of its intensity in the absence of acceptor. \(R_0\) can be defined as follows;

\[
R_0 = 0.218 \{k^2 n^{-4} \Phi_D(j(\lambda))\}^{1/6}
\]

\text{(25)}

\(k\) is the orientation factor of the transition dipole moment between the donor and acceptor. Optimization of this factor will yield the orientation of the donor and acceptor dipoles. \(n\) is the refractive index of the medium, \(\Phi_D\) is the fluorescence quantum yield of the donor in the absence of the acceptor and, \(j(\lambda)\) is the spectral overlap integral and is dependent on the degree of spectral overlap and extinction coefficients of the donor and acceptor spectra. Thus the efficiency of energy transfer can be defined as;
where is the relative fluorescence intensity of donor in the presence of acceptor and is the fluorescence intensity of donor in the absence of acceptor. The right most equality of equation (26) is used to relate the measured data to the orientation factor and the distance (r) between the fluorophores by nature of equality with the previous two representations of E.

By placing one fluorophore on TopN and the other on selected locations on the DNA construct, e.g. 5’ end of the scissile strand or the 5’ end of the non-scissile strand, it should be possible to determine both the average orientation and location of binding to dsDNA. For single site binding as is expected with spDNA one would expect to see a distance and orientation that is consistent with binding to the major groove at the location of the target site. In the case of nsDNA one may see a shorter distance between the donor and the acceptor if it is sliding across the surface of the DNA. If nsDNA is preferentially binding to one or more locations on the DNA this would also be detectable by comparison of the distance measurements from the 5’-scissile construct with those of the 5’-non-scissile construct.

The sparsity of the full length NMR spectra makes it difficult to draw conclusions with regard to the behavior of the NTD in comparison to TopN. The greater number of observed resonances in the spDNA : TopIY274F spectra suggests a higher degree of structural order in the protein in the presence of spDNA than with nsDNA. This supports the assertion that the active site is not pre-assembled until the specific sequence is encountered and may be evidence for a greater degree of dynamics in the presence of nsDNA.

A question left unanswered by this research is whether the behavior of TopN in the presence of nsDNA is representative of all non-specific DNA binding modes of this domain or
will different sequences elicit differing behavior? This could be addressed by $^1$H, $^{15}$N NMR titrations of TopN with DNA systematically substituted with different bases at each point in the CCCTT sequence. A titration with DNA completely lacking the consensus sequence would also be informative. As mentioned previously, different flanking bases to the consensus sequence have been shown to have higher occupancy in relaxation studies. What is not understood is whether this is due to binding interactions in the NTD or whether the flanking bases have a subtle effect on sequence recognition by the CTD. This may be answered by titrations of TopN with DNA containing different flanking sequences in a similar manner to the systematic investigations performed by Minka et. al. If TopN binding is altered, then one would expect to see divergent CSPs relative to the sequences used here. ITC could be used to measure changes, if any, in inherent binding strength when measured under low salt conditions. The influence of flanking bases on the importance of electrostatics in binding may be measured with ITC experiments under higher salt conditions.
References


