

**Structure Determination of Vascular Endothelial Growth Factor
Heparin-Binding Domain in Complex with a DNA Aptamer**

by

Sabrina R. Hunt

B.S., North Carolina State University, 2009

M.S., North Carolina State University, 2011

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
College of Arts & Sciences Department of Chemistry & Biochemistry
2017

This thesis entitled:
Structure Determination of Vascular Endothelial Growth Factor Heparin-Binding Domain in
Complex with a DNA Aptamer
written by Sabrina R. Hunt
has been approved for the College of Arts & Sciences Department of Chemistry & Biochemistry

Arthur Pardi

Deborah Wuttke

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Hunt, Sabrina R. (Ph.D., Biochemistry)

“Structure Determination of Vascular Endothelial Growth Factor Heparin-Binding Domain in Complex with a DNA Aptamer”

Thesis directed by Professor Arthur Pardi

Vascular endothelial growth factor is a cytokine that is required for the maintenance of healthy vasculature, and it is responsible for a variety of pathogenic conditions, such as age-related macular degeneration and tumor growth in numerous cancers. As an important therapeutic target, vascular endothelial growth factor has been the focus of many *in vitro* selection strategies to identify aptamers that bind the protein with high affinity and specificity. One of these is a 25-nucleotide DNA aptamer that targets the heparin-binding domain (HBD) of vascular endothelial growth factor. The work presented here focuses on determining the solution structure of the HBD-aptamer complex using NMR. The structure was determined using molecular dynamics simulations restrained with experimental NMR data. The structure of the complex is displayed as an ensemble of the 10 lowest energy structures with a heavy atom RMSD of $2.0 \pm 0.8 \text{ \AA}$ for the well-defined core of the complex. The HBD has two subdomains that retain a similar backbone conformation between the free and bound states, but the orientation of the two subdomains relative to one another changes. The aptamer forms two Watson-Crick base pairs and two tandem base triples that are proximal to a pyrimidine-rich triloop. The 5' and 3' ends of the aptamer base pair and form a tetraloop-like structure. The Watson-Crick base pairs and several unpaired nucleotides form the binding interface that spans both subdomains of the HBD. The structure presented here contributes to the larger goal of better understanding the molecular recognition between aptamers and their targets.

Acknowledgements

I would like to thank my advisor, Dr. Arthur Pardi, for his guidance and the role he has played in my growth as a scientist. Thank you also to the other members of my thesis committee: Dr. Robert Batey, Dr. Marcelo Sousa, Dr. Loren Hough, and Dr. Deborah Wuttke. They made it their mission to help me finish my Ph.D. in spite of a difficult project, and I will be forever grateful for their help and dedication.

Thank you to the members of the Wuttke lab, who adopted me as one of their own when the former members of the Pardi lab graduated. Nick and Meagan, thank you for keeping the coffee pot full and for the hundreds of troubleshooting conversations we have had over even more cups of coffee.

Finally I would like to thank the many friends I have made since beginning my graduate career in Boulder for their unwavering support and encouragement through the ups and downs of graduate school, most of whom have experienced it right alongside me. Kyle, Eric, Clara, Ben, Jen, Evan, and all the rest, you have made my decision to come to Boulder one of the best I have ever made. Last, but certainly not least, I would like to thank my family. Mom, Dad, and Dave, thank you for always believing in me...even when I didn't. I would never have made it this far without you. Thank you.

CONTENTS

Chapter 1 Introduction	1
1.1 Vascular endothelial growth factor	1
1.1.1 VEGF-A is a key regulator of angiogenesis	1
1.1.2 VEGF exists in multiple isoforms	2
1.1.3 VEGF initiates signaling pathways of binding VEGFR2.....	4
1.1.4 Pathogenic VEGF.....	7
1.1.5 Structure of VEGF ₁₆₅	8
1.2 Aptamers	
1.2.1 SELEX as a tool to isolate aptamers for a specific target	15
1.2.2 Protein-aptamer structures	18
1.2.3 Aptamers identified against VEGF ₁₆₅	24
1.2.4 Background on the DNA aptamer binding to VEGF ₁₆₅	28
1.2.5 Thesis overview	31
Chapter 2 Chemical shift assignments for the HBD bound to the aptamer	33
2.1 Introduction	33
2.1.1 Protein chemical shift assignments	33
2.1.2 <i>Pichia pastoris</i> to express recombinant proteins	35
2.1.3 Chapter overview	37
2.2 Materials and methods	37
2.2.1 HBD expression	37
2.2.2 HBD purification	41
2.2.3 NMR experiments	44

2.2.4	Dihedral angle predictions.....	45
2.3	Results and discussion	46
2.3.1	Backbone resonance assignments of the HBD bound to the aptamer	46
2.3.2	Dihedral angle calculations	55
2.3.3	Side chain resonance assignments of the HBD bound to the aptamer	56
2.3.4	HBD intramolecular NOEs	59
Chapter 3	Chemical shift assignments for the aptamer bound to the HBD	64
3.1	Introduction	64
3.1.1	Nucleic acid chemical shift assignments	64
3.1.2	Chapter overview	66
3.2	Materials and methods	66
3.2.1	Sample preparation for NMR experiments	66
3.2.2	NMR experiments	70
3.3	Results and discussion	71
3.3.1	Titration of the HBD into the aptamer	71
3.3.2	Assignment of the exchangeable protons	75
3.3.3	Assignment of the nonexchangeable protons	82
3.3.4	Base pair identification	90
3.3.5	Intramolecular and intermolecular NOEs	95
Chapter 4	Solution structure of the HBD-aptamer complex	99
4.1	Introduction	99
4.1.1	Solution structure determination	99
4.1.2	Using Xplor-NIH to determine the structures of macromolecules	101

4.1.3 Chapter overview	102
4.2 Materials and methods	103
4.2.1 Xplor-NIH structure determination	103
4.2.2 Nucleic acid structure analysis with 3DNA	105
4.2.3 Binding affinity measurements using fluorescence polarization	106
4.3 Results and discussion	108
4.3.1 Overview of the solution structure of the HBD-aptamer complex	108
4.3.2 Structure of the HBD in the HBD-aptamer complex	119
4.3.3 Structure of the aptamer in the HBD-aptamer complex	129
4.3.4 Binding interface of the HBD-aptamer complex	149
4.4 Future directions	159
References	161
Appendix A1 NMR experimental parameters	177
Appendix A2 Backbone connectivities for the HBD bound to the aptamer	181
Appendix A3 HBD resonance assignments	191
Appendix A4 TALOS+ dihedral angle predictions	198
Appendix A5 Aptamer sequential walk	200
Appendix A6 Aptamer resonance assignments	208
Appendix A7 Xplor-NIH scripts used for structure calculation	211
Appendix A8 MolProbity results from all structures in the ensemble	226
Appendix A9 Experiments performed with HBD ₁₀₅₋₁₆₅	235

TABLES

Table 1.1 Equilibrium dissociation constants for the RNA aptamers and VEGF ₁₆₅	26
Table 1.2 Equilibrium dissociation constants for the DNA aptamer and VEGF ₁₆₅	29
Table 2.1 <i>P. pastoris</i> media used for expression of unlabeled HBD	39
Table 2.2 <i>P. pastoris</i> media used for expression of ¹³ C, ¹⁵ N-labeled HBD	41
Table 2.3 Heparin column buffers and elution method	42
Table 3.1 Sequences for template DNA and primers used for site-specific labeling	68
Table 3.2 Buffers and elution method used for HPLC purification	70
Table 3.3 Summary of samples of site-specifically labeled aptamers	87
Table 3.4 Intermolecular NOEs	97
Table 4.1 Sequences of aptamer mutants	107
Table 4.2 Summary of NMR structural statistics for the 10 lowest energy structures of the HBD in complex with the aptamer	109
Table 4.3 Parameters for the base pairs in the hairpin	132
Table 4.4 Helical parameters for the base pairs in the hairpin	132
Table A1.1 NMR experimental parameters for HBD resonance assignments	177
Table A1.2 NMR experimental parameters for aptamer resonance assignments	178
Table A3.1 HBD backbone resonance assignments	191
Table A3.2 HBD side chain resonance assignments	196
Table A4.1 HBD dihedral angles predicted by TALOS+	198
Table A6.1 Aptamer aromatic resonance assignments	208
Table A6.2 Aptamer deoxyribose resonance assignments	209
Table A6.3 Aptamer resonance assignments for exocyclic groups	210

FIGURES

Figure 1.1 Alternative splicing of the VEGF-A gene	3
Figure 1.2 VEGFR2 signaling pathways	5
Figure 1.3 VEGF receptor-binding domain forms a cystine knot structure	10
Figure 1.4 VEGFRs are receptor tyrosine kinases responsible for VEGF signaling.....	11
Figure 1.5 Bevacizumab antibody fragment blocks VEGF signaling by binding the RBD	12
Figure 1.6 Cartoon representation of the HBD	13
Figure 1.7 Crystal structure of the HBD bound to the NRP b1 domain.....	14
Figure 1.8 Schematic of the SELEX process to identify aptamers against a specific target	16
Figure 1.9 Thrombin bound to two DNA aptamers, HD1 and HD22, that each contain a G- quartet scaffold	20
Figure 1.10 The von Willebrand factor bound to a DNA aptamer	22
Figure 1.11 Autotaxin bound to a DNA aptamer	23
Figure 1.12 Predicted secondary structures of representative RNA aptamers from the six families identified from SELEX	25
Figure 1.13 Sequence alignment of five truncated DNA aptamers from Family 1.....	28
Figure 1.14 Predicted secondary structure of the DNA aptamer	31
Figure 2.1 Magnetization transfer pathway for 1H-15N HSQC	34
Figure 2.2 HBD construct design	38
Figure 2.3 Representative trace of HBD eluting from the heparin affinity column	42
Figure 2.4 Representative trace of HBD eluting from the size exclusion column	43
Figure 2.5 ¹ H- ¹⁵ N HSQC spectra of free and bound HBD	47
Figure 2.6 Magnetization transfer pathways for the HNCACB and CBCA(CO)NH experiments	48
Figure 2.7 Average CA and CB chemical shifts	49

Figure 2.8 Strip plots of the HNCACB/CBCA(CO)NH spectra for Thr142-Arg145	51
Figure 2.9 Magnetization transfer pathways for the HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments	52
Figure 2.10 Strip plots of the HNCO/HN(CA)CO and HNCA/HN(CO)CA for residues Cys120, Glu122-Lys125, and Leu127-Val129	53
Figure 2.11 Chemical shift assignments of the backbone amides of the HBD bound to the aptamer	54
Figure 2.12 Magnetization transfer pathways for the TOCSY-HSQC and HCCH-TOCSY	57
Figure 2.13 Strip plots of Val129 from the HCCH-TOCSY	58
Figure 2.14 NOESY-HSQC strip plots of the side chain amino group of Asn141	60
Figure 2.15 Overlay of the addition and subtraction spectra of the SIM-NOESY-HSQC for the side chain amino group of Asn141	61
Figure 2.16 Number of intramolecular NOEs as a function of residue	63
Figure 3.1 Imino proton spectra of unlabeled aptamer during the titration with the HBD	72
Figure 3.2 Imino proton spectra of the free aptamer at pH 5.5 and pH 7.4	73
Figure 3.3 Imino proton spectra of the aptamer bound to the HBD at pH 5.5 and pH 7.4	74
Figure 3.4 Predicted secondary structure of the aptamer	76
Figure 3.5 Imino proton region of the 2D ^1H - ^1H NOESY spectrum	76
Figure 3.6 2D ^1H - ^{15}N imino-optimized HSQC spectrum at 25 °C	79
Figure 3.7 2D ^1H - ^{15}N amino-optimized HSQC spectrum at 10 °C	81
Figure 3.8 Strip plots of the aromatic-optimized and deoxyribose-optimized NOESY-HSQC spectra for nucleotides C1-T4	83
Figure 3.9 2D ^1H - ^{13}C aromatic-optimized HSQC spectrum	84
Figure 3.10 2D ^1H - ^1H HCCH-COSY spectrum	85
Figure 3.11 2D ^1H - ^{13}C aromatic-optimized HSQC spectrum of the aptamer bound to the HBD and specifically labeled with ^{13}C and ^{15}N at position A12	88

Figure 3.12 Recognition sequence of Nt. AlwI	89
Figure 3.13 2D ¹ H- ¹³ C deoxyribose-optimized HSQC spectrum	90
Figure 3.14 2D ¹ H- ¹⁵ N HNN-COSY correlating the hydrogen bond donors and acceptors of nucleotides in base pairs	92
Figure 3.15 NOEs observed that support the formation of the base triples	94
Figure 3.16 Number of aptamer intramolecular NOEs as a function of residue	95
Figure 3.17 Intermolecular NOEs between the HBD and the aptamer	98
Figure 4.1 Command line scripts used to run 3DNA	106
Figure 4.2 Solution structure of the HBD bound to the aptamer	110
Figure 4.3 Intramolecular NOEs by residue for the HBD and the aptamer	112
Figure 4.4 Ramachandran plot for the ensemble generated by MolProbity	114
Figure 4.5 MolProbity summary statistics for the structure closest to the mean	117
Figure 4.6. The ten lowest energy structures from calculations performed with and without imposing ideal restraints for disulfide bonds and base pairs	118
Figure 4.7 Cartoon representation of the bound HBD	120
Figure 4.8 Ribbon representation of the bound HBD	121
Figure 4.9 Interactions between the α -helix and β -sheet in the C-terminal subdomain	123
Figure 4.10 Overlay of the ribbon representation of free and bound HBD	125
Figure 4.11 The HBD has an extensive hydrophobic network spanning from the N-terminal subdomain to the C-terminal subdomain	127
Figure 4.12 Predicted secondary structure of the DNA aptamer	130
Figure 4.13 The aptamer forms a hairpin with four Watson-Crick base pairs and a three-nucleotide loop	131
Figure 4.14 Pyrimidine-rich triloop caps the aptamer hairpin	133
Figure 4.15 The aptamer forms two tandem major groove triples	136

Figure 4.16 Sequence alignment of the truncated aptamers from Family 1 of the SELEX experiment used to identify the aptamer described here	138
Figure 4.17 That aptamer forms a tetraloop-like structure	140
Figure 4.18 G20 may be involved in intramolecular or intermolecular interactions	142
Figure 4.19 C21 and A22 may participate in intramolecular or intermolecular interactions	144
Figure 4.20 G23 is likely to engage in π -stacking with G24	145
Figure 4.21 Unpaired nucleotides A17 and G16 contribute to the π -stacking interactions in the hairpin	148
Figure 4.22 Summary of the intermolecular interactions seen in more than 50% of the structures in the ensemble	150
Figure 4.23 Intermolecular interactions between the G3-C13 base pair and the amino acids His126, Leu127, Val129, and Lys140	151
Figure 4.24 Preliminary binding of the wild-type aptamer and two mutants: T4C and T4C/A12G	152
Figure 4.25 Intermolecular interactions between the T4-A12 base pair and the amino acids Leu127, Asn141, and Cys158	154
Figure 4.26 Intermolecular interactions between A15 and Val129, Asp131, and Lys136	156
Figure 4.27 Arg123 and Arg124 are proximal to the 5' end of the aptamer	157
Figure 4.28 Hydrogen bond between C21 (cyan) and Thr157 (green)	158
Figure A2.1 Backbone connectivities for Arg112 to Gly118	181
Figure A2.2 Backbone connectivities for Cys120 to Val129	182
Figure A2.3 HNCA/HN(CO)CA backbone connectivities for Cys120 to Val129	183
Figure A2.4 Backbone connectivities for Val129 to Cys135	184
Figure A2.5 Backbone connectivities for Cys135 to Lys140	185
Figure A2.6 Backbone connectivities for Lys140 to Arg145	186
Figure A2.7 Backbone connectivities for Arg145 to Gln150	187

Figure A2.8 Backbone connectivities for Gln150 to Glu155	188
Figure A2.9 Backbone connectivities for Glu155 to Cys160	189
Figure A2.10 Backbone connectivities for Cys160 to Arg164	190
Figure A5.1 Sequential walk from C1 to T4	200
Figure A5.2 Sequential walk from T4 to T7	201
Figure A5.3 Sequential walk from T7 to A10	202
Figure A5.4 Sequential walk from A10 to C13	203
Figure A5.5 Sequential walk from C13 to G16	204
Figure A5.6 Sequential walk from G16 to T19	205
Figure A5.7 Sequential walk from T19 to A22	206
Figure A5.8 Sequential walk from A22 to G24	207
Figure A9.1 HBD ₁₀₅₋₁₆₅ construct design	236
Figure A9.2 Representative ITC experiment of HBD ₁₀₅₋₁₆₅ binding to the aptamer	238
Figure A9.3 Representative ITC experiment of HBD ₁₁₁₋₁₆₅ binding to the aptamer	238
Figure A9.4 Representative BCA assay of the HBD ₁₀₅₋₁₆₅ heparin affinity column fractions	239
Figure A9.5 1H-15N HSQC of the misfolded HBD sample	240
Figure A9.6 1H-15N HSQC of the previously misfolded HBD sample after addition of 1 mM β - mercaptoethanol	240

Chapter 1: Introduction

1.1 Vascular endothelial growth factor

1.1.1 VEGF is a key regulator of angiogenesis

The vascular endothelial growth factor (VEGF) proteins are cytokines that are required for the maintenance of healthy vasculature and include VEGF-A, VEGF-B, VEGF-C, VEGF-D, and placenta growth factor. VEGF-A, from here on referred to as VEGF, is a key regulator of angiogenesis, which is the creation of new blood vessels from an existing source of blood vessels. The VEGF gene is essential; knockout of just one VEGF allele is embryonic lethal (Carmeliet et al. 1996; Ferrara et al. 1996). Angiogenesis is crucial during early development because the creation of the vasculature is necessary for the development of organs (Coultas et al. 2005). The main driver of VEGF expression in cells is hypoxia, a shortage of oxygen in the blood, during which the hypoxia inducible factors will target VEGF and other pro-angiogenic genes to induce expression and promote new blood vessel formation (Shweiki et al. 1992; Liu et al. 1995; Krock et al. 2011). Hypoxia can occur during normal physiological processes, such as development, wound healing, and exercising, all of which involve a temporarily low oxygen concentration reconciled by VEGF expression and subsequent angiogenesis (Brown et al. 1992; Nissen et al. 1998; Coultas et al. 2005; Bao et al. 2009; Gustafsson 2011; Krock et al. 2011). Hypoxia also occurs in disease states in which VEGF has been implicated. The best understood pathogenic roles of VEGF are in tumor growth and neovascular eye diseases such as diabetic retinopathy, retinopathy of prematurity, and the wet form of age-related macular degeneration (Section 1.1.4) (Witmer 2003; Goel & Mercurio 2013; Miller et al. 2013). These pathogenic roles have made VEGF a popular research topic and therapeutic target.

1.1.2 VEGF exists in multiple isoforms

Several isoforms of VEGF are known with different functions. The VEGF gene contains eight exons that encode the VEGF protein (Figure 1.1). The amino acids encoded by exon one and four amino acids from exon two make up the 26-amino acid signal sequence that directs the translation machinery to the endoplasmic reticulum so the nascent polypeptide enters the secretory pathway (Grünewald et al. 2010). The protein domain translated from exons two through five make up the receptor-binding domain, and the VEGF receptor binding sites are encoded by exons three and four (Vempati et al. 2014). All known naturally occurring isoforms of VEGF contain these first five exons (Vempati et al. 2014). The human VEGF gene is alternatively spliced into several isoforms that differ in their affinities for cellular receptors and heparan sulfate proteoglycans on the cell surface due to differential incorporation of exons 6 and 7 that encode the heparin-binding domains (Ferrara & Henzel 1989; Gospodarowicz et al. 1989; Soker et al. 1998). The currently known isoforms are VEGF₁₂₁, VEGF₁₄₅, VEGF₁₄₈, VEGF₁₆₂, VEGF₁₆₅, VEGF₁₈₃, VEGF₁₈₉, and VEGF₂₀₆, the most predominant of which are VEGF₁₂₁, VEGF₁₆₅, and VEGF₁₈₉ (Arcondéguy et al. 2013). The number associated with the VEGF isoform is the number of amino acids in the mature protein following cleavage of the signal sequence. The various isoforms are expressed at different levels and with different ratios in respect to one another depending on tissue type (Vempati et al. 2014). For example, VEGF₁₈₉ is highly expressed in the lung, whereas the brain mostly expresses VEGF₁₆₅ and the uterus has high levels of both VEGF₁₂₁ and VEGF₁₆₅ (Ng et al. 2001).

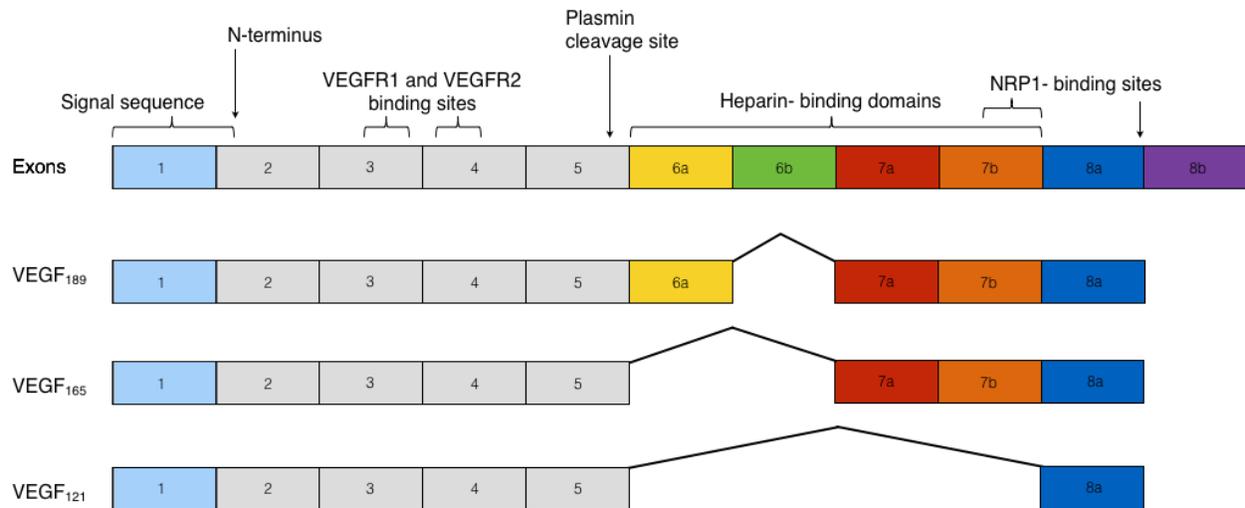


Figure 1.1 Alternative splicing of the VEGF gene. Exons of the VEGF gene and exon composition of the most prevalent isoforms of VEGF (VEGF₁₂₁, VEGF₁₆₅, and VEGF₁₈₉). Exons are not drawn to scale.

VEGF₁₂₁ is freely diffusible outside of the cell due to the lack of a heparin-binding domain and is therefore capable of longer range effects than those isoforms confined to the extracellular matrix (Vempati et al. 2014). The inclusion of the HBD causes the protein to associate with the extracellular matrix either intermittently, as is the case with VEGF₁₆₅, or constantly bound, as is the case of VEGF₁₈₉ and other long isoforms with multiple HBDs (Houck et al. 1992). The presence of VEGF isoforms attached to the surface and extracellular matrix of the cell from which it was secreted or a nearby cell may allow for rapid signaling events to occur without needing to reactivate VEGF expression (Park et al. 1993; Vempati et al. 2014). VEGFs can be released from the extracellular matrix by various mechanisms. Heparan sulfate proteoglycans can compete for binding with VEGF (Houck et al. 1992). Proteases can degrade the extracellular matrix, thereby liberating VEGF, or directly cleave VEGF, which allows it to then bind the receptors and participate in signaling. For example, plasmin will cleave VEGF₁₆₅ into a soluble fragment consisting of amino acids 1-110 (VEGF₁₁₀) that can bind the VEGFRs

(Houck et al. 1992). Urokinase will cleave VEGF₁₈₉, which is inactive due to obscuring the VEGFR2 binding site, and expose the VEGFR2 binding site (Plouët et al. 1997).

1.1.3 VEGF initiates signaling pathways by binding VEGFR2

VEGF expression is activated in response to hypoxic conditions. During hypoxia, hypoxia-inducing factors (HIF) will bind HIF-response elements to activate transcription of target genes such as VEGF (Krock et al. 2011). The various isoforms of VEGF will bind one of two receptors: VEGFR1 (also called fms-like-tyrosine kinase-1, Flt-1) or VEGFR2 (also called kinase insert domain-containing receptor, KDR). VEGF has a higher affinity for VEGFR1 than for VEGFR2, but the response initiated by the VEGF-VEGFR1 interaction is weaker than that resulting from binding of VEGF to VEGFR2 (Imoukhuede & Popel 2011). The signaling of heparin-binding VEGF isoforms via VEGFR2 is greatly enhanced by the co-receptors neuropilin-1 and heparan sulfate (Soker et al. 2002). Most angiogenesis signaling in vascular endothelial cells occurs when VEGF binds VEGFR2. This induces the dimerization of the VEGFR2 proteins that causes them to activate by trans-autophosphorylation (Lemmon & Schlessinger 2010). Two common VEGFR2 phosphorylation sites include Y1175 and Y951, and these different phosphorylation sites will result in activation of separate signaling pathways, including the Raf-MEK-ERK and Src-FAK-Paxillin pathways (Figure 1.2) (Simons et al. 2016).

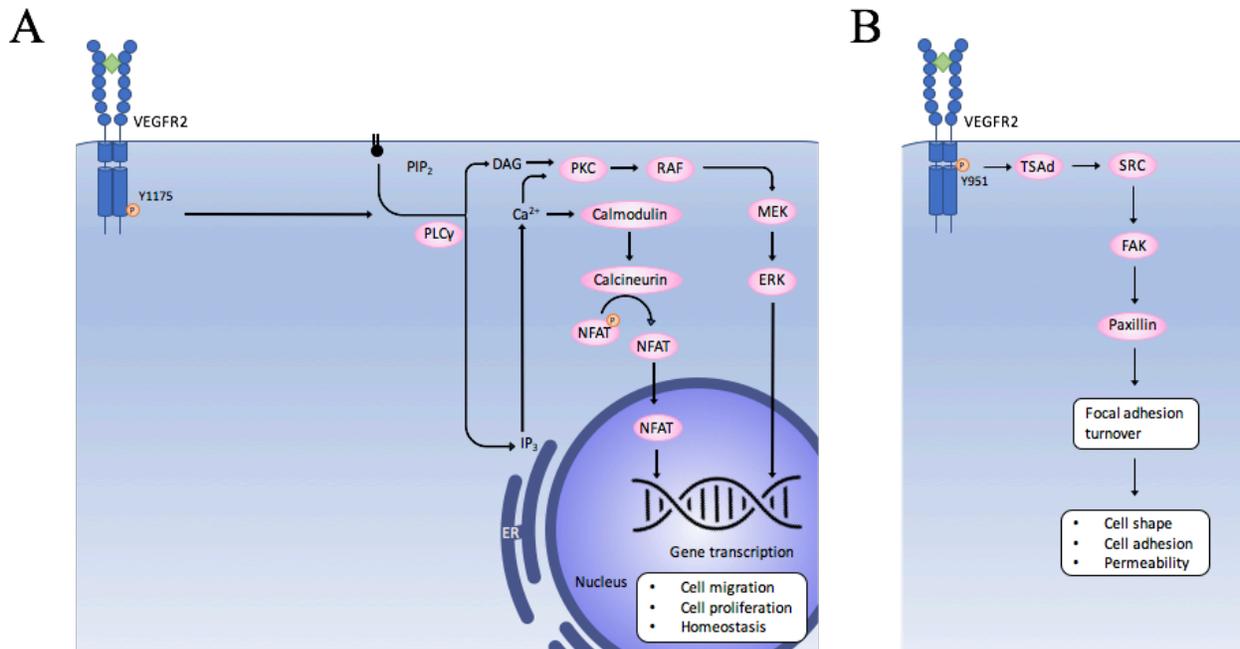


Figure 1.2 VEGFR2 signaling pathways. When VEGF binds the receptor, VEGFR2 dimerizes and autophosphorylates. **(A)** Phosphorylation of Y1175 initiates the RAF-MEK-ERK pathway that influences cell migration, cell proliferation, and vascular homeostasis. **(B)** Phosphorylation of Y951 initiates the SRC-FAK-Paxillin pathway that influences cell shape, cell adhesion, and vascular permeability. Adapted from Simons et al. 2016.

VEGFR2 phosphorylated in the C-terminal tail (Y1175) interacts with PLC γ , which converts PIP₂ into diacylglycerol and inositol 1,4,5-triphosphate (IP₃) (Figure 1.2A). Diacylglycerol activates protein kinase C β 2 (PKC) and the RAF-MEK-ERK cascade that leads to cell fate specification, proliferation, and migration (Lewis et al. 1998). This is achieved by activating the transcription factors responsible for many of the genes important for endothelial cell function—the E26 transformation-specific family—and by regulating the phosphorylation of histone deacetylase 7, which is responsible for activating the genes involved in endothelial cell proliferation and migration (Wang et al. 2008).

IP₃ is responsible for the release of calcium ions from the endoplasmic reticulum. Calcium signaling is important for PLC activation, as well as the activation and nuclear translocation of the nuclear family of activated T cell (NFAT) family of transcription factors by the calcium-dependent serine/threonine phosphatase calcineurin (Simons et al. 2016). NFAT proteins reduce the expression of VEGFR1, which results in more VEGF available to bind VEGFR2, so therefore enhances VEGFR2 signaling.

VEGFR2 phosphorylated in the kinase insert (Y951) interacts with the SH2 domain of T cell-specific adaptor (TSA_d), which then binds the SH3 domain of Src (Simons et al. 2016). The activated Src can then phosphorylate focal adhesion kinase (FAK) and form a Src-FAK complex to propagate the signal by phosphorylating Paxillin (Mitra et al. 2005). This pathway (Src-FAK-Paxillin) is important for focal adhesion turnover and cell migration, which are necessary for blood vessel growth.

1.1.4 Pathogenic VEGF

VEGF₁₆₅ is the predominant pathogenic isoform, and its most well-characterized roles are in cancer and intraocular diseases such as age-related macular degeneration (AMD); however, roles in additional diseases are continually being proposed for VEGF (Witmer 2003; Goel & Mercurio 2013; Amadio et al. 2016). Rheumatoid arthritis, ischemic cardiac and peripheral disease, inflammatory bowel disease, endometriosis, and preeclampsia have all been described as involving aberrant VEGF expression (Paleolog 2002; Carmeliet 2005; Scaldaferrri et al. 2009). This makes VEGF an important target of research and therapeutic development.

AMD is a leading cause of vision loss in people over the age of 50 (Miller et al. 2013). AMD involves the loss of vision due to the deterioration of the macula—the central part of the retina responsible for central and detailed vision. Most people with AMD have the dry form, which involves the thinning of the macular tissues and typically results in gradual vision loss, but the more damaging wet form of AMD that results in rapid vision loss (Nowak 2006). The wet form of AMD includes both proliferation of blood vessels and leakage from the new blood vessels (Miller et al. 2013). VEGF is responsible for both angiogenesis and vascular permeability, which exist in a cycle of new blood vessels leaking and the site of the leakage serving as a point of origin for new blood vessels (Drolet et al. 2016).

VEGF is also involved in cancer with an important role in tumorigenesis. Growing tumors will experience hypoxia, a condition that triggers the expression of VEGF. VEGF will then initiate angiogenesis to promote the growth of the tumor. VEGF has been implicated in roles beyond angiogenesis and vascular permeability in tumors, including affecting the function of immune cells and fibroblasts in tumors, as well as being involved in autocrine and paracrine signaling in tumor cells (Cao et al. 2012). In tumor cells, VEGF signaling is not limited to

signaling via VEGFR2. In fact, different receptors have been implicated in VEGF signaling in tumor cells, including neuropilin-1 and neuropilin-2 (NRP1/2) that normally function as co-receptors (Soker et al. 1998).

1.1.5 Structure of VEGF₁₆₅

VEGF₁₆₅ is a 46-kDa glycosylated and covalently linked homodimer with each protomer consisting of two domains: a receptor-binding domain (RBD) and a heparin-binding domain (HBD). VEGF₁₆₅ is N-glycosylated at Asn75, which is important for efficient secretion from the cell but does not in itself grossly alter the conformation of VEGF as determined by circular dichroism (Claffey et al. 1995; Brandner et al. 2006). Earlier studies suggested that glycosylation was not required for VEGF activity in cell-based assays (Peretz et al. 1992). However, Brander and colleagues showed that addition of heparin sulfate induced a conformational change as seen by circular dichroism in glycosylated VEGF₁₆₅. Since heparin sulfate proteoglycans have been shown to be important for VEGF₁₆₅ binding and VEGFR2 signaling, glycosylation may play a role in VEGF signaling (Ashikari-Hada et al. 2005).

The structures of the two separate domains have been solved using X-ray crystallography and NMR (Muller et al. 1997; Fairbrother et al. 1998; Stauffer et al. 2002). Currently, no structure exists for the full-length VEGF₁₆₅ protein. The two C-terminal residues of the RBD and the six N-terminal residues of the HBD have been shown to be highly mobile, so they may form a flexible linker connecting the two domains in the context of the full-length protein (Fairbrother et al. 1997; Fairbrother et al. 1998).

Receptor-binding domain

The crystal structure of the RBD showed that it takes the form of the cystine knot growth factor superfamily (Muller et al. 1997). The cystine knot motif is characterized by the presence of three disulfide bonds interwoven to create a stable ring-like structure (Iyer & Acharya 2011). The six cysteine residues involved in the cystine knot are typically numbered from I-VI moving from N- to C-terminus. The cystine knot found in growth factors includes disulfide bonds between Cys I and Cys IV, Cys II and Cys V, and Cys III and Cys VI (Figure 1.3) (Iyer & Acharya 2011). As is common with the cystine knot growth factor superfamily, an additional disulfide bond proximal to the cystine knot forms an intermolecular disulfide bond and the dimerization interface (Iyer & Acharya 2011; Muller et al. 1997). The three disulfide bonds of the cystine knot in the RBD are Cys57-Cys102, Cys61-C104, Cys26-Cys68, and the intermolecular disulfide bonds are from Cys51 and Cys60 on chain A to Cys60 and Cys51 on chain B respectively. Each protomer of the RBD contains seven beta-strands and two alpha helices. The central four beta strands form two pairs with only one hydrogen bond between the two pairs of antiparallel strands. The two protomers of the receptor-binding domain form a dimer by binding in an antiparallel manner.

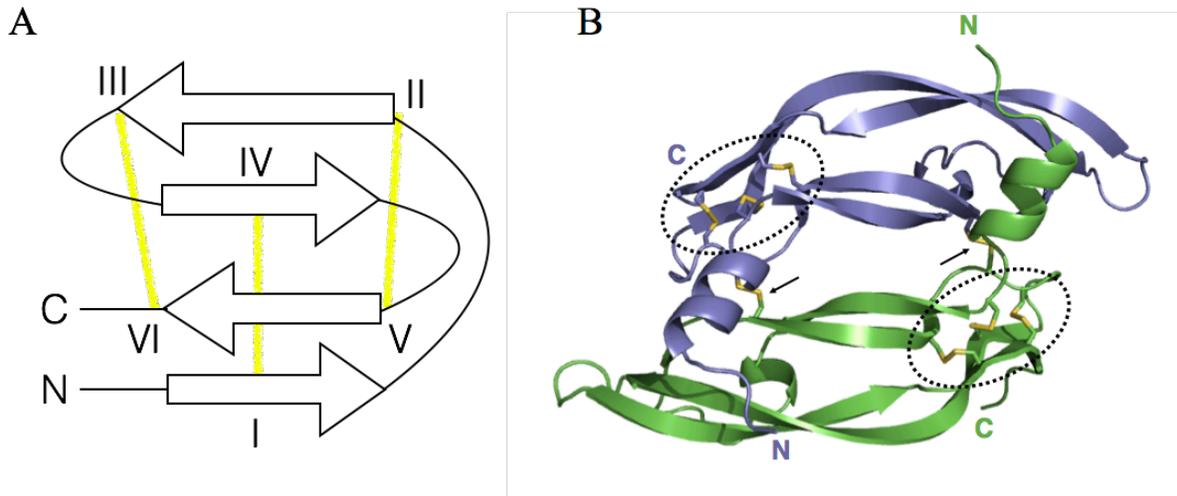


Figure 1.3 VEGF receptor-binding domain forms a cystine knot structure. (A) Schematic of the disulfide bonds forming the cystine knot in the cystine knot growth factor superfamily adapted from (Iyer & Acharya 2011). (B) Disulfide bonds involved in cystine knot are circled with dotted oval line. Intermolecular disulfide bonds forming the antiparallel homodimer are identified by the arrows.

The two main VEGFRs involved in VEGF signaling (VEGFR1 and VEGFR2) have a similar structure consisting of seven extracellular immunoglobulin homology domains, a single pass transmembrane domain, and cytosolic kinase domain (Figure 1.4 A) (Ferrara et al. 2003). Crystal structures of RBD with VEGFR1 and VEGFR2 show that VEGF binds both receptors with symmetrical binding sites at the poles of the dimer (Wiesmann et al. 1997; Brozzo et al. 2011; Markovic-Mueller et al. 2017). VEGF binds the receptors through three loops at the poles and its N-terminal alpha helix using both hydrophobic and electrostatic interactions with extracellular domains 2 and 3 of either VEGFR1 or VEGFR2 (Figure 1.4 B). The difference between these two complexes is minimal. The orientation of domains 2 and 3 of VEGFR1 is more twisted than in VEGFR2, but otherwise the interactions are similar (Markovic-Mueller et al. 2017). Markovic-Mueller and colleagues suggest instead that the difference in the activation of the two receptors is due to the co-factors involved and not the interactions between the receptors and VEGF.

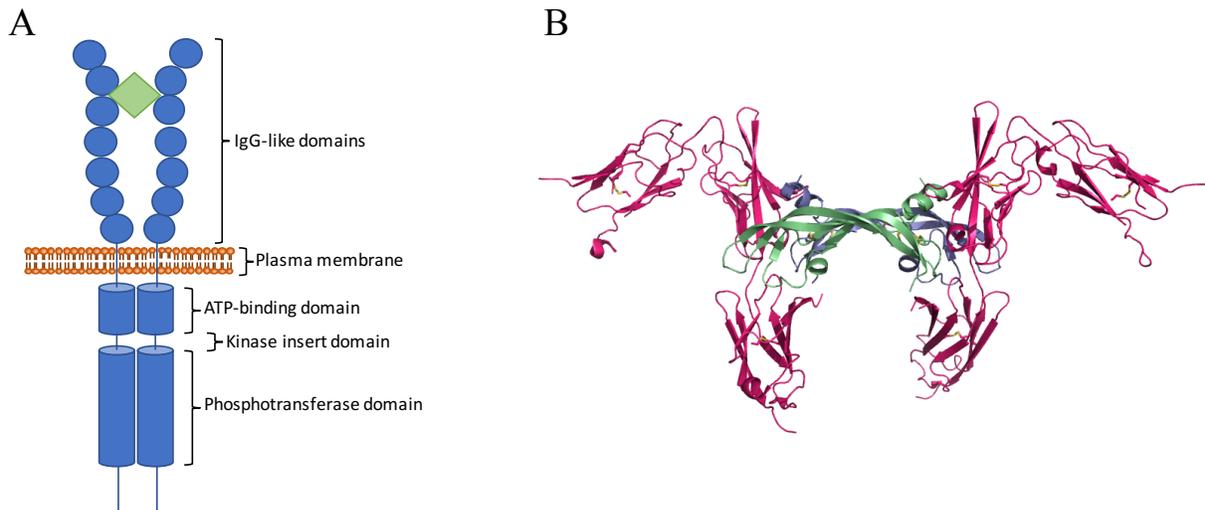


Figure 1.4 VEGFRs are receptor tyrosine kinases responsible for VEGF signaling. (A) Schematic of the domains of VEGFR. **(B)** Crystal structure of VEGF RBD (green) bound to domains two and three of VEGFR1 (blue) (PDB 5T89).

Therapeutics targeting the receptor-binding domain

Since VEGF is such an important therapeutic target, several drugs currently exist and are continuing to be developed to inhibit VEGF signaling. Various strategies are being used to accomplish this including inhibitors that target the VEGF receptors, the co-receptor neuropilin-1, bind all isoforms of VEGF, or are specific to VEGF₁₆₅.

Two antibody drugs are currently in clinical use: a humanized monoclonal antibody fragment bevacizumab (commercial name Avastin) and a humanized monoclonal antibody ranibizumab (commercial name Lucentis). They both bind the receptor-binding domain of VEGF₁₆₅, and therefore bind all isoforms of VEGF. The crystal structure of bevacizumab with the RBD shows that the antibody recognizes the same parts of the RBD that contact the receptor,

thereby preventing complex formation with the receptor and subsequent signaling events (Figure 1.5).

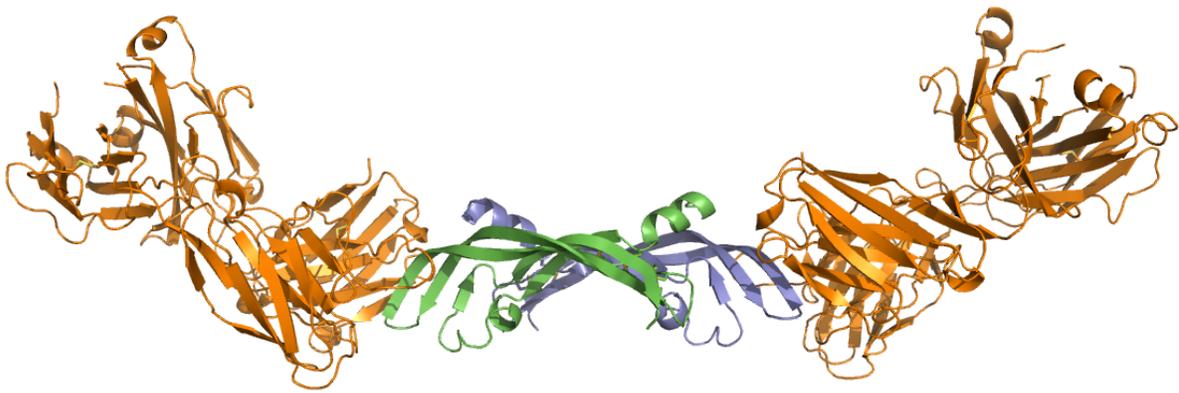


Figure 1.5 Bevacizumab antibody fragment blocks VEGF signaling by binding the RBD. Crystal structure of homodimeric VEGF RBD (green and blue) bound to bevacizumab fragment (orange) (PDB 1BJ1).

Heparin-binding domain

The HBD of VEGF₁₆₅ is a small domain consisting of amino acids 111-165 and is held together with four disulfide bonds: Cys117-Cys135, Cys120-Cys137, Cys139-Cys158, and Cys146-Cys160 (Figure 1.6). The NMR structure shows that the HBD contains two subdomains, each containing two disulfide bonds (Fairbrother et al. 1998; Stauffer et al. 2002). Each subdomain also contains a small β -sheet consisting of two antiparallel strands. The C-terminal subdomain also contains an α -helix that packs against the β -sheet.

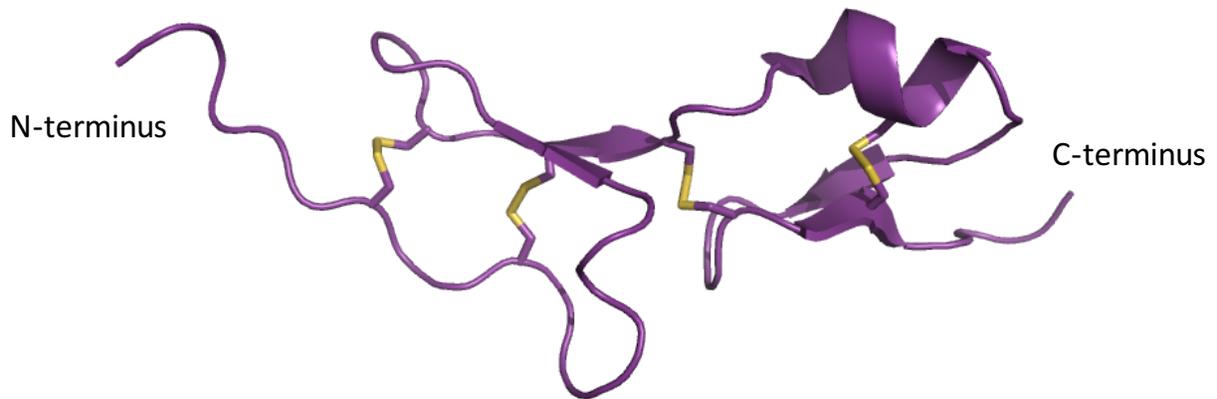


Figure 1.6 Cartoon representation of the HBD. The four disulfide bonds forming the core of the domain are shown as sticks with the sulfur atoms in yellow.

The surface of the HBD is dominated by positively charged amino acids, with 25% of the amino acids being lysines and arginines. This high concentration of positive charge allows the protein to bind negatively charged heparin in the form of heparan sulfate proteoglycans within the extracellular matrix and the cell membrane with an equilibrium dissociation constant of $0.16 \mu\text{M}$ ((Krilleke et al. 2007). The heparin-binding site on HBD was determined by mutating the basic residues on the surface of the HBD (Krilleke et al. 2007). The specific residues were chosen based on their suggested role in heparin binding by Robinson and colleagues who used molecular dynamics to model heparin binding to HBD (Robinson et al. 2006). This work by Krilleke and colleagues showed that residues Arg123, Arg124, and Arg149 were critical for heparin binding (Figure 1.7). Additionally, interaction with neuropilin-1 was disrupted when these residues were mutated suggesting that the presence of heparin is important for forming the neuropilin-1-VEGF-VEGFR2 ternary complex (Krilleke et al. 2007). neuropilin-1 binds VEGF₁₆₅ by contacting

residues located within the HBD, which include the C-terminal Arg165 as well as residues Lys147, Glu152, and Glu155 (Figure 1.7) (Parker et al. 2012). These interactions make the highest affinity interactions specific to isoforms of VEGF that contain the HBD.

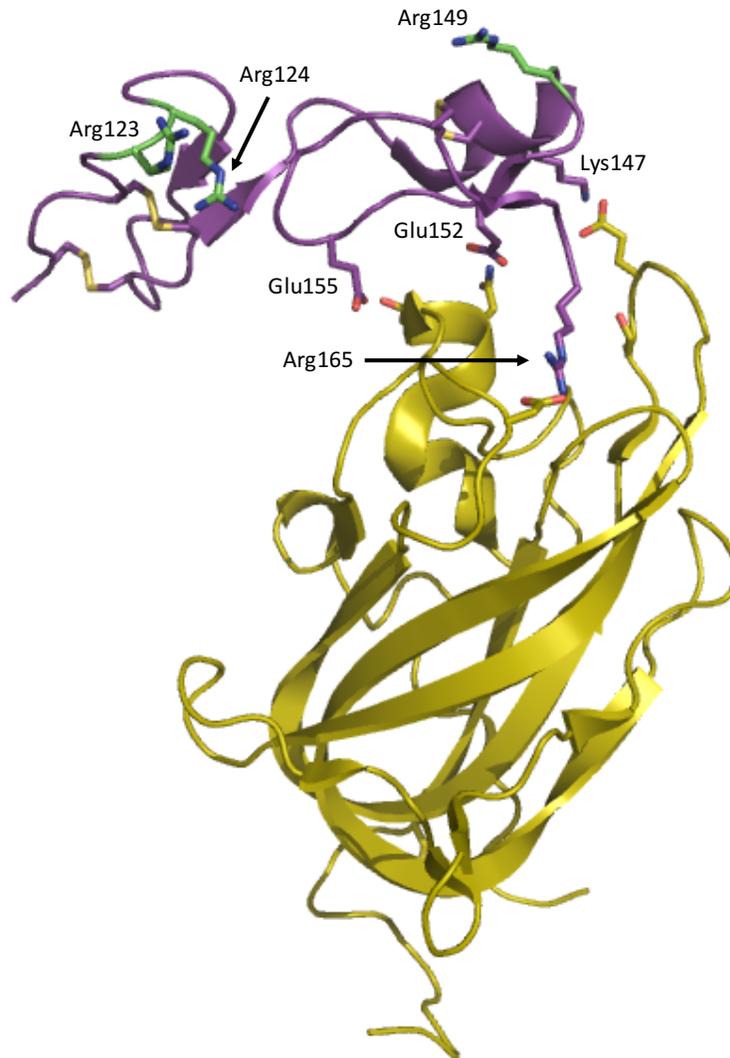


Figure 1.7 Crystal structure of the HBD bound to the neuropilin-1 b1 domain. The HBD is shown in purple and neuropilin-1 b1 domain is shown in gold. Amino acids important for the neuropilin-1-HBD interaction are shown as sticks. Arginine residues shown to be crucial for binding heparin as shown as green sticks. (PDB 4DEQ)

Therapeutics targeting the heparin-binding domain

In 2004, Pegaptanib sodium injection (commercial name Macugen) became the first FDA-approved aptamer and was used to treat the wet form of age-related macular degeneration. It specifically binds VEGF₁₆₅ by targeting the HBD, which was shown by identifying a photocrosslink between Cys137 in the HBD to U14 in Macugen (Ruckman et al. 1998). Additional experiments supported the model of Macugen binding the HBD including ¹H imino proton and ³¹P spectra NMR spectra of Macugen bound to the HBD and VEGF₁₆₅ that are nearly identical (Lee et al. 2005). Imino proton exchange rates were measured to show that the HBD-Macugen complex forms by induced fit, in which the aptamer has limited structure in the free state but takes on a well-ordered form when bound to the HBD (Lee et al. 2008).

1.2 Aptamers

1.2.1 SELEX as a tool to isolate aptamers for a specific target

In the 1990s, two papers were published simultaneously that described a method of *in vitro* selection of RNA molecules (Ellington & Szostak 1990; Tuerk & Gold 1990). Ellington and Szostak used *in vitro* selected RNAs to bind specific organic dyes. Tuerk and Gold selected RNAs against the bacteriophage T4 DNA polymerase (gp43) by randomizing an eight-nucleotide loop of the RNA known to bind gp43, and they coined the term SELEX (Systematic Evolution of Ligands by EXponential enrichment) for this technique. SELEX has become a powerful method of identifying nucleic acids that can bind a target molecule with high affinity and specificity. The process of SELEX involves the generation of an initial library containing 10¹⁴ sequences or more that sample a sequence space of typically 20 to 50 random nucleotides. The random region is flanked by fixed regions that encode sequences for amplification and regeneration of the aptamer

pool. For example, the fixed regions of an RNA library would contain the sequences necessary for transcription, reverse transcription, and PCR. The initial aptamer pool is generated, and the target molecule (e.g. protein) is incubated with the library (Figure 1.8). Aptamers that cannot bind the target protein or only bind it weakly are washed away. The remaining binders are eluted from the target molecule and amplified using the fixed regions (RT-PCR). Following amplification, the sequences can be analyzed via deep sequencing, or regenerated as a pool of aptamers for another round of selection.

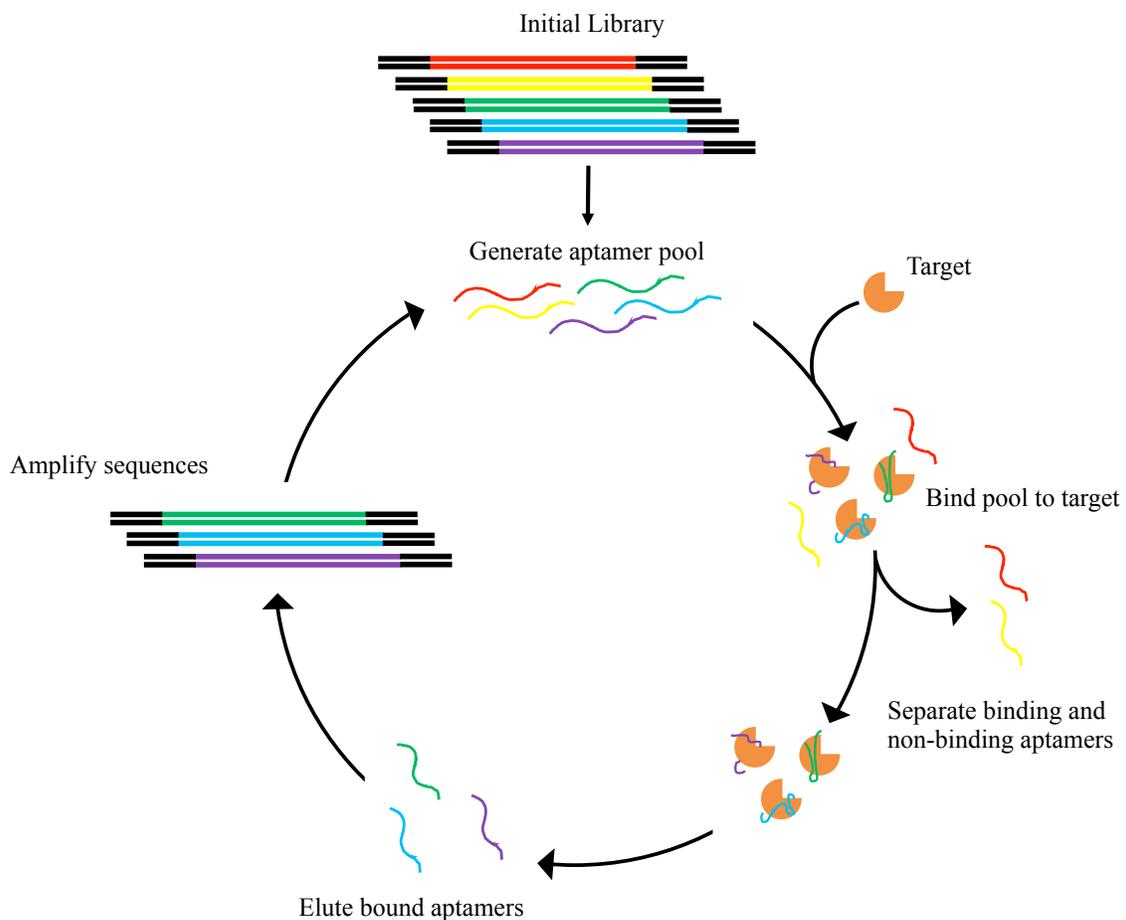


Figure 1.8 Schematic of the SELEX process to identify aptamers against a specific target. Adapted from (Schütze et al. 2011).

Due to the tendency of RNA to self-cleave using the 2' hydroxyl, modified sugars have been used in SELEX to produce aptamers that have the potential to be further developed into therapeutics. The aptamer that became Macugen was identified with SELEX using 2' fluoropyrimidines with additional 2'-O-methyl modifications added to all but two purines post-SELEX (Ruckman et al. 1998). Additional modifications were incorporated including a 5' PEG moiety, which was important for increasing the circulating half-life of the drug, and a 3'-3' dT linkage that helped Macugen resist exonucleases when it entered the human body (Ruckman et al. 1998).

Recently, nucleotides with more complex modifications have been incorporated into the SELEX method for the creation of slow off-rate modified aptamers (SOMAmers) (Davies et al. 2012; Gelinas et al. 2014; Jarvis et al. 2015). These modifications include hydrophobic groups like benzene attached to the 5' position of deoxyuridine. This gives the aptamers a more diverse chemical library than one that relies only on the four naturally occurring bases. Instead of binding interface being dominated by electrostatic and hydrogen-bonding interactions, the new modifications allow for a substantial amount of hydrophobic interactions to occur between the aptamers and the proteins (Gelinas et al. 2016).

Aptamers have a potential to be successful therapeutics. Since the advent of SELEX, modifications have been made to identify new aptamers using methods other than the original SELEX procedure. These include methods such as cell SELEX and Toggle SELEX. Cell SELEX selects for aptamers against an unknown target and has been successful for finding aptamers that target cancer cells, bacterial cells, or cells infected with viruses (Morris et al. 1998; Darmostuk et al. 2014). Toggle SELEX alternates rounds of selection with homologs from different organisms to promote the cross-reactivity that is required for preclinical testing but can be lost when

selecting for a highly specific aptamer (White et al. 2001; Nimjee et al. 2017). These different methods of selection provide new ways to find aptamers for therapeutic purposes. Aptamers have numerous advantages that could help them prevail as therapeutics over antibodies. Since aptamers are synthesized in the laboratory and not in an animal like antibodies, they are not subjected to the variability antibodies have between batches, and aptamers can be developed against toxins or targets that do not trigger an immune response (Nimjee et al. 2017).

Importantly, there exists the ability to rapidly reverse the effects of aptamers *in vivo*. These antidotes are complementary to the aptamer and once encountered, the aptamer will form stable Watson-Crick base pairs with the complementary strand and be rendered inactive (Rusconi et al. 2002). This strategy was tested in human patients using the anticoagulation factor IXa aptamer/antidote pair, but testing was halted during phase III trials due to several unexpected instances of PEG allergies (Nimjee et al. 2017). While the aptamers themselves are non-immunogenic, their carrier molecule modifications can be so further studies and optimization will be required before aptamers can be routinely used as therapeutics.

1.2.2 Protein-aptamer structures

Currently, there are only 19 published structures of different protein-aptamer complexes. Three of these are RNA aptamers bound to nucleic acid-binding proteins: bacteriophage MS2 coat protein, nuclear factor κ B, and ribosomal protein S8 from *Bacillus anthracis* (Convery et al. 1998; Huang et al. 2003; Davlieva et al. 2014). Six are RNA aptamers bound to proteins not known to bind nucleic acids: G-protein-coupled receptor kinase 2, lysozyme, thrombin, human immunoglobulin G, L-monocyte chemoattractant protein-1, which is bound to an L-RNA aptamer, and anaphylatoxin C5a, which is bound to a mixed L-DNA/L-RNA aptamer (Long et

al. 2008; Miyakawa et al. 2008; Tesmer et al. 2012; Padlan et al. 2014; Oberthür et al. 2015; Yatime et al. 2015). Seven structures include DNA aptamers bound to proteins: von Willebrand Factor, *Plasmodium falciparum* lactate dehydrogenase, *Plasmodium vivax* lactate dehydrogenase, autotaxin and three with thrombin (Huang et al. 2009; Krauss et al. 2012; Cheung et al. 2013; Krauss et al. 2013; Choi & Ban 2016; Kato et al. 2016; Pica et al. 2016). The remaining three protein-aptamer complex structures are of slow off-rate modified aptamers, SOMAmers, bound to proteins: platelet-derived growth factor B, interleukin-6, and nerve growth factor (Davies et al. 2012; Gelinas et al. 2014; Jarvis et al. 2015). RNA has long been known to take on complicated structural folds with various motifs and structural elements, such as tetraloops, noncanonical base pairs, base triples and quadruples, internal loops, and helix junctions (Batey et al. 1999; Leontis & Westhof 2003). The structures of the modified and unmodified DNA aptamers show that they are also able to take on structures just as complicated as RNA with many of these same motifs present. One common theme that emerges in the analysis of molecular recognition in protein-aptamer complex structures is the use of aptamer tertiary structures as scaffolds that orient specific regions of the aptamer to interact with its protein target.

Thrombin is an important protein in the process of blood clotting during wound healing. The two thrombin aptamers bind to different sites on thrombin that are interaction sites for many cofactors and substrates: the HD1 aptamer binds exosite I and the HD22 aptamer binds exosite II (Figure 1.9) (Krauss et al. 2012; Krauss et al. 2013; Pica et al. 2016). The two aptamers have been crystallized separately with thrombin and as a ternary complex. HD1 forms a canonical antiparallel G-quadruplex, and HD22 forms a G-quadruplex as well as a duplex connecting the 5' and 3' ends. Neither quadruplex is involved in recognizing thrombin. Instead, the quadruplex

structures in both thrombin aptamers act as scaffolds that bring the aptamers into the correct folds to present the necessary residues to recognize thrombin specifically. HD1 and HD22 bury 510 Å and 1118 Å of solvent accessible surface on the protein respectively. The exosites have numerous basic amino acids that give it an electropositive potential. The interface is dominated by electrostatic interactions between the protein and aptamer including ion pairs between the phosphate backbone and positively charged side chains.

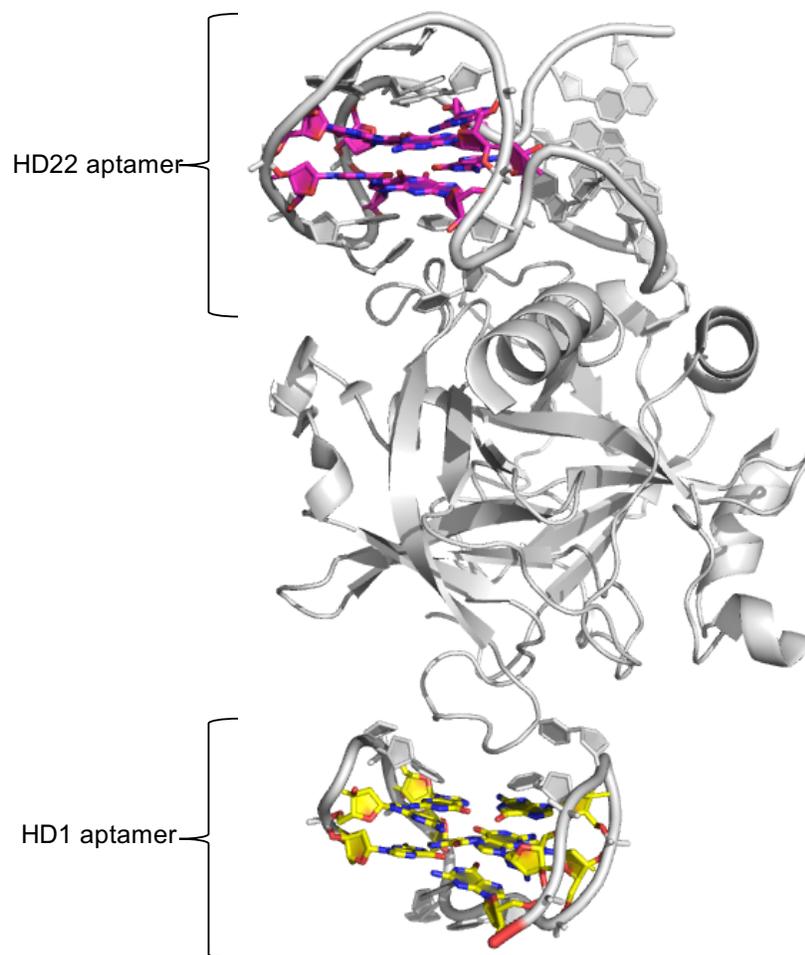


Figure 1.9 Thrombin bound to two DNA aptamers, HD1 and HD22, that each contain a G-quartet scaffold.

The von Willebrand Factor aptamer blocks the interaction of the von Willebrand Factor domain A1 with the platelet-receptor glycoprotein Ib α (Huang et al. 2009). The aptamer contains three helical duplex stems capped by two hairpin loops: one loop is a single-residue GCC loop, and the other is a typical pyrimidine-rich triloop (Figure 1.10). The helices are mainly B-form with Watson-Crick base pairs except for three noncanonical base pairs. Two of these noncanonical base pairs are located near the 3-way junction where they can stabilize the conformation of the junction. The third noncanonical base pair is a sheared type G(anti)-C(anti) closing the GCC loop of one hairpin. The majority of the bases are involved in base pairs, and those that are not are either involved in direct contact with the protein or stabilized by intramolecular interactions. Only one nucleotide in the aptamer is solvent exposed. The size of the interface is extensive, including 911 Å of solvent accessible surface on the aptamer and 1166 Å on the protein buried upon binding. The interface is dominated by hydrogen bonds, salt bridges, and cation- π interactions. Interestingly, the cation- π interactions between the protein and aptamer make up 65% of the binding interface on the protein. One thing that is evident when looking at the structure of this and other protein-aptamer complexes is how much of the aptamer does not contact the protein but often remains essential to binding and efficacy. This suggests that SELEX does not only identify nucleotides that can recognize a protein but also selects for a specific conformation of a nucleic acid that allows it to fold into a form to present those nucleotides to the protein for recognition.

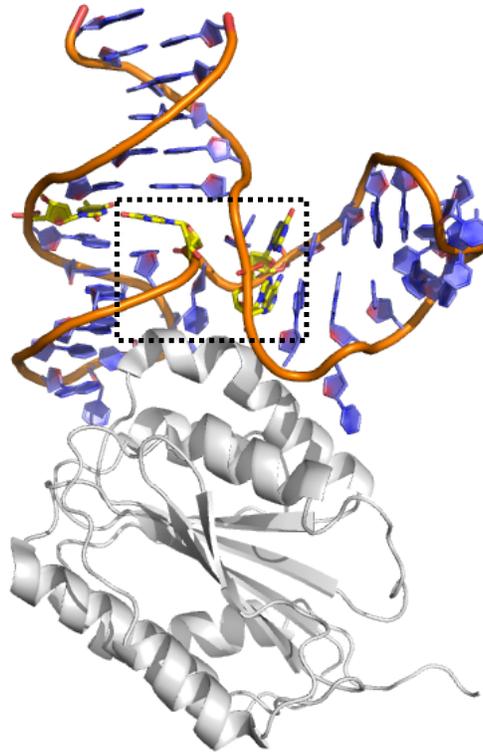


Figure 1.10 The von Willebrand factor bound to a DNA aptamer. The noncanonical base pairs are colored in yellow and are located near the junction of the three helices (boxed).

The activity of *Plasmodium falciparum* lactate dehydrogenase is a diagnostic marker for malaria infection (Makler & Hinrichs 1993). The *P. falciparum* lactate dehydrogenase protein forms a tetramer that binds the DNA aptamer with a stoichiometry of two protomers to one aptamer (Cheung et al. 2013). The aptamer forms a distorted hairpin, with a B form helix at the 5' and 3' ends, an asymmetrical internal loop, and a short hairpin capped by a CATA tetraloop. The interface of the complex has a large buried surface area with 1276 Å² buried by each aptamer. The interface of the complex involves base interactions between the protein and the internal loop of the aptamer, as well as electrostatic interactions with the phosphate backbone.

Autotaxin is a secreted enzyme that hydrolyzes lysophosphatidylcholine that has numerous physiological processes such as blood vessel and brain formation during development, and its aberrant expression has been implicated in idiopathic pulmonary fibrosis (Kato et al. 2016). The autotaxin aptamer is a 34 nucleotide DNA aptamer modified post-SELEX with 2' O-methyl modifications and 3' inverted dT (Kato et al. 2016). The aptamer folds into a conformation that includes a terminal stem, a corner junction, and a hairpin capped by a triloop. The L-shaped aptamer scaffold is formed by an adjacent base triple and base quadruple that stabilize the distorted conformation of the corner junction (Figure 1.11). The role of the corner junction is not limited to a scaffold but also forms part of the interface by binding a cleft in the insertion domain of the protein.

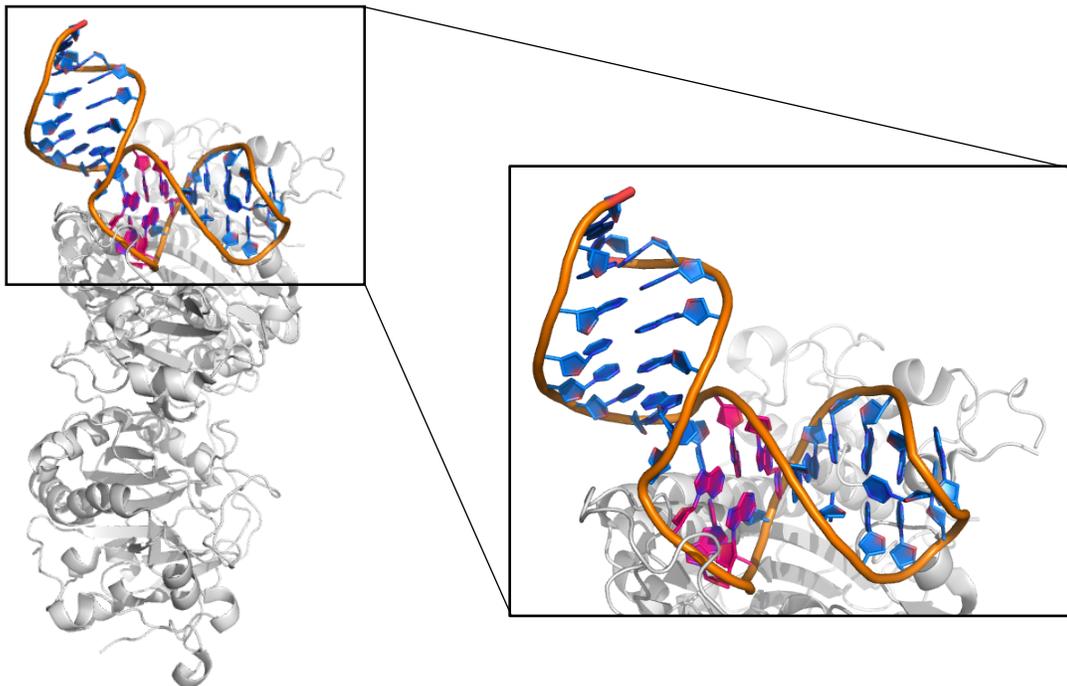


Figure 1.11 Autotaxin bound to a DNA aptamer. The nucleotides forming the base triple and quadruple in the corner junction are highlighted in pink.

1.2.3 Aptamers identified against VEGF₁₆₅

VEGF continues to be a popular target used in SELEX. Some of the first VEGF-targeting aptamers were published in the 1990s with more being published today. While the pool of available VEGF aptamers is large, our lab has focused on a subset that target the heparin-binding domain of VEGF₁₆₅ that includes five RNA aptamers and a DNA aptamer (Jellinek et al. 1994; Gold & Janjic 1998).

Unmodified RNA aptamers

The RNA aptamers were identified using SELEX with unmodified RNA nucleotides. The initial library consisted of 10^{14} RNA molecules each with a 30-nucleotide randomized region. Selections were done by incubating the RNA pool with VEGF₁₆₅ for 10-20 minutes at 37 °C in phosphate-buffered saline (10.1 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, and 2.7 mM KCl, pH 7.4). After 13 rounds of SELEX, the affinity of the RNA pool for VEGF₁₆₅ increased 100-fold. The 37 unique sequences that remained were grouped into six families based on sequence similarities, and the affinity for a representative aptamer from each family was measured with VEGF₁₆₅ (Figure 1.12).

boundaries of the minimal sequence. These sequences were labeled with a “t” for truncated, and the truncated aptamers from families 2-6 are the ones we investigated in our lab.

Table 1.1 Equilibrium dissociation constants* for the RNA aptamers and VEGF₁₆₅.

	Family 1	Family 2	Family 3	Family 4	Family 5	Family 6
Aptamer	100t	44t	12t	40t	84t	126t
K _D	0.5 nM	0.5 nM	1 nM	20 nM	2 nM	2 nM

*Values reported in (Jellinek et al. 1994).

To determine the efficacy of the aptamers, Jellinek and colleagues used ¹²⁵I-radiolabeled VEGF₁₆₅ in HUVEC cells to show that the represented aptamers from each of the families inhibited VEGF-VEGF receptors interaction. To determine the specificity of the aptamers, they chose two aptamers (100t and 44t) and measured the binding affinity to five other proteins with heparin-binding domains: basic fibroblast growth factor, platelet-derived growth factor, antithrombin III, thrombin, and plasminogen activator inhibitor I. These proteins bound the aptamers with K_D values greater than or equal to 0.6 μM. Using competition, they were able to show that all of the RNAs competed with one another and with heparin, which suggested that they all targeted the HBD. However, since the aptamers are large compared to the HBD (10 to 13 kDa aptamers binding to the 6.5 kDa HBD), competitive binding does not mean that they have the exact same binding site. Instead, there may simply be some overlap in the binding interfaces.

Unmodified DNA aptamers

The DNA aptamer that is the focus of this work was identified using the same SELEX method described for the RNA aptamers except with a 40-nucleotide randomized region (Gold & Janjic 1998). To ensure the aptamers were single-stranded during the selection, the reverse primer is tagged with biotin. The complementary strand could then be captured with streptavidin while the aptamers were dissociated from the duplex. The sequences from the SELEX experiments were grouped into four families or a fifth group of orphan sequences. The aptamer investigated in this work was grouped into Family 1 with 19 other sequences, plus two that were placed both in families 1 and 2. Two consensus sequences were reported for this family: AGACAAGAGTGCAGG and GGACTAGAGGGCAGT. Five sequences from family 1 were further analyzed to determine the 5' and 3' boundaries to find the minimal sequences required for binding VEGF₁₆₅. Figure 1.13 below shows these five truncated sequences aligned by the family consensus sequences. The 33t aptamer was tested in HUVEC cells to confirm inhibition of the VEGF₁₆₅-VEGFR interaction. Since the SELEX with the DNA aptamers was not published beyond the patent application, less analysis on the aptamers was provided by the authors.

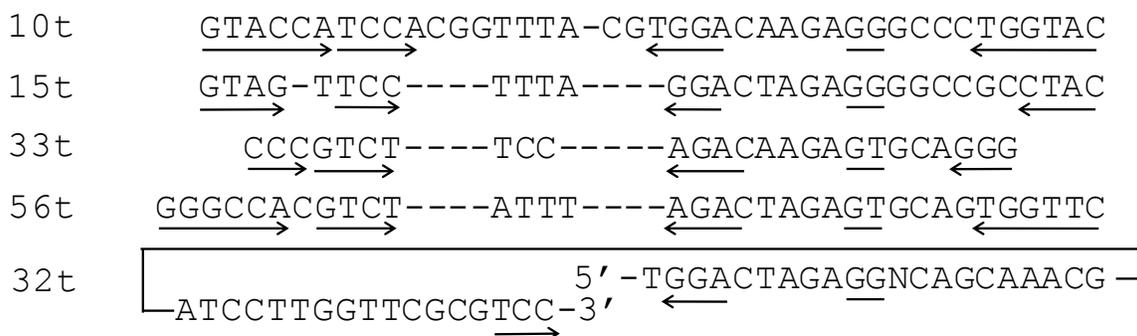


Figure 1.13 Sequence alignment of five truncated DNA aptamers from Family 1. Putative base paired regions are marked with arrows nucleotides hypothesized to participate in forming base triples are underlined.

1.2.4 Background on the DNA aptamer binding to VEGF₁₆₅

The binding of VEGF₁₆₅ to the DNA aptamer was characterized using isothermal titration calorimetry (ITC), fluorescence anisotropy, and surface plasmon resonance (Potty et al. 2009). Fluorescence anisotropy was performed using a fluorescently labeled DNA aptamer to measure the salt dependence of binding, which suggested that only two ion pairs were formed upon the aptamer binding VEGF₁₆₅. Potty and colleagues also found that the K_D increased when the binding was measured in phosphate-buffered saline (10.1 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, and 2.7 mM KCl, pH 7.4, Table 1.2) instead of Tris buffer. Addition of 5 mM MgCl₂ to the PBS buffer, increased the K_D 2.6-fold, which led them to develop two hypotheses: that the conformational flexibility of the free aptamer is important for forming the complex, and the presence of magnesium hinders that flexibility; or that the increase in ionic strength in general, or the presence of magnesium in particular, causes the aptamer to adopt a conformation that hinders binding to VEGF₁₆₅.

Table 1.2 Equilibrium dissociation constants for the DNA aptamer and VEGF₁₆₅.*

Buffer	Additional salt	K_D (nM)
20 mM Tris pH 7.4		1.9 ± 1.0
	50 mM NaCl	67 ± 9
	100 mM NaCl	150 ± 40
	150 mM NaCl	220 ± 50
	200 mM NaCl	860 ± 300
Phosphate-buffered saline		404 ± 70
	5 mM MgCl ₂	1100 ± 200

*Values reported in (Potty et al. 2009).
Mean ± standard deviation.

Potty and colleagues used ITC to measure the binding affinity and obtain thermodynamic information about the formation of the complex. The ITC measured a K_D of 40 ± 4 nM, which was measured in the phosphate-buffered saline that reported a 10-fold weaker affinity by fluorescence anisotropy. However, this discrepancy was not addressed. The ITC results further showed that the binding reaction was enthalpically favorable due to the probable formation of electrostatic interactions and hydrogen bonds and entropically unfavorable due to the hypothesized transition of the aptamer from a mostly disordered free state to a more ordered bound state. An unexpected finding from the ITC results was that the stoichiometry of the complex was one aptamer binding to one VEGF₁₆₅ homodimer, which led them to suggest that the aptamer was binding the dimer interface. This is in contrast to the expected stoichiometry of two aptamers binding to one VEGF₁₆₅ homodimer by way of the heparin-binding domains. Later chapters of this thesis will provide evidence that the latter hypothesis is correct, and the DNA aptamer recognizes VEGF₁₆₅ by binding the HBD.

The on and off rates of the binding reaction were determined using surface plasmon resonance. Corrected for mass-transfer limitations, Potty and colleagues determined a k_{on} of $5.14 \pm 0.03 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ and a k_{off} of $4.64 \pm 0.03 \times 10^{-4} \text{ s}^{-1}$. In addition to characterizing the binding reaction using SPR, they performed mutagenesis using a competition reaction between the mutated aptamers and the immobilized wild-type sequence to analyze the predicted secondary structure of the aptamer (Figure 1.14). Mutations of the nucleotides in the proposed loop region of the hairpin, nucleotides T7, C8, and C9, had a negligible effect on the binding. Significant effects were seen in the T4C and C5T mutants in the pyrimidine-rich strand of the proposed hairpin, but much less of an effect was seen in G3A and T6C. Interestingly, all four of the nucleotides in the purine-rich strand of the proposed hairpin had a 75% decrease in binding when mutated (A10G, G11A, A12G, C13T). Other nucleotides in the proposed consensus sequence showed a decrease in binding of 50% or more upon mutation (A14G, A15C/G/T, G16A, A17G, and G18A). Mutations of the remaining nucleotides were not reported. Using this information, they suggested that the predicted secondary structure from mFold was correct with the aptamer forming a short four base pair hairpin and additional base pairs between the 5' and 3' ends (Figure 1.14). Interestingly, the truncation of even a single nucleotide from the 5' end resulted in a complete loss of detectable binding.

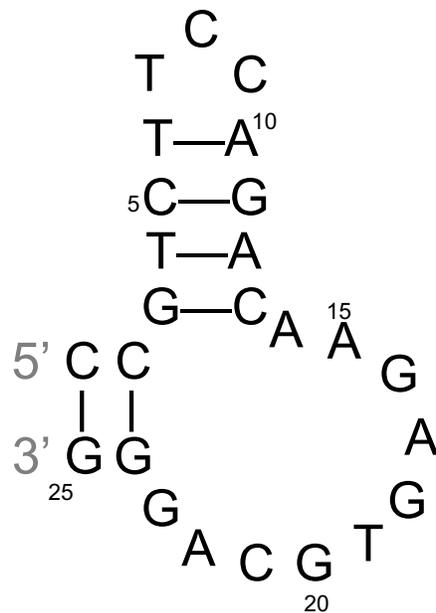


Figure 1.14 Predicted secondary structure of the DNA aptamer. Biochemical analysis of the 25-nucleotide DNA aptamer suggested that the aptamer forms a four-base hairpin capped by a triloop, and the 5' end forms two base pairs with the 3' end.

1.2.5 Thesis overview

The emphasis of this thesis is the structural characterization of a 25-nucleotide DNA aptamer bound to the heparin-binding domain (HBD) of VEGF₁₆₅. Chapter 2 describes the sample preparation for unlabeled and ¹³C, ¹⁵N-labeled HBD and the chemical shift assignments of the backbone and side chain resonances. Chapter 3 includes the method developed to label the aptamer at one position with a nucleotide labeled with ¹³C and ¹⁵N. It also describes the chemical shift assignments for the aptamer resonances and the identification of the two base triples. The intermolecular NOEs identified to define the HBD-aptamer interface are also described in Chapter 3. Chapter 4 describes the solution structure of the HBD-aptamer complex. The structure shows that the orientation of the two subdomains of the HBD changes upon binding the aptamer. The aptamer forms a hairpin with four base pairs capped by a pyrimidine-rich triloop. Two of the base pairs in the hairpin that are proximal to the triloop form a short triplex with two nucleotides,

G18 and T19, that were originally predicted to be unpaired. The 5' and 3' ends base pair and form a GNAR tetraloop-like structure, where nucleotides 20-23 form the tetraloop-like structure and nucleotides 3 to 19 act as an insertion in the tetraloop.

Chapter 2: Chemical shift assignments for the HBD bound to the aptamer

2.1 Introduction

2.1.1 Protein chemical shift assignments

NMR is a powerful technique that provides a method of determine the three-dimensional structure of macromolecules in solution. Structure determination using NMR requires manual inspection of numerous multi-dimensional spectra that provide the necessary information on the macromolecule. NMR is able to distinguish miniscule changes in the chemical environment experienced by a nucleus. Every NMR-active nucleus in a molecule displays a unique chemical shift, though sometimes the differences in chemical shifts are so small that the two peaks are indistinguishable due to the resolution of the acquired spectrum (Keeler 2002). Decades of research have resulted in the development of pulse sequences that utilize the through-bond and through-space connections between atoms (Cavanagh et al. 2007). It is through these various NMR pulse sequences that we are able to assign the chemical shifts from peaks in a spectrum to specific nuclei in a macromolecule. Due to the limited chemical diversity in proteins and the extensive overlap observed with increasing protein size, the field moved beyond the experiments that only assign the proton chemical shifts to multi-dimensional, heteronuclear experiments that also provide the chemical shifts of ^{13}C and ^{15}N incorporated into the protein (Ikura et al. 1990; Kay et al. 1990).

The ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum transfers magnetization from a proton to its attached nitrogen through single-bond J_{NH} coupling (Figure 2.1) (Bax, Ikura, et al. 1990; Bodenhausen & Ruben 1980). It provides a “fingerprint” spectrum unique to that protein under the experimental conditions. The number of peaks in this spectrum is equal to the total number of residues in the protein, excluding the N-terminal residue that is in

fast exchange with the solvent and the proline residues since they contain no amide protons. There are often also additional peaks in the spectrum from tryptophan, glutamine, asparagine, and arginine side chains. By doing a series of heteronuclear multi-dimensional NMR experiments, one can assign each peak in this spectrum to its associated residue in the protein.

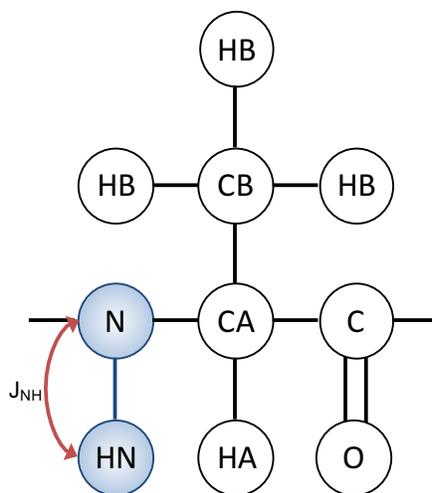


Figure 2.1 Magnetization transfer pathway for ^1H - ^{15}N HSQC. Magnetization is transferred from the amide proton (HN) to the amide nitrogen (N) through J-coupling (J_{NH}) and back to the hydrogen for detection.

The conventional strategy for assigning chemical shifts in a protein first involves assignment of the backbone nuclei (Cavanagh et al. 2007). The residues are connected sequentially using pulse sequences that identify both intra-residue and inter-residue correlations and comparing the resulting spectra to those obtained from pulse sequences that only identify inter-residue correlations. The proton chemical shifts of the side chains are also assigned because they are needed to determine the distance restraints between residues. Since there is often limited chemical shift dispersion between the side chains of residues with similar chemical properties or in similar chemical environments, the side chain heavy atoms are also assigned.

The most crucial piece of information required for structure determination is distance restraints in the form of nuclear Overhauser effects (NOEs). The NOE results from cross relaxation between two protons that are close in space and goes as $1/r^6$, where r is the distance between the nuclei, and therefore drops off quickly with distance (Keeler 2002). The NOE is typically seen between protons that are within five Ångstroms of one another. NOEs provide valuable information about the protons that are in proximity to one another in a folded macromolecule. However, the NOEs become powerful sources of information when the identities of the protons participating in the NOE are known, which requires chemical shift assignments.

2.1.2 *Pichia pastoris* to express recombinant proteins

Escherichia coli is a commonly used organism for recombinant protein expression. However, the *Pichia pastoris* expression system developed by Cregg and colleagues is an alternative method for the expression of more complex proteins that require post-translational modifications or disulfide bonds (Cregg et al. 1987). The HBD contains four disulfide bonds, and earlier studies that expressed HBD in *E. coli* required the protein to be refolded from inclusion bodies and cleaved from full-length VEGF with plasmin, which resulted in very poor yields (Christinger et al. 1996; Fairbrother et al. 1998). Therefore, *P. pastoris* was used for protein expression in more recent studies as well as the work described here (Lee et al. 2005; Lee et al. 2008). *P. pastoris* secretes recombinant proteins into the medium, and by sending the proteins through the secretory pathway, the disulfide bonds important for VEGF structure are formed (Macauley-Patrick et al. 2005).

P. pastoris is a methylotrophic yeast, which means that it can use methanol as its sole carbon source. When the cell is required to utilize methanol as a carbon source, the enzyme alcohol oxidase (AOX) oxidizes the methanol into formaldehyde and hydrogen peroxide within the peroxisome (Krainer et al. 2012). AOX has poor affinity for oxygen, so the cells compensate by making large amounts of AOX (Cregg et al. 1989). By attaching the gene encoding the HBD to the AOX1 promoter, we can induce high levels of HBD expression by treating the cells with methanol. The AOX gene is regulated by a two-step mechanism of repression/depression and induction. The gene is repressed by glucose and derepressed by glycerol, but AOX is only produced when the inducer methanol is present (Tschopp et al. 1987). Therefore, if the cells are grown with glucose as the sole carbon source instead of glycerol, which is done during the expression of ¹³C, ¹⁵N-labeled HBD, the cells must be given time in which glycerol is the sole carbon source. This ensures that the AOX1 promoter is derepressed prior to induction with methanol.

Several *P. pastoris* strains and plasmids are available, but for expression of the HBD, we used the KM71 strain and the pPICZαC plasmid (Invitrogen EasySelect™ *Pichia* Expression Kit). The KM71 strain has the Mut^S (methanol utilization slow) phenotype, which is reflected in its slow growth on methanol due to the disruption of one of the AOX genes, AOX1 (Macauley-Patrick et al. 2005). The cells still contain the AOX2 gene, so they can still metabolize methanol, just not as efficiently as cells containing functional AOX1. The AOX1 promoter is included in the pPICZαC plasmid to take advantage of the propensity of the cells to highly activate that promoter in the presence of methanol. However, instead of transcription of AOX1 being driven by the promoter, the cell transcribes the HBD gene at high levels. The HBD is targeted for secretion using the *Saccharomyces cerevisiae* α-factor prepro peptide secretion signal included in

the pPICZ α C plasmid. Since *P. pastoris* does not secrete many proteins into the medium, this method of expression serves as the first purification step (Macauley-Patrick et al. 2005). The pPICZ α C plasmid also contains a gene for zeocin antibiotic resistance, which provides a mechanism for screening transformants.

2.1.3 Chapter overview

This chapter focuses on the chemical shift assignments of the HBD in the complex with the aptamer. HBD was expressed in the *Pichia pastoris* methylotrophic yeast to ensure proper formation of the disulfide bonds and folding of the protein. A typical preparation of unlabeled HBD from 2 L of *P. pastoris* yielded 6 mg of protein, and 1 L of *P.pastoris* yielded 1.5 mg of ^{13}C , ^{15}N -labeled HBD. Regardless of the presence of stable isotopes, approximately 65% of the HBD was properly folded with correctly formed disulfide bonds. Standard NMR experiments were collected to assign the backbone and side chain resonances of the HBD.

2.2 Materials and methods

2.2.1 HBD expression

The HBD, residues 111 to 165 of the human VEGF₁₆₅, was expressed using the *P. pastoris* strain KM71 transformed with the pPICZ α C plasmid using the EasySelect™ *Pichia* Expression Kit (Invitrogen) (Figure 2.2).

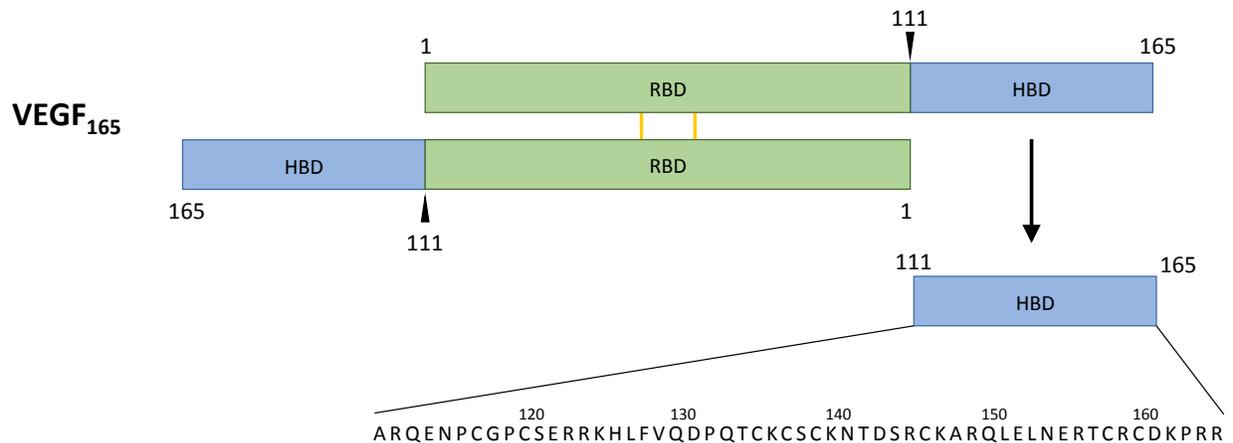


Figure 2.2 HBD construct design. Full-length human VEGF₁₆₅ is an antiparallel homodimer with receptor-binding domains (RBD, green) held together with two intermolecular disulfide bonds (yellow lines) and two heparin-binding domains (HBD, blue). The construct used in the work described here is residues 111 to 165 from VEGF₁₆₅.

Unlabeled HBD expression

The HBD was expressed as described in the EasySelect™ *Pichia* Expression Kit manual (Invitrogen), which was based on the work by Cregg and Higgins (Cregg & Higgins 1995). Starter cultures were grown in 250 mL baffled flasks containing 50 mL buffered minimal glycerol with histidine (BMGH) medium (Table 2.1) and 50 μL of a glycerol stock containing between 2.5×10^9 and 5×10^9 cell/mL. Cell concentration was determined by measuring the OD₆₀₀ and using the conversion one OD₆₀₀ is equal to approximately 5×10^7 cell/mL stated in the *P. pastoris* expression kit manual (Invitrogen). Glycerol stocks were prepared by first streaking the cells on a YPD agar plate [1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, and 2% (w/v) agar]. Plates were incubated at 30 °C for two days, then a single colony was picked and cultured in YPD medium [1% (w/v) yeast extract, 2% (w/v) peptone, and 2% (w/v) dextrose] until they reached log phase (OD₆₀₀ = 2-6). Cells were harvested by centrifugation at 3400 x g using a Hermle 2300K tabletop centrifuge (Labnet) equipped with a 220.97 V02 rotor. The

supernatant was removed, and the cells were resuspended in YPD containing 15% glycerol to an OD₆₀₀ between 50 and 100 (2.5×10^9 to 5×10^9 cell/mL). The glycerol stocks were flash frozen in liquid nitrogen and stored at -70 °C in 50 µL aliquots.

Table 2.1 *P. pastoris* media used for expression of unlabeled HBD.

Buffered Minimal Glycerol with Histidine (BMGH)	Buffered Minimal Methanol with Histidine (BMMH)
100 mM potassium phosphate, pH 6.0 0.34% (w/v) yeast nitrogen base without ammonium sulfate or amino acids 1% (w/v) ammonium sulfate 4×10^{-5} % (w/v) biotin 1% (v/v) glycerol 4×10^{-3} % (w/v) L-histidine	100 mM potassium phosphate, pH 6.0 0.34% (w/v) yeast nitrogen base without ammonium sulfate or amino acids 1% (w/v) ammonium sulfate 4×10^{-5} % (w/v) biotin 4×10^{-3} % (w/v) L-histidine 0.1% (w/v) casamino acids 1×10^{-3} % (v/v) antifoam 1% (v/v) methanol 2X daily

The starter culture was grown overnight at 28 °C shaking at 250 rpm until the cells reached log phase (OD₆₀₀ = 2-6). The culture was then divided into four equal volumes and used to inoculate four cultures of 500 mL BMGH medium in 3 L baffled wide-mouth flasks. Cells were grown for 24 h (OD₆₀₀ = 10-15) at 28 °C shaking at 250 rpm. Cells were then harvested by spinning at 3,000 rpm for 15 min in a Beckman model J2-21 floor centrifuge using a Thermo Scientific FIBERLite F10BA- 6 x 500y rotor. Supernatant was removed, and the cells were resuspended by vortexing in 800 mL Buffered Minimal Methanol with Histidine (BMMH) medium without methanol (Table 2.1). The resuspended cells were divided into two 400 mL cultures in 3 L baffled flasks and allowed to recover in the incubator for 30 min before the first methanol induction. Each methanol induction consisted of 8 mL of filtered 50% methanol in water. The 50% methanol solution was added twice a day (morning and evening) for two days. The cells were then spun down at 10,000 rpm for 25 min using a Thermo Scientific FIBERLite

F14BA- 6 x 250y rotor in the Beckman J2-21 centrifuge. The medium was collected and filtered through a 0.2 µm polyethersulfone membrane filter. Purification was typically performed immediately, but the filtered medium could be stored overnight at 4 °C with 1 mM EDTA to discourage any proteases present in the medium.

¹³C, ¹⁵N-labeled HBD

The expression of ¹³C, ¹⁵N-labeled HBD used the same protocol as described for unlabeled HBD with a few differences. The starter culture and two 500 mL cultures all contained BMDH medium containing ¹³C-glucose and ¹⁵N-ammonium chloride (Table 2.2). Glucose was chosen over glycerol because of the high price of ¹³C- glycerol. ¹³C, ¹⁵N- ISOGRO® supplement (Sigma-Aldrich) was included at a concentration of 0.5 g/L to enhance HBD expression. Once the cells reached an OD₆₀₀ between 10 and 15, they were harvested and resuspended in 400 mL BMMH medium lacking methanol but including 0.1% ¹³C-glycerol. The glycerol was necessary to derepress the AOX promoter, which had been repressed in the presence of glucose.

Proceeding immediately from glucose as a carbon source to methanol as a carbon source would put unnecessary stress on the cells since even in the methylotrophic yeast, methanol can be toxic (Macauley-Patrick et al. 2005). The cells were incubated at 28 °C and 250 rpm for 2 to 3 hours to consume the glycerol and depress the AOX promoters. Protein expression was induced by adding 8 mL of ¹³C-methanol diluted to 50% in water twice (morning and evening) since the majority of the HBD expression occurs within the first 24 hours. Following the expression, the medium containing the secreted HBD was collected and loaded onto a heparin affinity column for HBD purification.

Table 2.2 *P. pastoris* media for expression of ¹³C, ¹⁵N- labeled HBD.

Buffered Minimal Dextrose with Histidine (BMDH)	Buffered Minimal Methanol with Histidine (BMMH)
100 mM potassium phosphate, pH 6.0 0.34% (w/v) yeast nitrogen base without ammonium sulfate or amino acids 1% (w/v) ¹⁵ N- ammonium chloride (10 g/L) 4 x 10 ⁻⁵ % (w/v) biotin 1.5% (w/v) ¹³ C- dextrose (glucose) (15 g/L) 4 x 10 ⁻³ % (w/v) L-histidine 0.05% (w/v) ¹³ C, ¹⁵ N- ISOGRO (0.5 g/L)	100 mM potassium phosphate, pH 6.0 0.34% (w/v) yeast nitrogen base without ammonium sulfate or amino acids 1% (w/v) ¹⁵ N- ammonium chloride (10 g/L) 4 x 10 ⁻⁵ % (w/v) biotin 4 x 10 ⁻³ % (w/v) L-histidine 0.1% (w/v) casamino acids 1 x 10 ⁻³ % (v/v) antifoam 0.1% (v/v) ¹³ C- glycerol (1 mL/L) 1% (v/v) ¹³ C- methanol 2X daily (10 mL/L)

2.2.2 HBD purification

The filtered *P. pastoris* medium was loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare Life Sciences) per the instructions provided by the supplier at 5 mL/min using a BioRad BioLogic LP system. HBD was eluted using a step gradient of increasing concentration of NaCl (Table 2.3). Two peaks eluted from the heparin column: the peak that eluted at 0.5 M NaCl contained misfolded HBD, and the peak that eluted at 0.6 M NaCl contained correctly folded HBD (Figure 2.3; Appendix A1). Fractions containing correctly folded HBD were concentrated using an Amicon Ultra concentrator MWCO 3 K (Millipore) and run over a size exclusion column (HiLoad 16/60 Superdex 75, GE Healthcare Life Sciences) in sodium phosphate buffer (10 mM sodium phosphate pH 7.4, 100 mM NaCl, 3 mM KCl, 0.1 mM EDTA) (Figure 2.4). HBD fractions were concentrated and buffer exchanged into sodium phosphate NMR buffer (10 mM sodium phosphate pH 7.4, 100 mM NaCl, 3 mM KCl, 0.1 mM EDTA, 0.15 mM TSP, 10% D₂O/90%H₂O).

Table 2.3 Heparin column buffers and elution method

Buffer A	10 mM sodium phosphate pH 7.0	
Buffer B	10 mM sodium phosphate pH 7.0 2 M NaCl	
Elution method		
10 min	100 % Buffer A	5 mL/min
10 min	10 % Buffer B	2 mL/min <i>Begin fraction collection</i>
20 min	25 % Buffer B	2 mL/min
30 min	30 % Buffer B	2 mL/min

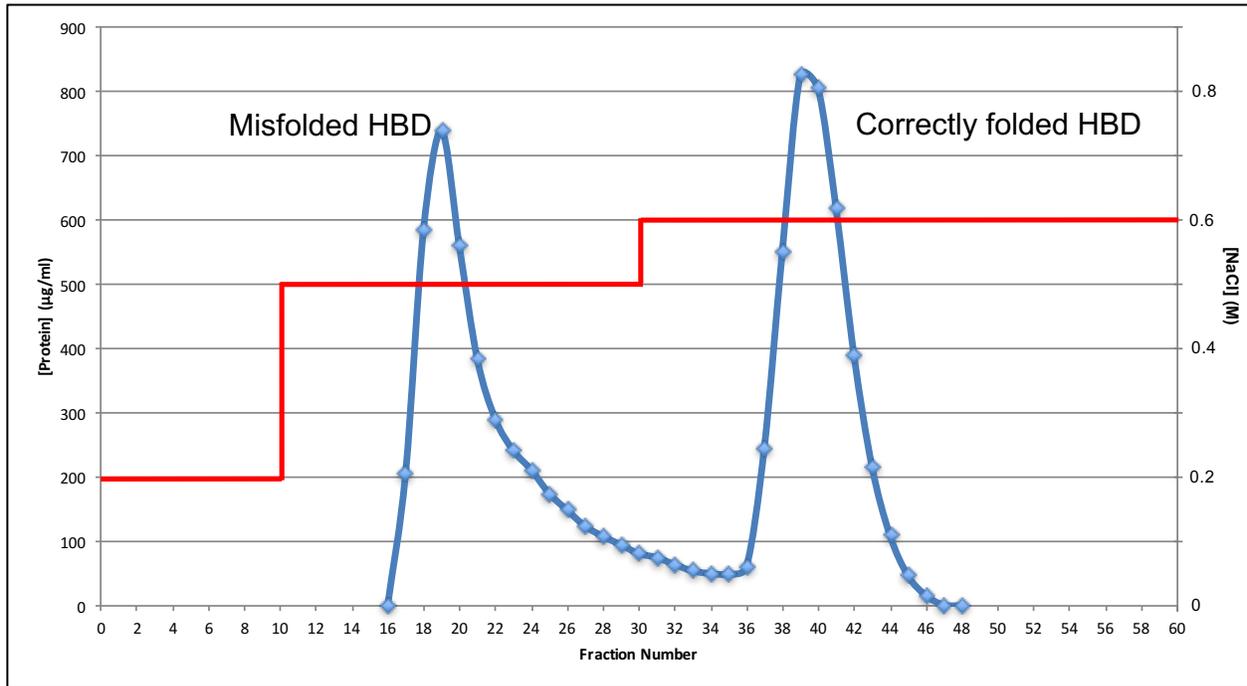


Figure 2.3 Representative trace of HBD eluting from heparin affinity column. HBD is eluted from the heparin affinity column using a step gradient of increasing NaCl concentration (shown as the red line and y-axis on the right side of the graph). Two peaks elute during the gradient. Misfolded HBD elutes during the 0.5 M NaCl step, while correctly folded HBD elutes during the 0.6 M NaCl step (Appendix A1). Protein concentration was determined for each fraction using the BCA assay.

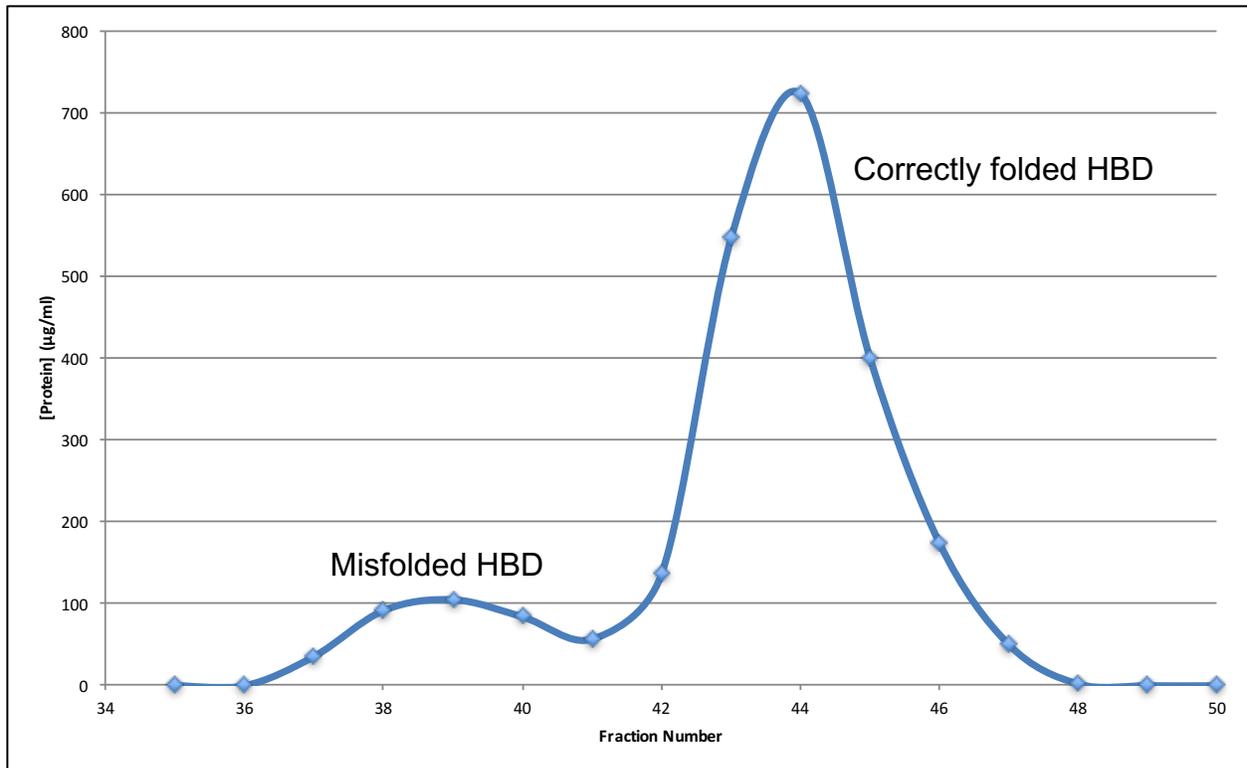


Figure 2.4 Representative trace of relevant fractions eluting from the size exclusion column. Misfolded HBD elutes from the size exclusion column first if it is present in the sample. The correctly folded HBD can be cleanly separated from the misfolded protein by taking only fractions from the second peak that do not overlap with the first peak. Protein concentration was determined for each fraction using the BCA assay.

HBD contains one phenylalanine as its only aromatic residue, which means its extinction coefficient is extremely low and unreliable for determining the concentration by absorbance at 280 nm. Instead, the bicinchoninic acid (BCA) assay (Thermo Scientific) was used to determine the concentration of HBD. The peptide bond, cysteine, tryptophan, and tyrosine are all capable of reducing copper from Cu^{2+} to Cu^{1+} (Wiechelman et al. 1988). The BCA reagent chelates the Cu^{1+} and forms a deep purple color that absorbs at 562 nm (Smith et al. 1985). The increase in A_{562} tracks linearly with increasing protein concentration between 20 $\mu\text{g}/\text{mL}$ to 2 mg/mL (Sigma-Aldrich 2012). The assay is set up with a standard curve using BSA, and the A_{562} of HBD is compared to the standard curve to determine the protein concentration. A typical preparation of HBD from *P. pastoris* yielded 6 mg of unlabeled protein from a 2 L growth and 1.5 mg of ^{13}C , ^{15}N -labeled HBD from a 1 L growth. Regardless of the presence of stable isotopes, approximately 65% of the HBD was properly folded with correctly formed disulfide bonds.

2.2.3 NMR experiments

All NMR experiments were performed on a Varian Inova 600 MHz or VNMRS 800 MHz spectrometer equipped with a triple resonance, z-axis gradient cold probe. All experiments were performed at 25 °C using the Biopack pulse sequences unless otherwise stated. NMR spectra were processed using NMRPipe and analyzed using CCPNMR Analysis (Delaglio et al. 1995; Vranken et al. 2005).

The ^{15}N - or ^{13}C , ^{15}N -HBD was prepared in 10 mM sodium phosphate pH 7.4, 100 mM NaCl, 3 mM KCl, 0.1 mM EDTA, 0.15 mM TSP, and 10% $\text{D}_2\text{O}/90\% \text{H}_2\text{O}$ and titrated into the aptamer until approximately 90% of the aptamer was bound (Section 3.3.1). The HBD-aptamer complex was then buffer exchanged using a 0.5 mL ZebaTM spin desalting column 7K MWCO

(Thermo Scientific) into 20 mM deuterated succinate pH 5.5, 50 mM NaCl, 0.1 mM EDTA, 0.05% NaN₃, 0.15 mM TSP, and 10% D₂O/90% H₂O. A sample of 0.3 mM ¹³C, ¹⁵N-HBD bound to unlabeled aptamer or 0.6 mM ¹⁵N-HBD bound to unlabeled aptamer, both in the succinate buffer, were used to assign the protein resonances.

The following 3D spectra were collected to assign the backbone: HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, and CBCA(CO)NH. Side chain assignments were done using the following 3D experiments: HCCH-total correlation spectroscopy (TOCSY), ¹⁵N TOCSY-HSQC, H(CCO)NH-TOCSY, and C(CO)NH-TOCSY. Intramolecular and intermolecular NOEs were identified using the Biopack 3D ¹⁵N NOESY-HSQC and the 3D ¹³C, ¹⁵N SIM-NOESY-HSQC (Eichmüller et al. 2001). A list of the experimental parameters is shown in Appendix A2.

2.2.4 Dihedral angle predictions

Ranges for the ϕ and ψ dihedral angles were obtained using the program Torsion Angle Likelihood Obtained from Shift and Sequence Similarity plus (TALOS+) using the Janus supercomputer operated by the University of Colorado Boulder (Cornilescu et al. 1999; Shen et al. 2009). The chemical shifts of the backbone atoms (HN, N, HA, CA, CB, and C) were used as input for the program. The output file contained the predicted ϕ and ψ dihedral angles and standard deviations of the angles. The structure calculation program CYANA was used to convert the TALOS+ output file into a dihedral angle restraints file with the range on the accepted dihedral angles equal to twice the standard deviation (Güntert 2004). PdbStat was used to convert the dihedral angle restraint file from CYANA format to Xplor-NIH format (Tejero et al. 2013).

2.3 Results and discussion

2.3.1 Backbone resonance assignments of the HBD bound to the aptamer

The HBD construct is 55 residues long, but 49 peaks are in the ^1H - ^{15}N HSQC spectrum expected because HBD contains four proline residues, an unlabeled histidine, and the N-terminal amide group does not typically show up in the spectrum due to fast exchange with the solvent (Cavanagh et al. 2007). Only forty-eight peaks from backbone amide groups are seen in the ^1H - ^{15}N HSQC spectrum of the HBD bound to the aptamer (Figure 2.5). The amide group of Ser121 does not appear in the spectrum of the aptamer-bound HBD and has low intensity in the HSQC of the free HBD, perhaps due to exchange with the solvent or increased dynamics causing line broadening (Mittermaier & Kay 2009; Kleckner & Foster 2011). Several residues C-terminal to Ser121 (Glu122-Lys125) have peaks in the HSQC that are lower intensity than those from the rest of the protein, which may be due to unfavorable relaxation properties. The chemical shifts of the free HBD were previously assigned by Fairbrother and colleagues (Fairbrother et al. 1998). The chemical shifts were reproduced here. The chemical shift of nearly every peak in the spectrum changes when bound to aptamer (Figure 2.5). Titration of the HBD into the aptamer showed that the complex is in the slow exchange regime, so every peak in the spectrum had to be reassigned through conventional methods instead of tracking peak movement during a titration (Section 3.3.1) (Cavanagh et al. 2007).

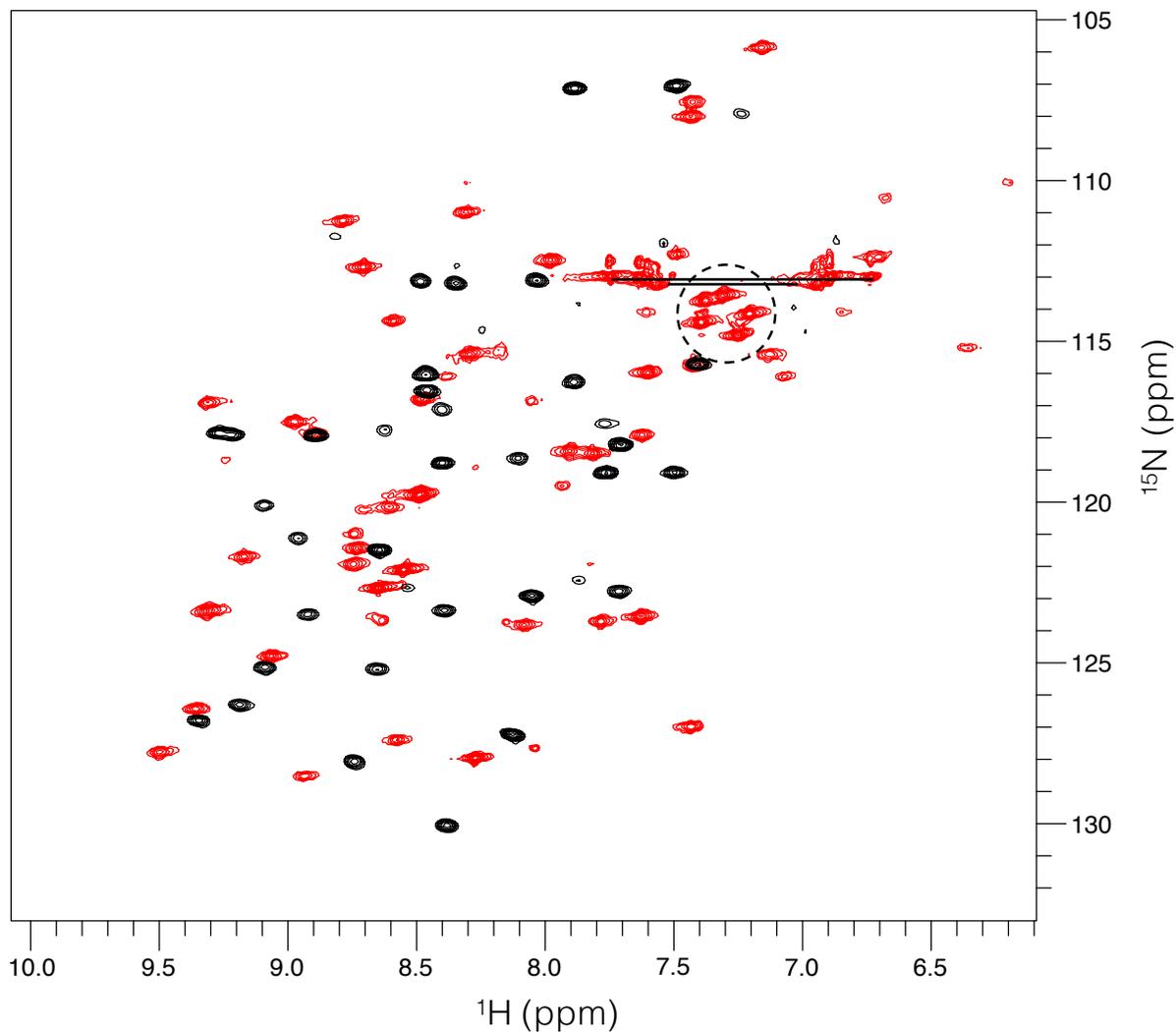


Figure 2.5 ^1H - ^{15}N HSQC spectra of free and bound HBD. The spectra of free HBD (black) and HBD bound to aptamer (red) were both collected in 10 mM sodium phosphate pH 7.4, 100 mM NaCl, 3 mM KCl, 0.1 mM EDTA, 0.15 mM TSP, and 10% D_2O . Most of the peaks between 6.7 and 7.7 ppm in the proton dimension and between 112 and 115 ppm in the nitrogen dimension belong to side chain protons. The asparagine and glutamine amino groups are connected by a solid line. The arginine amide protons are clustered by a dashed circle and are folded in the spectrum from the true position of the peaks around 95 ppm.

Backbone resonance assignments were started by first identifying specific residues in the HNCACB spectrum. The HNCACB experiment transfers magnetization from the amide proton to its attached nitrogen as well as the CA and CB (Figure 2.6A) (Wittekind & Mueller 1993). While the higher intensity peak in the spectrum corresponds to the intra-residue atoms, magnetization is also transferred to the preceding residue CA and CB (Cavanagh et al. 2007). This happens because the one-bond coupling constant between the intra-residue amide nitrogen to CA and two-bond coupling constant between the inter-residue amide nitrogen to CA have similar values of 7-11 Hz and 4-9 Hz respectively (Delaglio et al. 1991). Both the intra-residue and inter-residue CA and CB could often be seen in the spectrum of HBD in the complex. The inter-residue peaks were confirmed by using the complementary experiment, CBCA(CO)NH, which transfers magnetization from the preceding residue CA and CB, through the carbonyl carbon without labeling it, and to the amide group (Figure 2.6B) (Grzesiek & Bax 1992).

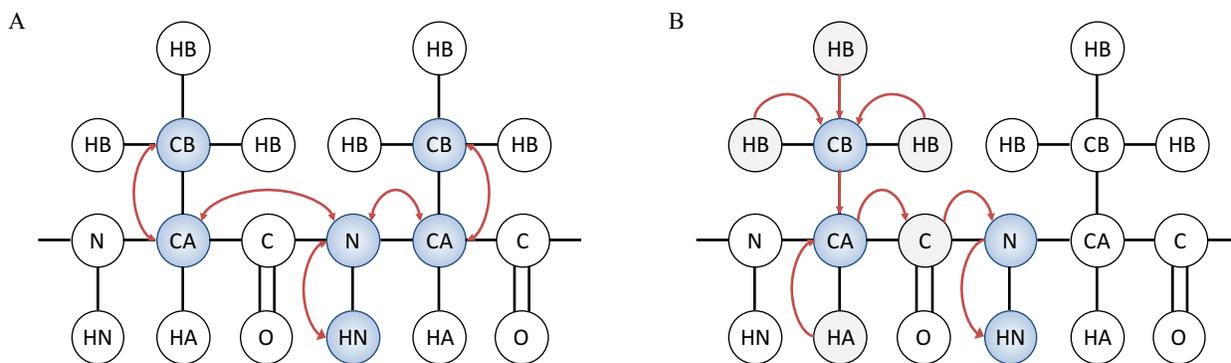


Figure 2.6 Magnetization transfer pathways for the HNCACB and CBCA(CO)NH experiments. Nuclei in blue circles are frequency-labeled; nuclei in grey circles are involved in the coherence transfer but no frequency information is collected on them. Red arrows show the transfer of magnetization. (A) HNCACB is an out-and-back experiment that transfers magnetization from the HN to the N, from the N to the intra-residue and inter-residue CA, then to the CB and back to the CA, N, and finally the HN for detection. (B) CBCA(CO)NH magnetization is transferred from the HA and HB protons to the CA and CB respectively, and from the CB to the CA. Magnetization is then transferred from the CA to the N and then to the HN.

Amino acids with unique chemical shifts can be identified using the chemical shifts of the CA and CB. Amino acids with distinctive chemical shifts (e.g. serine, threonine, alanine, and glycine) are common residues with which to begin backbone assignments. This is because serine and threonine both have CB chemical shifts between 60 and 70 ppm, which are significantly different from the other amino acids (Figure 2.7). Alanine has an upfield CB chemical shift around 15 ppm that is also well separated from the other amino acids. Glycine has an upfield CA chemical shift around 40 ppm, whereas the other amino acids have typical CA chemical shifts between 50 and 60 ppm. These characteristics allow for easy identification of the glycine, alanine, serine, and threonine residues.

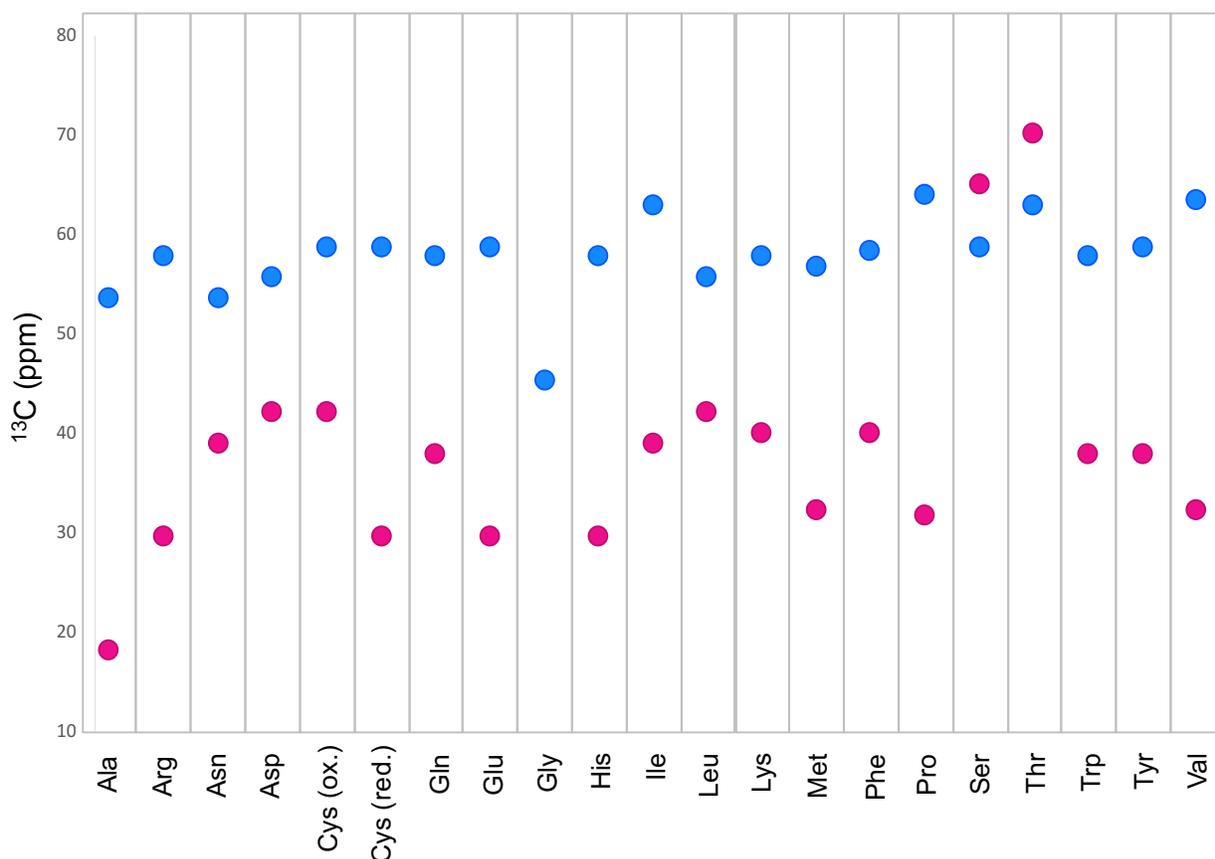


Figure 2.7 Average CA and CB chemical shifts. Average chemical shifts for alpha carbons (blue) and beta carbons (pink). Adapted from Cavanagh et al. 2007.

The best place to begin assignments of the HBD was Thr142-Asp143-Ser144 because it contained two of the four amino acid types with easily identifiable chemical shifts (Figure 2.8). From there, the residues were connected sequentially using the CA and CB. The four proline residues (Pro116, Pro119, Pro132, and Pro163) broke up the connections. The HNCACB had too poor of signal to noise due to unfavorable relaxation properties for residues Ser121 to His126 to allow for sequential assignment. The HNCA and the complementary HN(CO)CA correlated the amide group to the attached CA and that of the preceding residue (Figures 2.9 A and B). The HNCO and HN(CA)CO provided chemical shifts of the carbonyl carbons (Figures 2.9 C and D). With these experiments, connections could be made between Glu122, Arg123, Arg124, and Lys125 (Figure 2.10). Cys120 could not be connected to Glu122 because the amide group from Ser121 did not show up in any 3D spectra. Lys125 also could not be connected to Leu127, but that is because during the *P. pastoris* growth, unlabeled histidine is supplemented into the medium, so His126 is unlabeled and therefore does not show up in these spectra. There is a low intensity peak in the HNCO from Leu127 that could be from natural abundance ^{13}C in His126, but it may also be an artifact. Additional strip plots of backbone connectivities can be found in Appendix A3. Following the completion of the backbone assignments, the fingerprint HSQC spectrum was assigned (Figure 2.11). The backbone assignments are nearly complete, with 94% of the amide protons, 87% of the amide nitrogens, 96% of the HA, 93% of the CA, and 98% of the CO assigned (Appendix A4).

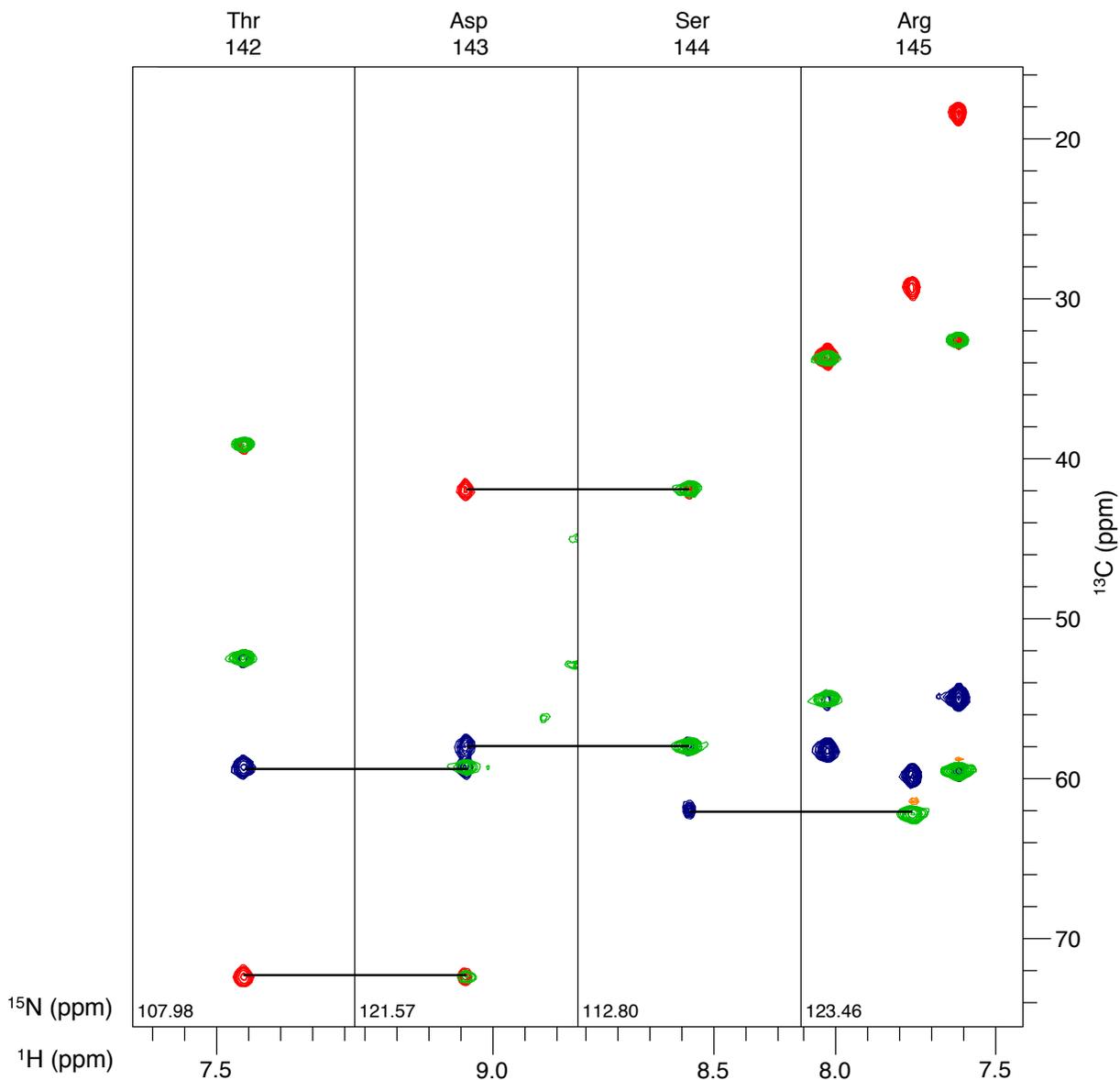


Figure 2.8 Strip plots of the HNCACB/CBCA(CO)NH spectra for Thr142- Arg145. The HNCACB spectrum shows the alpha carbons with positive contours in blue and beta carbons with negative contours in red for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. There is no peak corresponding to the beta carbon of Ser144, but the chemical shift of the alpha carbon was indicative of the amino acid type.

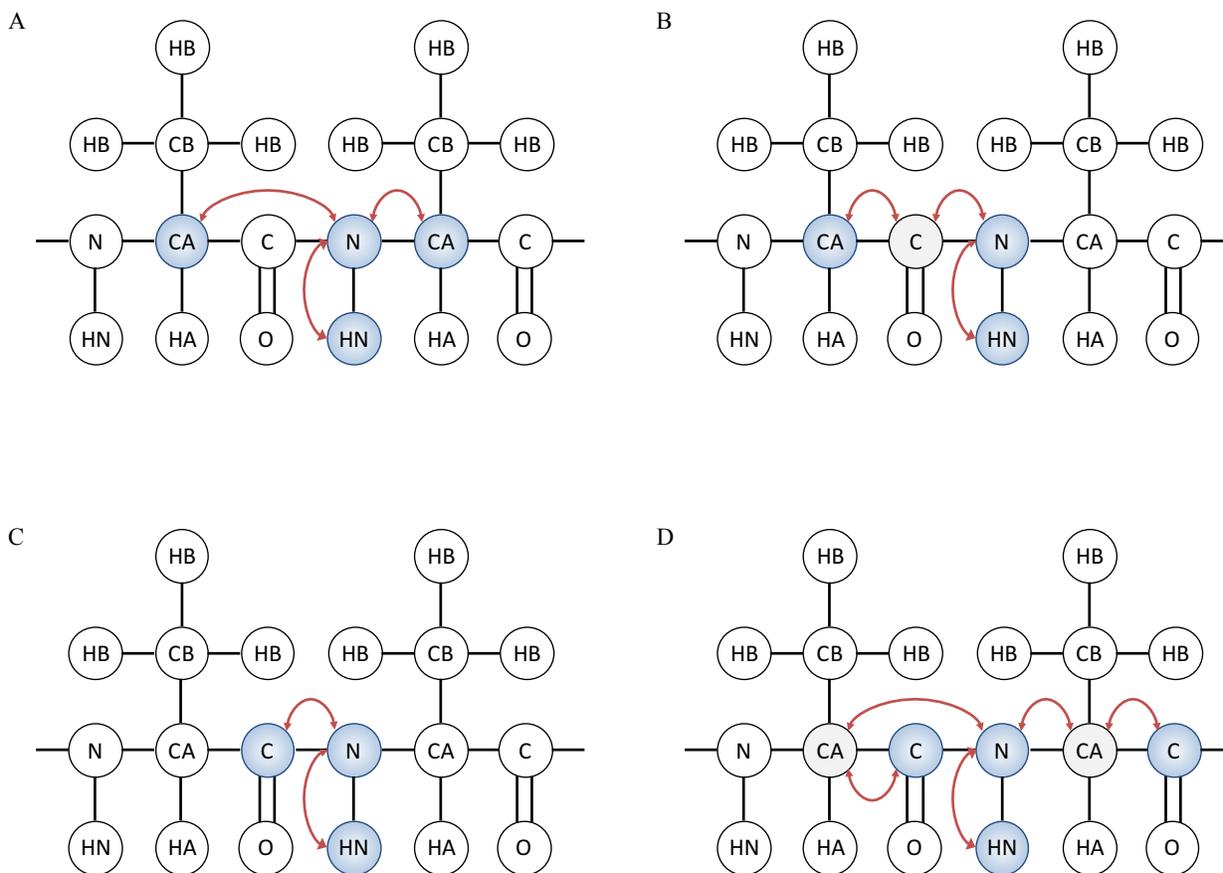


Figure 2.9 Magnetization pathways for HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments. (A) In the HNCA, magnetization is transferred from the HN to the N. From there, the magnetization is transferred to both the intra-residue and inter-residue CA, back through the N to the HN. (B) In the HN(CO)CA, magnetization is transferred from the HN through the N, C, and CA of the previous residue and back to the HN through the C and N. (C) In the HNCO, magnetization is transferred from the HN to the N and C, then back to the N and HN. (D) In the HN(CA)CO, magnetization is transferred from the HN to the N. The magnetization is then transferred through the intra-residue and inter-residue CA to the C and back through the CA and N to the HN.

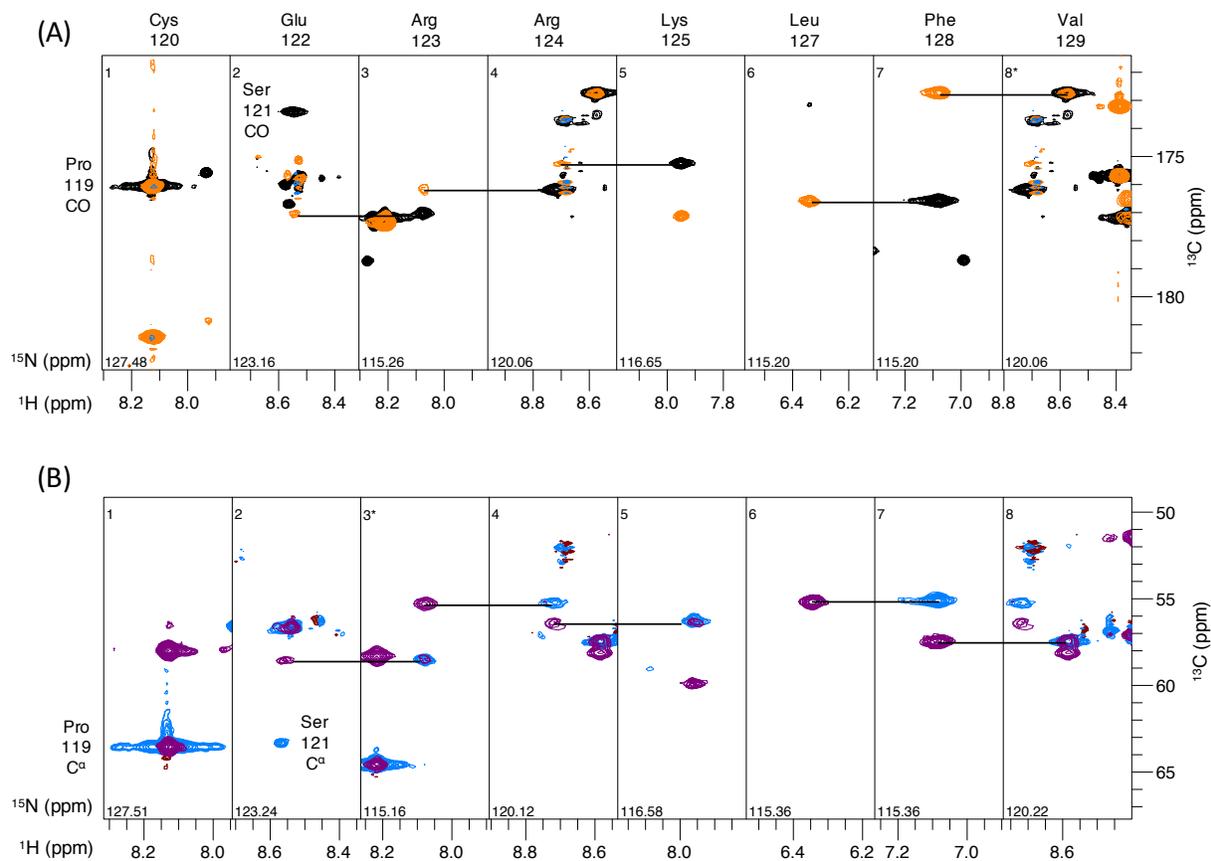


Figure 2.10 Strip plots of HNCO/HN(CA)CO and HNCA/HN(CO)CA for residues Cys120, Glu122-Lys125, and Leu127-Val129. (A) Overlay of the HNCO (black) and HN(CA)CO (orange) spectra. (B) Overlay of the HNCA (purple) and HN(CO)CA (blue) spectra.

(A) ¹¹¹ARQENPCGP ¹²⁰CSERRKHLFV ¹³⁰QDPQTCKCSC ¹⁴⁰KNTDSRCKAR
¹⁵⁰QLELNERTCR ¹⁶⁰CDKPRR

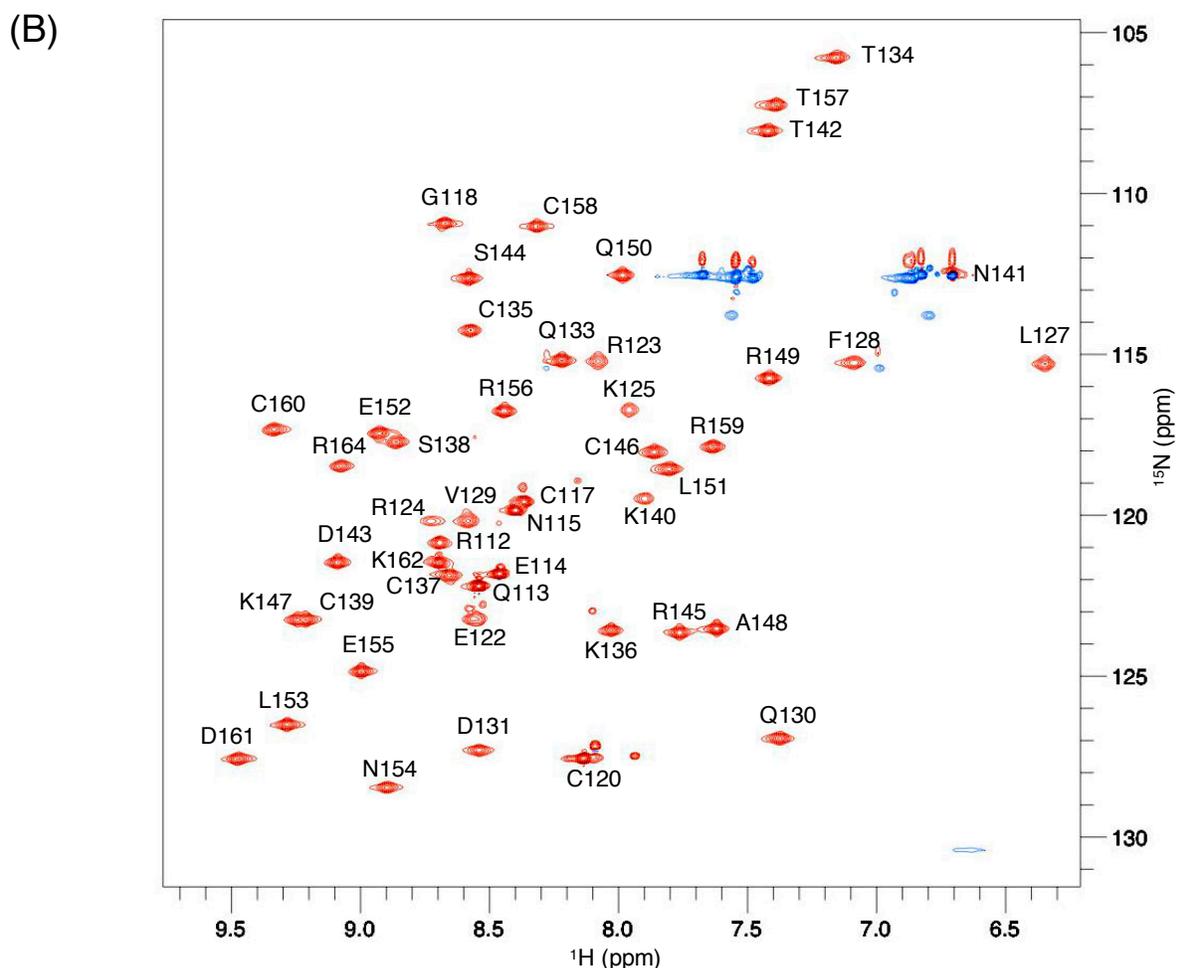


Figure 2.11 Chemical shift assignments of the backbone amides of the HBD bound to the aptamer. (A) Amino acid sequence of the HBD. Residues with backbone amide assignments are black and those without backbone amide assignments are in gray. (B) The ¹H-¹⁵N HSQC spectrum of the HBD-aptamer complex was collected in succinate NMR buffer pH 5.5 at 25 °C. Peaks are assigned with the one letter amino acid code using the construct numbering scheme 111 to 165.

2.3.2 Dihedral angle calculations

Dihedral angles were calculated using the program TALOS+ (Cornilescu et al. 1999; Shen et al. 2009). Chemical shifts of the backbone nuclei are sensitive to conformation (Pardi et al. 1983; Williamson 1990; Wishart et al. 1991; Ösapay & Case 1994). It was found that protein homologues had similar secondary chemical shifts, the amount by which the observed chemical shift deviates from the chemical shift of the amino acid in a random coil, which suggested that the backbone conformation information was included in the chemical shift values (Redfield & Robertson 1991; Mielke & Krishnan 2009). The program TALOS+ takes the chemical shift input from the protein of interest and breaks it up into fragments of three amino acids. These fragments are searched against a database looking at the secondary chemical shift and the amino acid sequence. Ten similar triples are selected, and if they have similar ϕ and ψ angles, then the values can be reliably averaged and reported as the predicted dihedral angles for the protein of interest. The program has a low reported error rate of $< 2.0\%$, which is usually from amino acids that are not in secondary structure elements (Shen et al. 2009). TALOS+ also attempts to estimate the order parameters of the residues from the backbone chemical shifts (Berjanskii & Wishart 2005). Since the program has a higher rate of error in flexible regions of proteins, the predicted order parameters allow one to avoid potentially erroneous predictions in a protein for which there is no experimentally obtained data about the dynamics.

TALOS+ labels predicted dihedral angles as “good,” “warn,” or “dynamic.” For a prediction to be labeled “good,” there must be a majority consensus in the database triplet matches. The “warn” label is given to a residue if there is no consensus in the database matches, and it is not recommended that these dihedral angle predictions be used. A residue is listed as “dynamic” if the predicted order parameter is less than 0.5. The dihedral angles for residues

listed as “dynamic” (Arg112-Pro116) and those listed as “warn” (Cys117, Gly118, Arg123, Arg124, Cys139, and Arg164) were not included in the dihedral angle restraint file used in the structure calculation. The dihedral angles predicted for the HBD in the complex are shown in Appendix A5.

2.3.3 Side chain resonance assignments of the HBD bound to DNA aptamer

The type of amino acid was determined by the unique chemical shifts of some amino acids, such as alanine, glycine, serine, and threonine, and by analysis of the HNCACB/CBCA(CO)NH spectra. The 3D ^1H - ^{15}N TOCSY-HSQC connects the amide group to the intra-residue side chain protons, which are correlated by magnetization transferred through ^1H isotropic mixing (Figure 2.12A) (Marion, Driscoll, et al. 1989). Additional side chain assignments were done using the 3D HCCH-TOCSY transfers magnetization from a proton to its attached carbon and to all of the carbons in the side chain through isotropic ^{13}C mixing (Figure 2.12B) (Bax, Clore, et al. 1990). For studies here, the additional carbons were not frequency-labeled and the magnetization was transferred to the attached protons that were frequency-labeled during the experiment. An example plot using the fully assigned Val129 side chain is shown below in Figure 2.13. By convention, the detected ^1H dimension is shown on the y-axis instead of the x-axis.

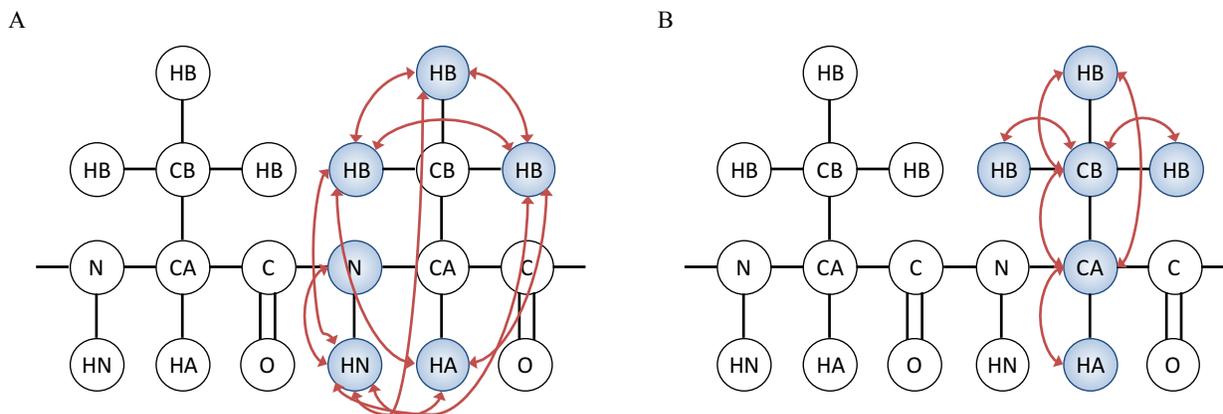


Figure 2.12 Magnetization transfer pathways for the TOCSY-HSQC and HCCH-TOCSY. (A) In the TOCSY-HSQC, magnetization is transferred between the protons using isotropic mixing, then from the HN to the N and back for detection. (B) In the HCCH-TOCSY, magnetization is transferred from the protons to their attached carbons. The magnetization is transferred among the carbons through isotropic mixing and back to the protons for detection.

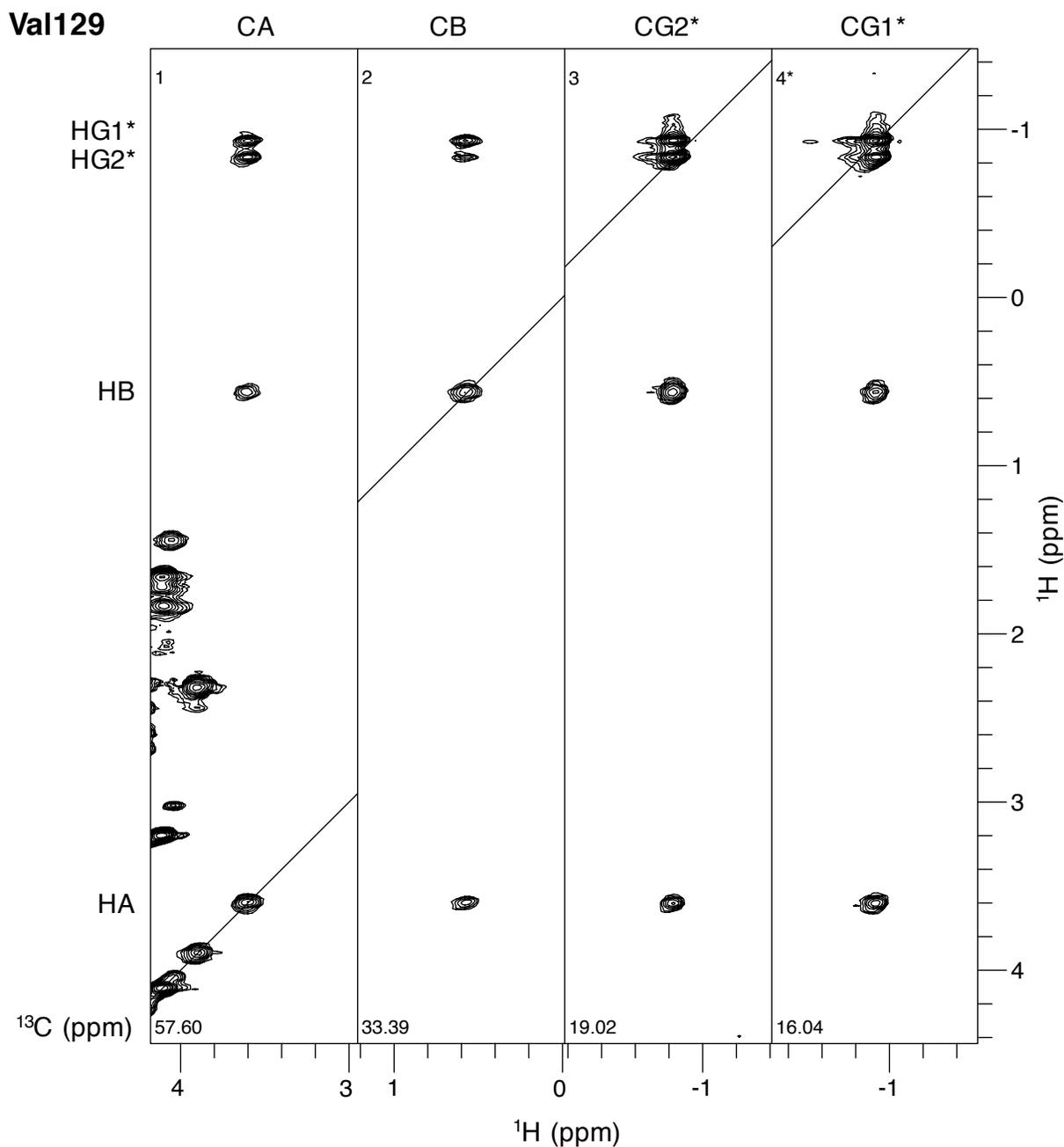


Figure 2.13 Strip plots of Val129 from the HCCH-TOCSY. The diagonal peak in each strip corresponds to a specific CH or CH₃ group in Val129, but the other protons in the side chain appear in each strip. This spectrum allows for correlation of the side chain atoms.

Two additional 3D spectra were collected: the H(CCO)NH-TOCSY correlated the amide group to the side chain protons of the previous residue, and the CC(CO)NH-TOCSY correlated the amide group to the side chain carbons of the previous residue (Grzesiek et al. 1993; Montelione et al. 1992). These spectra resulted in a significant increase in the number of assigned side chains. Of the 54 labeled amino acids in the HBD construct, 81% of the HB, 94% of the CB, 42% of the HG, 83% of the CG, 53% of the HD, 63% of the CD or ND, 65% of the HE or HZ and 60% of the CE, CZ, or NE were assigned (Appendix A4). Assignment of the side chains was important for properly identifying intra-residue and inter-residue NOEs within the protein, as well as intermolecular NOEs between the protein and aptamer.

2.3.4 HBD intramolecular NOEs

NOEs provide critical information for structure determination because they are seen between protons that are close in space. For the structure determination of the HBD-aptamer complex, both intramolecular and intermolecular NOEs were collected and identified. Intramolecular NOEs were collected with 3D ^{15}N NOESY-HSQC and 3D ^{13}C , ^{15}N SIM-NOESY-HSQC experiments. The ^{15}N NOESY-HSQC displays NOEs from a proton attached to a ^{15}N -labeled amide nitrogen to other protons in the HBD and the aptamer (Figure 2.14) (Marion, Kay, et al. 1989; Zuiderweg & Fesik 1989). The ^{13}C , ^{15}N SIM-NOESY-HSQC is able to distinguish between intramolecular and intermolecular NOEs (Figure 2.15). The experiment collects two sets of data: in one spectrum, all of the NOEs have the same sign, and in the second spectrum, the sign of the intramolecular NOEs is opposite the first spectrum (Eichmüller et al. 2001). Upon addition of the two data sets following acquisition, the intramolecular NOEs cancel out, and only the intermolecular NOEs are seen in the spectrum. Upon subtraction of the data,

only the intramolecular NOEs are present. Only NOEs that were present in the addition spectra and not in the subtraction spectrum were treated as intermolecular NOEs, and the remainder were identified as intramolecular or treated as ambiguous. Inter-residue NOEs were identified either with confidence based on unique chemical shifts or left as ambiguous due to similar chemical shifts until preliminary structure calculations resolved the ambiguity (Section 4.2.1).

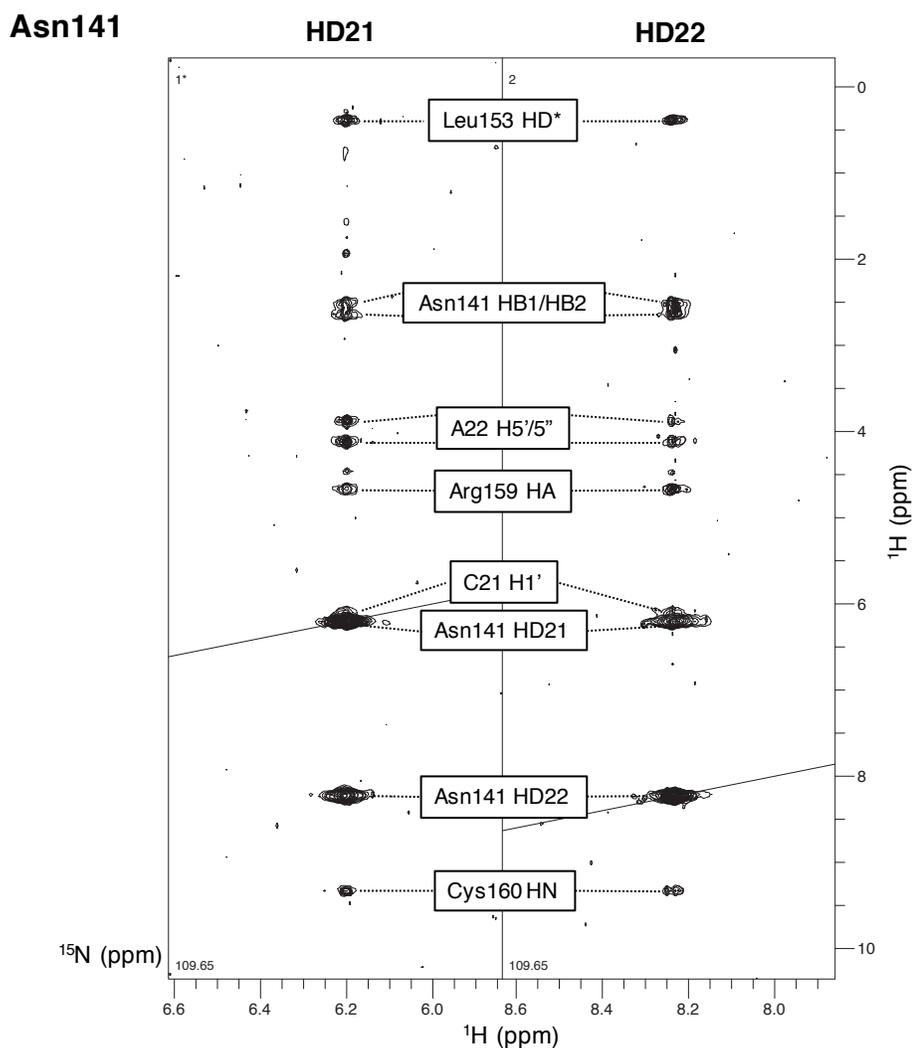


Figure 2.14 NOESY-HSQC strip plots of the side chain amino group of Asn141. The NOESY-HSQC shows NOEs from the amide proton of Asn141 to other protons in this residue and other nearby protons in the HBD and the aptamer protons. However, the experiment does not provide a way to discriminate between intramolecular and intermolecular NOEs.

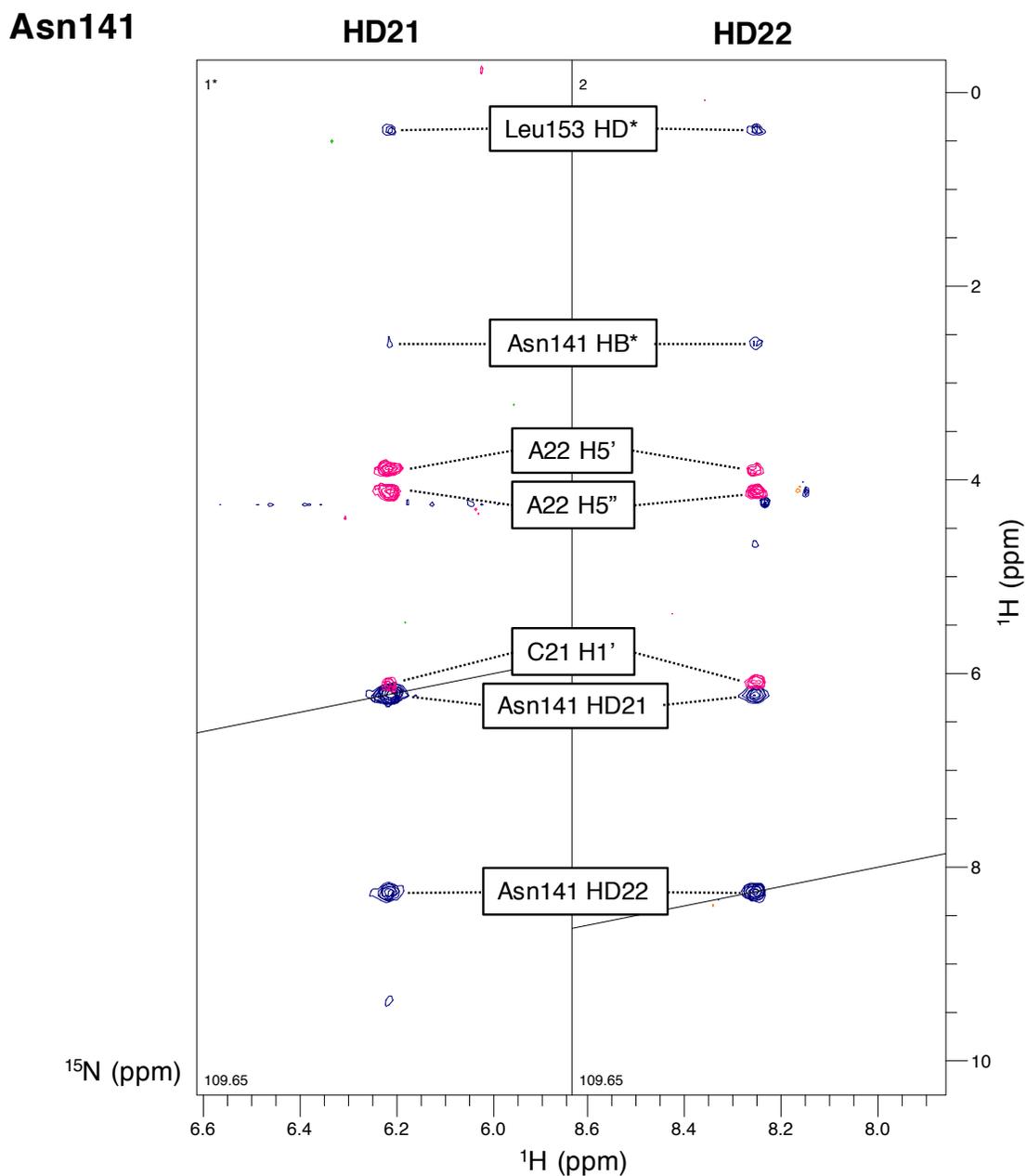


Figure 2.15 Overlay of addition and subtraction spectra of the SIM-NOESY-HSQC for the side chain amino group of Asn141. Intramolecular NOEs (blue) and intermolecular NOEs (pink) appear in separate spectra following subtraction and addition, respectively, of the collected data.

From the ^{15}N NOESY-HSQC and ^{13}C , ^{15}N SIM-NOESY-HSQC spectra, 548 intramolecular NOEs were identified. This included 150 intra-residue, 146 sequential, 113 medium-range (between protons belonging to amino acids that are separated by two to four residues in the sequence), and 139 long-range NOEs (between protons belonging to amino acids that are separated by five or more residues in the sequence). The number of NOEs is shown as a function of residue in Figure 2.16. More NOEs are concentrated in the C-terminal subdomain (59% of the NOEs, amino acids 139-165) than in the N-terminal subdomain (41% NOEs, amino acids 111-138). This is similar to what was seen in the structure determination of the free HBD, in which the first 17 residues of the HBD had limited NOEs (Fairbrother et al. 1998). The authors measured the heteronuclear $\{^1\text{H}\}$ - ^{15}N NOEs for the backbone amides and found that only the first six residues exhibited motion on the picosecond to nanosecond timescale, and they hypothesized that the remaining residues with limited NOEs were experiencing dynamics on other timescales. The NMR structure of the free HBD included 459 total non-redundant NOEs: 163 intra-residue, 108 sequential, 66 medium-range, and 122 long-range NOEs (Fairbrother et al. 1998). The larger number of NOEs identified for the aptamer-bound HBD resulted in a well-defined protein structure (Chapter 4).

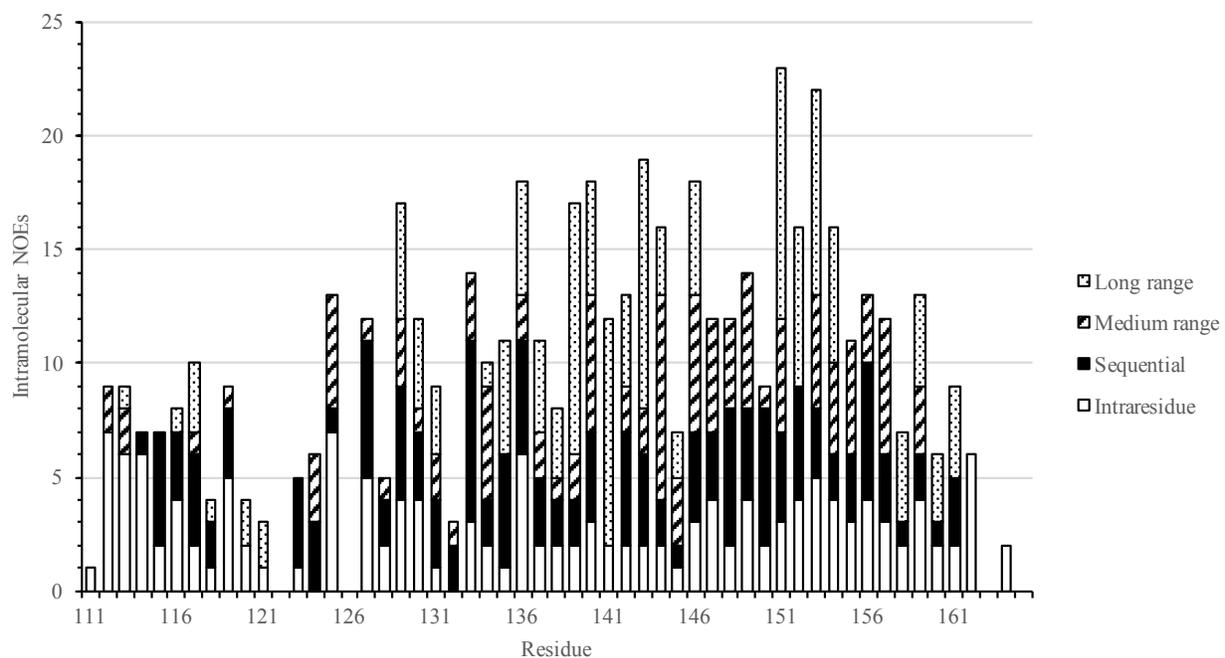


Figure 2.16. Number of HBD intramolecular NOEs as a function of residue. NOEs are divided into four categories: intra-residue (white), sequential (black), medium-range (two to four residues apart; stripes), and long-range (five or more residues apart; dots).

Chapter 3: Chemical shift assignments for the aptamer bound to the HBD

3.1 Introduction

3.1.1 Nucleic acid chemical shift assignments

NMR is a powerful tool for studying nucleic acids and protein-nucleic acid complexes in solution. Similar to the strategy for studying proteins by NMR, ^{13}C and ^{15}N can be incorporated into the nucleic acids to assist in better resolution and dispersion both to make assignments and define structural restraints. RNA and DNA oligonucleotides up to about 100 nucleotides long can both be purchased synthesized with ^{13}C - and ^{15}N -labeled nucleotides. However, *in vitro* transcription using ^{13}C - and ^{15}N -labeled nucleotides harvested from bacteria is a popular and cost-effective method for obtaining milligram quantities of RNA (Batey et al. 1992; Nikonowicz et al. 1992). Similarly, there are strategies for obtaining isotopically labeled DNA that do not rely on the expensive custom orders from a company. Double stranded DNA can be prepared with uniform labeling using a plasmid and growing *E. coli* in isotope-supplemented medium or through PCR (Louis et al. 1998). One method for obtaining labeled single-stranded DNA is described in this chapter.

Labeling nucleic acids with ^{13}C and ^{15}N has not completely resolved the problems of limited chemical shift dispersion in the NMR spectra. To overcome some of these problems, several strategies can be employed depending on the system being investigated. Specialized NMR pulse sequences have been developed that are base-specific (Simorre et al. 1995; Simorre et al. 1996a; Simorre et al. 1996b). The use of non-uniform sampling has made it possible to acquire multi-dimensional NMR spectra in a fraction of the time and with higher resolution than is possible with traditional linear sampling methods (Barna et al. 1987; Hyberts et al. 2014). Large nucleic acids can be broken up into smaller, more manageable domains or minimal

binding sequences, or specifically labeled with only a subset of the nucleotides (Lu et al. 2010; Liu et al. 2015).

Even small nucleic acids, like the 25-nucleotide DNA aptamer investigated here, can benefit from site-specific labeling to reconcile ambiguity. The labeling strategy developed here is based on the site-specific labeling of RNA described by Liu and colleagues (Liu et al. 2015). The protocol involved coupling a biotinylated double-stranded DNA template to streptavidin beads, and adding T7 RNA polymerase with an incomplete mixture of nucleotides. For example, if cytosine 5' triphosphate was not included, the polymerase would be forced to pause at the first site where cytosine needed to be incorporated. Addition of labeled cytosine 5' triphosphate would then resume transcription. Liu and colleagues successfully used this method for both labeling a single site on the RNA and for labeling specific regions such as a terminal loop in a hairpin. As is described later in this chapter, this strategy was modified to label the aptamer at one adenine site with ^{13}C and ^{15}N to complete assignment of the adenine H2 protons.

The strategy for assigning the chemical shifts of the aptamer typically involved first identifying the nucleotides participating in base pairing by following the imino proton NOE pattern from one base pair to the next in a helix. The exchangeable protons—the guanine H1 and thymine H3 imino protons, and the amino groups of adenine, guanine, and cytosine—and the attached nitrogens were assigned using imino- and amino-optimized HSQC experiments as well as various NOESY-type spectra. The nonexchangeable protons in the aromatic and deoxyribose rings were assigned using the aromatic- and deoxyribose-optimized HSQCs in combination with the aromatic- and deoxyribose-optimized NOESY-HSQC. Both the aromatic- and deoxyribose-optimized NOESY-HSQC spectra were used to connect the bases to the intra-residue sugars. The

remaining protons and carbons in the deoxyribose ring were assigned using HCCH-COSY and HCCH-TOCSY experiments.

3.1.2 Chapter overview

This chapter focuses on the chemical shift assignments of the aptamer in the complex with the HBD, and these assignments are shown in Appendix A6. NMR titrations showed that the HBD and the aptamer formed a 1:1 complex that was in the slow exchange regime. A site-specific labeling strategy was designed for the aptamer to confidently assign the H2 and H8 protons of the adenines. Following assignment of the aptamer resonances, intermolecular NOEs were identified to define the interface of the complex.

3.2 Materials and methods

3.2.1 Sample preparation for NMR experiments

Preparation of unlabeled HBD used for NMR samples was described in Section 2.2.1. The preparation of samples containing unlabeled aptamer was described in Section 2.2.2. Aptamer uniformly labeled with ^{13}C and ^{15}N was purchased from Silantes (by the Willson lab, University of Houston). The lyophilized aptamer was resuspended in 10 mM sodium phosphate pH 7.4, 100 mM NaCl, 3 mM KCl, 0.1 mM EDTA, 0.15 mM TSP, and 10% $\text{D}_2\text{O}/90\%$ H_2O . The HBD was also prepared in this buffer and titrated into the aptamer until the aptamer was fully bound. The sample was then buffer exchanged using a 0.5 mL spin column (Thermo Scientific) into succinate NMR buffer (20 mM deuterated succinate pH 5.5, 50 mM NaCl, 0.1 mM EDTA, 0.05% NaN_3 , 0.15 mM TSP, and 10% $\text{D}_2\text{O}/90\%$ H_2O). Aptamer assignments were performed using a sample of 0.6 mM ^{13}C , ^{15}N -labeled aptamer bound to unlabeled HBD in succinate NMR

buffer. To assign the nonexchangeable protons, many of which fall near the chemical shift of water, the sample was buffer exchanged into the succinate NMR buffer containing 99% D₂O using a 0.5 mL spin column.

Site-specific labeling of aptamer

Aptamers labeled at specific positions with ¹³C and ¹⁵N were prepared to help establish or confirm resonance assignments. A single-stranded DNA template was purchased from Integrated DNA Technologies (5'-CCCTGCACTCTTGTCTGGAAGACGGTTTTGATCCACCTGATCG-ACTTAGTCT-3'). The underlined 5' end of the template DNA corresponds to the complementary sequence of the 25-nucleotide aptamer. The primer lengths were different depending on the nucleotide that was going to be labeled (Table 3.1). To anneal the primer to the template DNA, 200 μL of 1 mM primer and 200 μL of 1 mM template DNA in 1X NEB2 (10 mM Tris-HCl pH 7.9, 50 mM NaCl, 10 mM MgCl₂, and 1mM DTT) were combined in a 1:1 ratio and heated for 5 min at 90 °C followed by incubation on ice. The primer-template mixture was mixed with 2 mM ¹³C, ¹⁵N-dATP and 0.8 units of Klenow polymerase (New England Biolabs) per nanomole of template and diluted to a final concentration of 200 μM. The Klenow reaction to add a single ¹³C, ¹⁵N-dATP proceeded at 37 °C for 1 h followed by heat inactivation of the enzyme by incubating the sample at 75 °C for 20 min. The sample was spun down using a microcentrifuge to pellet the denatured polymerase that precipitated during the heat inactivation. Excess ¹³C, ¹⁵N-dATP was removed using a 0.5 mL Millipore Amicon Ultra concentrator (molecular weight cutoff 3,000 Da) and washing the sample twice with 0.5 mL 1X NEB2 containing 1 mM unlabeled dATP. A second Klenow reaction was performed using 1 mM unlabeled dNTPs, 0.4 units of Klenow polymerase per nanomole of template, and a final total

volume of 1 mL. The reaction proceeded at 37 °C for 90 min followed by heat inactivation of the enzyme. For samples that required alkaline cleavage, those that labeled A10 and A12, an equal volume of 1 M NaOH was added, and the samples were incubated at 95 °C for 20 min.

Table 3.1: Sequences for template DNA and primers used for site-specific labeling.

Template strand sequence	
5'-CCCTGCACTCTTGTCTGGAAGACGGTTTTGATCCACCTGATCGACTTAGTCT-3'	
Primer sequences	
A10	5'-TCAGGTGGATCAAACCGTCTTCC-3'
A12	5'-TCAGGTGGATCAAACCGTCTTCCAG-3'
A15	5'-CCGTCTTCCAGACA-3'
A17	5'-CCGTCTTCCAGACAAG-3'
A22	5'-CCGTCTTCCAGACAAGAGTGC-3'

The desired 25-nucleotide DNA was purified using denaturing polyacrylamide gel electrophoresis. The sample was mixed with an equal volume of loading buffer (4 M urea, 50 % formamide, and trace bromophenol blue) and boiled for 5 min at 95 °C. The sample was loaded on a 15 % denaturing polyacrylamide gel (7M urea, 1X TBE, 15 % acrylamide 19:1). The gel was run at 30 W until the bromophenol blue reached the bottom of the gel. The DNA was visualized using UV shadowing and cut out of the gel using a sterile razor blade. The crush and soak method was used to extract the DNA from the gel. The crushed gel pieces were soaked in 10 mM Tris pH 7.5 and 1 mM EDTA rotating overnight at 4 °C. The sample was then centrifuged at 3000 g for 15 min at 4 °C to pellet the gel chunks. The supernatant was removed from the sample, filtered with a 0.2 µm filter, and concentrated to about 250 µL.

The reaction to label the aptamer at A22 used a modified protocol in which the single-stranded template DNA was biotinylated at the 3' end and conjugated to streptavidin agarose beads (Life Technologies). The reaction to produce aptamer labeled at A22 was the first one

performed, and it followed the protocol established by Liu and colleagues, where the template strand was biotinylated and attached to streptavidin beads (Liu et al. 2015). Streptavidin agarose beads were washed using 100 mM sodium phosphate pH 7.2 and 150 mM NaCl. Biotinylated template in the same buffer was added as 500 μ L of 140 μ M template to 400 μ L of streptavidin agarose beads for a final concentration of 78 μ M. The 200 μ L annealing reaction was performed using 25 nmol of biotinylated DNA bound to the streptavidin beads and 25 nmol primer that were heated to 90 °C for 5 min followed by incubation on ice. The template-primer samples were washed with 1X NEB2, then 1.2 mM of each unlabeled dGTP and ^{13}C , ^{15}N -dATP and 0.2 units Klenow polymerase per nanomole of template were added to the reaction. The reaction proceeded at 37 °C for 90 min, at which point the enzyme was heat inactivated by incubation at 75 °C for 1 h. To further promote strand dissociation, the reaction was heated to 90 °C for 25 min. The supernatant was removed and an equal volume of new 1X NEB2 buffer was added to the beads. The samples were heated to 90 °C for an additional 10 min before the supernatant was removed. This protocol was performed twice, and the samples were combined. The aptamer was purified using a C18 column on a 1260 Infinity high-performance liquid chromatography (HPLC) system from Agilent Technologies. The buffers and method used for HPLC purification are shown below. Fractions containing the aptamer were then ethanol precipitated. Each fraction (0.2 mL in a 2 mL tube) was combined with 200 μ L of 3 M sodium acetate, 1.6 mL 100% ethanol, and 1 μ L glycogen. The samples were stored at -20 °C overnight and spun down at 16,000 g for 10 min to pellet the precipitated DNA. The pellet was briefly washed with ethanol and dried before it was resuspended in water.

Table 3.2: Buffers and elution method used for HPLC purification

Buffer A	0.1 M triethylammonium acetate in H ₂ O	
Buffer B	100% acetonitrile	
Method		
Min	% buffer B	Flow rate
0-20	0 % B	2 mL/min
20-40	0-15 % B	2 mL/min
40-150	15% B	0.2 mL/min

3.2.2 NMR experiments

All NMR experiments with the aptamer were performed on a Varian Inova 600 MHz, VNMRS 800 MHz, or VNMRS 900 MHz spectrometers equipped with a triple resonance, z-axis gradient cold probes. Experiments were performed at 25 °C using the Biopack pulse sequences unless otherwise stated. The NMR spectra were processed using NMRPipe and analyzed using CCPNMR Analysis (Delaglio et al. 1995; Vranken et al. 2005).

The imino protons were assigned using a 2D ¹H - ¹H NOESY and 2D ¹H - ¹⁵N imino-optimized HSQC experiments (Kay et al. 1992). The amino protons were assigned using the 2D ¹H - ¹⁵N amino-optimized HSQC and 2D ¹H - ¹H amino-optimized NOESY-HSQC experiments (Kay et al. 1992; Zhang et al. 1994). The nonexchangeable protons were assigned using 2D ¹H - ¹³C aromatic-optimized and deoxyribose-optimized HSQC experiments and 3D aromatic-optimized and deoxyribose-optimized NOESY-¹³C-HSQC experiments (Vuister & Bax 1992). The two NOESY-HSQCs allowed for intra-residue assignment from the deoxyribose ring to its base and from that base to the deoxyribose of the previous (i - 1) residue. The deoxyribose protons and attached carbons were assigned using the 3D HCCH-COSY (Pardi & Nikonowicz 1992). The Biopack pulse sequences used for the NOESY-¹³C-HSQC and HCCH-COSY

experiments were all for ribose chemical shifts in RNA. Since the deoxyribose H2'/H2'' and C2' have different chemical shifts ranges than the H2' and C2' of ribose, the carbon offset frequency and sweep widths were adjusted. The carbon offset frequency was set to 65 ppm and the sweep width was set to 40 ppm (45 to 85 ppm in the carbon dimension) to allow for adequate spectral resolution by forcing the C2' shifts, normally around 35 ppm to fold in the spectrum.

Intramolecular NOEs were identified using the 2D ^1H - ^1H NOESY, 3D deoxyribose-optimized NOESY-HSQC, and 3D aromatic-optimized NOESY-HSQC experiments. Intermolecular NOEs were identified using the 3D deoxyribose-optimized NOESY-HSQC, 3D aromatic-optimized NOESY-HSQC, and the 3D ^{13}C , ^{15}N SIM-NOESY-HSQC experiments (Eichmüller et al. 2001). A table with select experimental parameters is shown in Appendix A2.

3.3 Results and discussion

3.3.1 Titration of the HBD into the aptamer

A titration was set up by collecting NMR spectra of the aptamer with increasing amounts of ^{15}N -labeled HBD to determine the chemical exchange regime and stoichiometry the complex. The imino proton peak at 12.6 ppm from the free aptamer was monitored through the titration until it was obscured by an overlapping peak from the bound aptamer (Figure 3.1). Titration past the 1:1 ratio did not result in further changes in the spectrum. Therefore, it was determined that the complex was forming a 1:1 complex, which has been further supported using isothermal titration calorimetry (Appendix). The titration also showed that the complex was in the slow exchange regime, which occurs when the rate of exchange is less than the difference between the resonance frequencies of a nucleus in the free and bound states (Lian et al. 1994). Instead of the proton peaks moving as the population weighted average from the chemical shift of the free

proton to the chemical shift of the bound proton, the peaks from the free aptamer decreased in intensity as the peaks from the bound aptamer increased in intensity through the titration (Cavanagh et al. 2007).

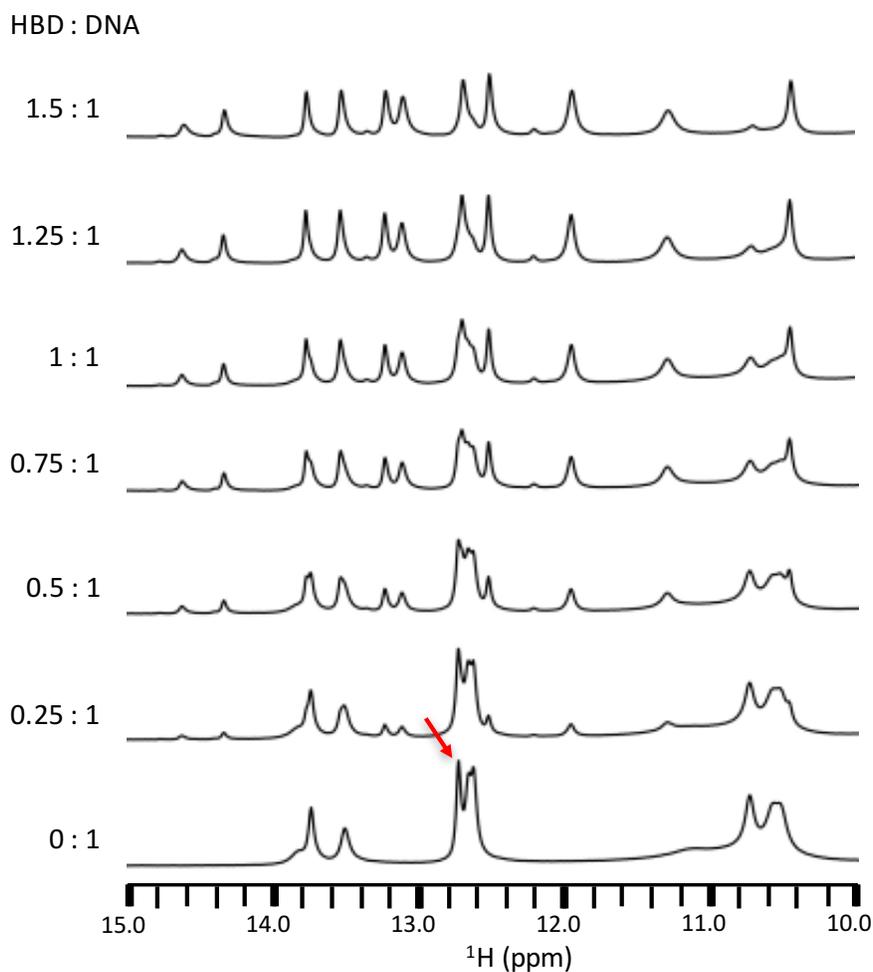


Figure 3.1 Imino proton spectra of unlabeled aptamer during the titration with the HBD. The red arrow specifies the peak from the free aptamer that was monitored through the titration. It is clearly seen in the free aptamer spectrum and is obscured in the later titration points due to overlap with peaks from the bound aptamer. Spectra were collected at 10 °C.

Once the complex was formed, the sample was buffer exchanged into succinate buffer (20 mM deuterated succinate pH 5.5, 50 mM NaCl, 0.1 mM EDTA, 0.05% NaN₃, 0.15 mM TSP, and 10% D₂O). This buffer matches the one used in the published NMR structure of free HBD (Fairbrother et al. 1998). The sample was not originally prepared in this buffer due to concern about the appearance of the imino proton spectrum of the aptamer at pH 5.5. Compared to the aptamer at pH 7.4, the aptamer at pH 5.5 contains several peaks between 10 and 11.5 ppm, where imino protons from nucleotides involved in noncanonical base pairs are seen (Figure 3.2) (Fürtig et al. 2003). Additionally, there were two peaks with chemical shifts around 15.5 ppm, which are often indicative of protonated cytosines (Leitner et al. 1998). We were concerned that the aptamer was taking on a different and incorrect fold that would not bind the HBD. Therefore, the complex was formed at pH 7.4 and buffer exchanged into pH 5.5.

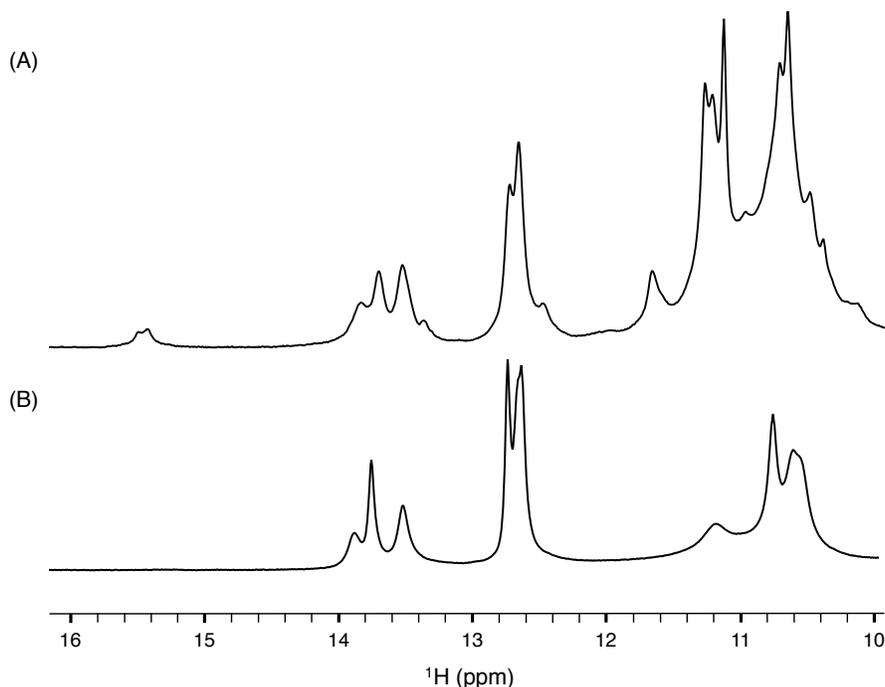


Figure 3.2 Imino proton spectra of free aptamer at pH 5.5 and pH 7.4. (A) Spectrum of the free aptamer collected in succinate buffer pH 5.5. (B) Spectrum of the free aptamer collected in sodium phosphate buffer pH 7.4. Both spectra were collected at 5°C.

The imino proton spectra of the bound aptamer in the two different buffers are nearly identical in the region where Watson-Crick base paired imino protons appear (12 to 14.5 ppm), but there are some differences in those protons with chemical shifts between 10 and 12 ppm (Figure 3.3). Chemical exchange between the imino proton and the solvent decreases with lower pH, so the succinate buffer may be allowing us to see resonances from protons that are exchanging too rapidly at pH 7.4 to be seen (Matthew & Richards 1983). There is, however, a noticeable difference in the peak at 12 ppm that is sharp at pH 7.4 and completely broadened out at pH 5.5.

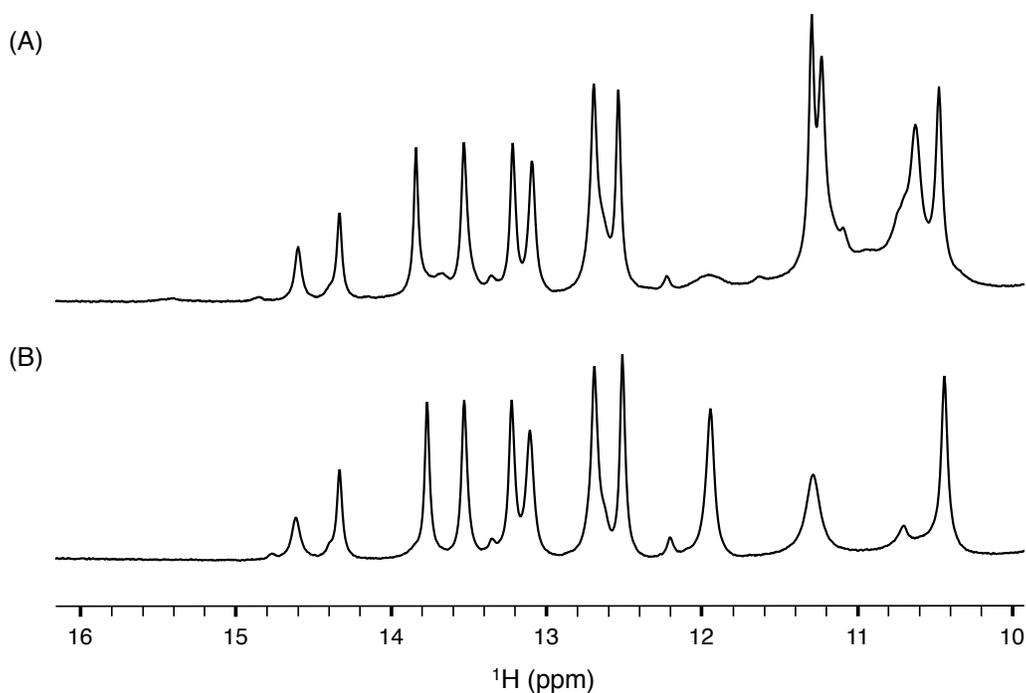


Figure 3.3 Imino proton spectra of aptamer bound to the HBD at pH 5.5 and pH 7.4. (A) Spectrum of the bound aptamer collected in succinate buffer, pH 5.5. (B) Spectrum of the bound aptamer collected in sodium phosphate buffer, pH 7.4. Both spectra were collected at 10°C.

3.3.2 Assignment of the exchangeable protons

Assignment of the imino protons

The starting place for the resonance assignments of the aptamer is the helical walk using the base paired imino protons. Since the distance between adjacent base pairs is 3.4 Å in a B-form DNA helix, which is within the optimal range for NOE transfer, the NOEs can be used to “walk” up a helix. The chemical shift of the imino proton is indicative of its identity: a peak with a proton chemical shift between 12 and 13.5 ppm typically belongs to a guanine in a Watson-Crick base pair; a chemical shift between 13.5 and 15 ppm typically belongs to a thymine in a Watson-Crick base pair; peaks with chemical shifts between 10 and 12 ppm are usually imino protons involved in noncanonical base pairs (e.g. GT base pairs) or not base paired but protected from exchange with the solvent (Fürtig et al. 2003).

A sample of unlabeled aptamer bound to ^{15}N -HBD was used to collect a 2D NOESY spectrum at 5 °C in H_2O to obtain the NOEs between inter-residue imino protons. Instead of a definitive, canonical helical walk, the NOESY spectrum suggested that the aptamer may be forming an interesting tertiary structure. The nucleic acid folding program mFold predicted a hairpin: either a four base pair stem with a three-nucleotide loop or a three base pair stem with a five-nucleotide loop (Figures 3.4A and B) (Zuker 2003). The biochemical data published by Potty and colleagues supported the four base pair stem capped by a triloop, but it also predicted two additional base pairs connecting the 5' and 3' ends (Figure 3.4C and Section 1.2.5) (Potty et al. 2009). These six predicted base pairs were not enough to reconcile the data from the NOESY. Ten distinct imino protons are seen along the diagonal of the spectrum, nine of which fall into the region of the spectrum where imino protons involved in Watson-Crick base pairs appear (12-15 ppm) (Figure 3.5).

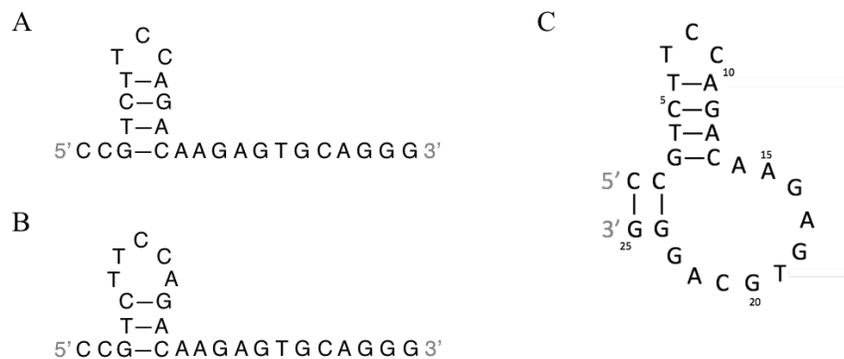


Figure 3.4 Predicted secondary structures of the aptamer. Two mFold predicted secondary structures with (A) three- and (B) five-base loops. (C) The secondary structure predicted by Potty et al. based on mutagenesis.

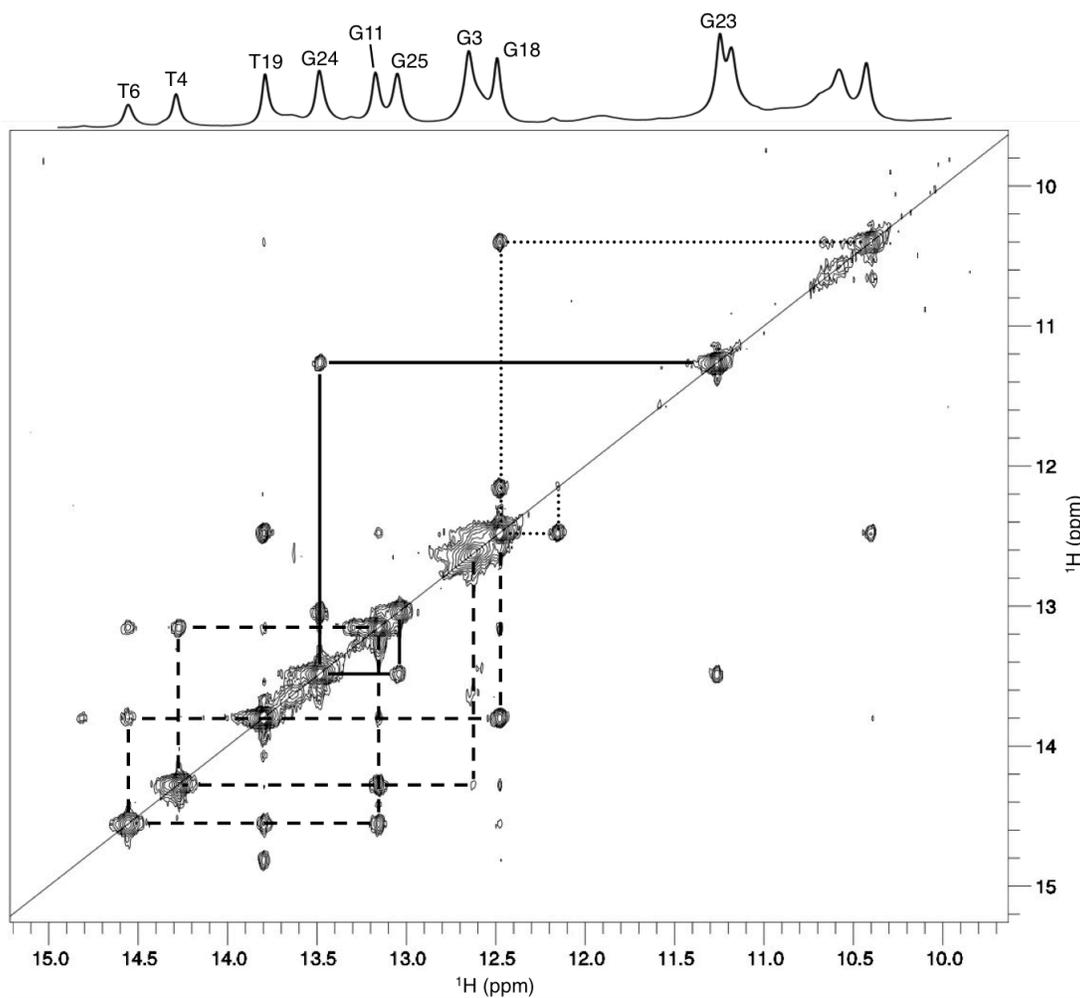


Figure 3.5 Imino proton region of the 2D ^1H - ^1H NOESY spectrum. The three helical walks suggested by the data were: GGG (solid line), GGG (dotted line), and GTTGTG (dashed lines).

The three helical walks suggested by the data were: GGG, GGG, and GTTGTG. One GGG segment (Figure 3.5, solid line) contains a guanine with an imino proton chemical shift outside of the typical range for Watson-Crick base pairs (11.14 ppm). This guanine could either be in a non-canonical base pair or not base paired but protect from rapid exchange with the solvent. These three guanine nucleotides were assigned as G23-G24-G25. G24 (13.44 ppm) and G25 (13.01 ppm) are in Watson-Crick base pairs with C2 and C1 respectively, and G23 does not form a base pair. Further support for the assignment of the G25 imino proton is the observation that the peak for the imino proton in the imino HSQC is has very weak intensity at 25 °C but has strong intensity at 10 °C, which is common for the terminal base pair and suggests that it is “breathing” more at the higher temperature and allowing for a higher exchange rate with the solvent.

Based on the sequence, the identity of the TGTG imino proton NOE pattern was easily identified as the four base pair hairpin with G3 H1 at 12.68 ppm, T4 at 14.30 ppm, G11 at 13.18 ppm, and T6 at 14.47 ppm. Additional experiments showed that the other imino protons in this walk (Figure 3.5, dashed lines) also included G18 (12.49 pm) and T19 (13.73 ppm).

The remaining imino pattern of the second GGG sequence (Figure 3.5, dotted line) also NOEs to T19 H3, and so they were preliminarily assigned as G16-G18-G20. G18 H1 at 12.49 ppm was not confidently assigned until preliminary structure calculations suggested a solution to the unassigned base pair. The peaks at 10.46 and 12.23 ppm are hypothesized to belong to G16 and G20, but the lack of NOEs to or from these protons leaves them unassigned. Overall, 75% of the guanine imino protons and 75% of the thymine imino protons have been assigned.

Assignment of the imino nitrogens

2D ^1H - ^{15}N imino HSQC spectra were collected at 25 °C as well as 10 °C to slow down the imino proton exchange with the solvent, and these spectra were used to assign the guanine N1 and thymine N3. The nitrogen chemical shift ranges for guanine and thymine are distinct enough to be able to confirm the preliminary assignment of the imino protons. The typical chemical shift range for guanine N1 is 140 to 150 ppm, and 157 to 165 ppm for thymine N3. With the exception of G18 N1, which has a chemical shift of 151 ppm, the imino nitrogens of the aptamer fall within these standard ranges. The HSQC spectrum is particularly useful for identifying the peaks with proton chemical shifts around 13.5 ppm, which could belong to either guanine or thymine.

The guanine N1 and thymine N3 chemical shifts were assigned as shown in Figure 3.3. The imino proton of G25 has weak intensity at 25 °C but has intensity comparable to the other guanine imino protons at 10 °C, which supported the assignment. The imino proton of T7 (14.73 ppm) and the imino proton at 12.23 ppm (belonging to either G16 or G20) did not appear in the HSQC in succinate buffer at 10 °C, but they did appear as low intensity peaks in the imino HSQC prior to buffer exchange when the sample was in sodium phosphate buffer. Perhaps the loss in sample during buffer exchange caused the peaks to dip below the detection limit or the exchange properties of the protons changed moving from sodium phosphate buffer at pH 7.4 to the succinate buffer at pH 5.5. Overall, 75% of the guanine imino nitrogens and 75% of the thymine imino nitrogens have been assigned.

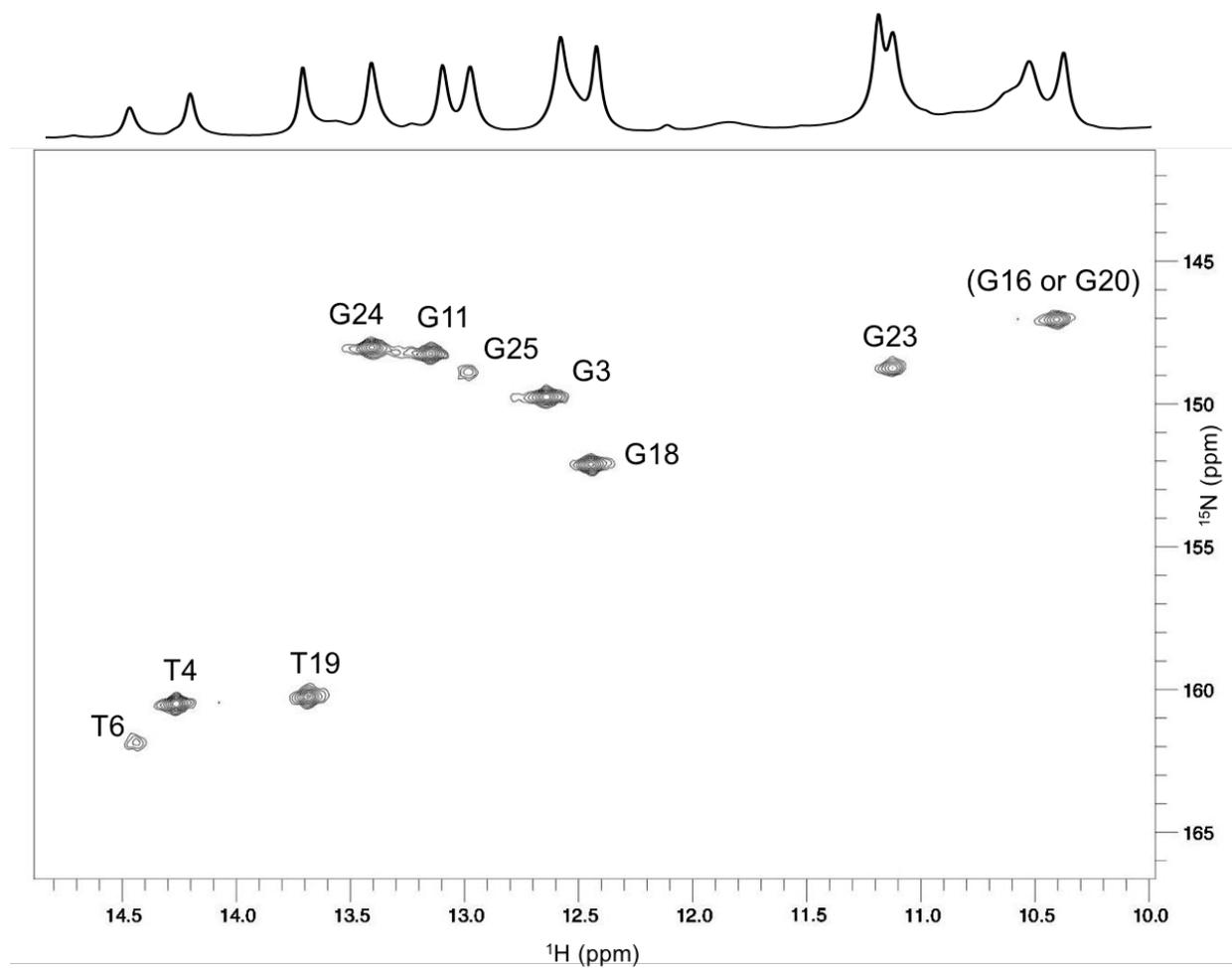


Figure 3.6 2D ^1H - ^{15}N imino-optimized HSQC spectrum at 25 °C. Assignments for the imino NH peaks are shown in the spectrum. The peak at 10.46 ppm belongs to either G16 or G20.

Assignment of the amino protons and nitrogens

The 2D amino HSQC was used to correlate the amino protons to the amino nitrogen on adenine, cytidine, and guanine. Nitrogen chemical shifts are different between the bases, which allowed us to preliminarily assign the base type. The typical chemical shift ranges for the amino nitrogens are 74 to 76 ppm for guanine, 77 to 82 ppm for adenine, and 95 to 100 ppm for cytosine. With the exception of G23, the amino nitrogens of the aptamer fall within these standard ranges. Various NOE patterns are used to assign the amino protons for six of the 21 bases with amino groups. NOEs between the amino H and imino H correlate the amino group to the base pair partner. NOEs between the amino protons and the H5 can be used to assign cytosine amino protons.

The amino-optimized HSQC and 2D NOESY were used to assign the amino groups. The amino protons of C1, C2, and C5 were assigned based on NOEs from the amino protons to the H5. C1 amino protons at 7.00 and 8.37 ppm show NOEs to the C1 H5 at 6.04 ppm. The C2 amino protons are at 7.33 and 8.95 ppm, and they show NOEs to the H5 at 5.68 ppm. The C5 amino protons at 7.67 and 9.16 ppm show NOEs to the H5 at 6.08 ppm. NOEs between the amino protons and imino protons involved in base pairs were also used to make assignments. The C13 amino protons (6.73 and 7.41 ppm) show NOEs to the imino proton of the guanine (G3) to which it is base paired. The A12 amino protons (7.86 and 8.23 ppm) shows an NOE to T4 H3. Both the T6 and T19 imino protons show NOEs to the A10 amino protons at 7.03 and 7.34 ppm. The other nine peaks in the amino-optimized HSQC, belonging to one cytosine, two or three adenines, and two or three guanines, remain unassigned (Figure 3.7).

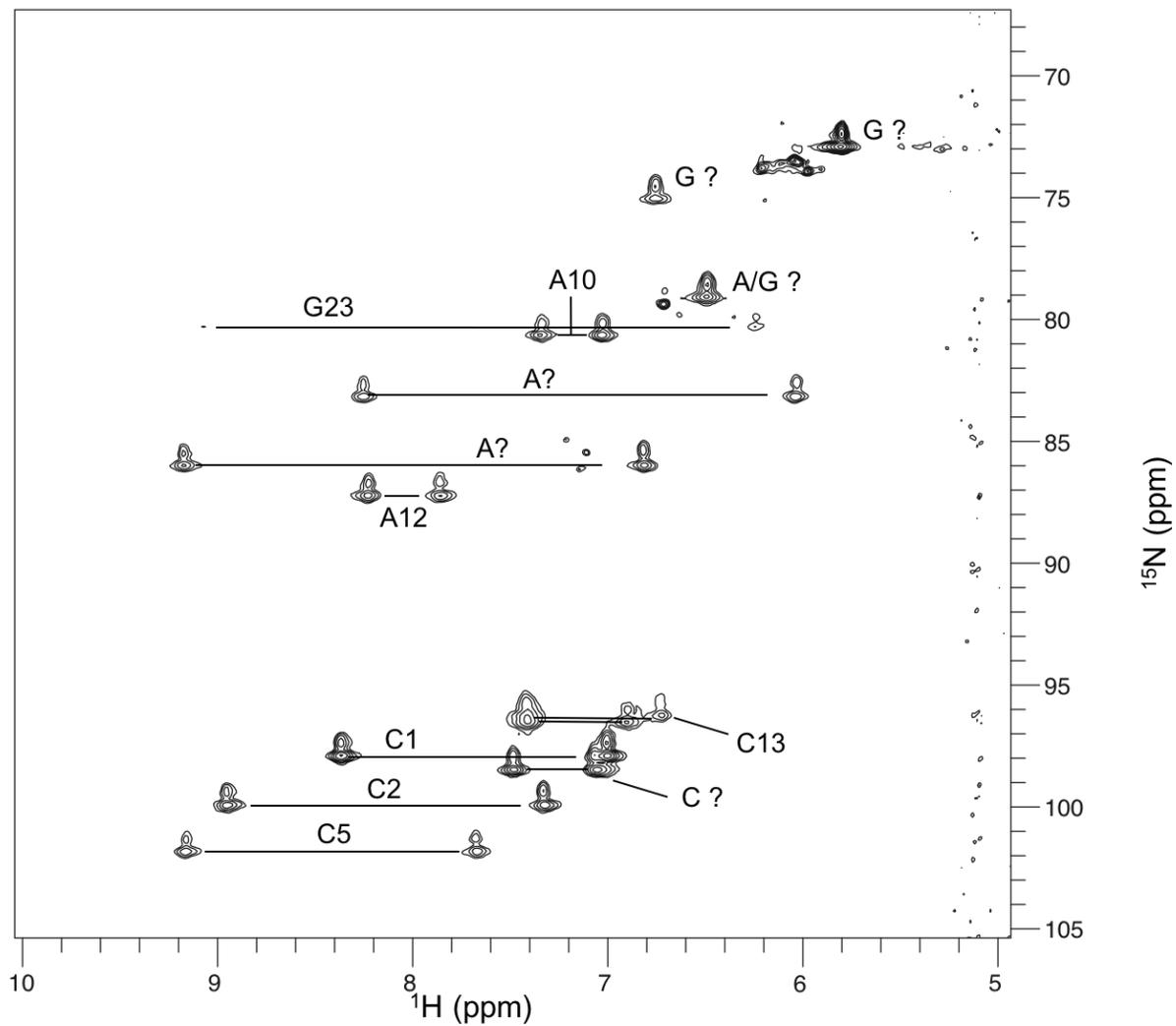


Figure 3.7 2D ^1H - ^{15}N amino-optimized HSQC at 10 °C. Assignments for the amino NH_2 groups are shown in the spectrum. Those that remain unassigned are labeled with the base type (determined by the nitrogen chemical shift) and a question mark.

3.3.3 Assignment of the nonexchangeable protons

Assignment of the aromatic protons and carbons

The 2D ^1H - ^{13}C aromatic HSQC, 3D aromatic-optimized NOESY-HSQC, and 3D deoxyribose-optimized NOESY-HSQC were used to assign the aromatic CH groups: H8-C8 (purines), H2-C2 (adenine), and H6-C6 (pyrimidines). The C2 chemical shift range is well separated from C6 and C8: 150 to 160 ppm for C2 versus 133 to 143 ppm for C6 and C8. The C6 can be distinguished from C8 in the aromatic HSQC with these experimental parameters by the appearance of a doublet in the carbon dimension caused by 68 Hz C5-C6 coupling (Ippel et al. 1996). H5-C5 peaks were assigned using the 2D ^1H - ^{13}C HSQC and 3D NOESY-HSQC.

The 3D NOESY-HSQC were used to sequentially assign the nucleotides. The NOE pattern allows us to walk from the purine H8 or pyrimidine H6 to the sequential nucleotide H1'/H2'/H2'' and from there to the intra-residue H6/H8 (Figure 3.8 and Appendix A5). This sequential walk allowed for assignment of the H6 and H8 protons and C6 and C8 carbons. G25 did not appear in the NOESY-HSQC spectra.

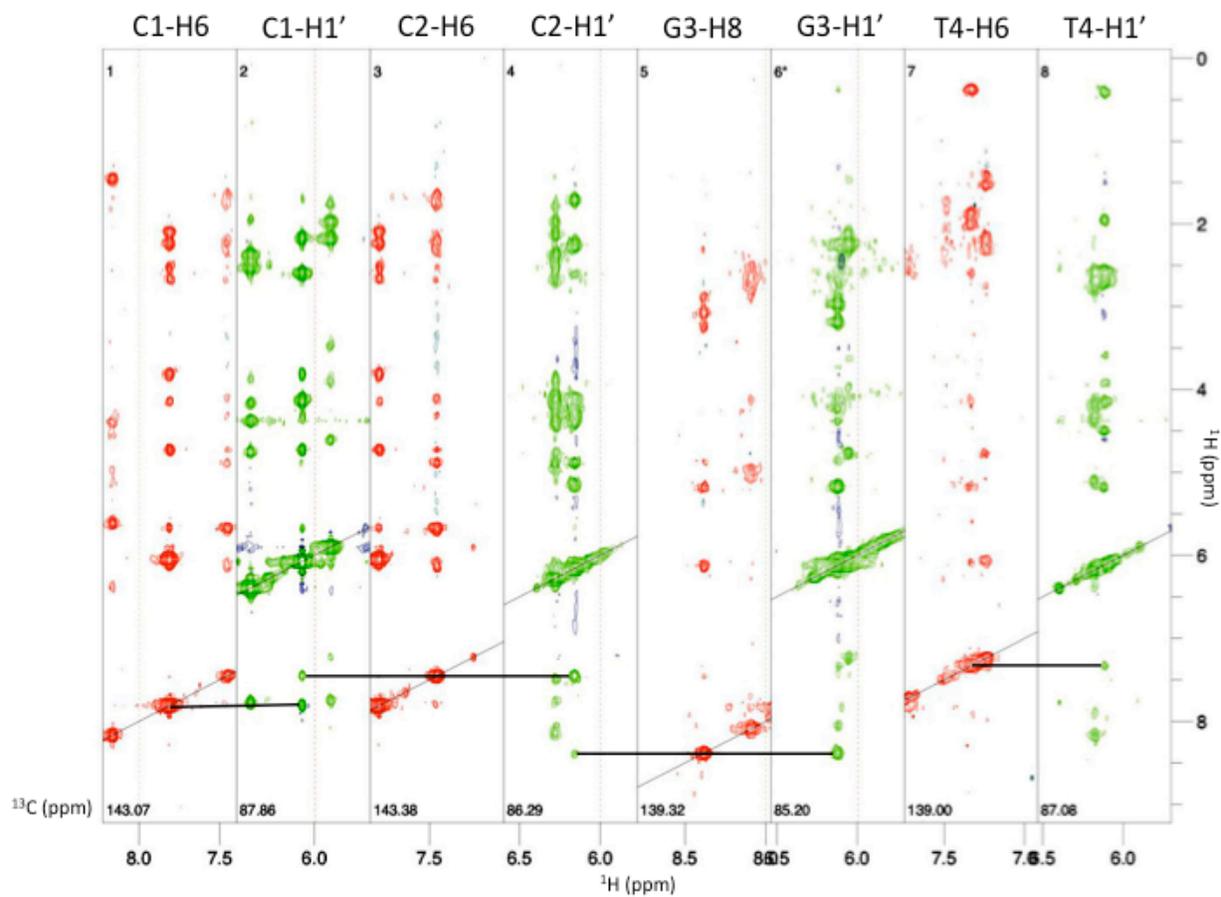
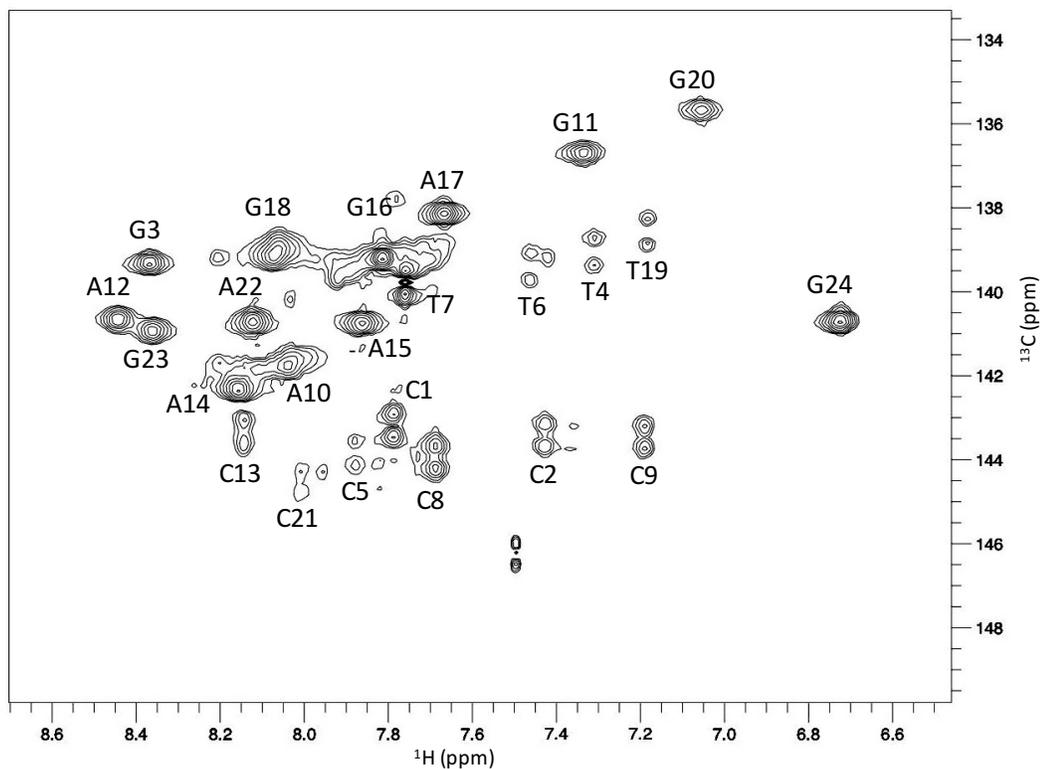


Figure 3.8 Strip plots of the aromatic-optimized and deoxyribose-optimized NOESY-HSQC spectra for nucleotides C1-T4. The aromatic-optimized (red) and deoxyribose-optimized (green) spectra show the sequential connection of nucleotides C1, C2, G3, and T4.

A



B

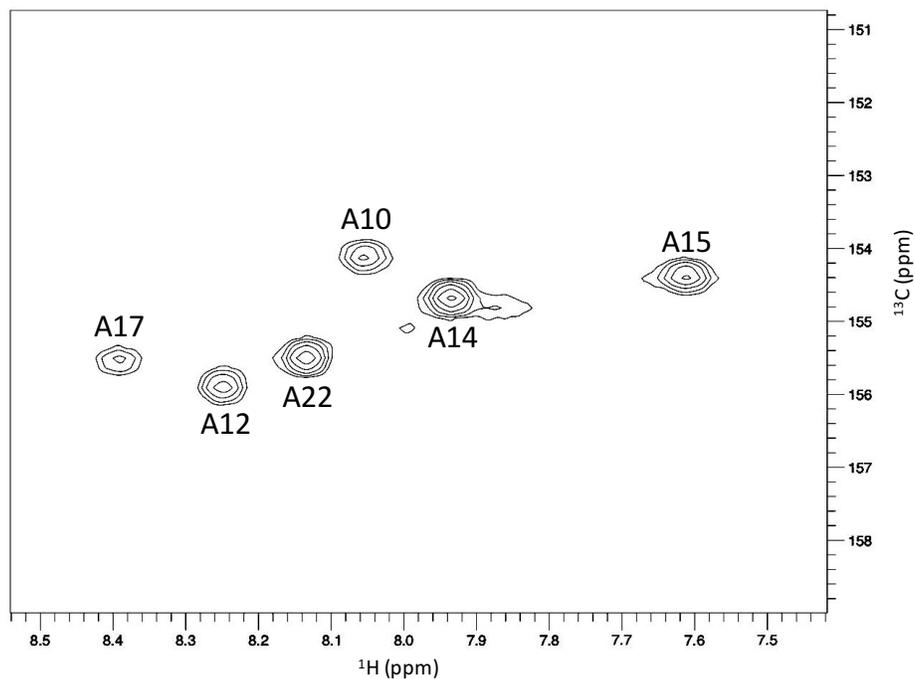


Figure 3.9 2D ^1H - ^{13}C aromatic-optimized HSQC spectrum. The spectrum shows the assignments for the resonances in the (A) H6-C6 and H8-C8 region and (B) H2-C2 region. The unlabeled peak in the H6-C6 region may be an artifact or from a minor population of aptamer that is unbound or in an alternate conformation.

2D HCCH-COSY and 2D ribose (H5-C5) and aromatic (H6-C6) ^1H - ^{13}C HSQCs were used to correlate the H5 of a cytidine base to the H6. The H5/C5 groups of cytidines show up in the deoxyribose-optimized 2D HSQC and 3D NOESY-HSQC. All seven H5-H6 pairs are well-defined in the 2D (^1H - ^1H) HCCH-COSY, which provided correlation between the H5 and H6. The peaks were dispersed enough to only need the H6 chemical shift to assign the H5, but NOEs from the 3D aromatic- and deoxyribose-optimized NOESY-HSQCs were used to confirm the assignments. The assignments of the H5-C5 groups of the seven cytidines are shown in Figure 3.10. Overall, 100% of the H6-C6 groups and 93% of the H8-C8 groups were assigned, and the chemical shifts are shown in Figure 3.9 and are listed in Appendix A6.

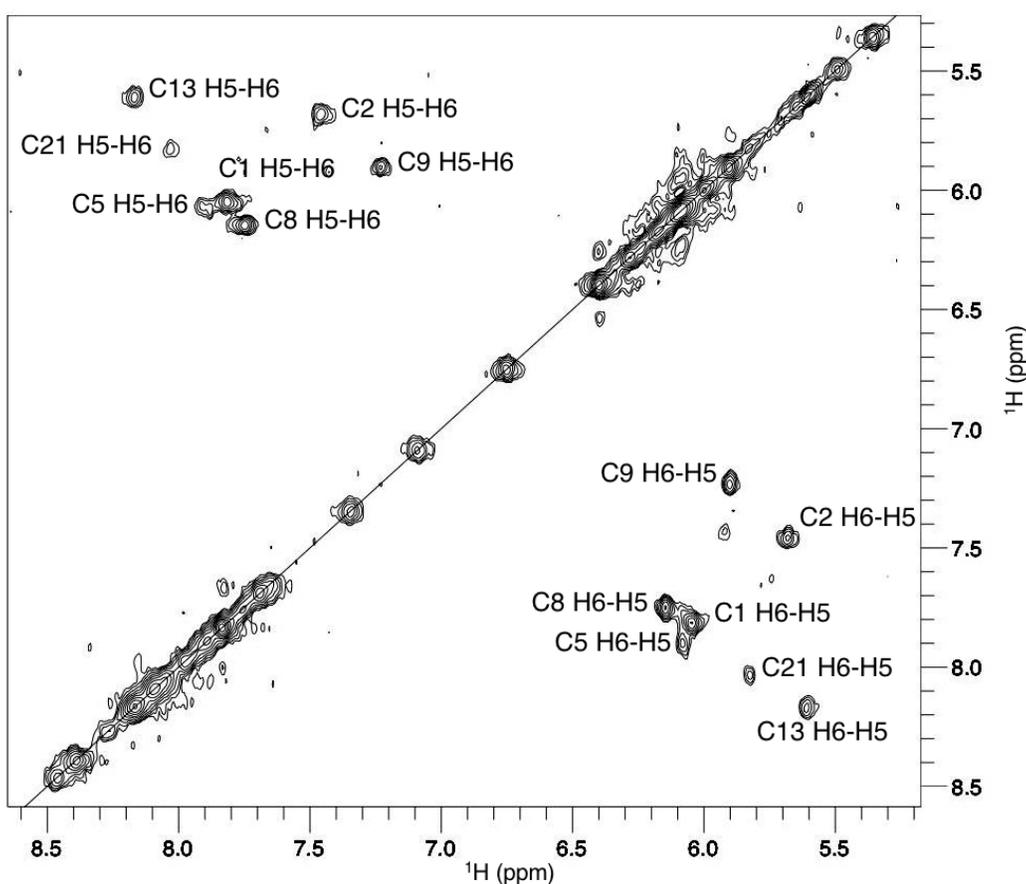


Figure 3.10 2D ^1H - ^1H HCCH-COSY spectrum. The spectrum shows the assignments for the H5 and H6 protons of the cytosines.

Assignment of adenine H2 and C2 resonances

Aptamers with ^{13}C and ^{15}N incorporated a specific nucleotide were used to assign or confirm assignments of the adenine H2 protons. There were very few NOEs to or from the adenine H2 protons in the 3D NOESY- ^{13}C -HSQC spectra, which made it difficult to correlate the H2's to the H8's that were assigned using the sequential walk. Since the H2 is part of the Watson-Crick face of adenine, it was necessary to confidently assign the H2 in order to identify the nucleotides involved in base pairs. Since the aptamer bound to the HBD did not show any crosspeaks in the 2D (^1H - ^1H) HCCH-TOCSY, which can be used to correlate the H2 and H8, we decided to use site-specific labeling to assign the adenine resonances (Marino et al. 1994). A method developed recently by Liu and colleagues was modified to incorporate ^{13}C , ^{15}N -dATP into the aptamer at specific positions and confidently assign the adenine H2 and H8 (Liu et al. 2015).

Aptamers were made that included a ^{13}C , ^{15}N -labeled A at one specific site on the aptamer (e.g. A22). Labeling position A22 was the most straightforward because the remainder of the sequence was 5'-GGG-3' so the only steps required to label A22 were annealing of the primer to the template DNA and adding ^{13}C , ^{15}N -dATP, unlabeled dGTP, and Klenow polymerase. Since the entire sequence 3' of the primer was being added at once, the quantities of dNTPs and enzyme used in the reaction were different than what was required for the other labeled aptamers. Labeling position A10 or A12 required additional steps because the primers to add the labeled dATP to these positions had too low of a melting temperature (29 °C and 38.8 °C respectively). To increase the melting temperature of the primer, additional nucleotides were added to the 5' end of the aptamer sequence that are complementary to the template DNA. Since the presence of additional nucleotides could alter the binding behavior and chemical shifts of the

aptamer in the NMR sample, we chose to cleave off the additional 5' nucleotides not part of the original aptamer sequence. This was accomplished using primers that included a single ribonucleotide at the position immediately 5' to the start of the aptamer sequence. By incubating this sample at high temperature and high pH, the ribonucleotide cleaved the strand into two fragments of DNA: the 15-nucleotide primer and the 25-nucleotide aptamer. The different length of the two fragments, as well as the 52-nucleotide DNA template, made it easy to isolate the aptamer with purification using denaturing polyacrylamide gel electrophoresis.

Site-specifically labeled oligonucleotides were used to form a complex with unlabeled HBD, and a 2D aromatic-optimized HSQC was collected to assign the H8-C8 and H2-C2 for that specifically labeled adenine (Figure 3.7). We also extended this method to make two aptamers that were labeled at one adenine and one thymine (T7A10 and A12T19) to confirm imino proton assignments. A summary of the site-specifically labeled samples is shown in Table 3.3.

Table 3.3 Summary of samples of site-specifically labeled aptamers

Site(s) labeled	Concentration of NMR sample	Resonance assigned or assignment confirmed	Method used to produce sample
T7/A10	40 μ M	A10 H2: Confirmed A10 H8: Confirmed T7 H3: No signal T7 H6: Confirmed	Template attached to beads; HPLC purification
A12/T19	60 μ M	A12 H2: Confirmed A12 H8: Confirmed T19 H3: Assigned T19 H6: Confirmed	Template attached to beads; HPLC purification
A15	300 μ M	H2: Confirmed H8: Assigned	Template free in solution; PAGE purification
A17	100 μ M	H2: Assigned H8: Confirmed	Template free in solution; PAGE purification
A22	120 μ M	H2: Assigned H8: Confirmed	Template free in solution; PAGE purification

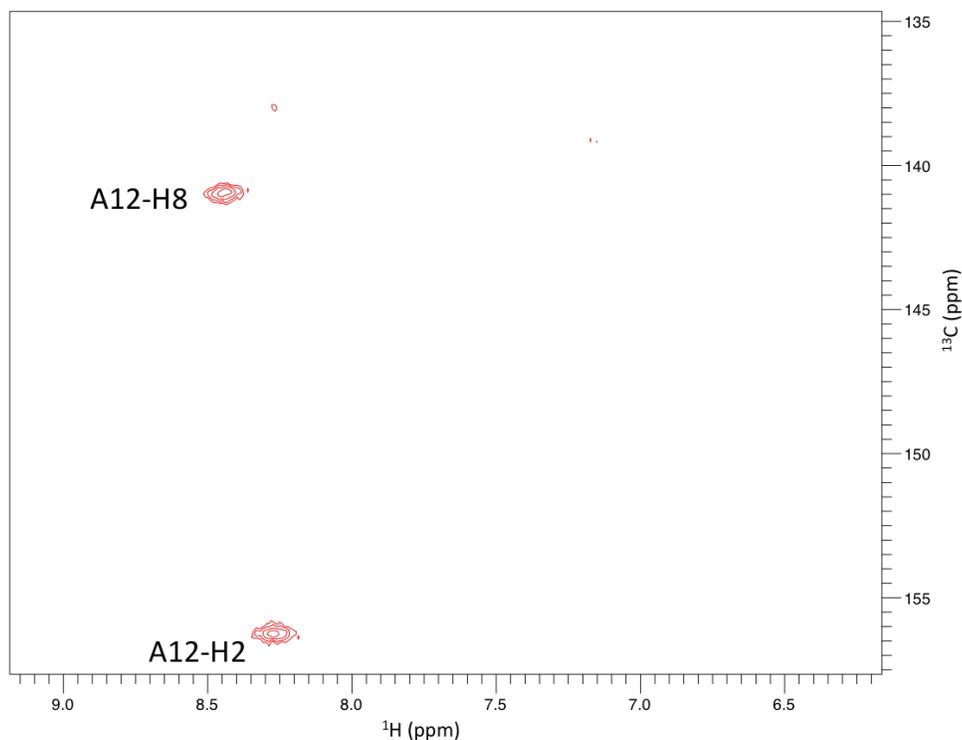


Figure 3.11 2D ^1H - ^{13}C aromatic-optimized HSQC spectrum of the aptamer bound to the HBD and specifically labeled with ^{13}C and ^{15}N at position A12.

Alternate version of the protocol

The protocol to site-specifically label the aptamer was originally developed using a biotinylated DNA template attached to streptavidin beads, which is similar to the method described by Liu et al (2015). This proved to be an inefficient setup for our system for two reasons. First, using the beads is more appropriate for obtaining smaller amounts of product, and the NMR sample requires milligram quantities of the aptamer. Second, it was extremely difficult to dissociate the aptamer from the DNA template. Various strategies to dissociate the strands were tried, but instead of efficient duplex denaturation, there was minimal denaturation and considerable streptavidin dissociation from the agarose beads that rendered the beads useless.

The aptamers that included labeling at A10 and A12 were originally separated from the primer by nickase enzyme treatment. The enzyme Nt. AlwI recognized the sequence GGATC and cleaved four bases beyond the recognition sequence (Figure 3.12), which was engineered into the primer and template designs. The enzyme was not very efficient for our system, and a large quantity of expensive enzyme was needed for moderate cleavage of the primer from the aptamer.



Figure 3.12 Recognition sequence of Nt. AlwI. The nickase enzyme cuts the top strand four bases 3' of the recognition sequence (denoted by the black triangle).

Assignment of deoxyribose protons and attached carbons

The deoxyribose protons and carbon resonances were also assigned. The 2D deoxyribose-optimized HSQC spectrum was collected with the center of the carbon frequency set to 65 ppm and spectral width in the indirect dimension set to 70 ppm to cover the entire range of deoxyribose carbon chemical shifts. The range is larger due to the chemical shifts of the C2', which are around 35 ppm and well separated from the remaining sugar carbons that fall in the range of 55 to 90 ppm. As seen in Figure 3.13, the chemical shifts of most of the deoxyribose CH/CH₂ groups are poorly resolved due to severe overlap. The chemical shifts of the deoxyribose protons and carbons were assigned using the deoxyribose-optimized NOESY-HSQC and 3D HCCH-COSY. Of the 25 nucleotides in the aptamer, 84% of the H1' and C1', 80% of the H2'* (the H2' and H2'' do not have stereospecific assignments), 84% of the C2', 64% of the H3' and C3', 76% of the H4' and C4', 36% of the H5'*, and 50% of the C5' have

been assigned. Assignments for the deoxyribose protons and carbons can be found in Appendix A6.

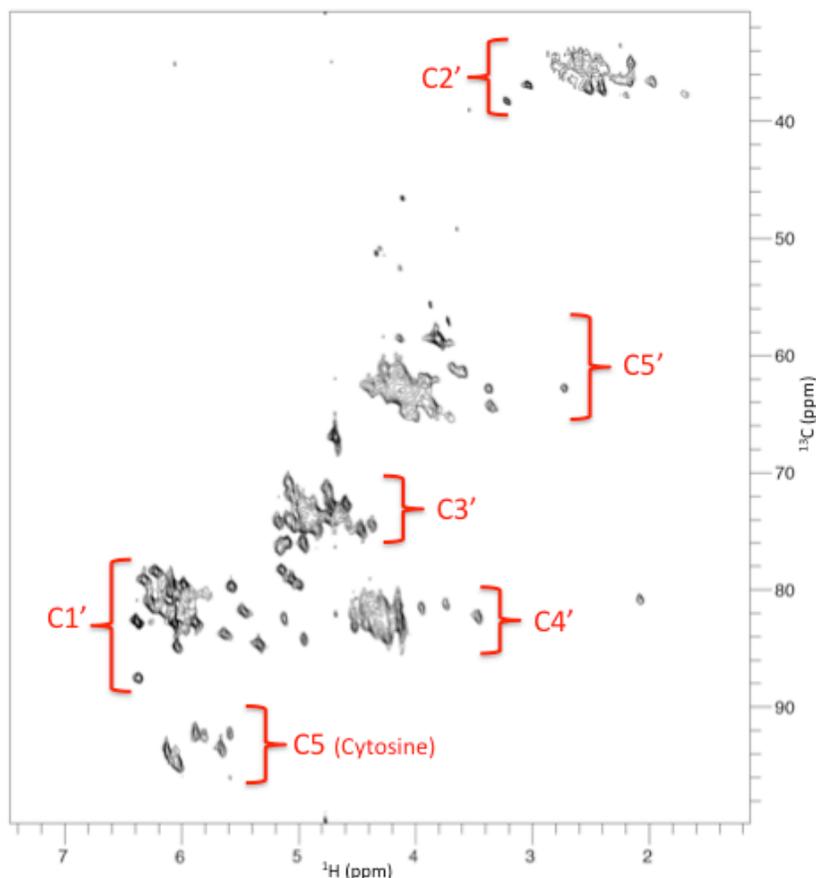


Figure 3.13 2D ¹H-¹³C deoxyribose-optimized HSQC spectrum. The protons and carbons within the deoxyribose ring experience severe overlap. The H5-C5 groups of cytosine also appear in the deoxyribose HSQC.

3.3.4 Base pair identification

Base pairs were identified using both through-bond and through-space correlations. The 2D ¹H-¹⁵N HNN-COSY spectrum was used to identify hydrogen bond acceptor nitrogens in base pairs. The HNN-COSY experiment transfers magnetization from the imino proton to its bound nitrogen, and then it transfers to the acceptor nitrogen through J_{NN} coupling (Dingley & Grzesiek

1998). The spectrum shows two peaks in the nitrogen dimension for each imino proton resonance (Figure 3.14). The upfield nitrogen resonances (around 155 ppm) correspond to the N3 of thymine and the N1 of guanine. The downfield nitrogen resonances (around 210 ppm) have the opposite sign and correspond to the hydrogen bond acceptor: N3 of cytidine or N1 of adenine in Watson-Crick base pairs. The chemical shift of the acceptor nitrogen is indicative of the base involved in the base pair. The N3 of cytosine is typically between 182 and 205 ppm. The N1 of adenine typically has a chemical shift between 215 and 230 ppm. The N7 of adenine and guanine, which would be involved in Hoogsteen base pairs, overlap with the N1 of adenine, and have chemical shifts ranging from 220 to 240 ppm.

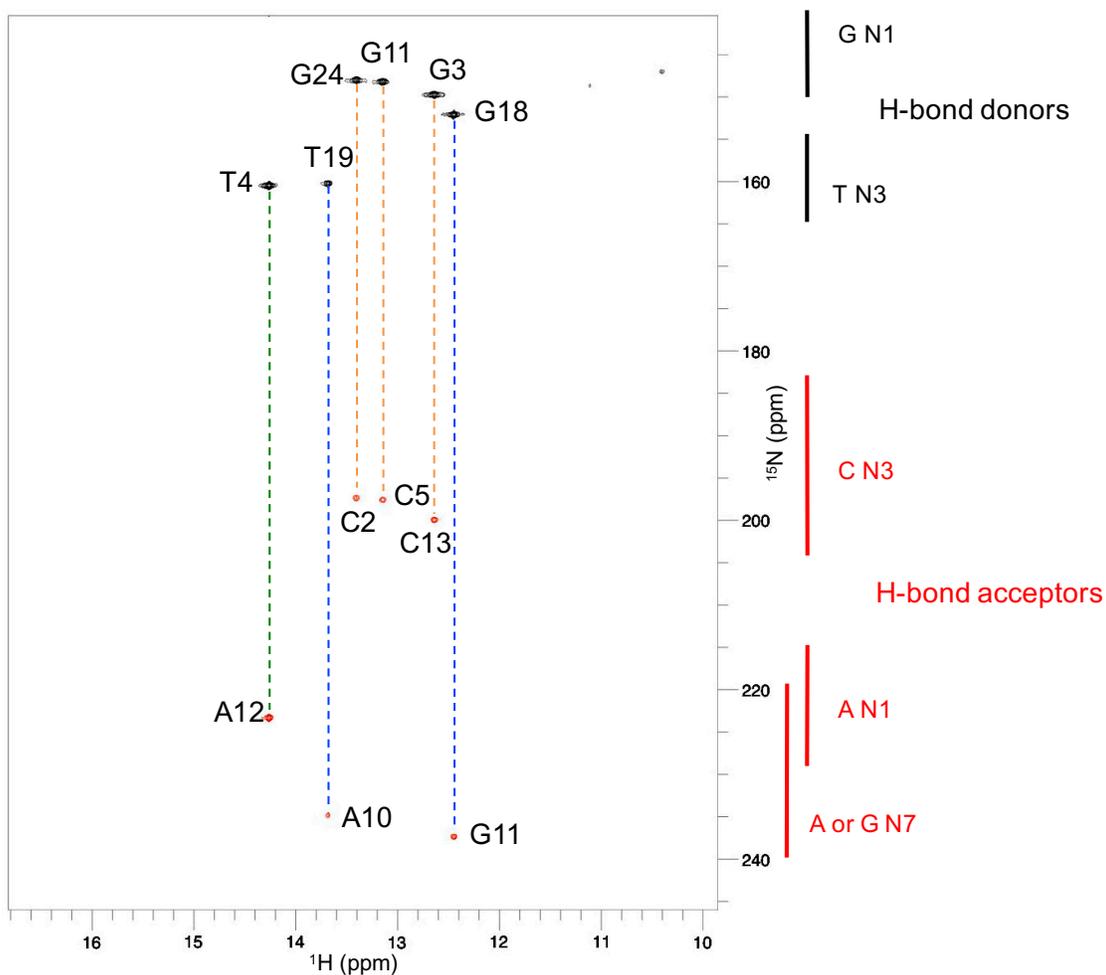


Figure 3.14 2D ^1H - ^{15}N HNN-COSY correlating the hydrogen bond donors and acceptors of nucleotides in base pairs. GC Watson-Crick base pairs are designated by an orange dotted line, AT Watson-Crick base pairs by a green dotted line, and Hoogsteen AT and GG base pairs by a blue dotted line.

The imino-imino NOEs showed the helical walk of G-T-G-T from G3, T4, G11, and T6 that were base paired to C13, A12, C5, and A10 respectively. These preliminary assignments were supported by the 2D NOESY and 2D ^1H - ^1H amino-optimized NOESY-HSQC that connected imino protons to amino groups involved in the base pair. The HNN-COSY also supports Watson Crick base pairs for G3, T4, and G11 (Figure 3.14). The signal of the T6 imino proton is too low to appear in the spectrum.

The imino-imino NOESY spectrum also showed a helical walk of G-G-G, which was assigned as G23-G24-G25. Secondary structure predictions proposed that the 3' terminal guanines base pair with the 5' terminal cytosines, which led to the preliminary assignment of the C1-G25 base pair. The HNN-COSY supported a Watson-Crick base pair for G24 and C2, but G25 did not appear in the spectrum. The C2-G24 base pair was confirmed by NOEs between the G24 H1 and the C2 amino protons, and the C1-G25 base pair was supported by NOEs between G25 H1 and C1 amino protons.

Unexpectedly, G18 and T19 were also involved in base pairs. The HNN-COSY showed that they were both involved in Hoogsteen base pairs with adenine or guanine because of the downfield chemical shifts of the hydrogen-bond acceptors within the N7 chemical shift range. T19 H3 formed a hydrogen bond with a nitrogen with a chemical shift of 233.9 ppm, and G18 H1 formed a hydrogen bond with a nitrogen at 236.5 ppm. The preliminary Hoogsteen base pair assignments were later determined to be base triples between T19·A10-T6 and G18·G11-C5. The assignment of the base triples is supported by NOEs between T19 H3 and A10 H8 and between G18 H1 and the A10 H2'/H2'' sugar protons as well as the C5 amino protons (Figure 3.15).

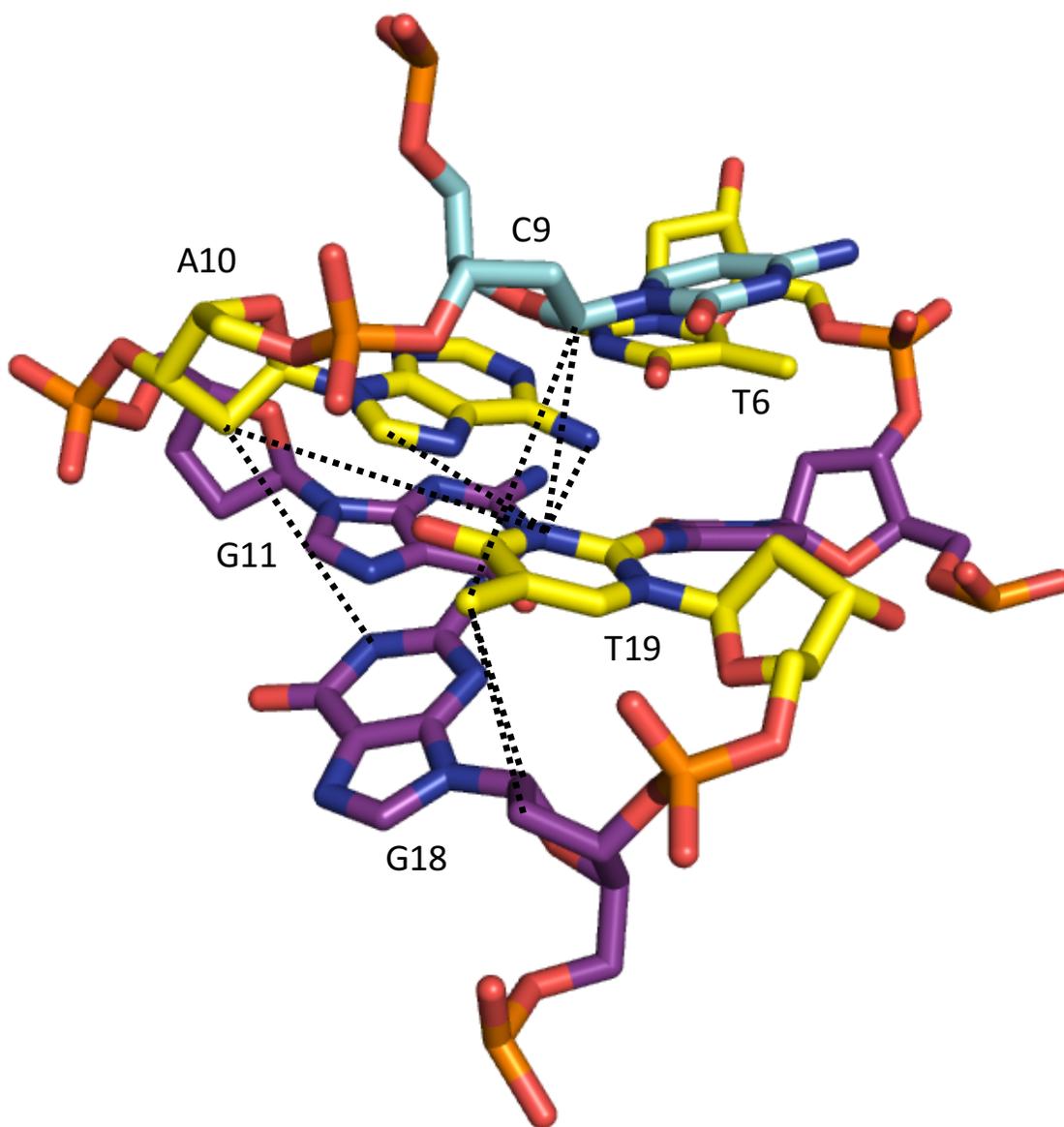


Figure 3.15 NOEs observed that support the formation of the base triples. This representative structure shows the tandem base triples formed by G18·G11·C5 (purple) and T19·A10·T6 (yellow). C9 (cyan) is part of a triloop connecting T6 and A10, but the remainder of the triloop is not shown for clarity. NOEs that were observed and support the base triples are shown as dotted lines between the heavy atoms.

3.3.5 Intramolecular and intermolecular NOEs

Intramolecular NOEs were identified for both assignments and structural restraints using the aromatic-optimized and deoxyribose-optimized NOESY-HSQC experiments with the understanding that resonances from the HBD would also appear in the spectrum. NOEs that could belong to either the HBD or the aptamer were treated as ambiguous until preliminary structure calculations could determine the correct NOE. A total of 306 intramolecular NOEs were identified for the aptamer, and the number of NOEs is shown as a function of residue number in Figure 3.16. The total number of intramolecular NOEs included 177 intra-residue, 92 sequential, 13 medium-range (between protons belonging to nucleotides that are separated by two to four residues in the sequence), and 24 long-range (between nucleotides that are separated by five or more residues in the sequence).

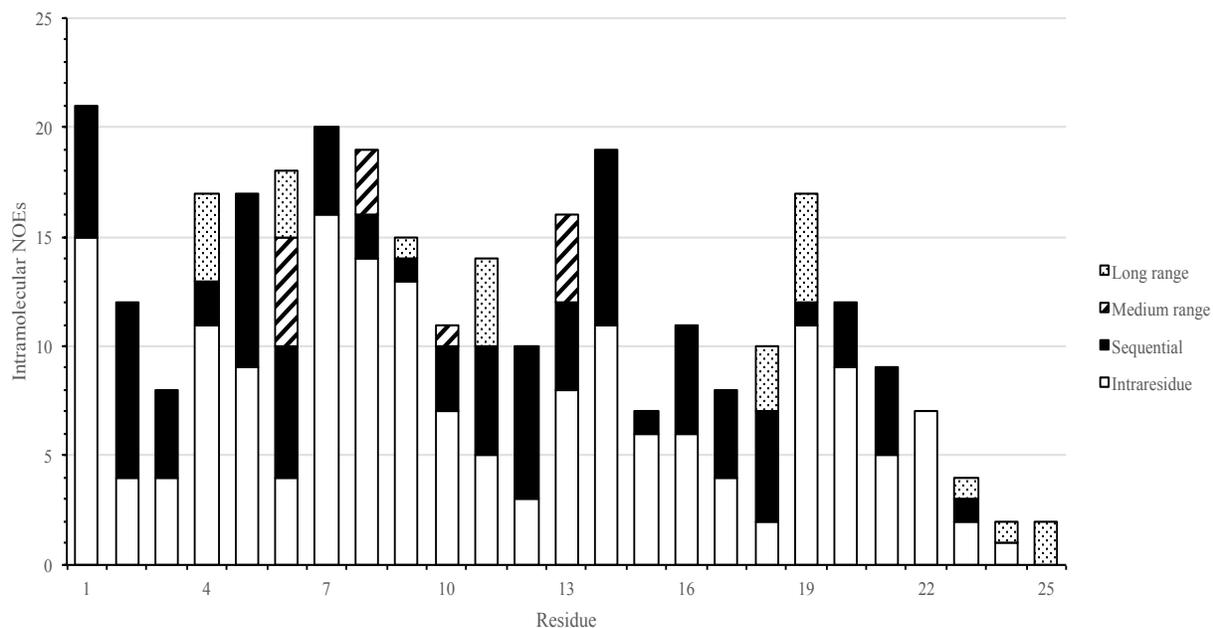


Figure 3.16. Number of aptamer intramolecular NOEs as a function of residue. NOEs are divided into four categories: intra-residue (white), sequential (black), medium-range (two to four residues apart; stripes), and long-range (five or more residues apart; dots).

The SIM-NOESY-¹³C, ¹⁵N-HSQC was invaluable for identifying intermolecular NOEs due to its ability to distinguish intermolecular NOEs from intramolecular NOEs by the sign of the peaks, but several experiments were used in the process. The 2D NOESY was used to identify imino to protein NOEs. The NOESY-¹⁵N-HSQC provided NOEs from the HBD amide proton, but it was carefully analyzed with the knowledge that intramolecular and intermolecular NOEs were indistinguishable in the spectrum. The labeled aptamer bound to unlabeled HBD was also used to assign intermolecular NOEs. The aromatic and deoxyribose NOESY-¹³C-HSQC spectra were both used to identify the intermolecular NOEs.

Fifteen amino acids from HBD were found to interact with eight nucleotides of the aptamer (Table 3.4 and Figure 3.17). Leu127 shows NOEs to the deoxyribose of T4 as well as the Watson-Crick and non-Watson Crick faces of the base. Additional NOEs are seen from Leu127 to the imino proton of G3 and the H8 and H4' of A17. Val129 also shows NOEs to the imino protons of G3 and T4, but it also shows NOEs to the sugar protons of A14 and A15, as well as the H8 of A15 and the amino group of C13.

Lys125 NOEs to the H2'* group of G3 and the H8 of A17. Lys136 also shows NOEs to A15-H1' and the H2 of both A14 and A15. Asp131 and Ser138 amide protons NOEs to A15-H2 as well. Lys140, Asn141, and Thr142 all NOE to the H2 of A12. Asn141 also NOEs to T4-H1' along with Leu153, and Thr142 also NOEs to A12-H1'.

Asn154, Thr157, Cys158, and Arg159 all share NOEs to the H5 of C21. Of the 51 NOEs listed in Table 3.4, 18 are between an amino acid and deoxyribose sugar and 33 are between an amino acid and a base.

Table 3.4 Intermolecular NOEs.

Amino acid	Atom	NOE to aptamer
Arg124	HE	C1 H5, C2-(H6, H2'*)
Lys125	HN	G3-H2'*, A17-H8
Leu127	HN	A17-H8, A17-H4'
	HA	G3-H1, T4-(H3, H5#)
	HG	T4-(H6, H1', H2'*, H3', H4', H5'*)
	HD*	G3-H1, T4-(H3, H6, H5#, H5'*)
Val129	HB	G3-H1
	HG*	G3-H1, T4-H3, A14-H1', C13-H41/H42, A15-(H8, H1', H2'*, H4')
Asp131	HN	A15-H2
Lys136	HB*	A14-H2, A15-H2
	HE*	A14-H2, A15-(H2, H1')
Ser138	HN	A15-H2
Cys139	HN	T4-H1'
Lys140	HE*	A12-H2
Asn141	HA	A12-H2
	HB*	T4-H1'
Thr142	HN	A12-H2
	HG*	A12-H1'
Leu153	HD*	T4-H1'
Asn154	HB*	C21-H5
Thr157	HN	C21-H5
	HB	C21-H5
	HG*	C21-H5
Cys158	HN	T4-H1', C21-H5
Arg159	HN	C21-H5
	HB*	C21-H5
	HD*	C21-H5

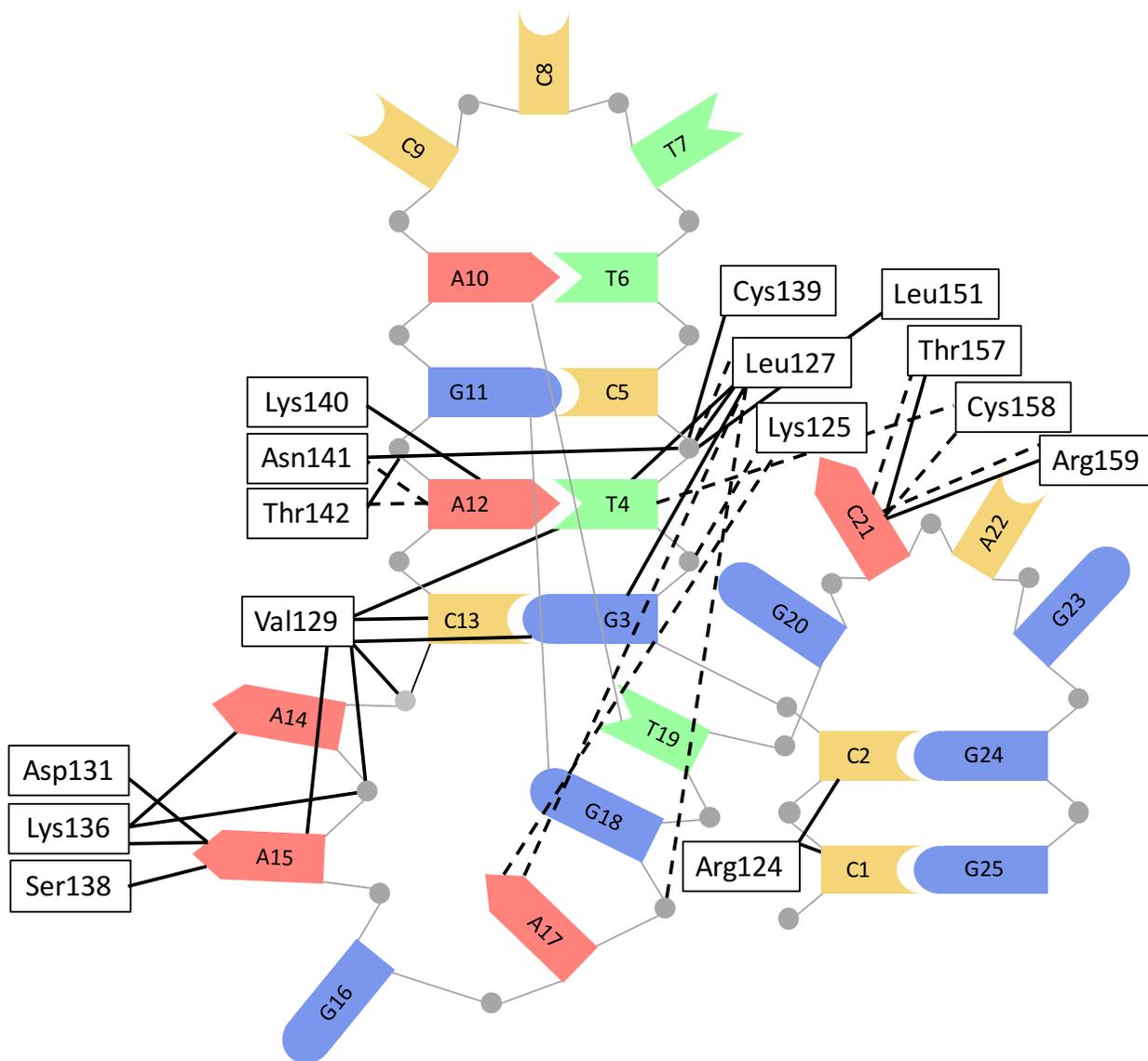


Figure 3.17 Intermolecular NOEs between HBD and the aptamer. The gray circles connected by lines between the bases represent the deoxyribose. The intermolecular NOEs between amino acid side chains and the aptamer are displayed as solid black lines. The intermolecular NOEs between the amino acid backbone and the aptamer are shown as black dashed lines.

Chapter 4 Solution structure of the HBD-aptamer complex

4.1 Introduction

4.1.1 Solution structure determination

NMR provides a method for determining the three-dimensional structures of macromolecules. Unlike the widely used technique of X-ray crystallography, which serves as a snapshot of a biomolecule trapped in a static crystallizable conformation, NMR structures capture the structure of the molecule in solution. NMR structures are commonly calculated by incorporating experimental data into a simulated annealing protocol using restrained molecular dynamics (Cavanagh et al. 2007). The structure calculation typically begins with the protein or oligonucleotide as an extended chain at high temperature. During the process of simulated annealing, the temperature decreases while the weights of the experimental restraints increase. The resulting structures should satisfy the experimental data included in the calculations.

When solving an NMR structure, the goal is to sample a large conformational space and calculate a structure that fits the experimental restraints (Cavanagh et al. 2007). The calculations are complicated by the fact that the experimental restraints are not included as one specific value but as a range of possible acceptable values. The NOEs only provide qualitative distance information because NOEs are affected by local motion, and post-acquisition processing of the data can affect the peak intensities (Roberts 1993; Evans 1995). Structure calculations are repeated many times to find an unbiased structure that fits the experimental restraints. It is therefore more accurate to display an NMR structure as an ensemble of 10-20 structures. This ensemble of structures should have no violations, which means it satisfies the experimental restraints, and it should be converged meaning it has a low root-mean-square deviation (RMSD) of atomic positions when the structures are superpositioned. These regions of low RMSDs are

defined with high confidence, whereas regions with higher RMSDs that are not converged either display regions of bona fide flexibility in the molecule or a lack of definition in that area of the structure. The quality of an NMR ensemble is typically judged on how well the structures agree with the experimental data included in the calculations and how well the geometry conforms to expected values. The geometric quality can be assessed using programs such as MolProbity that analyzes multiple parameters, such as dihedral angles, bond lengths, and the number of steric clashes (Davis et al. 2004; Davis et al. 2007; Chen et al. 2010).

Various experimental NMR data can be incorporated into the structure calculations such as dihedral angles, residual dipolar couplings, hydrogen bonds, and, most importantly, NOE distance restraints. Dihedral angles, such as those predicted with the program TALOS using backbone chemical shifts, are often included because the backbone chemical shifts of a protein are indicative of secondary structure elements (Cornilescu et al. 1999; Shen et al. 2009). Nucleic acid dihedral angles can also be included, but no program currently exists to predict dihedral angles from chemical shifts. Instead, they are either included as a range of values expected for A- or B-form helices, if the nucleotides are expected to form such a conformation, or directly measured with through-bond experiments (Wuttke et al. 1997; Fürtig et al. 2003). Additionally, residual dipolar couplings are often included because they provide long-range information that complements the short-range NOE data (Tjandra et al. 1997). Hydrogen bond restraints are also included for both nucleic acids, in the form of base pair restraints, and proteins, as backbone hydrogen bonds identified empirically using hydrogen-deuterium exchange or as a knowledge-based restraint that compares the calculated structure to the PDB to identify sites of likely hydrogen bonds within secondary structure elements (Wagner 1983; Grishaev & Bax 2004; Schwieters et al. 2006).

4.1.2 Using Xplor-NIH to determine the structures of macromolecules

We are using Xplor-NIH to calculate the structure of the HBD-aptamer complex (Schwieters et al. 2003; Schwieters et al. 2006). Xplor-NIH employs restrained molecular dynamics in torsion angle space to minimize a target function that includes potentials for experimental data as well as covalent geometry and non-bonded contacts (Schwieters et al. 2003; Schwieters et al. 2006). The geometry sampled by the structure can be confined to a set of accepted values obtained from high resolution crystal structures, such as bond lengths, bond angles, and planarity of aromatic groups. By restricting the size to these so-called internal coordinates, the conformational space that is searched during the calculations is much smaller but still sufficiently large for determining the structures of biomolecules (Schwieters & Clore 2001).

Xplor-NIH is one program that is used for determining the solution structures of protein-nucleic acid complexes from NMR data. However, it has recently been shown that the many NMR structures of RNA in the PDB are of lower quality than structures determined by X-ray crystallography based on analysis of steric clashes and geometric outliers (Bermejo et al. 2016). Bermejo and colleagues analyzed X-ray and NMR structures deposited into the PDB between 2010 and 2014 using the MolProbity validation software (Davis et al. 2004; Davis et al. 2007; Chen et al. 2010). For example, structures solved by X-ray crystallography at a resolution of at least 2.0 Å, the average “clashscore”, a term used by MolProbity to describe the number of steric clashes per 1,000 atoms, was 3.8, whereas for NMR structures this number was 39.0 (Bermejo et al. 2016). Recent improvements to the Xplor-NIH nucleic acid force field—the fundamental chemical information, such as bond lengths, bond angles, chirality, planarity, and interatomic distances—seek to improve the quality of nucleic acid NMR structures (Bermejo et

al. 2016). The parameters of the covalent information, such as the atomic radii, were adjusted to match those used by the MolProbity validation software (Bermejo et al. 2016). Bermejo and colleagues also included two additional knowledge-based, or statistical, potentials to the Xplor-NIH software. The first is a base-base positioning potential that uses high resolution crystal structures from the Nucleic Acid Database to provide a statistical description of the position of the bases relative to others in the structure (Kuszewski et al. 2001). This decreases the conformational space to be sampled during the calculations. The second knowledge-based potential uses the Nucleic Acid Database and Protein Data Bank to find physically reasonable torsion angles to further limit the conformational space sampled during the calculations (Bermejo et al. 2012; Bermejo et al. 2016). While these potentials bias the conformational search, Xplor-NIH does not limit the calculated structures to ones that are similar to those seen in the databases but instead calculates structures that remain consistent with the NMR data included in the calculations (Kuszewski et al. 2001; Bermejo et al. 2012). Published structures that were recalculated using the new force field had better geometry and atom contacts and fewer unfavorable backbone conformations and steric clashes (Bermejo et al. 2016). This new forcefield, which is incorporated in recent versions of Xplor-NIH (version 2.41 and later), was used to calculate the structure of the HBD-aptamer complex.

4.1.3 Chapter overview

This chapter describes the solution structure of the HBD bound to the aptamer. The complex has a heavy atom RMSD of 2.0 ± 0.8 Å using the residues making up the well-defined core of the complex: nucleotides 3-6 and 10-13 and amino acids 128-162. The protein backbone RMSD for amino acids 128-162 is 1.4 ± 0.5 Å. The aptamer contacts both subdomains of the

HBD, and the orientation of the two subdomains relative to one another changes upon aptamer binding. The aptamer has three major structural units: a four base pair hairpin capped by a triloop, a tetraloop-like structure, and two tandem base triples. The binding interface contains a variety of interactions, including hydrogen bonds, hydrophobic interactions, and potential salt bridges.

4.2 Materials and methods

4.2.1 Xplor-NIH structure determination

Structure calculations were performed using Xplor-NIH version 2.44 and the Janus supercomputer operated by the University of Colorado Boulder, and the scripts used in the calculations are provided in Appendix A3. NOEs were identified as described in previous chapters. NOEs that were included as ambiguous restraints during preliminary calculations were reconciled by analyzing the 20 lowest energy structures and determining which NOEs were satisfied. The distance restraints were defined using CYANA 2.0, which converts the peak volumes to a calculated distance using a continuous calibration curve (Güntert & Buchner 2015). The calibration constants are determined automatically so that the median of the upper distance limits equals a user-defined value, here 4.5 Å. Upper and lower distance bounds were set to +20% and the van der Waals distance respectively. Dihedral angle restraints for the HBD were predicted using the program TALOS+ (Sections 2.2.4 and 2.3.2). For the aptamer, generic B-form angles obtained from Xplor-NIH version 2.44 example scripts were included for nucleotides forming the Watson-Crick pairs in the hairpin residues 3-6 and 10-13 (α : $-70^\circ \pm 50^\circ$, β : $180^\circ \pm 50^\circ$, γ : $70^\circ \pm 35^\circ$).

Two hundred structures were calculated using restrained molecular dynamics and simulated annealing starting from an extended chain. High temperature dynamics were performed at 3500 K for 15 ps. Then, simulated annealing was performed by decreasing the temperature to 25 K in steps of 7 K and 0.7 picoseconds per step. Experimental NOEs, protein dihedral angles from TALOS+, and generic dihedral angles for the Watson-Crick base pairs in the aptamer (nucleotides 3-6 and 10-13) were included during simulated annealing. Since the protein contains four disulfide bonds (Cys117-C135, Cys139-Cys158, and Cys146-Cys160), additional NOE restraints of 2.0 ± 0.1 Å were added to all sulfur atoms involved in the disulfide bonds. Covalent disulfide bonds were added during refinement.

Since the refinement used a single structure as the starting model, the ten lowest energy structures from the simulated annealing were each used in an execution of the refinement script, but the ensemble described here originated from the refinement that started with the lowest energy structure calculated during simulated annealing. The refinement included additional NOE and planarity restraints for the base pairs and base triples, as well as hydrogen bonds for the protein backbone and covalent disulfide bonds. Since NOEs were seen between nucleotides involved in Watson-Crick base pairs, and preliminary calculations supported formation of those base pairs, additional idealized base pair distance restraints and planarity restraints were added to the structure calculations during refinement. NOEs were also seen between the nucleotides that form the base triples, so idealized base pair distance restraints were added, but no planarity restraints were included for the Hoogsteen base pairs in the triples since preliminary calculations suggested that these deviated from planarity. Hydrogen bond restraints were added to the protein backbone during refinement by using a knowledge-based hydrogen bonding potential. During refinement, the starting model was heated to 1000 K and slowly cooled to 25 K in 7 K

increments. The ensemble described here contains the 10 lowest energy structures from the 200 structures during refinement. PyMOL Molecular Graphics System (Version 1.8.4.2 enhanced for Mac OS X, Schrödinger, LLC) was used to analyze the ensemble, calculate the pairwise RMSD, and prepare figures.

4.2.2 Nucleic acid structure analysis with 3DNA

The 3DNA software suite version 2.3 was used to calculate the base pair, base step, and helical parameters of the aptamer bound to the HBD (Lu & Olson 2003). Two programs were used on individual members of the ensemble: **find_pair** and **analyze** (Lu & Olson 2008).

Find_pair was used to identify base pairs and helical regions in the structure, and the **-p** option was included to obtain detailed information on the base pairs and base triples (Figure 4.1).

Analyze calculated the base pair, base step, and helical parameters, as well as torsion angles, sugar puckers, helix radius, helical and base pair vectors, and bond distances. The programs were executed using the Unix/Linux-like command line scripts shown in Figure 4.1. The parameters obtained for each member of the ensemble were averaged and are reported in the text as the mean \pm standard deviation.

```

A
  find_pair -OPTION INPUT_FILE OUTPUT_FILE
  analyze -OPTION INPUT_FILE OUTPUT_FILE

B
  find_pair complex.pdb find_pair.inp
  find_pair -p complex.pdb find_all.inp
  analyze find_pair.inp analyze.out

```

Figure 4.1 Command line scripts used to run 3DNA. (A) Generic script used to run the **find_pair** and **analyze** programs. (B) Example script used to run programs. In the top two lines, `complex.pdb` is a standard PDB file containing one model from the ensemble of the complex. The output file generated in the first line (`find_pair.inp`) is used as the input file for **analyze**. The output file that describes higher base associations, which was generated in the second line (`find_all.inp`), is not compatible with **analyze**.

4.2.3 Binding affinity measurements using fluorescence polarization

Fluorescence polarization was used to measure the binding affinity for the aptamer mutants for the HBD. Wild type and mutant DNA aptamers were synthesized with a fluorescein isothiocyanate (FITC) modification at the 5' end by Integrated DNA Technologies (Table 4.2). The lyophilized aptamer was resuspended in water to a concentration of 100 μ M. For binding assays, the aptamer was diluted to a concentration of 20 nM in binding buffer (20 mM Tris HCl pH 7.4, 10 mM NaCl, 3% glycerol, and 0.01% IGEPAL) and annealed by heating at 95 $^{\circ}$ C for 2 minutes followed by incubation on ice. Reactions were done in duplicate in a black polystyrene 384-well flat bottom assay plate (Corning). The HBD titration was made by serial dilutions starting from the highest concentration of 10.5 μ M HBD in binding buffer and diluting 1:1.9 (11 μ L to 21 μ L) in binding buffer. The aptamer was added to each well as 10 μ L of 20 nM aptamer in binding buffer for a final reaction volume of 20 μ L. Reactions were incubated at room temperature for 1 hour prior to measuring polarization. Plates were read in a CLARIOstar

microplate reader (BMG Labtech) using an excitation wavelength of 482 nm and emission wavelength of 530 nm with an LP 504 dichroic filter. Data were collected as the intensities of parallel and perpendicular polarized light.

Table 4.1 Sequences of aptamer mutants

Aptamer	Sequence
Wild type	5'[FITC]-CCGTCTTCCAGACAAGAGTGCAGGG 3'
T4C	5'[FITC]-CCG C CTTCCAGACAAGAGTGCAGGG 3'
T4C/A12G	5'[FITC]-CCG C CTTCCAG G CAAGAGTGCAGGG 3'

Fluorescence anisotropy was calculated using Equation 4.1:

Equation 4.1

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

where anisotropy (A) is defined as the difference in the intensity of parallel (I_{\parallel}) and perpendicular (I_{\perp}) polarized light divided by the total fluorescence. Anisotropy is converted to fraction bound using Equation 4.2:

Equation 4.2

$$\frac{[RL]}{[R]_T} = \frac{A - A_f}{A_b - A_f}$$

where the $[R]_T$ is the total receptor concentration, $[RL]$ is the concentration of bound receptor, A is the measured anisotropy, A_b is the anisotropy of bound fluorescent aptamer, and A_f is the anisotropy of free fluorescent aptamer. The receptor (R) represents the fluorescent aptamer and the ligand (L) represents the HBD. The equilibrium dissociation constant was calculated using the Microsoft Excel solver plug-in and Equation 4.3:

Equation 4.3

$$\frac{[RL]}{[R]_T} = \frac{([L]_T + n[R]_T + K_D) - \sqrt{(-[L]_T - n[R]_T - K_D)^2 - 4[L]_T n[R]_T}}{2[R]_T}$$

where $[L]_T$ is the total ligand concentration, K_D is the equilibrium dissociation constant, and n is the number of binding sites.

4.3 Results and discussion

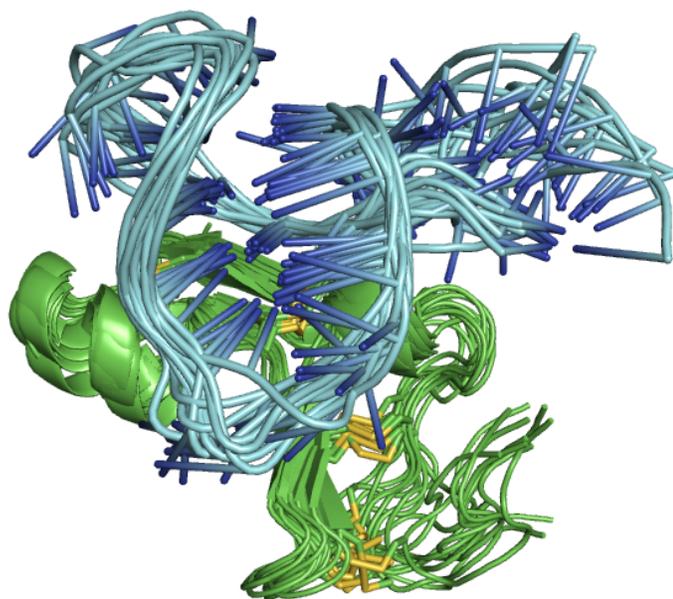
4.3.1 Overview of the solution structure of the HBD-aptamer complex

The solution structure of the complex is shown in Figure 4.2 as an ensemble created by a superposition of the 10 lowest energy structures. The core of the complex, which includes the base pairs in the duplex region of the hairpin (nucleotides 3-6 and 10-13) and amino acids 128-162, has an average heavy atom RMSD of 2.0 ± 0.8 Å. The protein backbone has an RMSD of 1.4 ± 0.5 Å for amino acids 128-162. The restraints included in the calculations and the violations present in the solution structure are shown in Table 4.2.

Table 4.2 Summary of NMR structural statistics for the 10 lowest energy structures of the HBD in complex with the aptamer

	Protein	Aptamer
Total NOE restraints	548	306
Intra-residue	150	177
Sequential	146	92
Medium range	113	13
Long range	139	24
Intermolecular		55
Dihedral restraints	72	20
Structural statistics	10 Lowest-Energy Structures	
Number of NOE violations >0.5 Å	2.2 ± 1.6	
Number of dihedral violations > 5°	3.5 ± 1.4	
Pairwise RMSD (Å)	All heavy atoms	
Protein (128-162)	2.06 ± 0.80 Å	
Aptamer (3-6, 10-13)	1.52 ± 0.59 Å	
Complex (3-6, 10-13, 128-162)	1.99 ± 0.77 Å	
Ramachandran statistics		
Favored regions (%)	94.7%	
Allowed regions (%)	4.2%	
Disallowed regions (%)	1.1%	

A



B

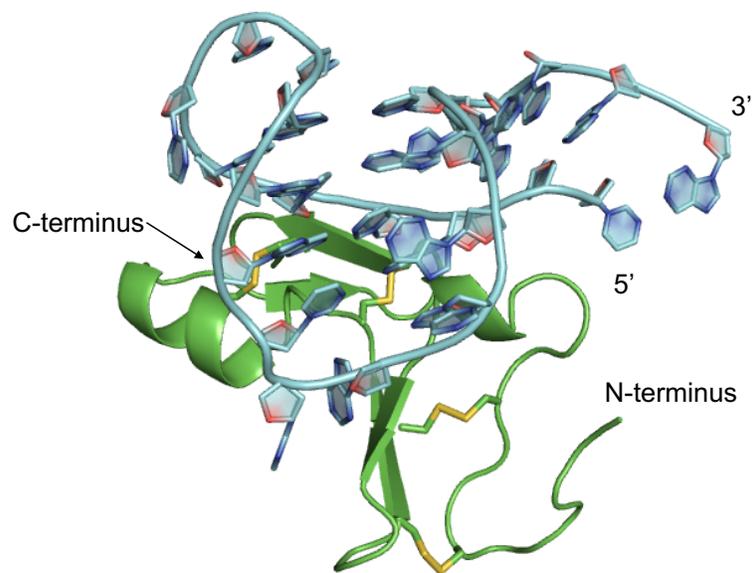


Figure 4.2 Solution structure of the HBD bound to the aptamer. The HBD is shown in green with the disulfide bonds shown as sticks with yellow sulfur atoms. The aptamer is shown in pale cyan. **(A)** Superposition of the ten lowest energy structures. The bases are portrayed as rods for clarity. **(B)** Lowest energy structure from the ensemble with bases shown explicitly.

The HBD and the aptamer form a 1:1 complex, as was determined by NMR titrations and isothermal titration calorimetry (Section 3.3.1 and Appendix A4). The aptamer blocks the heparin-binding site and contacts both subdomains of the HBD. Upon binding the aptamer, the HBD retains the secondary structures present in the solution structure of the free HBD. The N-terminal subdomain has limited NOEs, so it shows more possible conformations in the ensemble that are consistent with the data, whereas the C-terminal subdomain has a larger number of intramolecular NOEs and is much better defined with a larger number of intramolecular NOEs. A total of 854 NOEs were included as distance restraints (Table 4.1). The number of intramolecular NOEs per residue for the HBD and for the aptamer are shown in Figure 4.3.

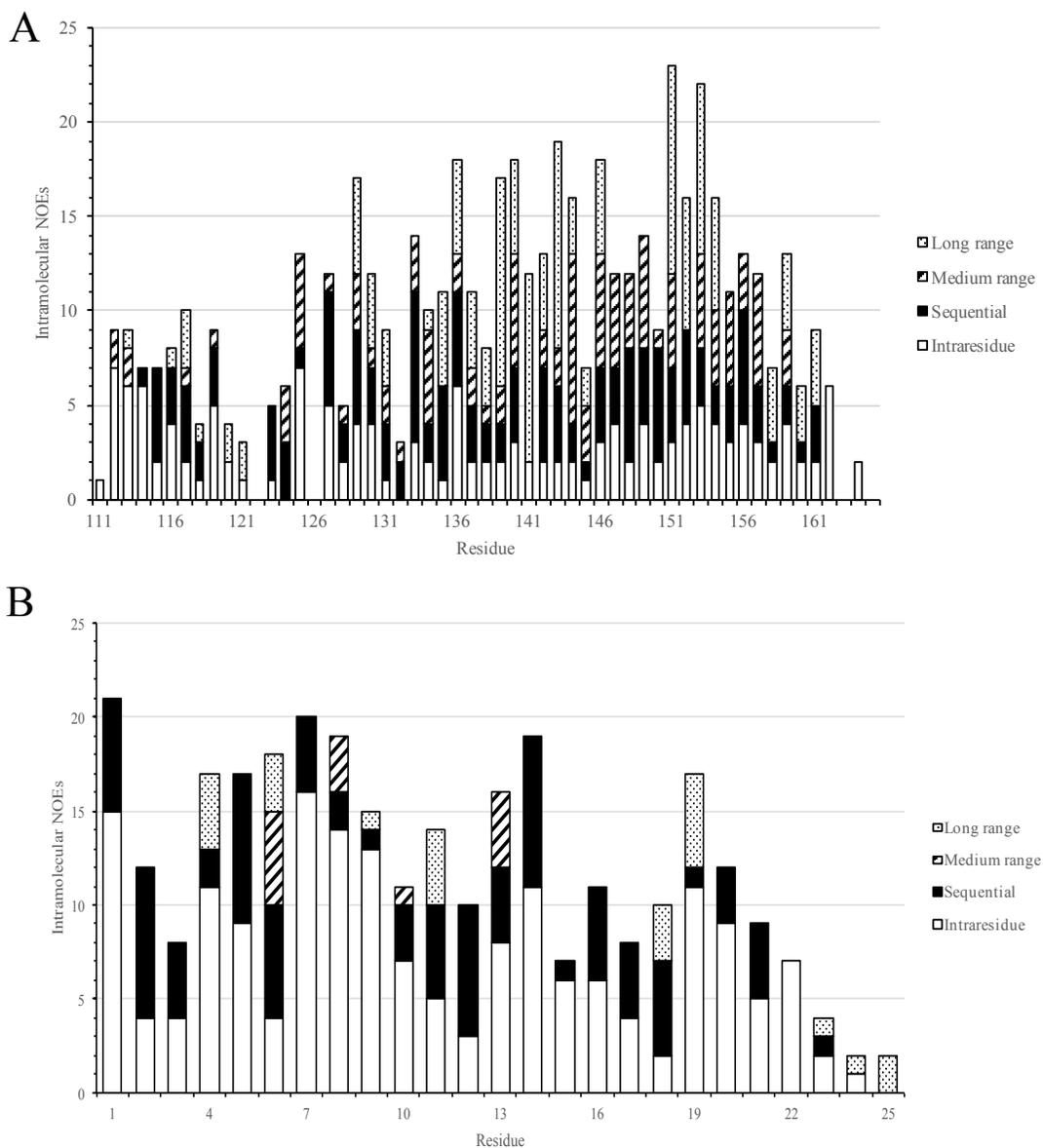


Figure 4.3 Intramolecular NOEs by residue for the HBD and the aptamer. Intermolecular NOEs for (A) the HBD and (B) the aptamer are divided into four categories: intraresidue (white), sequential (black), medium-range (two to four residues apart; stripes), and long-range (five or more residues apart; dots).

The individual subdomains of the bound HBD overlay well with those of the free HBD, but there is a change in orientation between the two subdomains, which causes the conformation of the protein as a whole to change upon binding the aptamer. This change in subdomain orientation is facilitated by the aptamer binding and stabilized by a hydrophobic network running through both subdomains of the HBD. The aptamer has three notable structural elements: a tetraloop-like structure, a hairpin capped by a pyrimidine-rich triloop, and a short triplex of two tandem base pairs. The aptamer shows significant variability between the structures in the ensemble, particularly in unpaired regions and in the tetraloop-like structure where there are limited NOEs. Nucleotides in the hairpin, tetraloop-like structure, and outside of the major structural motifs all make direct contact with the HBD. The base triples do not contact the protein, and are therefore acting as a scaffold to anchor the aptamer in the correct conformation to bind the HBD.

The quality of the structures in the ensemble was assessed using the MolProbity validation web server, which is unique among validation programs in its ability to analyze NMR and crystal structures containing nucleic acids (Chen et al. 2010; Davis et al. 2004; Davis et al. 2007). MolProbity is able to construct a Ramachandran plot for the ensemble, but in depth analysis is done on the individual models in an ensemble separately. The Ramachandran plot for the ensemble created by MolProbity shows that 94.7% of the residues are in favored regions, 98.9% are in allowed regions (Figure 4.4). Gly118 showed up as an outlier in four of the ten structures. Glu114 and Pro119 were each an outlier in one structure. No structures had more than one outlier, and four of the ten structures had no Ramachandran outliers. Additional results from the MolProbity analysis of the structure closest to the average are shown in Figure 4.5, and results from the other structures in the ensemble are shown in Appendix A5.

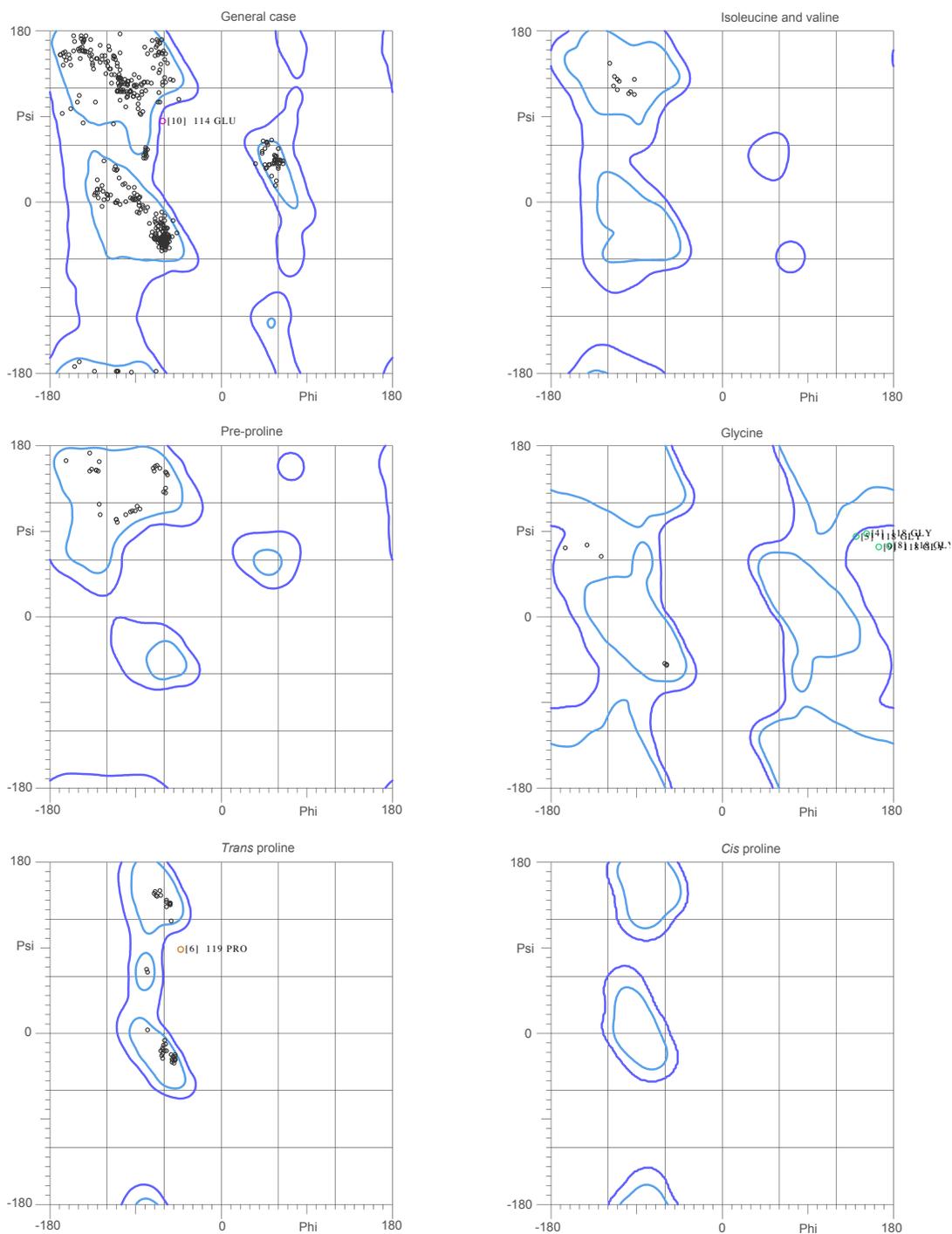


Figure 4.4 Ramachandran plot for the ensemble generated by MolProbity. The light blue and dark blue borders specify the favored and allowed regions respectively. The “general case” plot shows the dihedral angles for all residues except glycine, proline, and “pre-proline” (residues immediately preceding prolines, except glycine and proline).

MolProbity analyzes the all-atom contacts to find steric clashes between non-bonded atoms. The clashscore is the number of steric overlaps of 0.4 Å or greater per 1000 atoms, and it is determined using a program called Probe that uses a 0.25 Å-radius probe along the van der Waals radii to find areas of overlap between non-bonded atoms (Chen et al. 2010; Davis et al. 2004; Davis et al. 2007; Word et al. 1999). The average clashscore for the members of ensemble is 9.0 ± 2.7 . While each member of the ensemble has different atoms involved in clashes, the clash between HA of Gln130 and the Lys136 O is labeled as severe in seven of the ten structures. The NOE restraint between the HB* of Gln130 and the HN of Lys136 is likely playing a role in driving the backbone atoms too close in some structures, so the NOE may instead be an artifact caused by spin diffusion.

The protein geometry is evaluated and compared to the expected bond lengths and bond angles established for protein structures (Engh & Huber 1991). Bond lengths or bond angles that are more than four standard deviations away from the expected values are considered outliers (Chen et al. 2010). No members of the ensemble have protein bond lengths flagged as outliers, and each member has the CA-CB-CG bond angles of His126 and Phe128 considered outliers at 4.3 standard deviations away from the expected values. His126 is unlabeled in the protein, so no restraints involving that residue were included in the calculations. However, it sits at the interface with the aptamer, so the position of the aptamer may be affecting the side chain angle of His126. Phe128 aromatic ring and the side chain of Leu127 are involved in hydrophobic packing interactions in all of the structures in the ensemble. This interaction may lead to the distortion in the bond angle reported by MolProbity. The side chain rotamers are also evaluated and compared to specific rotamer definitions (Lovell et al. 2000). Those that fall outside of the expected values are considered outliers (Chen et al. 2010). The ensemble has an average of $1.5 \pm$

1.1 side chain rotamers labeled as outliers. Two of the structures in the ensemble have no rotamer violations. The residues that show up as outliers in the rest of the ensemble are different between structures, but the most common outliers are Val129 (five of the eight structures) and Cys158 (four of the eight structures). MolProbity also analyzes the position of the CB and compares it to the ideal value calculated by the geometry of the backbone atoms (Lovell et al. 2003). Outliers are marked if the CB is 0.25 Å or more away from the ideal position (Davis et al. 2007; Chen et al. 2010). None of the structures have residues with outlier CB positions. The conformations of the peptide bonds of each residue are also evaluated as *cis* or *trans* with the expectation that the peptide bonds are all in the *trans* conformation with the exception of proline residues, which have a higher frequency of *cis* conformations than the other amino acids (MacArthur & Thornton 1991; Stewart et al. 1990). None of the structures have *cis* peptide bonds. The MolProbity score provides a quick summary of the quality of the structure by combining the clashscore, percent of residues falling in the unfavorable regions of the Ramachandran plot, and the percent of side chain rotamer outliers (Davis et al. 2007; Chen et al. 2010). The average MolProbity score for the ensemble is 2.1 ± 0.3 . This score can be thought of as the X-ray crystallographic resolution of a structure that would be expected to have similar results from a MolProbity analysis (Chen et al. 2010).

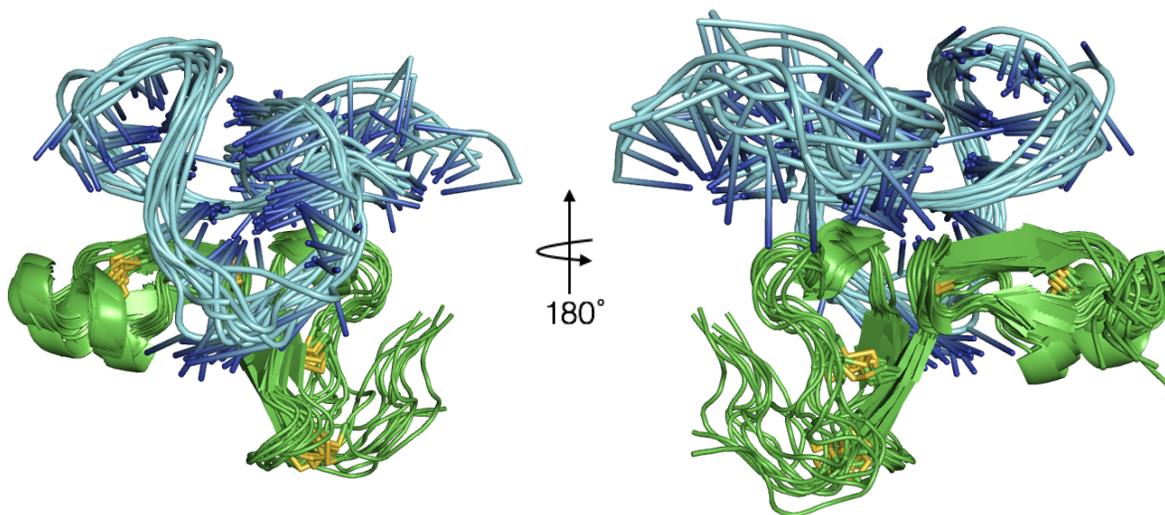
All-Atom Contacts	Clashscore, all atoms:	7.16		86 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	1.92%	Goal: <0.3%
	Favored rotamers	51	98.08%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	50	94.34%	Goal: >98%
	MolProbity score [^]	1.99		76 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	16	76.19%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

Figure 4.5 MolProbity summary statistics for the structure closest to the mean.

Disulfide bond NOE restraints were included during simulated annealing, and base pair NOEs and base pair planarity restraints were included during refinement. To ensure that these additional restraints were not biasing the structure, simulated annealing was repeated using the NOE distance restraints and dihedral angle restraints but without the additional disulfide bond or base pair restraints (Figure 4.6). The heavy atom RMSD of the core of the complex—nucleotides 3-6 and 10-13 and amino acids 128-162—is 3.2 ± 1.3 Å, and the protein backbone has an RMSD of 1.9 ± 0.8 Å. These values are higher, as expected, than the RMSDs for the complex with the additional restraints imposed, which were 2.0 ± 0.8 Å for the heavy atom RMSD for the core of the complex and 1.4 ± 0.5 Å for the protein backbone. Instead of trying to compare two ensembles, the ensemble calculated without additional restraints was compared to the average structure from the ensemble calculated with the additional NOE and planarity restraints by calculating the pairwise RMSD of each member of the ensemble to that average structure. This

average RMSD is $2.0 \pm 0.3 \text{ \AA}$, which shows that it is the experimental restraints and not the additional imposed restraints that are driving the structure during the calculations.

A



B

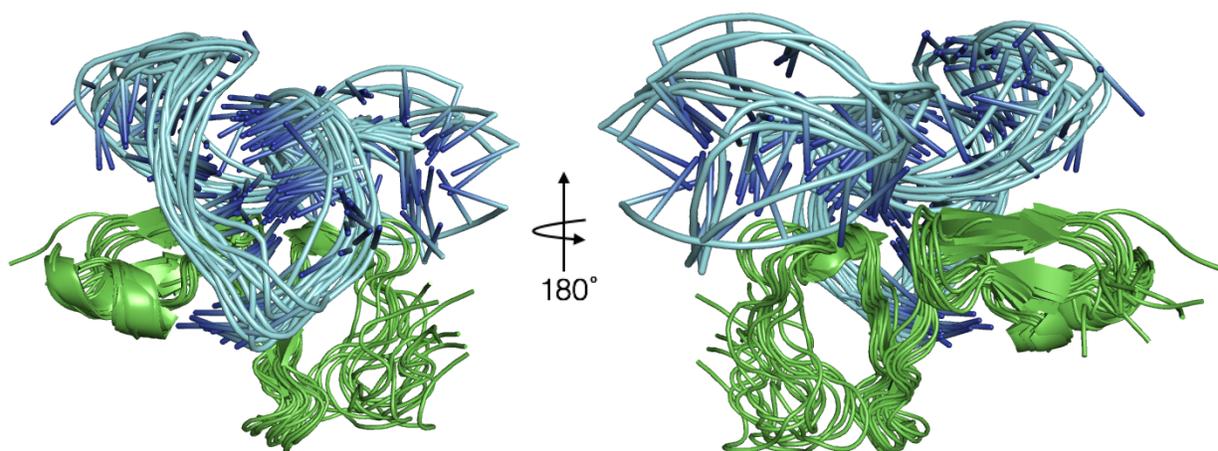


Figure 4.6 Ten lowest energy structures from calculations performed with and without imposing ideal restraints for disulfide bonds and base pairs. The HBD is in green with sulfur atoms of disulfide bonds shown in yellow, and the aptamer in cyan. (A) Ensemble described in this chapter, which was calculated with additional restraints to form covalent disulfide bonds and ideal base pairs. (B) Ensemble from simulated annealing calculation performed without covalent disulfide bonds and without base pair NOEs or planarity restraints.

4.3.2 Structure of the HBD in the HBD-aptamer complex

The 55-amino acid HBD contains two subdomains that are described as the N-terminal and C-terminal subdomains (Figure 4.7). The structure of the protein bound to the aptamer is well defined. The HBD has an average backbone RMSD of 2.4 ± 0.9 Å using all residues, or 1.5 ± 0.6 Å using only amino acids 118-162, which exclude the poorly defined N- and C-termini. The two subdomains are not equally converged due to difference in the number of NOEs defining each subdomain. The N-terminal subdomain (amino acids 111-138) has 8.3 NOE restraints per residue and an average backbone RMSD of 2.6 ± 1.0 Å (Figure 4.8A). The C-terminal subdomain (amino acids 139-165) has an average backbone RMSD of 1.2 ± 0.4 Å and 12.4 NOE restraints per residue, which includes more medium-range and long-range NOE restraints. The N-terminal subdomain has 1.5 medium-range NOE restraints per residue and 1.3 long-range NOE restraints per residue, whereas the C-terminal subdomain has 3.8 medium-range and 3.0 long-range NOE restraints per residue. This leads to better definition of the C-terminal subdomain and translates to a lower backbone RMSD (Figure 4.8B).

The N-terminal subdomain contains a two-strand antiparallel β -sheet. Residues Val129, Gln130, and Asp131 form one strand, and Lys136, Cys137, and Ser138 form the second strand. Two disulfides bonds hold the subdomain together: one between Cys117 and Cys135, and the other between Cys120 and Cys137. Residues 125-127 form a half turn of an α -helix, but the remainder of the N-terminal subdomain has no defined secondary structure. Fairbrother and colleagues suggested that this region of the protein exhibited motion on various timescales, but we cannot make a similar observation with the HBD bound to the aptamer without explicitly studying the HBD dynamics (Fairbrother et al. 1998).

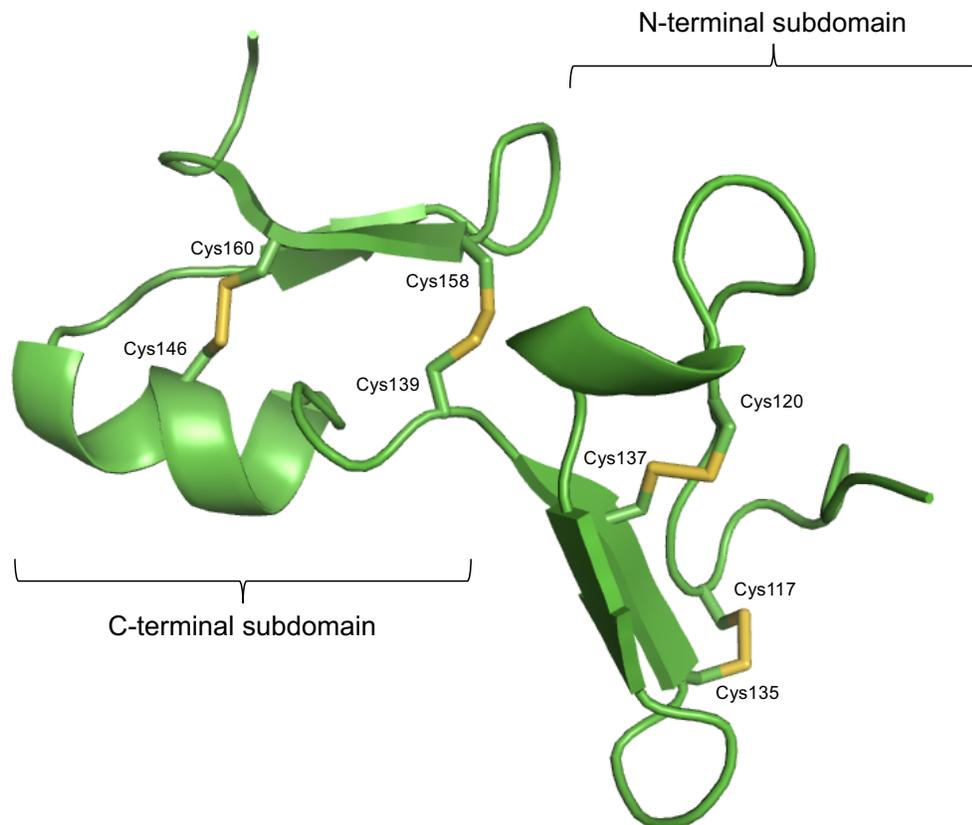
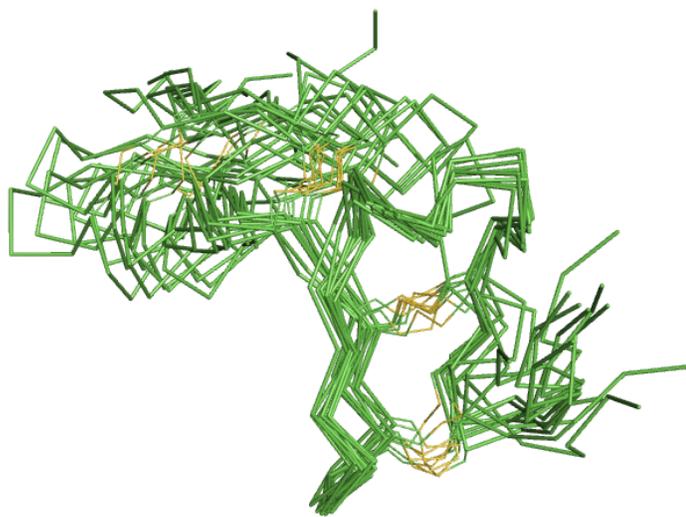


Figure 4.7 Cartoon representation of the bound HBD. Each subdomain of HBD contains two disulfide bonds shown as stick with the sulfur atoms in yellow. The aptamer is not shown.

A



B

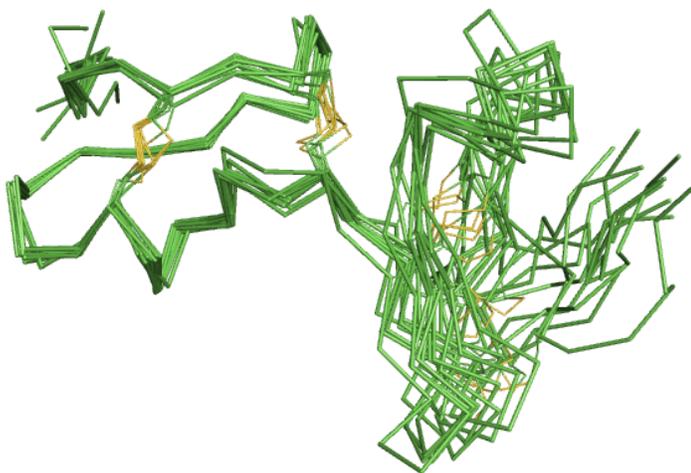


Figure 4.8 Ribbon representation of the bound HBD. (A) Superposition of the ensemble was optimized for the backbone atoms of the N-terminal subdomain (amino acids 111-138). (B) Superposition of the ensemble was optimized for the backbone atoms of the C-terminal subdomain (amino acids 139-165). The aptamer is not shown.

The C-terminal subdomain contains the other two disulfide bonds subdomain (Cys139-C158 and Cys146-Cys160), as well as an α -helix packed against an antiparallel β -sheet. The two-turn α -helix contains seven amino acids: Asp143, Ser144, Arg145, Cys146, Lys147, Ala148, and Arg149. The two antiparallel β -strands making up the short sheet include Leu151, Glu152, and Leu153 on one strand and Arg159, Cys160, and Asp161 on the other. The structure of the subdomain is stabilized by hydrophobic packing between Leu151, Leu153, and the Cys146-Cys160 disulfide bond, as well as hydrogen bonding between the amide proton of Leu151 that forms a hydrogen bond with the carbonyl oxygen of Cys146 (Figure 4.9). In four of the structures in the ensemble, there is a second hydrogen bond between the side chain carboxylate group of Asp143 and the amide H of Leu153. The secondary structure elements as well as the tertiary interaction between the α -helix and β -sheet are preserved between the structures of free and bound HBD.

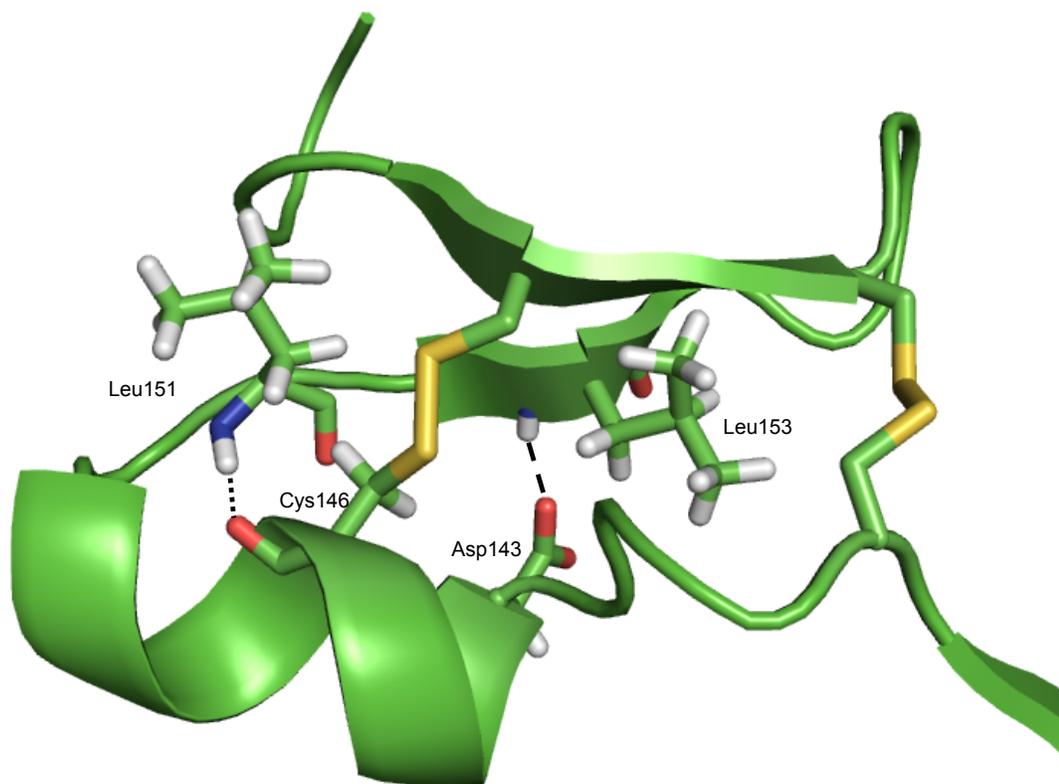
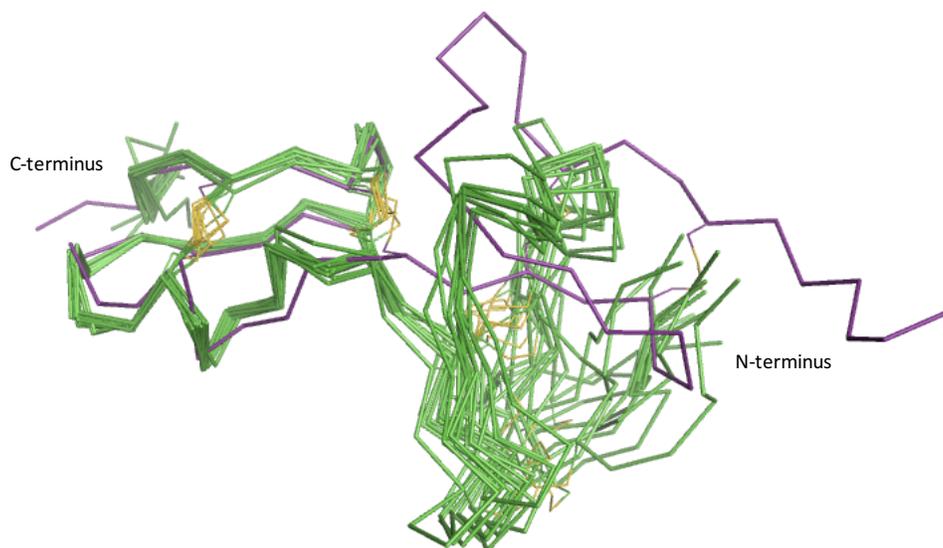


Figure 4.9 Interactions between the α -helix and β -sheet in the C-terminal subdomain. All structures in the ensemble show a hydrogen bond (dotted line) between the Leu151 amide hydrogen and the Cys146 carbonyl oxygen. Four of the structures, including the one shown here, also show the hydrogen bond between the Asp143 carboxylate group and the Leu153 amide hydrogen (dashed line).

Overall, the backbone conformations of the individual subdomains of the bound HBD are very similar to those of the free HBD due to the preservation of the secondary structure elements discussed. The average backbone RMSD of amino acids 139-162 of the C-terminal subdomain between the ensemble of the bound HBD and the structure closest to the mean from the ensemble of the free HBD is 1.3 ± 0.1 Å (Figure 4.10A). The N-terminal subdomain experiences slightly more deviation from the free HBD with an average backbone RMSD of 1.8 ± 0.3 (using amino acids 118-138; Figure 4.10B). The N-terminal seven residues and C-terminal three residues were not included in the RMSD calculations because the backbone conformations of these residues are variable in both the free and bound HBD ensembles. Despite the similarity of the subdomains between free and bound HBD, the orientation of the two subdomains relative to one another changes upon binding the aptamer. The two subdomains become closer together, forming a more “L” shape than the “linear” orientation seen in the structure of the free HBD.

A



B

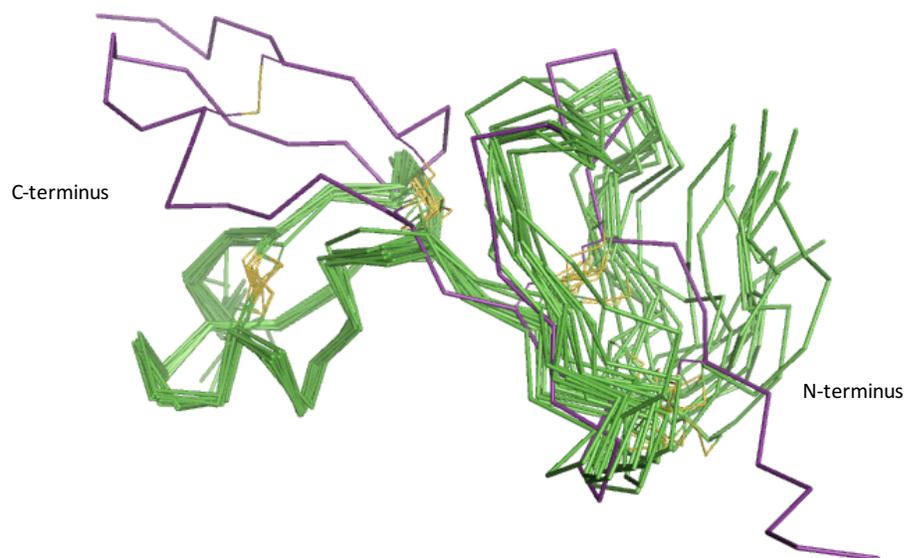
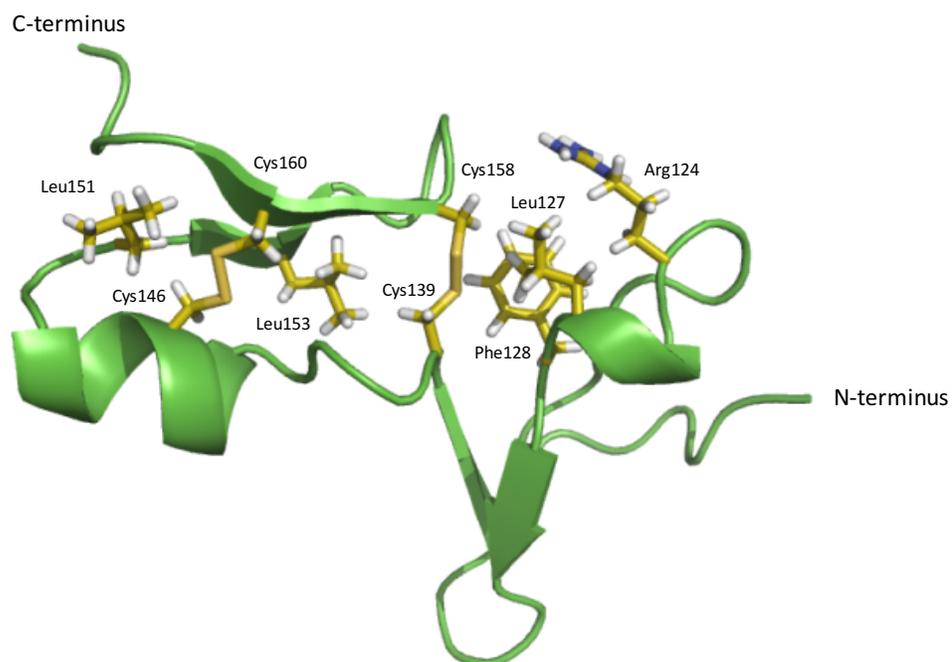


Figure 4.10 Overlay of the ribbon representation of free and bound HBD. A representative structure of free the HBD is shown in purple, and the ensemble of the bound HBD is shown in green. (A) Superposition optimized for the backbone atoms of the C-terminal subdomain. (B) Superposition optimized for the backbone atoms of the N-terminal subdomain. The aptamer is not shown.

In addition to secondary structure elements and disulfide bonds stabilizing the conformation, the HBD has a network of hydrophobic interactions involving residues from both subdomains (Figure 4.11A). The methylene protons from Arg124 pack against the side chains of Leu127 and Phe128, which pack against the Cys139-Cys158 disulfide bond and form the interface between the N- and C-terminal subdomains. Leu153 packs between the two disulfide bonds in the C-terminal subdomain: Cys139-Cys158 and Cys146-Cys160. Leu151 packs against the other side of the Cys146-Cys160 disulfide bond. A hydrophobic network is also seen in the structure of the free HBD, where Phe128 instead interacts with the Cys137-Cys120 disulfide bond (Figure 4.11B) (Fairbrother et al. 1998). The additional hydrophobic interactions from the methylene protons of the Arg124 side chain in the bound HBD may help stabilize the new orientation of the subdomains since they do not participate in the hydrophobic network in the free HBD.

A



B

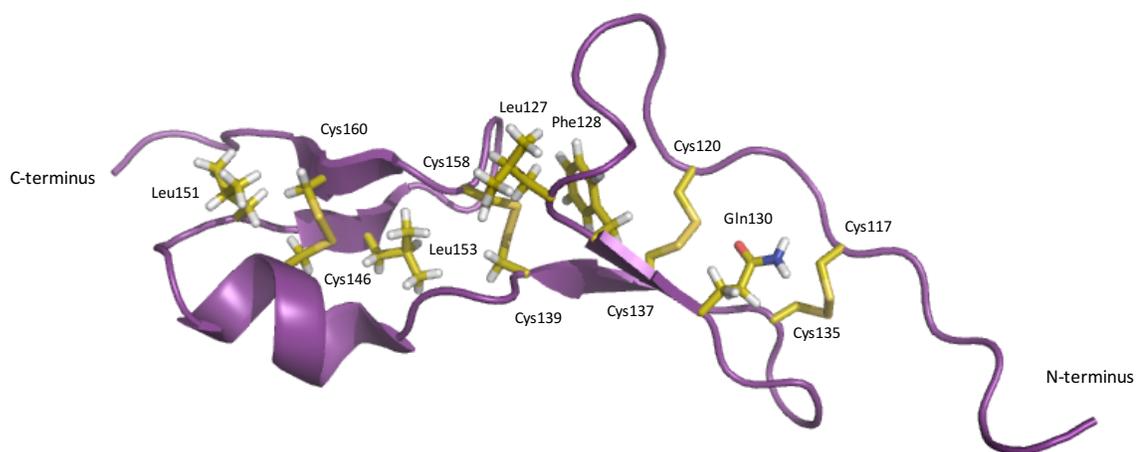


Figure 4.11 The HBD has an extensive hydrophobic network spanning from the N-terminal subdomain to the C-terminal subdomain. The residues involved in the hydrophobic interactions are shown as gold sticks in representative structures from (A) the bound HBD and (B) the free HBD.

The change in subdomain orientation contrasts with most other published protein-aptamer complexes in which the backbone conformation of the protein shows little to no change upon aptamer binding (Gelinias et al. 2016). In these published structures, the RMSDs of the backbone atoms between free protein and bound protein are between 0.3 and 1.8 Å (Gelinias et al. 2016). The exception to this is the nuclear factor κ B (NF κ B) protein-aptamer complex (Huang et al. 2003). The structure of the free NF κ B has not been determined, but the conformation of the protein bound to the RNA aptamer is quite different from that of NF κ B bound to its canonical DNA ligand (RMSD 5.1 Å). The NF κ B DNA-binding domain has two domains that interact with the DNA, and while the backbone conformation of each domain changes little between binding the DNA and binding the aptamer (1.2 Å for the N-terminus and 1.8 Å for the C-terminus), the orientation of the two domains relative to one another changes. While this is not something seen in all structures of aptamers that bind multi-domain proteins, such as the immunoglobulin G aptamer, it may provide an alternate mechanism for aptamers to use when binding to multi-domain proteins or proteins like the HBD with subdomains able to take on multiple orientations (Nomura et al. 2010).

4.3.3 Structure of the aptamer in the HBD-aptamer complex

In section 3.3.2, the predicted secondary structure of the aptamer was presented. The mFold webserver, an internet-based software that does *in silico* folding of nucleic acids, predicted that the aptamer folds only into a four base pair hairpin capped by a three nucleotide loop (Figure 4.12A) (Zuker 2003). Potty and colleagues used surface plasmon resonance to determine the effect of mutations of the aptamer on its ability to compete the wild-type aptamer off of the HBD. Mutations to the nucleotides predicted to form base pairs resulted in decreased binding and was interpreted as evidence of base pairing. This data supported the hairpin predicted by mFold, but it also suggested that the 5' and 3' ends formed two additional base pairs (Figure 4.12B, Section 1.2.5) (Potty et al. 2009). However, the solution structure reported here shows that the aptamer is taking on a complicated fold that was not previously predicted. The aptamer forms two tandem major groove triples (Figure 4.12C). G18 forms a reverse Hoogsteen base pair with G11, which is in a Watson-Crick base pair with C5. The closing base pair of the triloop—a Watson-Crick base pair between T6 and A10—forms a base triple with T19 via a reverse Hoogsteen base pair with A10. The three main structural motifs identified in the solution structure of the aptamer bound to the HBD are a hairpin capped by a triloop, a short triplex, and a tetraloop-like structure.

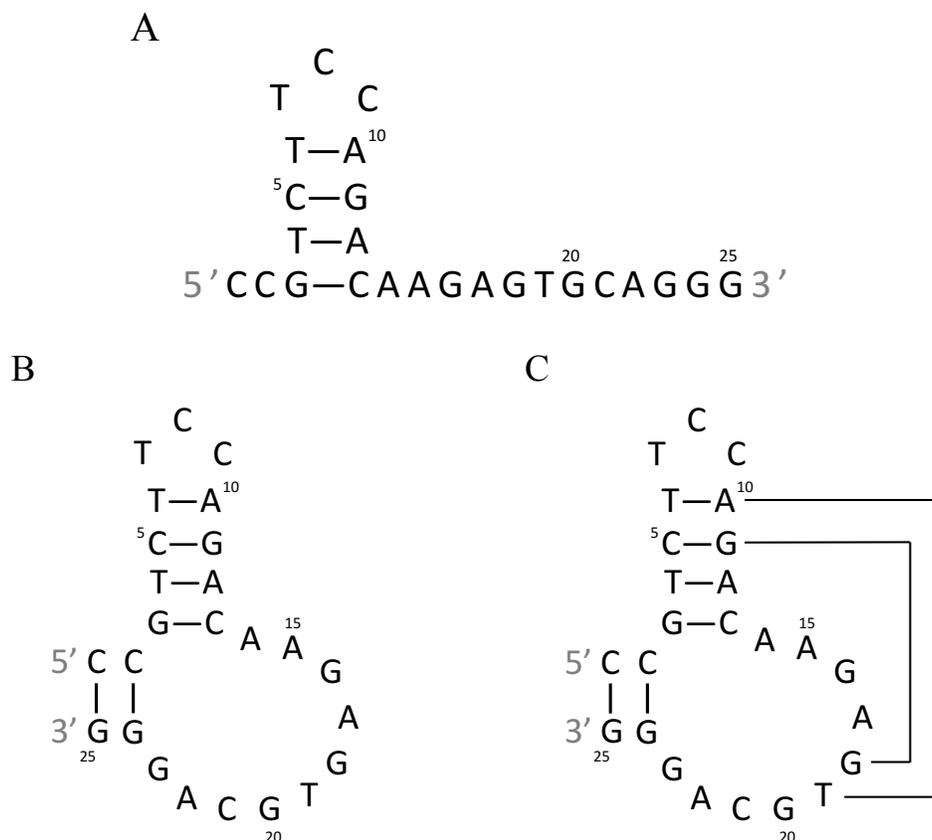


Figure 4.12 Predicted secondary structure of the DNA aptamer. (A) mFold secondary structure prediction. (B) Secondary structure prediction based on biochemical analysis (Potty et al. 2009). (C) Secondary and tertiary structures of the aptamer based on the solution structure presented here.

Four base hairpin

The hairpin consists of four Watson Crick base pairs—G3-C13, T4-A12, C5-G11, and T6-A10—and a pyrimidine-rich triloop including nucleotides T7, C8, and C9 (Figure 4.13). The two pairs at the base of the stem loop, G3-C13 and T4-A12, contribute to the binding interface with the HBD. The two pairs proximal to the triloop form additional hydrogen bonds with G18 and T19 to form a short triplex. The hairpin therefore has roles in both constructing the aptamer tertiary structure and recognizing the HBD.

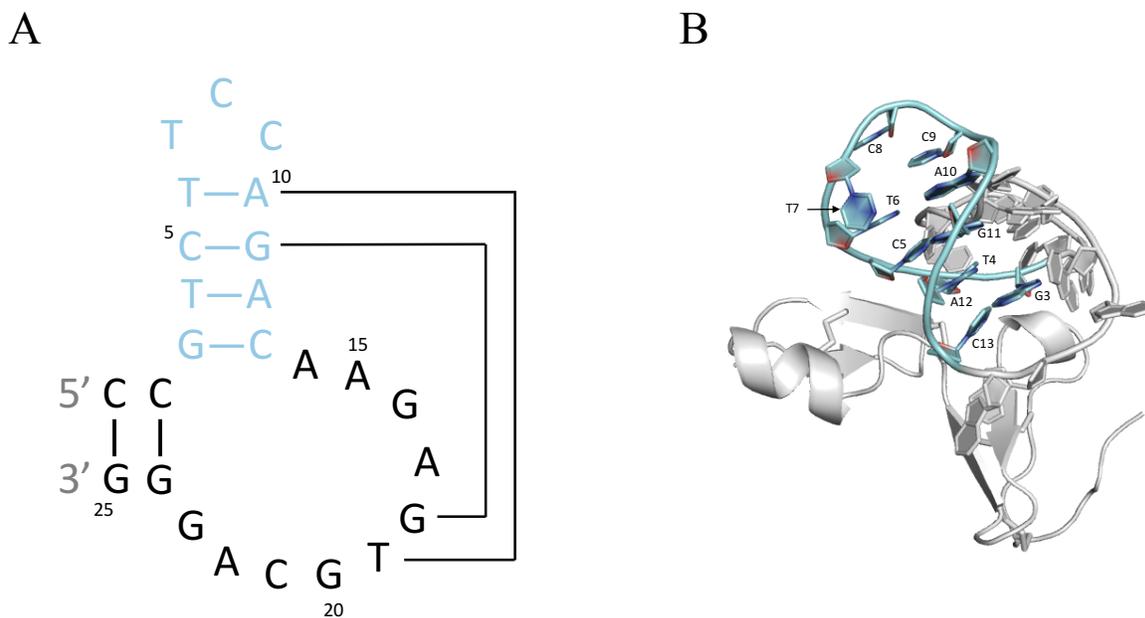


Figure 4.13 The aptamer forms a hairpin with four Watson-Crick base pairs and a three-nucleotide loop. (A) Secondary structure of the aptamer with hairpin bases highlighted cyan. (B) Representative structure of the aptamer with the hairpin bases highlighted in cyan.

The base pairs in the hairpin are distorted from planarity based on the values for buckle, propeller, and opening measured by the nucleic acid analysis software 3DNA (Table 4.3) (Lu & Olson 2003). The degree of propeller twist varies between members of the ensemble, but the majority of structures have values outside of those typically seen in a duplex, -15° to 11° (Richardson 2015). Unlike the other base pairs, the propeller twist for G3-C13 is outside the duplex range in all ten structures in the ensemble. This is likely due to the interactions between the base pair and the protein. Deviations of base pairs from planarity is often seen in regions of higher order structure and ligand binding sites (Richardson 2015). It is therefore likely that the base pair parameters measured for the aptamer reflect two levels of distortion: the protein interface distorting the geometry of the G3-C13 and T4-A12 base pairs, and the formation of the base triples distorting the C5-G11 and T6-A10 base pairs. The hairpin also deviates from ideal

B-form geometry based on the local base helical parameters such as helical rise and helical twist, which have typical values of 3.3 Å and 36.5° respectively (Table 4.4) (Olson et al. 2001). The helical parameters show that the hairpin has a more extended conformation than a standard B-form helix.

Table 4.3 Parameters for the base pairs in the hairpin.

Base pair	Buckle (°)	Propeller Twist (°)	Opening (°)
G3-C13	-3.4 ± 4.8	35.5 ± 9.9	-18.0 ± 4.7
T4-A12	-3.4 ± 4.6	4.6 ± 13.0	-9.3 ± 4.8
C5-G11	-0.8 ± 4.4	-0.4 ± 20.0	-2.5 ± 3.0
T6-A10	0.3 ± 4.5	6.2 ± 17.7	-10.4 ± 5.3

Table 4.4 Helical parameters for the base pairs in the hairpin.

Base pair step	Helical Rise (Å)	Helical Twist (°)
G3-C13/ T4-A12	2.4 ± 1.2	40.9 ± 13.5
T4-A12/ C5-G11	3.4 ± 0.3	42.2 ± 5.3
C5-G11/ T6-A10	2.9 ± 0.5	35.6 ± 5.3

The hairpin is capped by a pyrimidine-rich three-nucleotide loop consisting of nucleotides T7, C8, and C9 (Figure 4.14). These nucleotides have various conformations in the ensemble due to limited NOEs. In eight of the ten structures, the base of C8 is stacked on the base of T6, and in the remaining two structures, the bases of C8 and C9 are involved π -stacking but C8 is too far from T6 for a productive interaction. Stacking was identified by bases that overlap and are separated by 3.4 Å between the two aromatic planes (Kool 2001). In two of the structures with a C8-T6 pi-stacking interaction, C9 is in the *syn* conformation and stacks with C8. In the remaining eight structures, C9 has slight overlap with A10 to allow for some stacking interactions. In nine of the ten structures, T7 is oriented toward the minor groove side of the base

pairs in the hairpin and forms hydrophobic contacts with its methyl group and the sugar protons of T6. Based on its conformation, it is possible that the T7 is interacting with the protein. In the structure closest to the mean, the HH21 and HE protons of the Arg149 guanidinium group are 2.8 Å and 3.1 Å away, respectively, from the exocyclic O4 of T7. These residues could rearrange to maximize hydrogen bonding or π -stacking interactions. However, in the other structures of the ensemble, they are not arranged optimally for intermolecular interactions. Instead, eight structures have the T7 methyl group contacting the T6 deoxyribose methylene protons.

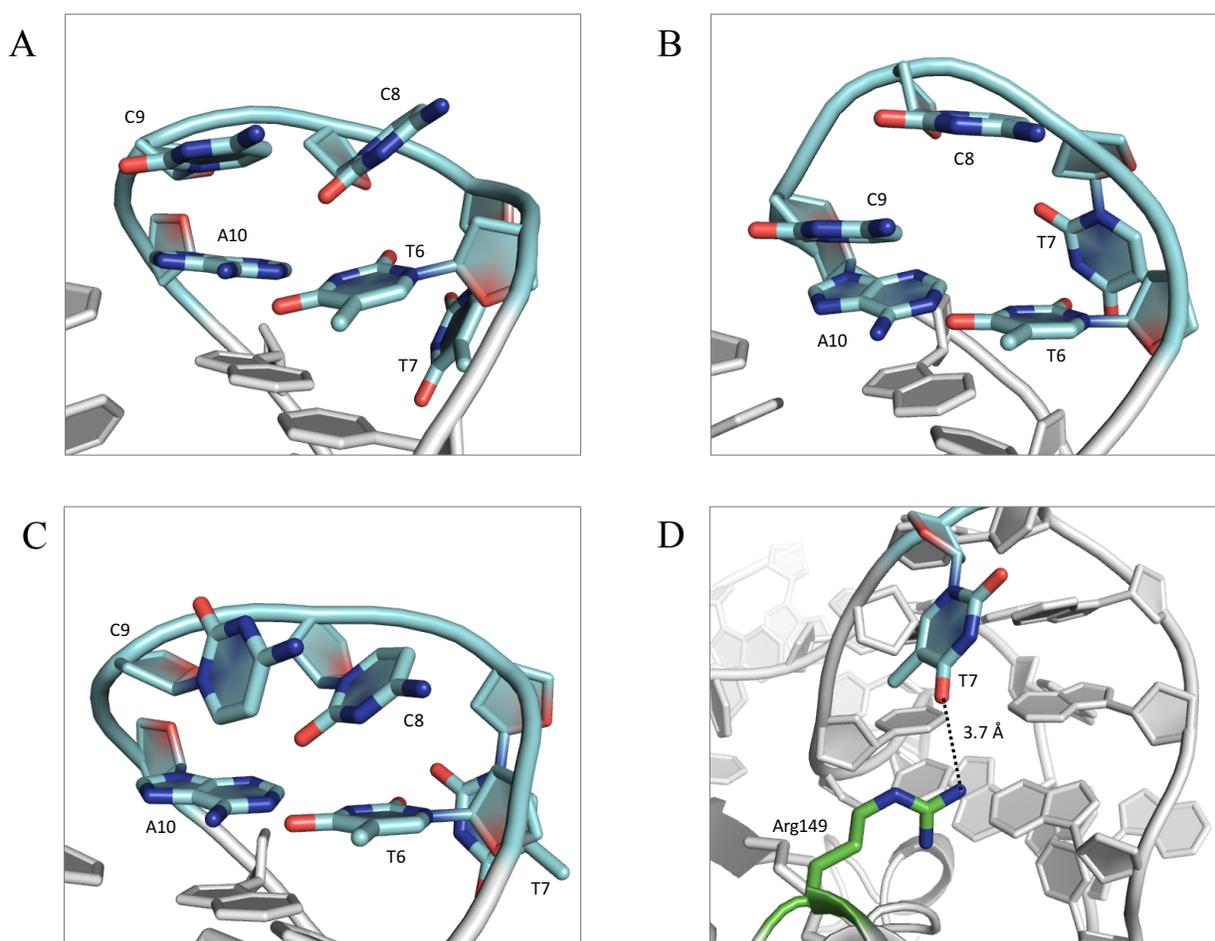


Figure 4.14 Pyrimidine-rich triloop caps the aptamer hairpin. The bases in the triloop show numerous conformations such as (A) *anti* C9 π -stacking with A10 and C8 π -stacking with T6, (B) C8 too far to stack away to stack with T6, (C) *syn* C9 π -stacking with C8, and (D) intermolecular interaction between Arg149 and T7.

The T7-C8-C9 triloop takes on the pyrimidine-rich triloop structure seen in the cruciform promoter for the bacteriophage N4 virion RNA polymerase, which also has a TCC triloop, and the von Willebrand factor (vWF) DNA aptamer, which has a TTC triloop (Chou et al. 1999; Huang et al. 2009). In these triloops, the first nucleotide of the loop, generically numbered T₁, inserts into the minor groove side of a hairpin. In the cruciform promoter triloop, this thymine forms hydrogen bonds with two CG Watson-Crick base pairs proximal to the loop. In the vWF aptamer and the majority of structures in the HBD-aptamer ensemble, T₁ interacts with the deoxyribose of the pyrimidine of the closing base pair. The other two pyrimidines in the triloops, C₂/T₂ and C₃, are able to stack on the closing base pair. In the cruciform DNA, both C₂ and C₃ are involved in stacking—C₂ stacks on the 5' cytosine of the CG base pair, and C₃ stacks on the 3' guanine. The vWF aptamer is similar in that T₂ stacks on the 5' cytosine of the CG base pair, but C₃ doesn't stack on the 3' guanine. The two cytosines (C8 and C9) in the VEGF aptamer take on conformations similar to the cruciform promoter as seen in the stacking interactions involving the pyrimidines at loop positions two and three (Figure 4.14). However, unlike these examples, the VEGF aptamer triloop has a TA closing base pair instead of the CG in the vWF aptamer and cruciform promoter. Changing the CG closing base pair to GC in vRNAP decreased the stability of the structure based on melting curves and emphasizes the importance of the sequence of the closing base pair (Chou et al. 1999). The presence of a second CG base pair in the VEGF aptamer may be a more favorable interaction since the additional amino group from the guanine would provide another hydrogen-bonding partner for the T7 in the minor groove.

Tandem base triples

The aptamer was predicted by mFold and Potty and colleagues to have a second loop in the form of a large bulge loop consisting of nucleotides 14-23 (Figure 4.12B) (Zuker 2003; Potty et al. 2009). Instead, two of these nucleotides—G18 and T19—form base triples with the major groove faces of the two closing Watson-Crick base pairs of the hairpin proximal to the triloop: C5-G11 and T6-A10 (Figures 4.12C and 4.15A, B, and C). G18 forms a reverse Hoogsteen base pair with G11, which is supported by the H1-N7 hydrogen bond detected in the ^1H - ^{15}N HNN-COSY (Section 3.3.3). Additional evidence for this triple includes NOEs between the G18 H1 and C5 amino protons, and between the G18 H1 and A10 H2'/H2'' protons (Figure 4.15D). T19 forms a reverse Hoogsteen base pair with A10, which was also detected in the HNN-COSY. NOEs that support the triple include the T19 H3 to the C9 H1', A10 H2'/H2'', A10 amino protons, and A10 H8. The triples are distant from the binding interface, which suggests that their role may be purely structural.

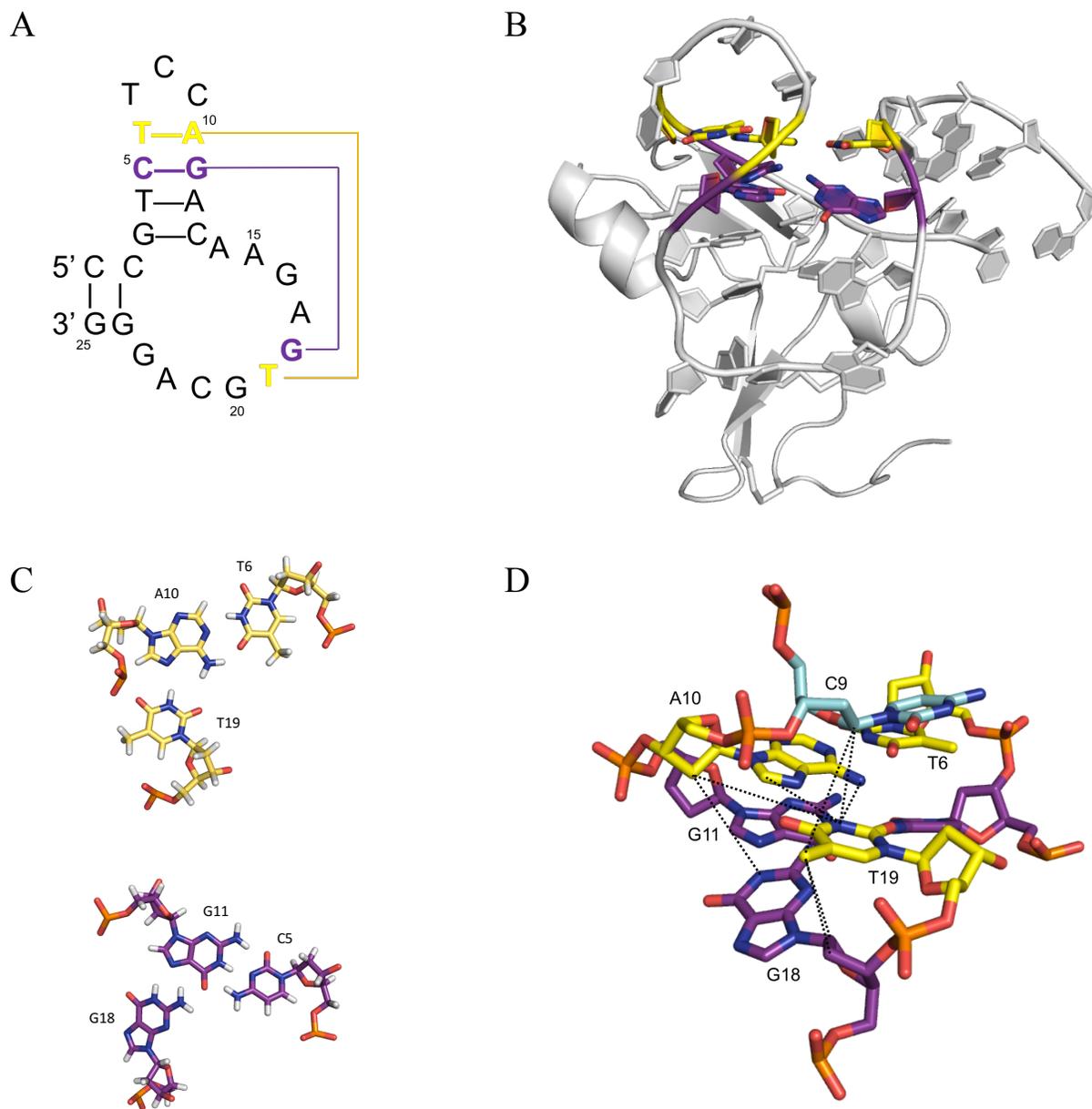


Figure 4.15 The aptamer forms two tandem major groove triplexes. The G18·G11·C5 triple is shown in purple, and the T19·A10·T6 triple is shown in yellow. (A) Secondary structure of the aptamer with the triples highlighted. (B) Representative structure of the aptamer with the triples highlighted. (C) Stick representation of the base triples. (D) NOEs that were observed and support the base triples are shown as dotted lines between the heavy atoms.

Finally, the SELEX sequences provided in the original patent describing the identification of the aptamer support the formation of the base triples. Five of the 22 sequences grouped into Family One were further analyzed to determine the minimal binding construct and the binding affinity. The sequences of these five aptamers are aligned in Figure 4.15 by predicted base pairing regions. Through this alignment, there is clear evidence for the base triples. All sequences show the equivalent of the G18·G11-C5 base triple in the aptamer (Figure 4.16, blue text). When the AT base pair proximal to the hairpin loop is mutated to GC, there are compensatory mutations from a T (T·A-T triple) to a G (second G·G-C triple) (Figure 4.16, orange text). The alignment also shows that the TCC triloop is not required. The loop can be a tetraloop or a pentaloop, but there are still three consecutive thymines. The pyrimidine-rich loop may be adopting a triloop-like structure with the additional nucleotides bulged out, or the pyrimidines may be important for interactions with the protein that are not well-defined in the current ensemble. The base pairing of the 5' and 3' ends is common and typically longer than that seen in 33t or the aptamer investigated here.

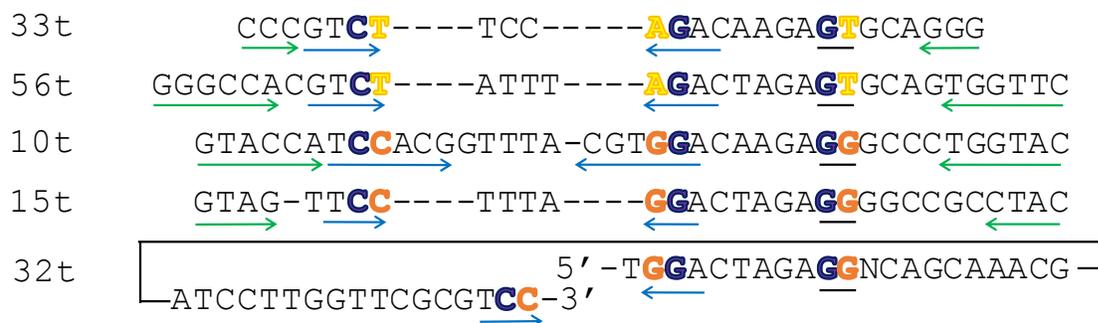


Figure 4.16 Sequence alignment of the truncated aptamers from Family 1 of the SELEX experiment used to identify the aptamer described here. Regions predicted to form base pairs are marked with an arrow. The arrows in the same color that are pointing in opposite directions form base pairs. Bases predicted to form the third base in the triplets are underlined. The bases corresponding to the G18·G11·C5 in the aptamer presented in this chapter is shown in purple text, and those corresponding to the T19·A10·T6 are in yellow text. The aptamers that have a second G·G·C triple in place of the T·A·T triple have those nucleotides displayed in orange text.

Base triples are a prevalent structural motif in both synthetic and naturally occurring nucleic acids. It has long been known that both DNA and RNA are able to form triplexes (Morgan & Wells 1968; Beal & Dervan 1991; Duca et al. 2008). Base triples form when a third base interacts with a Watson-Crick base pair by hydrogen bonding with the Hoogsteen face of the purine in the Watson-Crick pair. Two major classes of triplexes exist: R·R·Y and Y·R·Y, where R is a purine (adenine or guanine) and Y is a pyrimidine (cytosine or thymine/uracil). By convention, the first nucleotide listed forms the third strand by interaction with the Hoogsteen face of the second nucleotide listed, which is in a Watson-Crick base pair with the third nucleotide listed. Y·R·Y includes the isomorphous triples T·A·T and C⁺·G·C but the cytosine must be protonated to form the C⁺·G·C triple, so these triplexes are most stable at low pH (Gilbert & Feigon 1999). The R·R·Y triplexes are stable at physiological pH and contain nonisomorphous triples—G·G·C, A·A·T, and T·A·T—that cause distortion in the triplex compared to the Y·R·Y triplex (Gilbert & Feigon 1999; Leontis et al. 2002). The third strand in a

Y·R·Y triplex is parallel to the purine strand, and in the R·R·Y triplex it is antiparallel (Gilbert & Feigon 1999). Depending on the strand direction, T·A·T fits into either triplex class.

Base triples have been found in numerous nucleic acid structures, including structures of aptamers and physiological RNAs, and they have been shown to have both structural and functional roles. Several RNAs, such as the *Tetrahymena* ribozyme, purine riboswitches, and the HIV TAR RNA use base triples to organize ligand binding sites (Tao et al. 1997; Guo et al. 2004; Noeske et al. 2005). The HIV TAR RNA that interacts with the Tat protein has a U·A·U triple that forms upon binding Tat (Tao et al. 1997). The triple is responsible for forming the binding pocket for the arginine from Tat. Tao and colleagues showed that the uridine forming the Hoogsteen base pair in the base triple was essential for forming the binding site for Tat.

Base triples also have important structural roles. They appear in several aptamers, including the autotaxin DNA aptamer and the nerve growth factor slow off-rate modified DNA aptamer, or SOMAmer (Jarvis et al. 2015; Kato et al. 2016). The base triple in the autotaxin aptamer stabilizes a corner junction that takes on a distorted structure, which is necessary because the corner junction interacts with the autotaxin protein (Kato et al. 2016). The nerve growth factor SOMAmer takes the form of a non-helical triplex (Jarvis et al. 2015). The third strand of aptamer only plays a structural role and is not directly involved in protein recognition. The VEGF aptamer triples act similarly to those in the other aptamers and appear to play a purely structural role. They do not interact with the protein but instead provide the necessary scaffolding for the aptamer to have a high affinity interaction with the protein.

Tetraloop-like structure

The aptamer also appears to form a tetraloop-like structure, in which residues 1, 2, and 20-25 form the stem loop and residues 3-19 are an insertion in the tetraloop (Figure 4.17). The two cytosines at the 5' end (C1 and C2) form Watson-Crick base pairs with the last two guanines at the 3' end (G24 and G25); however, due to limited NOEs in this region of the aptamer, it is not well defined. The four nucleotides in the loop are G20-C21-A22-G23, which fit into the less common GNAR tetraloop family, which are typically involved in intermolecular interactions (Zwieb 1992; De Guzman 1998; Wild et al. 2001; Hainzl et al. 2002).

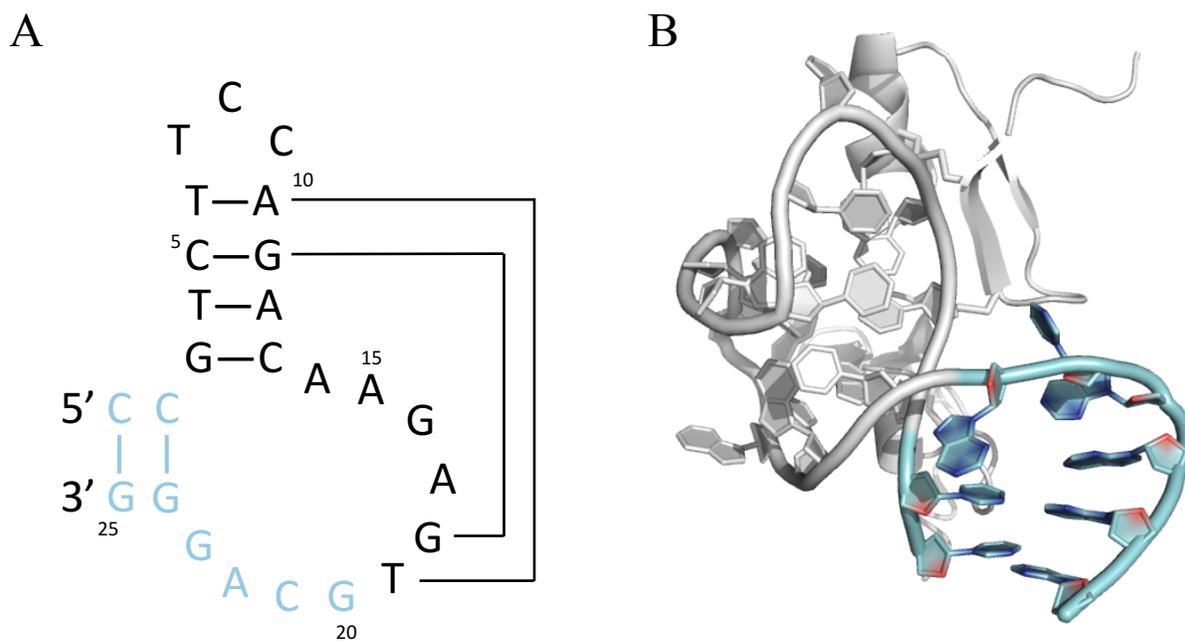


Figure 4.17 The aptamer forms a tetraloop-like structure. (A) Secondary structure of the aptamer with nucleotides forming the tetraloop-like structure highlighted cyan. (B) Representative structure of the aptamer with nucleotides forming the tetraloop-like structure highlighted in cyan.

G20 has several conformations in the ensemble and could be participating in intramolecular or intermolecular interactions. The G20 base stacks on the C2 sugar in two structures (Figure 4.18A). The G20 base has edge to face stacking with the C2 base in one structure. G20 and C2 are poised for base to base stacking, but too far apart in the structure for actual productive stacking, in two structures. One of these structures has a large enough distance between the bases of C2 and G20 (approximately 7.5 Å) that the guanidinium group of Arg124 could intercalate and participate in π - π stacking with both C2 and G20 bases, which would create a cation- π stair motif seen in other protein-aptamer complexes (Figure 4.18B) (Huang et al. 2009; Oberthür et al. 2015). A different structure has the G20 base stacking with the edge of the Arg124 guanidinium group. Another structure has G20 involved in both intramolecular and intermolecular hydrogen bonds: the G20 amino group with the G23 carbonyl, and the G20 carbonyl with Arg123 guanidinium group (Figure 4.18C). One structure is completely different from the rest and has the G20 imino proton in a hydrogen bond with the C8 amino group (Figure 4.18D).

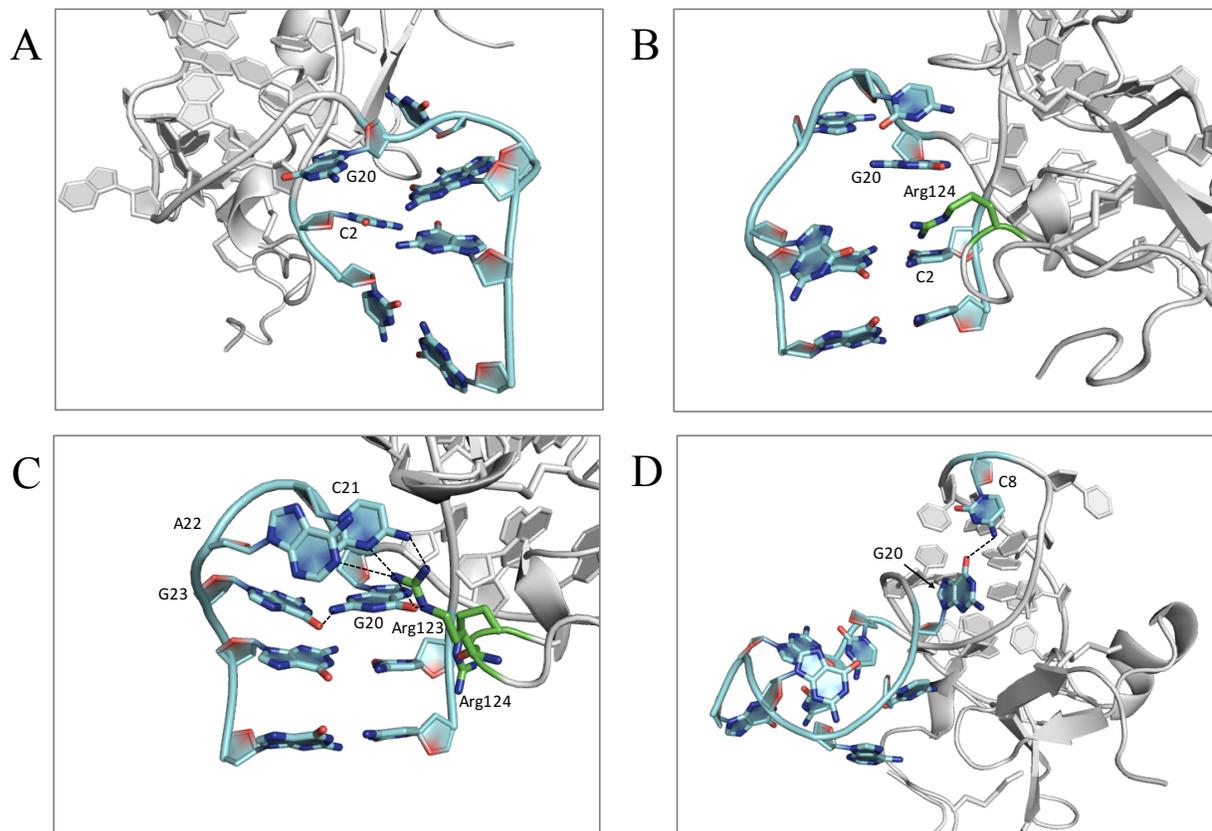


Figure 4.18 G20 may be involved in intramolecular or intermolecular interactions. G20 has several conformations in the ensemble. Some of these include (A) G20 π -stacking with C2, (B) G20 π -stacking with Arg124, (C) G20 in an intramolecular hydrogen bond (dashed line) with G23 and an intermolecular hydrogen bond with Arg124, and (D) G20 in an intramolecular hydrogen bond with C8.

The base of C21 is oriented toward the protein in all members of the ensemble. The amino group of C21 is able to make either an intramolecular hydrogen bond with the O1P of C5 (five structures) or an intermolecular dipole-dipole interaction with the Thr157 side chain hydroxyl (one structure) or hydrogen bond with the backbone carbonyl of Thr157 (three structures) (Figures 4.19A and B). Stacking interactions between C21 and A22 are seen in six of the ten structures in the ensemble. Two structures have edge-to-face stacking between the edge of C21 and the face of A22. Three structures have face-to-face stacking between the bases of C21 and A22. One structure shows base-sugar contacts between the base of A22 and the sugar of

C21. One structure has A22 in edge-to-face π -stacking with G20 and face-to-face stacking with G23. An additional interaction seen in the ensemble includes a possible dinucleotide platform with the Hoogsteen face of G23 adjacent to the sugar face of A22, but this interaction would need to be water-mediated due to the lack of hydrogen bond donors (Figure 4.19A). In two structures, A22 interacts with Arg156: one in which the edge of the A22 base participates in edge-to-face π -stacking with the guanidinium group of Arg156, and one in which the guanidinium group is able to hydrogen bond with the N1 of A22, as well as G20 O6 and C21 N3 (Figure 4.19B).

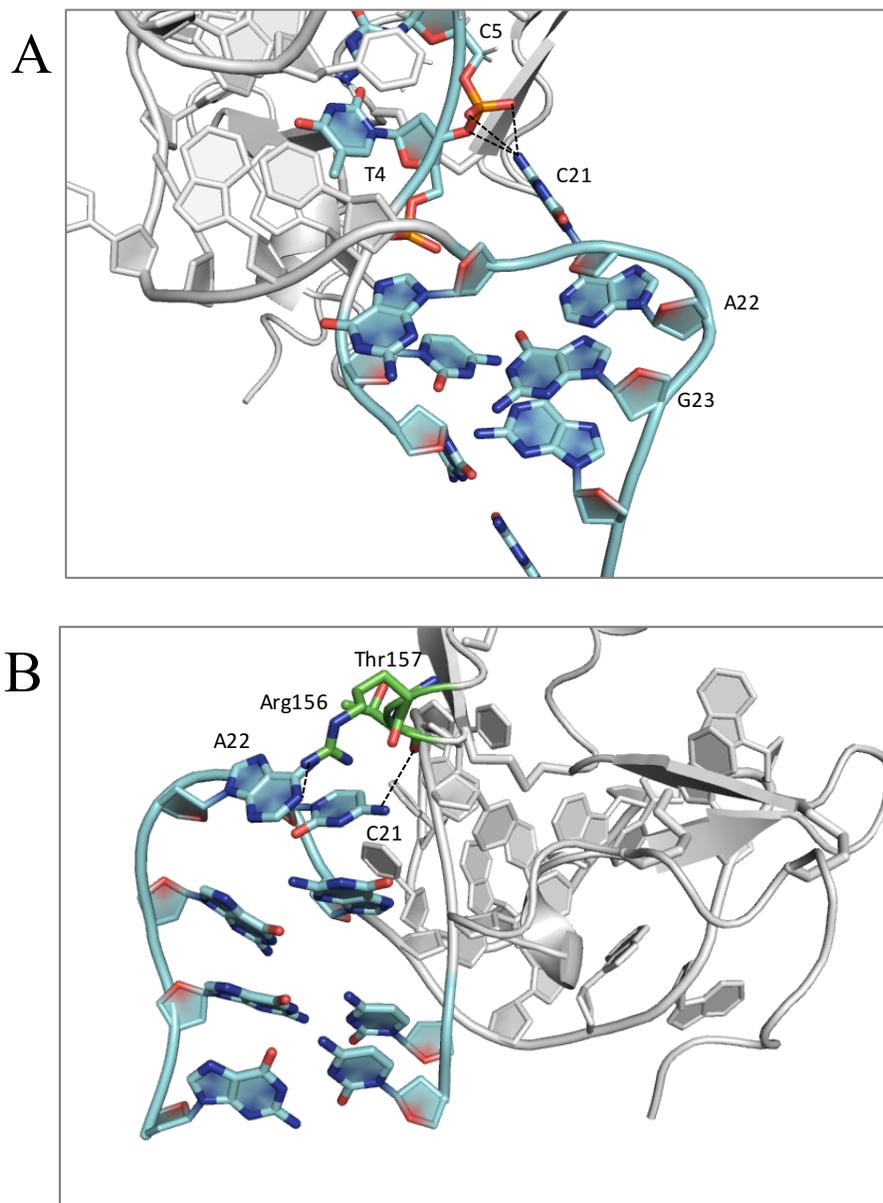


Figure 4.19 C21 and A22 may participate in intramolecular or intermolecular interactions. (A) C21 forms an intramolecular hydrogen bond with the phosphate group between T4 and C5. A22 forms a dinucleotide platform with G23 through what would likely be a water-mediated interaction. (B) The C21 amino group forms an intermolecular hydrogen bond with the backbone carbonyl of Thr157. The A22 N1 forms a hydrogen bond with the guanidinium group of Arg156.

G23 is involved in π -stacking with G24 in eight of the structures, six of which are face-to-face stacking, one is edge-to-face stacking, and one is the G23 sugar stacking with the G24 base. One structure also has a hydrogen bond between the O6 of G23 and the amino group of G20 (Figure 4.18C). Half of the structures have further π -stacking at the 5' and 3' ends between the bases of G24 and G25, as well as between C2 and C1. The other half of the structures have backbone distortions that prevent π -stacking between the terminal Watson-Crick base pairs.

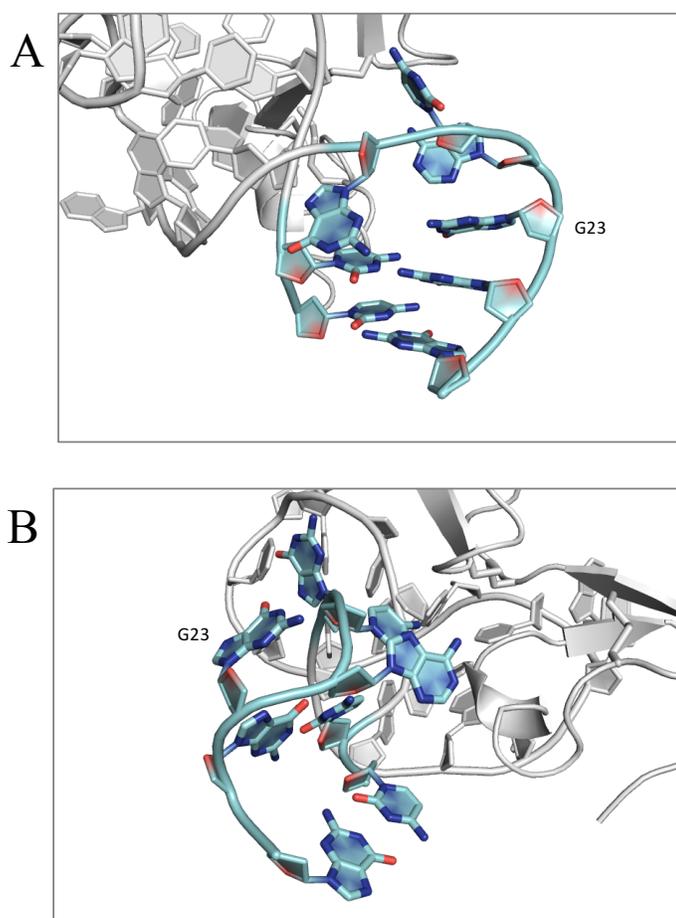


Figure 4.20 G23 is likely to engage in π -stacking with G24. (A) G23 π -stacking with G24. (B) The backbone of the aptamer takes on an unusual conformation that disrupts the stacking interaction between G23 and G24.

Tetraloop insertions have been seen previously in many ribosomal RNAs. The 23S rRNA from the halophilic archaeon *Haloarcula marismortui* has a pentaloop that forms a GNRA tetraloop-like structure with a bulged base (Ban et al. 2000). This structure was supported by comparing the *H. marismortui* stem loop to the corresponding stem loop in the 23S rRNA from the extremophilic bacterium *Deinococcus radiodurans*, which forms a tetraloop (Schlünzen et al. 2001; Leontis & Westhof 2003). Since then, the motif has appeared in rRNAs from numerous organisms, as well as the lysine riboswitch and severe acute respiratory syndrome coronavirus viral RNA (Robertson et al. 2005; Garst et al. 2008). While these tetraloop-like structures typically mimic the GNRA family of tetraloops, where N is any nucleotide and R is a purine (adenine or guanine), a UNCG tetraloop-like structure was seen in the U4 snRNA structure, which suggests that the formation of a tetraloop-like structure is not limited to GNRA tetraloops (Liu et al. 2011). If the VEGF aptamer is taking on this tetraloop-like structure with G20, C21, A22, and G23, then it would take the form of the less common GNAR tetraloop family. GNAR tetraloops have been seen in the human and archaeal signal recognition particle (SRP) ribonucleoprotein complexes and human immunodeficiency virus stem loop three Ψ RNA (Zwieb 1992; De Guzman 1998; Wild et al. 2001; Hainzl et al. 2002). One of the proteins in the SRP complex, SRP19, from the archaeon *Methanococcus jannaschii* interacts with two helices of the SRP RNA: helix six capped by a GUAG tetraloop and helix eight that has a GAAA tetraloop (Hainzl et al. 2002). The 5' guanine of the GUAG tetraloop stacks on the helical base pairs, and forms a sheared-type GG base pair with a hydrogen bond between the 5' guanine amino group and the 3' guanine N7. The 5' guanine and the looped out U both have extensive contacts with the SRP19 protein. The A stacks on the 3' guanine, and they both hydrogen bond with nucleotides in the GAAA tetraloop in helix eight. The A forms a base triple with the sheared GA

base pair of the GAAA tetraloop and the 3' guanine forms a base quadruple with the sheared 5' guanine and the closing GC base pair of the GAAA tetraloop.

The HIV GGAG tetraloop that participates in intermolecular interactions with the protein is quite different (De Guzman 1998). The 5' guanine still stacks on the helix closing base pair, but the other three nucleotides are sticking out to contact the protein and are not involved in intramolecular stacking interactions (De Guzman 1998). There is, however, a hydrogen bond between the amino group of the 3' guanine and the O6 of the 5' guanine. The nucleotides contacting the protein are mainly involved in hydrophobic interactions as well hydrogen bonds formed between the protein and the Watson-Crick faces of the bases. Due to the limited NOEs defining the tetraloop-like structure in the VEGF aptamer, it is currently unclear which GNAR tetraloop best describes the tetraloop-like structure in the VEGF aptamer, or if it has a different pattern of intramolecular and intermolecular interactions than what is seen in the HIV and SRP tetraloops. Nevertheless, characteristics of both tetraloops are seen in the aptamer ensemble.

The unpaired nucleotides A14, A15, G16, and A17 are involved in mostly intramolecular interactions, but some have interactions with the protein that are discussed in detail in Section 4.3.4. In all structures in the ensemble, A14 and A15 are in a face-to-face π -stacking interaction. In six structures, this stacking interaction continues to C13 sugar, which packs against the A14 base. A15 interacts with amino acids in the N-terminal β -sheet in the HBD. G16 is not well defined in the ensemble due to limited NOEs. Five of the structures have G16 flipped out away from the rest of the complex, where it appears to have no interactions with the protein or other nucleotides in the aptamer. However, the SELEX data suggest that this nucleotide is important in the aptamer because it is invariant among the sequences, so it is likely involved in an important intermolecular or intramolecular interaction. In one structure, G16 is flipped in toward the N-

terminal subdomain of the HBD where it could interact with nearby Lys125. Four of the structures have G16 stacking with A17 but still not interacting with the protein (Figure 4.21). A17 stacks with the G3-C13 base pair to close the base of the hairpin where it also may be able to interact with the HBD.

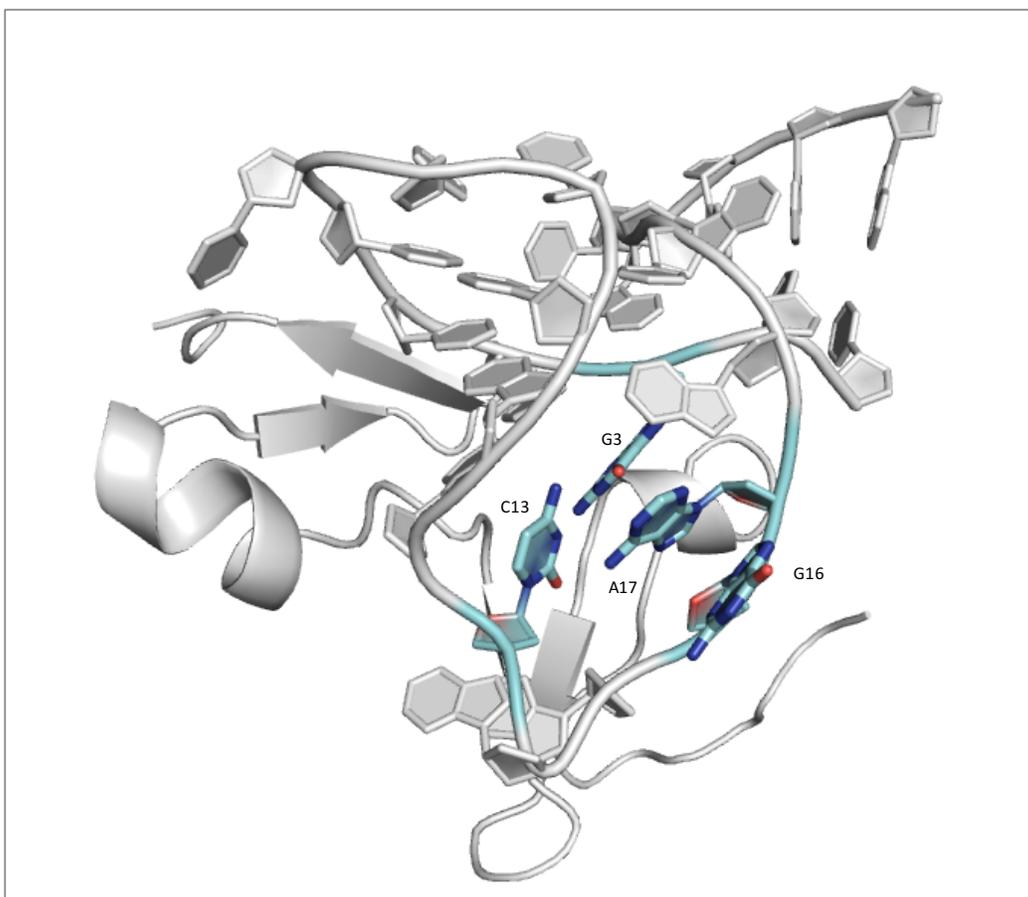


Figure 4.21 Unpaired nucleotides A17 and G16 contribute to the π -stacking interactions in the hairpin.

4.3.4 Binding interface of the HBD-aptamer complex

The HBD-aptamer interface includes 877 Å² of buried solvent-accessible surface area calculated using Pymol and Equation 4.4 (Gelinas et al. 2016):

Equation 4.4

$$SASA_{buried} = \frac{(SASA_{HBD} + SASA_{aptamer}) - SASA_{complex}}{2}$$

where SASA is the solvent-accessible surface area. The two base triples do not contact the protein. Instead, they act as scaffold to properly orient the aptamer on the protein. The aptamer contacts both subdomains of the protein in a way that alters the orientation of the two subdomains relative to their linear orientation in the structure of the free HBD (Section 4.3.1). The base of the hairpin sits on the protein surface, and several paired and unpaired nucleotides are involved in forming the binding interface (Figure 4.22). The interface includes several hydrogen bonds, van der Waals interactions, and potential salt bridges.

The van der Waals interactions appear to the interface, but that may be the result of some of the NOEs biasing these amino acids. The methyl groups on the side chains of valine and leucine have strong signal in the NMR spectra due to three degenerate methyl protons contributing to the signal and the fast rotation around the carbon-carbon bond leading to slow relaxation and therefore sharper signal (Nicholson et al. 1992; Tugarinov & Kay 2005). The charge-charge and hydrogen bonding interactions with exchangeable protons are typically difficult to detect in the NMR spectra, especially if these protons are also exchanging with the solvent. Therefore, even though the highly positive HBD and negatively charged aptamer are forming a complex, the details of the electrostatic interactions and many hydrogen bonds are absent in the NMR data.

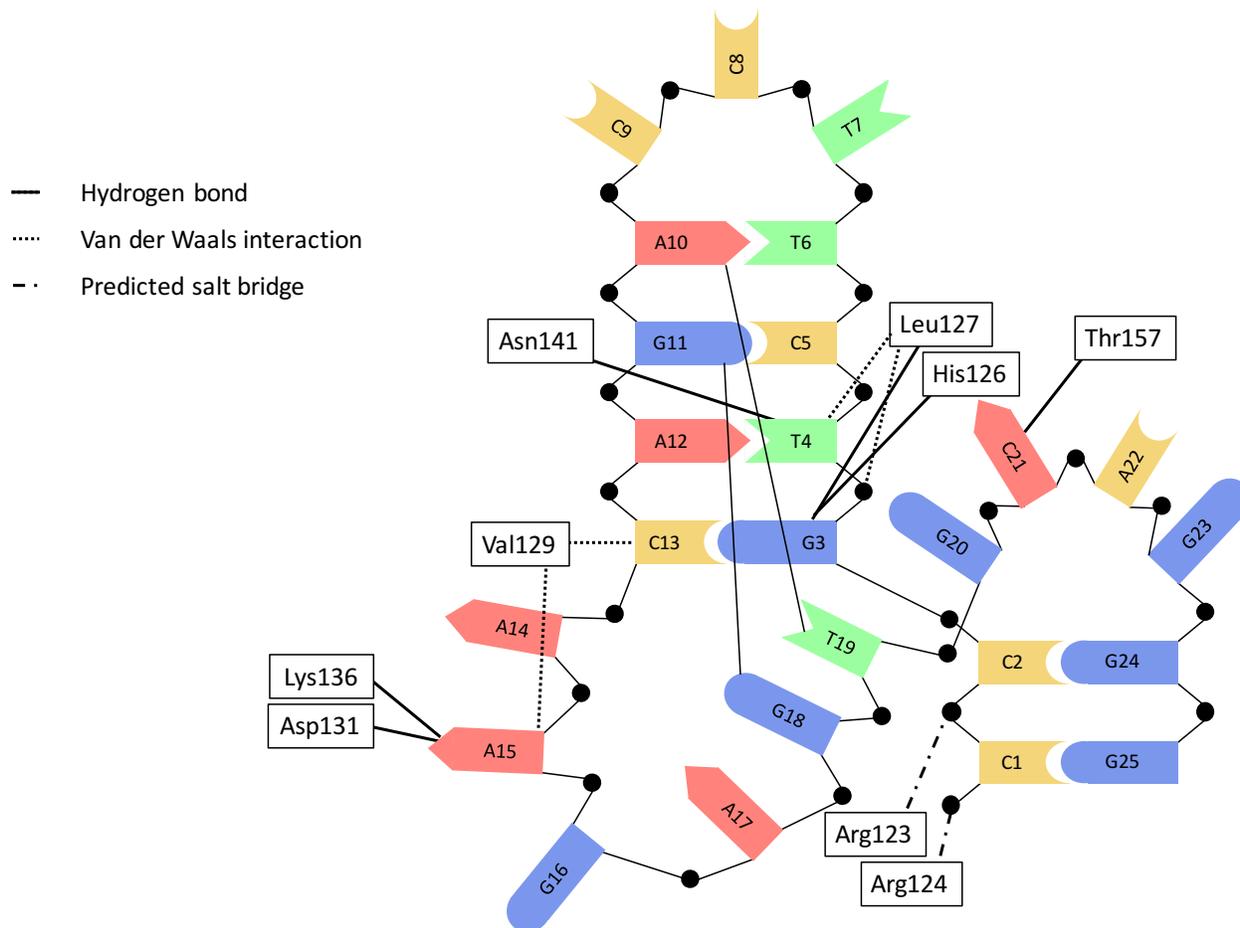


Figure 4.22 Summary of the intermolecular interaction seen in more than 50% of the structures in the ensemble. Hydrogen bonds are marked by the solid black line. The dotted line specifies van der Waals interactions, and the line that alternates dashes and dots shows predicted salt bridges.

Interactions between the HBD and the aptamer hairpin

The minor groove faces of the Watson-Crick G3-C13 base pair sit on the N-terminal subdomain of the HBD. The deoxyribose of G3 has van der Waals contacts with the Leu127 side chain in all of the structures in the ensemble (Figure 4.23). The amino group of G3 is proximal to both the His126 and Leu127 backbone carbonyl oxygens. In six of the ten structures, the G3 amino group is well positioned to hydrogen bond with both oxygens. In two structures, it is only close enough to hydrogen bond with Leu127 carbonyl oxygen, and in the remaining two structures, G3 is oriented such that the amino group is outside of hydrogen bonding range. The C13 carbonyl oxygen is involved in van der Waals interactions with one of the methyl groups of Val129. Additional interactions are seen in four structures between the Lys140 amino group and the O4' of C13 or between Arg145 guanidinium group and C13 O1P in three structures.

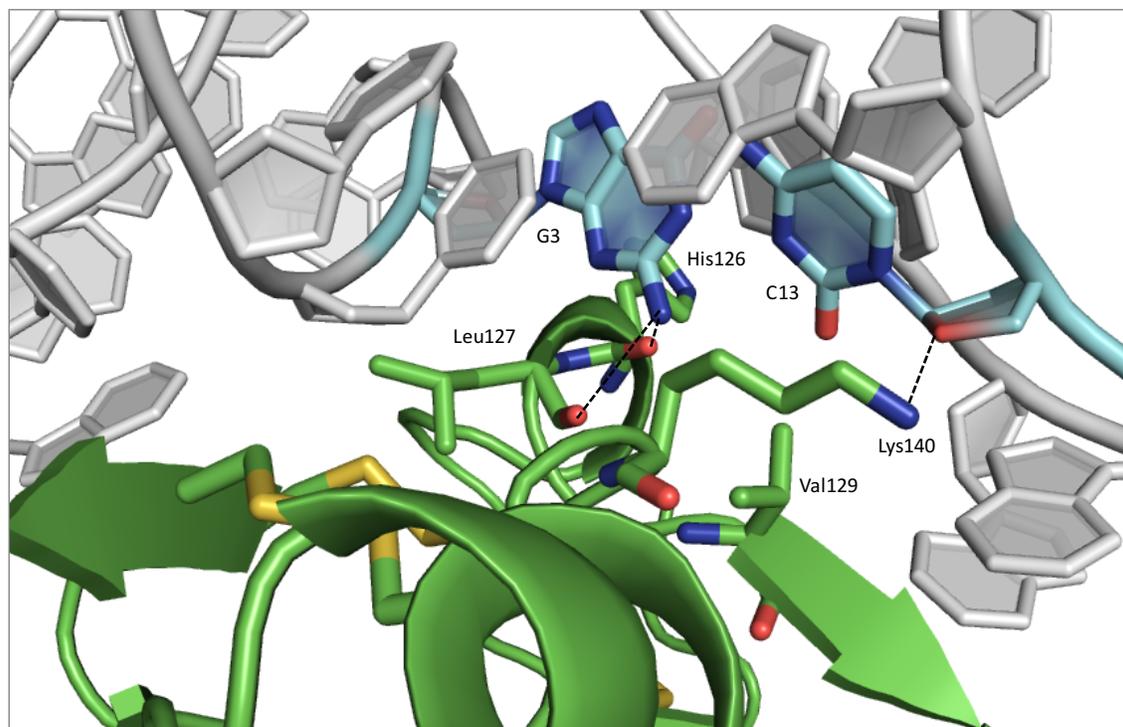


Figure 4.23 Intermolecular interactions between the G3-C13 base pair (cyan) and the amino acids His126, Leu127, Val129, and Lys140 (green).

The T4-A12 base pair also interacts with the protein, and to test the importance of this interaction, fluorescence anisotropy was used to measure the binding affinities between the HBD and two mutants. The T4C mutant disrupts the Watson-Crick base pair, and the compensatory mutant T4C/A12G restores the base pair. Initial results suggest that preservation of the Watson-Crick base pair between T4 and A12 is necessary for the aptamer to bind the HBD. The T4C mutant showed no detectable binding by fluorescence anisotropy (Figure 4.23). When the compensatory mutation was made (T4C/A12G), the binding was restored and had a tighter binding affinity ($K_D = 4.5$ nM) than wild type does for the HBD ($K_D = 20$ nM). The minor groove faces of cytosine and thymine are identical, so the intermolecular contacts to the nucleotide at position 4 were not disrupted. Instead, this suggests that the preservation of this base pair within the hairpin is important for the aptamer binding and may be stabilizing the structures of the aptamer and the binding interface.

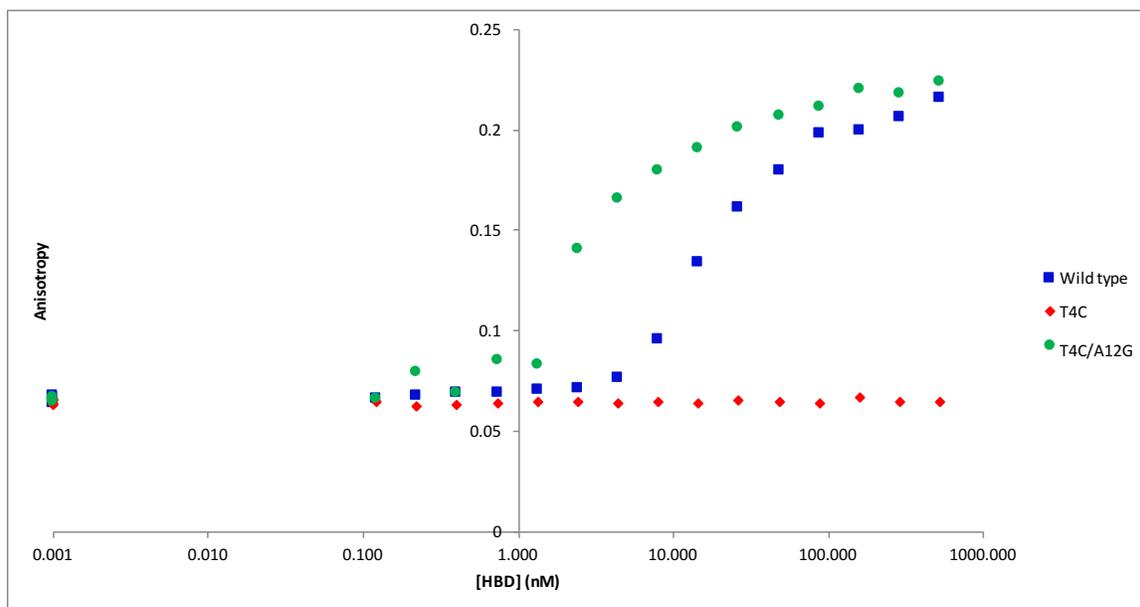


Figure 4.24 Preliminary binding of the wild-type aptamer and two mutants: T4C and T4C/A12G. The plot shows fluorescence anisotropy versus protein concentration.

The T4-A12 base pair interacts with the HBD C-terminal subdomain and the interface of the two subdomains. The side chain of Leu127 and the methylene protons of Cys158 have van der Waals contacts with the deoxyribose of T4. Additional interactions between the HBD and the hairpin include hydrogen bonding interactions. The amino group of Asn141 is hydrogen bonded to the O2 of T4 in seven of the ten structures (Figure 4.25). Two structures have the Asn141 amino group hydrogen bonded to Cys158 carbonyl oxygen instead of interacting with the aptamer. The HBD contacts the base or deoxyribose of C5 in five structures. Four structures have the Asn141 carboxamide protons close enough to hydrogen bond with the O4' of the C5 deoxyribose. One of these structures has an additional interaction with Lys140 close enough to hydrogen bond with C5 O2 while Asn141 hydrogen bonds to the C5 O4'. One other structure in the ensemble has the Asn141 carboxamide protons close enough to form a hydrogen bond with C5 O2. There are no NOEs supporting an interaction between Asn141 and C5, but such an interaction would suggest that one of the base triples (G18·G11-C5) has a role beyond forming a scaffold and could have a role in recognizing the protein.

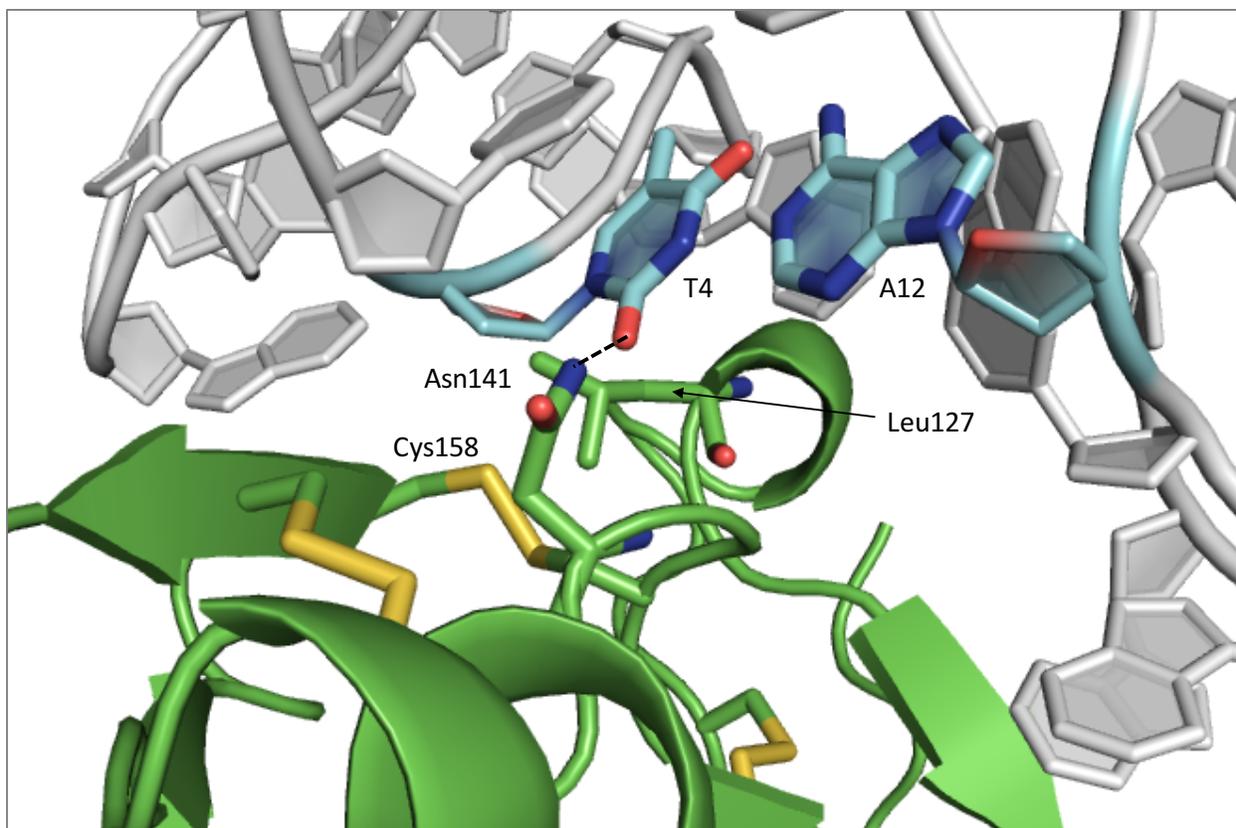


Figure 4.25 Intermolecular interactions between the T4-A12 base pair (cyan) and the amino acids Leu127, Asn141, and Cys158 (green). A hydrogen bonds between Asn141 and T4 is identified with a black dashed line. Leu127 and Cys158 are have van der Waals contacts with the deoxyribose of T4.

Interactions between the HBD and unpaired nucleotides

Two unpaired nucleotides contribute to the binding interface: A15, which forms a π -stacking interaction with the base of A14, and A17, which sits at the base of the hairpin and forms a π -stacking interaction with the G3-C13 base pair. A15 is positioned against one of the strands of the antiparallel β -sheet in the N-terminal subdomain of the HBD (Figure 4.26). It has two different conformations—one in which the major groove face points out toward the solvent, and one where the major groove face points in toward to the HBD. When the major groove face is pointed out, which is does in six of the ten structures, the amino group is able to hydrogen

bond with the carboxylate group of Asp131. Three of those structures, as well as three additional structures, have the amino group of Lys136 within, or just outside, typical hydrogen bonding distance, which suggests that it could make a hydrogen bond with the N1 of A15 with optimal side chain orientation. Additionally, all structures have van der Waals contact between the side chain of Val129 and A15: three with the base, five with the deoxyribose, and two with both. The intermolecular interactions between the HBD and A15 are extensive and suggest that this nucleotide may be very important for the integrity of the interface. Since both the N1 and exocyclic amino group of the Watson-Crick face of A15 are recognized by the protein, a mutation from adenine to guanine would be expected to negatively impact binding.

In four of the ten structures, His126 HE2 interacts with A17 through either the O5' or O1P. His126 has numerous possible orientations in the structure due to the absence of NOEs. Supplemental unlabeled histidine was added to the *P. pastoris* medium during protein expression, so His126 is invisible in the NMR spectra. Additional NMR spectra could be acquired using a sample of HBD expressed with supplemented ^{13}C , ^{15}N -labeled histidine to assign His126 and identify intramolecular and intermolecular NOEs, however if the intermolecular interactions with His126 are limited to hydrogen bonding with the phosphate bound oxygens, then additional NOESY spectra would have limited use due to the low number of protons in the sugar-phosphate backbone.

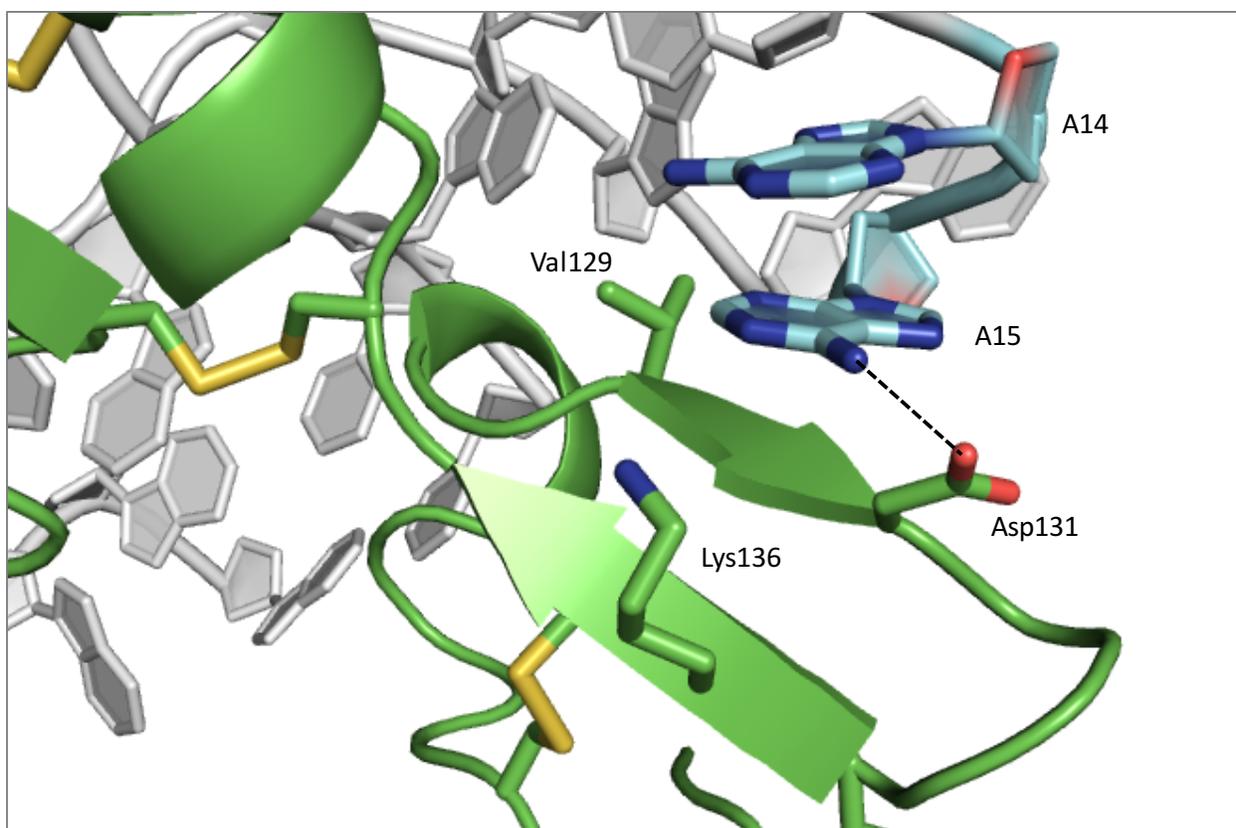


Figure 4.26 Intermolecular interactions between A15 (cyan) and Val129, Asp131, and Lys136 (green). Potential sites of hydrogen bonds—between Lys136 and A15 N3 and between Asp131 and A15 amino group—are identified with a black dashed line.

Interactions between the HBD and the tetraloop-like structure

The tetraloop-like structure formed by the base pairs connecting the 5' and 3' ends (C1-G25 and C2-G24) and four unpaired nucleotides—G20, C21, A22, and G23—was described earlier as having numerous orientations in the solution structure. In fact, only three nucleotides—C1, C2 and C21—have NOEs to the HBD. The remaining nucleotides of the loop have the potential to interact with the protein, and they each do so in at least one structure in the ensemble. The two cytosines at the 5' end of the aptamer contact Arg123 and Arg124 in the HBD N-terminal subdomain (Figure 4.27). The NOEs between these residues are limited, so the arginines and cytosines all have numerous conformations in the ensemble. However, due to the

proximity of the charges, it is likely that Arg123 and Arg124 form salt bridges with the phosphate backbone of C1 and C2 at the 5' end of the aptamer.

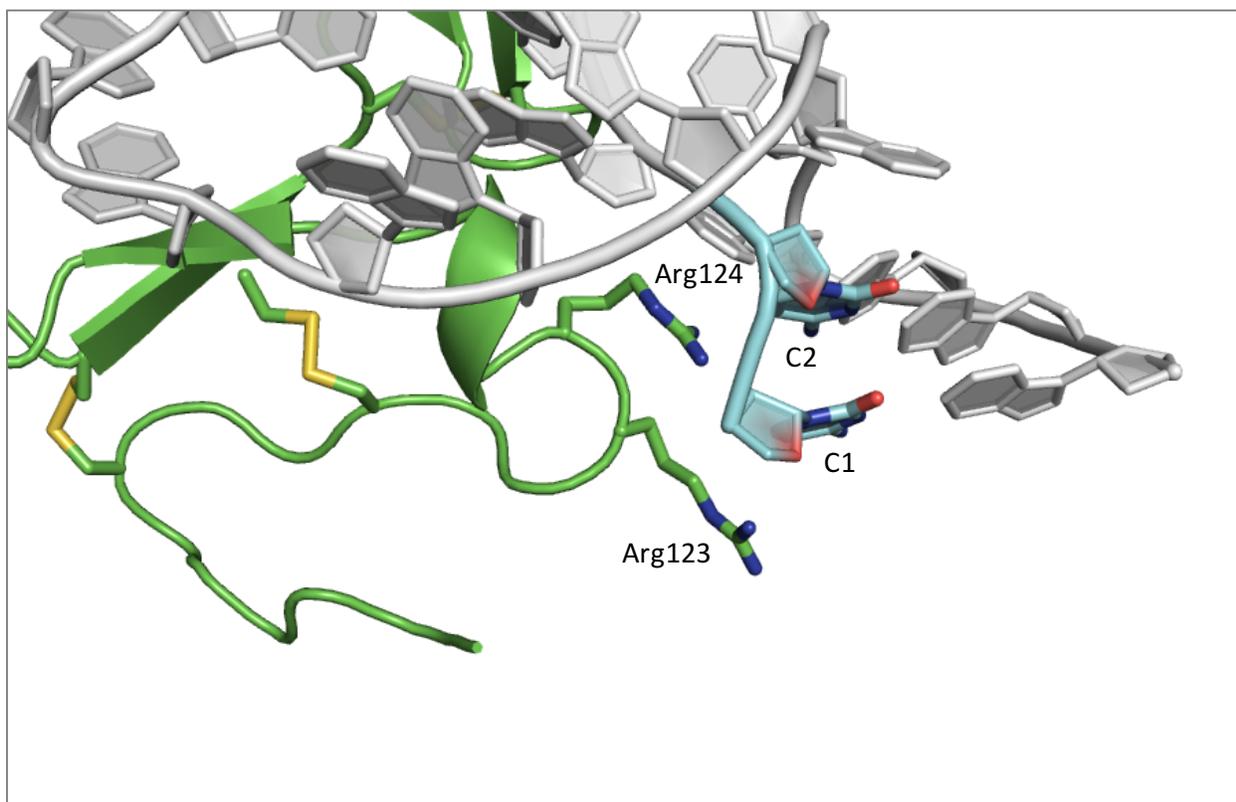


Figure 4.27 Arg123 and Arg124 are proximal to the 5' end of the aptamer. The two arginines (green) may be involved in salt bridges or π -stacking interactions with C1 and C2 (cyan).

C21 contributes to the binding interface by interacting with the loop connecting the two antiparallel strands of the β -sheet in the C-terminal subdomain (Figure 4.28). Nine of the structures show van der Waals contacts between C21 and Thr157. Three of the ten structures have a hydrogen bond between the amino group and the backbone carbonyl oxygen of Thr157, and in one of those structures also with the carbonyl of Arg156. Four of the ten structures have C21 in an intramolecular instead of an intermolecular interaction with a hydrogen bond between

the amino group and T4 O3' or C5 O1P. In nearly all of the structures in the ensemble, the C21 Watson-Crick face points away from the protein. While this could be an accurate description of what is happening in solution, this could also be an artifact because the C21 amino protons are not assigned, and there are no other protons on the Watson-Crick face available for NOEs.

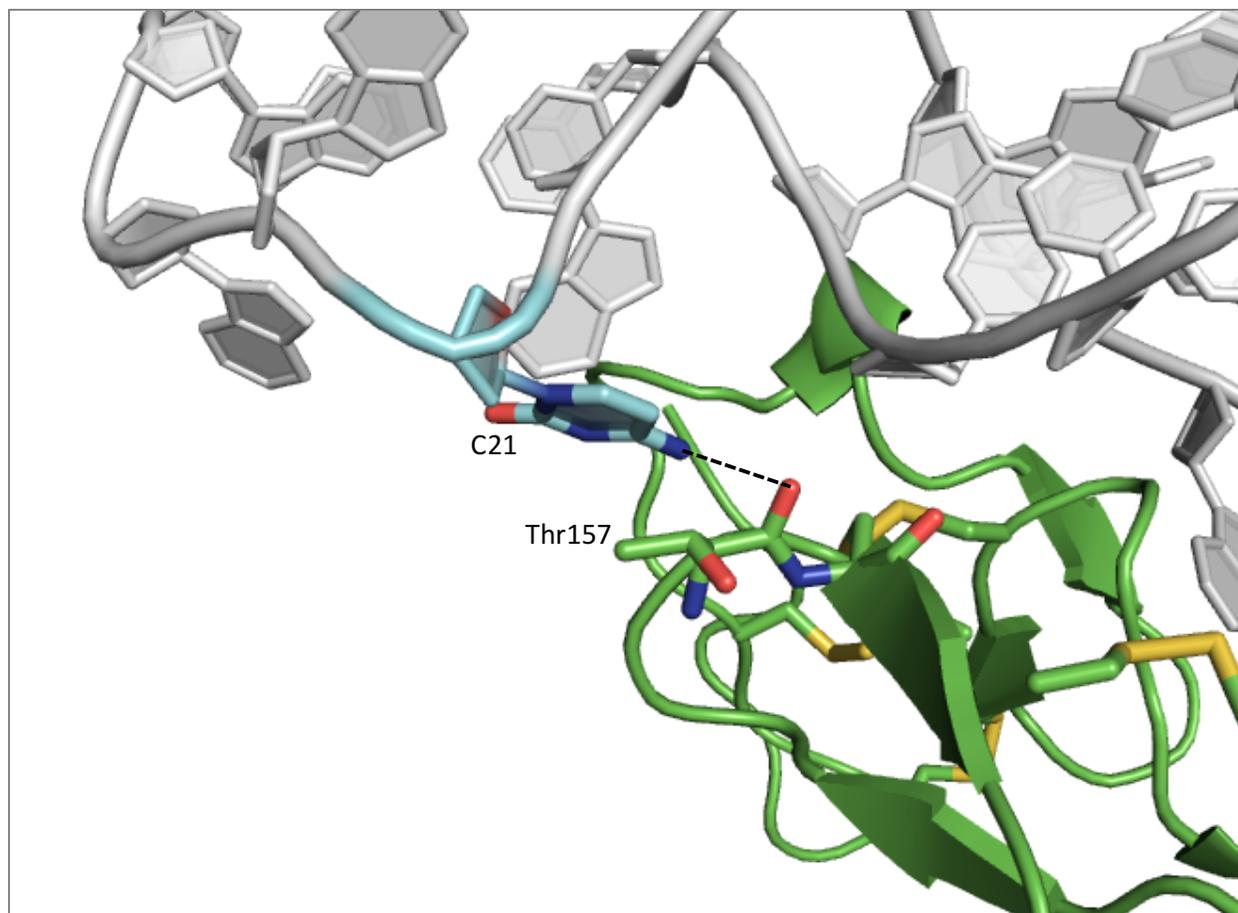


Figure 4.28 Hydrogen bond between C21 (cyan) and Thr157 (green). Potential hydrogen bond between the C21 amino group and the Thr157 backbone carbonyl oxygen is shown with a black dashed line.

4.4 Future directions

The solution structure of the heparin-binding domain of VEGF with a DNA aptamer was presented in this chapter. Additional biochemical experiments are in progress to better understand the interactions that are occurring in the complex. First among these are mutations to disrupt the base triples and compensatory mutations to restore them. Additional mutations include those nucleotides that interact with the HBD and those that may be important in π -stacking interactions. The salt dependence of the equilibrium dissociation constant will also be determined, as this will also reveal the number of salt bridges formed in the complex. Several regions of the complex are not well defined because of too few NOEs. Additional NMR experiments may increase the number of intermolecular and intramolecular aptamer NOEs and may help better define the aptamer. The nucleotides particularly in need of more NOEs include G16 and those in the tetraloop-like structure (nucleotides 20-23).

The aptamer studied here is only one of several aptamers that have been identified that target the HBD of VEGF (Jellinek et al. 1994; Gold & Janjic 1998). One question that remains is how these different aptamers target the same small domain with high affinity and specificity. Additional high-resolution structures of these different RNA and DNA aptamers would provide insight into what is likely multiple distinct mechanisms of binding. It would be interesting to see if any of the other aptamers also form base triples when binding the HBD or if they contact a larger percentage of the surface area of the HBD. The aptamer presented here binds only one side of the protein, so other larger aptamers may contact both sides of the protein, or they may also interact with the receptor-binding domain of VEGF. While the number of publications investigating the structures of protein-aptamer complexes is growing, there is yet to be a comparison between two aptamers that contact the same protein surface. Thrombin is currently

the only protein that has crystal structures that been solved with multiple aptamers, but the two aptamers are non-competitive and bind different sites on thrombin (Pica et al. 2016). The aptamers that target the HBD, on the other hand, are competitive with one another, which would make their structures particularly interesting to compare and provide further insight about the various mechanisms aptamers use to target specific proteins.

References

- Amadio, M., Govoni, S. & Pascale, A., 2016. Targeting VEGF in eye neovascularization: What's new?: A comprehensive review on current therapies and oligonucleotide-based interventions under development. *Pharmacological Research*, 103, pp.253–269.
- Arcondéguy, T. et al., 2013. VEGF-A mRNA processing, stability and translation: A paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucleic Acids Research*, 41(17), pp.7997–8010.
- Ashikari-Hada, S. et al., 2005. Heparin regulates vascular endothelial growth factor165-dependent mitogenic activity, tube formation, and its receptor phosphorylation of human endothelial cells. Comparison of the effects of heparin and modified heparins. *The Journal of biological chemistry*, 280(36), pp.31508–15.
- Ban, N. et al., 2000. The complete atomic structure of the large ribosomal subunit at 2. *Science*, 289(5481), pp.905–920.
- Bao, P. et al., 2009. The Role of Vascular Endothelial Growth Factor in Wound Healing. *Journal of Surgical Research*, 153(2), pp.347–358.
- Barna, J.C. et al., 1987. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of Magnetic Resonance (1969)*, 73(1), pp.69–77.
- Batey, R.T. et al., 1992. Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Research*, 20(17), pp.4515–4523.
- Batey, R.T., Rambo, R.P. & Doudna, J.A., 1999. Tertiary motifs in RNA structure and folding. *Angewandte Chemie - International Edition*, 38(16), pp.2326–2343.
- Bax, A., Ikura, M., et al., 1990. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *Journal of Magnetic Resonance (1969)*, 86(2), pp.304–318.
- Bax, A., Clore, G.M. & Gronenborn, A.M., 1990. ¹H¹H correlation via isotropic mixing of ¹³C magnetization, a new three-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *Journal of Magnetic Resonance (1969)*, 88(2), pp.425–431.
- Beal, P. a & Dervan, P.B., 1991. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science (New York, N.Y.)*, 251(4999), pp.1360–1363.
- Berjanskii, M. V. & Wishart, D.S., 2005. A simple method to predict protein flexibility using secondary chemical shifts. *Journal of the American Chemical Society*, 127(43), pp.14970–14971.

- Bermejo, G.A., Clore, G.M. & Schwieters, C.D., 2016. Improving NMR Structures of RNA. *Structure*, 24(5), pp.806–815.
- Bermejo, G.A., Clore, G.M. & Schwieters, C.D., 2012. Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures. *Protein Science*, 21(12), pp.1824–1836.
- Bodenhausen, G. & Ruben, D.J., 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1), pp.185–189.
- Brandner, B. et al., 2006. Investigating the effect of VEGF glycosylation on glycosaminoglycan binding and protein unfolding. *Biochemical and Biophysical Research Communications*, 340(3), pp.836–839.
- Brown, L.F. et al., 1992. Expression of Vascular Permeability Factor (Vascular Endothelial Growth Factor) by Epidermal Keratinocytes during Wound Healing. *The Journal of Experimental Medicine*, 176, pp.1375–1379.
- Brozzo, M.S. et al., 2011. Thermodynamic and structural description of allosterically regulated VEGF receptor 2 dimerization. *Blood*, 119(7), pp.1781–1788.
- Cao, Y. et al., 2012. VEGF exerts an angiogenesis-independent function in cancer cells to promote their malignant progression. *Cancer Research*, 72(16), pp.3912–3918.
- Carmeliet, P. et al., 1996. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature*, 380(6573), pp.435–439.
- Carmeliet, P., 2005. Angiogenesis in life, disease and medicine. *Nature*, 438 (December), pp.932–936.
- Cavanagh, J. et al., 2007. *Protein NMR spectroscopy. Principles and Practice*. Academic Press.
- Chen, V.B. et al., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp.12–21.
- Cheung, Y.-W. et al., 2013. Structural basis for discriminatory recognition of Plasmodium lactate dehydrogenase by a DNA aptamer. *Proceedings of the National Academy of Sciences of the United States of America*, 110(40), pp.15967–72.
- Choi, S. & Ban, C., 2016. Crystal structure of a DNA aptamer bound to Pv LDH elucidates novel single-stranded DNA structural elements for folding and recognition. *Nature Publishing Group*, (September), pp.1–13.
- Chou, S.-H., Tseng, Y.-Y. & Chu, B.-Y., 1999. Stable formation of a pyrimidine-rich loop hairpin in a cruciform promoter. *Journal of Molecular Biology*, 292(2), pp.309–320.

- Christinger, H.W. et al., 1996. Crystallization of the receptor binding domain of vascular endothelial growth factor. *Proteins*, 26(3), pp.353–357.
- Claffey, K.P., Senger, D.R. & Spiegelman, B.M., 1995. Structural requirements for dimerization, glycosylation, secretion, and biological function of VPF/VEGF. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1246(1), pp.1–9.
- Convery, M.A. et al., 1998. Crystal structure of an RNA aptamer protein complex at 2.8 angstrom resolution. *Nat Struct Biol*, 5(2), pp.133–139.
- Cornilescu, G., Delaglio, F. & Bax, A., 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, 13(3), pp.289–302.
- Coultas, L., Chawengsaksophak, K. & Rossant, J., 2005. Endothelial cells and VEGF in vascular development. *Nature*, 438(7070), pp.937–945.
- Cregg, J.M. et al., 1989. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and cellular biology*, 9(3), pp.1316–23.
- Cregg, J.M. et al., 1987. High-level expression and efficient assembly of hepatitis B surface antigen in the methylotrophic yeast, *Pichia pastoris*. *Nature Biotechnology*, 5, pp.479–485.
- Cregg, J.M. & Higgins, D.R., 1995. Production of foreign proteins in the yeast *Pichia pastoris*. *Canadian Journal of Botany*, 73(S1), pp.891–897.
- Darmostuk, M. et al., 2014. Current approaches in SELEX: An update to aptamer selection technology. *Biotechnology Advances*, 33(6), pp.1141–1161.
- Davies, D.R. et al., 2012. Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proceedings of the National Academy of Sciences of the United States of America*, 109(49), pp.19971–6.
- Davis, I.W. et al., 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35, pp.375–383.
- Davis, I.W. et al., 2004. MolProbity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research*, 32, pp.615–619.
- Davlieva, M. et al., 2014. Structure analysis of free and bound states of an RNA aptamer against ribosomal protein S8 from *Bacillus anthracis*. *Nucleic Acids Research*, 42(16), pp.10795–10808.
- Delaglio, F. et al., 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), pp.277–293.

- Delaglio, F., Torchia, D.A. & Bax, A., 1991. Measurement of ^{15}N - ^{13}C J couplings in staphylococcal nuclease. *J Biomolec NMR*, 1, pp.439–446.
- Dingley, A.J. & Grzesiek, S., 1998. Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide 2JNN Couplings. *Journal of the American Chemical Society*, 120(33), pp.8293–8297.
- Drolet, D.W. et al., 2016. Fit for the Eye: Aptamers in Ocular Disorders. *Nucleic Acid Therapeutics*, 0(0), p.nat.2015.0573.
- Duca, M. et al., 2008. The triple helix: 50 years later, the outcome. *Nucleic Acids Research*, 36(16), pp.5123–5138.
- Eichmüller, C. et al., 2001. Simultaneous measurement of intra- and intermolecular NOEs in differentially labeled protein-ligand complexes. *Journal of Biomolecular NMR*, 21(2), pp.107–116.
- Ellington, a D. & Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818–22.
- Engh, R.A. & Huber, R., 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A*, 47(4), pp.392–400.
- Evans, J.N.S., 1995. *Biomolecular NMR Spectroscopy*, Oxford: Oxford University Press.
- Fairbrother, W.J. et al., 1997. ^1H , ^{13}C , and ^{15}N backbone assignment and secondary structure of the receptor-binding domain of vascular endothelial growth factor. *Protein science : a publication of the Protein Society*, 6(10), pp.2250–60.
- Fairbrother, W.J. et al., 1998. Solution structure of the heparin-binding domain of vascular endothelial growth factor. *Structure*, 6(5), pp.637–648.
- Ferrara, N. et al., 1996. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature*, 380(6573), pp.439–442.
- Ferrara, N., Gerber, H.P. & LeCouter, J., 2003. The biology of VEGF and its receptors. *Nat Med*, 9(6), pp.669–676. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12778165>.
- Ferrara, N. & Henzel, W.J., 1989. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochemical and Biophysical Research Communications*, 161(2), pp.851–858.
- Fürtig, B. et al., 2003. NMR spectroscopy of RNA. *ChemBioChem*, 4(10), pp.936–962.

- Garst, A.D. et al., 2008. Crystal structure of the lysine riboswitch regulatory mRNA element. *Journal of Biological Chemistry*, 283(33), pp.22347–22351.
- Gelinas, A.D. et al., 2014. Crystal structure of interleukin-6 in complex with a modified nucleic acid ligand. *Journal of Biological Chemistry*, 289(12), pp.8720–8734.
- Gelinas, A.D., Davies, D.R. & Janjic, N., 2016. Embracing proteins: Structural themes in aptamer-protein complexes. *Current Opinion in Structural Biology*, 36, pp.122–132.
- Gilbert, D.E. & Feigon, J., 1999. Multistranded DNA structures Dara E Gilbert and Juli Feigon*. *Current opinion in structural biology*, 9, pp.305–314.
- Goel, H.L. & Mercurio, A.M., 2013. VEGF targets the tumour cell. *Nature reviews. Cancer*, 13(12), pp.871–82.
- Gold, L. & Janjic, N., 1998. High-affinity oligonucleotide ligands to vascular endothelial growth factor (VEGF). Available at: <https://www.google.com/patents/US5811533>.
- Gospodarowicz, D., Abraham, J. a & Schilling, J., 1989. Isolation and characterization of a vascular endothelial cell mitogen produced by pituitary-derived folliculo stellate cells. *Proceedings of the National Academy of Sciences of the United States of America*, 86(19), pp.7311–7315.
- Grishaev, A. & Bax, A., 2004. An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *Journal of the American Chemical Society*, 126(23), pp.7281–7292.
- Grünewald, F.S. et al., 2010. Structure-function analysis of VEGF receptor activation and the role of coreceptors in angiogenic signaling. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(3), pp.567–580.
- Grzesiek, S., Anglister, J. & Bax, A., 1993. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in ¹³C/¹⁵N-Enriched Proteins by Isotropic Mixing of ¹³C Magnetization. *Journal of Magnetic Resonance, Series B*, 101(1), pp.114–119.
- Grzesiek, S. & Bax, A., 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society*, 114(16), pp.6291–6293.
- Güntert, P., 2004. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)*, 278(1980), pp.353–378.
- Güntert, P. & Buchner, L., 2015. Combined automated NOE assignment and structure calculation with CYANA. *Journal of Biomolecular NMR*, 62(4), pp.453–471.

- Guo, F., Gooding, A.R. & Cech, T.R., 2004. Structure of the Tetrahymena Ribozyme. *Molecular Cell*, 16(3), pp.351–362.
- Gustafsson, T., 2011. Vascular remodelling in human skeletal muscle. *Biochemical Society Transactions*, 39(6), pp.1628–1632.
- De Guzman, R.N., 1998. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 -RNA Recognition Element. *Science*, 279(5349), pp.384–388.
- Hainzl, T., Huang, S. & Sauer-Eriksson, A.E., 2002. Structure of the SRP19 RNA complex and implications for signal recognition particle assembly. *Nature*, 417(6890), pp.767–771.
- Houck, K.A. et al., 1992. Dual Regulation of Vascular Endothelial Growth Factor Bioavailability by Genetic and Proteolytic Mechanisms. , 267(36), pp.26031–26037.
- Huang, D.-B. et al., 2003. Crystal structure of NF-kappaB (p50)₂ complexed to a high-affinity RNA aptamer. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16), pp.9268–73.
- Huang, R.H. et al., 2009. A Structural Explanation for the Antithrombotic Activity of ARC1172, a DNA Aptamer that Binds von Willebrand Factor Domain A1. *Structure*, 17(11), pp.1476–1484.
- Hyberts, S.G. et al., 2014. Perspectives in magnetic resonance: NMR in the post-FFT era. *Journal of Magnetic Resonance*, 241(1), pp.60–73.
- Ikura, M., Kay, L.E. & Bax, a, 1990. A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, 29(19), pp.4659–4667.
- Imoukhuede, P.I. & Popel, A.S., 2011. Quantification and cell-to-cell variation of vascular endothelial growth factor receptors. *Experimental Cell Research*, 317(7), pp.955–965.
- Ippel, J.H. et al., 1996. Heteronuclear Scalar Couplings in the Bases and Sugar Rings of Nucleic Acids: Their Determination and Application in Assignment and Conformational Analysis. *Magnetic Resonance in Chemistry*, 34(13), pp.S156–S176.
- Iyer, S. & Acharya, K.R., 2011. Tying the knot: The cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *FEBS Journal*, 278(22), pp.4304–4322.
- Jarvis, T.C. et al., 2015. Non-helical DNA Triplex Forms a Unique Aptamer Scaffold for High Affinity Recognition of Nerve Growth Factor. *Structure*, 23(7), pp.1293–1304.
- Jellinek, D. et al., 1994. Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry*, 33(34), pp.10450–10456.

- Kato, K. et al., 2016. Structural basis for specific inhibition of Autotaxin by a DNA aptamer. *Nature Structural & Molecular Biology*, 23(October 2015), pp.1–9.
- Kay, L., Keifer, P. & Saarinen, T., 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society*, 114(26), pp.10663–10665.
- Kay, L.E. et al., 1990. 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. *Journal of Magnetic Resonance*, 89(3), pp.496–514.
- Keeler, J., 2002. *Understanding NMR Spectroscopy* 1st ed.,
- Kleckner, I.R. & Foster, M.P., 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1814(8), pp.942–968.
- Kool, E.T., 2001. Hydrogen bonding, base stacking, and steric effects in dna replication. *Annual review of biophysics and biomolecular structure*, 30, pp.1–22.
- Krainer, F.W. et al., 2012. Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. *Microbial Cell Factories*, 11(1), p.22.
- Krauss, I.R. et al., 2013. Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Crystallographica Section D: Biological Crystallography*, 69(12), pp.2403–2411.
- Krauss, I.R. et al., 2012. High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Research*, 40(16), pp.8119–8128.
- Krilleke, D. et al., 2007. Molecular mapping and functional characterization of the VEGF164 heparin-binding domain. *Journal of Biological Chemistry*, 282(38), pp.28045–28056.
- Krock, B.L., Skuli, N. & Simon, M.C., 2011. Hypoxia-Induced Angiogenesis: Good and Evil. *Genes & Cancer*, 2(12), pp.1117–1133.
- Kuszewski, J., Schwieters, C. & Clore, G.M., 2001. Improving the accuracy of NMR structures of DNA by means of a database potential of mean force describing base-base positional interactions. *Journal of the American Chemical Society*, 123(17), pp.3903–3918.
- Lee, J.-H. et al., 2005. A therapeutic aptamer inhibits angiogenesis by specifically targeting the heparin binding domain of VEGF165. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52), pp.18902–7.

- Lee, J.H., Jucker, F. & Pardi, A., 2008. Imino proton exchange rates imply an induced-fit binding mechanism for the VEGF165-targeting aptamer, Macugen. *FEBS Letters*, 582(13), pp.1835–1839.
- Leitner, D., Schröder, W. & Weisz, K., 1998. Direct Monitoring of Cytosine Protonation in an Intramolecular DNA Triple Helix. *Journal of the American Chemical Society*, 120(28), pp.7123–7124.
- Lemmon, M.A. & Schlessinger, J., 2010. Cell signaling by receptor tyrosine kinases. *Cell*, 141(7), pp.1117–1134.
- Leontis, N.B., Stombaugh, J. & Westhof, E., 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic acids research*, 30(16), pp.3497–531.
- Leontis, N.B. & Westhof, E., 2003. Analysis of RNA motifs. *Current Opinion in Structural Biology*, 13(3), pp.300–308.
- Lewis, T.S., Shapiro, P.S. & Ahn, N.G., 1998. Signal Transduction through MAP Kinase Cascades. *Advances in Cancer Research*, 74(February), pp.49–139.
- Lian, L.Y. et al., 1994. [23] Protein-ligand interactions: Exchange processes and determination of ligand conformation and protein-ligand contacts. *Methods in Enzymology*, 239, pp.657–700.
- Liu, S. et al., 2011. Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. *RNA (New York, N.Y.)*, 17(9), pp.1655–63.
- Liu, Y. et al., 1995. Hypoxia Regulates Vascular Endothelial Growth Factor Expression in Endothelial Cells. *Circulation Research*, 77, pp.638–643.
- Liu, Y. et al., 2015. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature*, 522(7556), pp.368–72.
- Long, S.B. et al., 2008. Crystal structure of an RNA aptamer bound to thrombin. *RNA (New York, N.Y.)*, 14(12), pp.2504–12.
- Louis, J.M. et al., 1998. Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *Journal of Biological Chemistry*, 273(4), pp.2374–2378.
- Lovell, S.C. et al., 2003. Structure validation by C α geometry: ϕ, ψ and C β deviation. *Proteins: Structure, Function and Genetics*, 50(3), pp.437–450.
- Lovell, S.C. et al., 2000. The penultimate rotamer library. *Proteins: Structure, Function and Genetics*, 40(3), pp.389–408.

- Lu, K., Miyazaki, Y. & Summers, M.F., 2010. Isotope labeling strategies for NMR studies of RNA. *Journal of Biomolecular NMR*, 46(1), pp.113–125.
- Lu, X.-J. & Olson, W.K., 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols*, 3(7), pp.1213–27.
- Lu, X.J. & Olson, W.K., 2003. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research*, 31(17), pp.5108–5121.
- MacArthur, M.W. & Thornton, J.M., 1991. Influence of proline residues on protein conformation. *Journal of Molecular Biology*, 218(2), pp.397–412.
- Macauley-Patrick, S. et al., 2005. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, 22(4), pp.249–270.
- Makler, M.T. & Hinrichs, D.J., 1993. Measurement of the lactate dehydrogenase activity of *Plasmodium falciparum* as an assessment of parasitemia. *American Journal of Tropical Medicine and Hygiene*, 48(2), pp.205–210.
- Marino, J.P., Prestegard, J.H. & Crothers, D.M., 1994. Correlation of Adenine H2/H8 Resonances in Uniformly ¹³C Labeled RNAs by 2D HCCH-TOCSY: A New Tool for ¹H Assignment. *Journal of the American Chemical Society*, 116(5), pp.2205–2206.
- Marion, D., Driscoll, P.C., et al., 1989. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: *Biochemistry*, 28(15), pp.6150–6156.
- Marion, D., Kay, L.E., et al., 1989. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American ...*, 111(4), pp.1515–1517.
- Markovic-Mueller, S. et al., 2017. Structure of the Full-length VEGFR-1 Extracellular Domain in Complex with VEGF-A. *Structure*, pp.1–12.
- Matthew, J.B. & Richards, F.M., 1983. The pH dependence of hydrogen exchange in proteins. *Journal of Biological Chemistry*, 258(5), pp.3039–3044.
- Mielke, S.P. & Krishnan, V. V., 2009. Characterization of protein secondary structure from NMR chemical shifts. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 54(3–4), pp.141–165.
- Miller, J.W. et al., 2013. Vascular endothelial growth factor a in intraocular vascular disease. *Ophthalmology*, 120(1), pp.106–114.

- Mitra, S.K., Hanson, D. a & Schlaepfer, D.D., 2005. Focal adhesion kinase: in command and control of cell motility. *Nature reviews. Molecular cell biology*, 6(1), pp.56–68.
- Mittermaier, A.K. & Kay, L.E., 2009. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences*, 34(12), pp.601–611.
- Miyakawa, S., Nomura, Y. & Sakamoto, T., 2008. Structural and molecular basis for hyperspecificity of RNA aptamer to human immunoglobulin G. *Rna*, pp.1154–1163.
- Montelione, G.T. et al., 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *Journal of the American Chemical Society*, 114(27), pp.10974–10975.
- Morgan, A.R. & Wells, R.D., 1968. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *Journal of molecular biology*, 37(1), pp.63–80.
- Morris, K.N. et al., 1998. High affinity ligands from in vitro selection : Complex targets. *Rna*, 95(March), pp.2902–2907.
- Muller, Y.A. et al., 1997. The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 Å resolution: multiple copy flexibility and receptor binding. *Structure*, 5(10), pp.1325–1338.
- Ng, Y.S. et al., 2001. Differential expression of VEGF isoforms in mouse during development and in the adult. *Developmental dynamics : an official publication of the American Association of Anatomists*, 220(2), pp.112–121.
- Nicholson, L.K. et al., 1992. Dynamics of methyl groups in proteins as studied by proton-detected carbon-13 NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry*, 31(23), pp.5253–5263.
- Nikonowicz, E.P. et al., 1992. Preparation of ¹³C and ¹⁵N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Research*, 20(17), pp.4507–4513.
- Nimjee, S.M. et al., 2017. Aptamers as Therapeutics. *Annual Review of Pharmacology and Toxicology*, 57(1), pp.61–79.
- Nissen, N.N. et al., 1998. Vascular endothelial growth factor mediates angiogenic activity during the proliferative phase of wound healing. *The American journal of pathology*, 152(6), pp.1445–52.
- Noeske, J. et al., 2005. An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5), pp.1372–7.

- Nomura, Y. et al., 2010. Conformational plasticity of RNA for target recognition as revealed by the 2.15 Å crystal structure of a human IgG-aptamer complex. *Nucleic Acids Research*, 38(21), pp.7822–7829.
- Nowak, J.Z., 2006. Age-related macular degeneration (AMD): pathogenesis and therapy. *Pharmacological reports : PR*, 58(3), pp.353–63.
- Oberthür, D. et al., 2015. Crystal structure of a mirror-image L-RNA aptamer (Spiegelmer) in complex with the natural L-protein target CCL2. *Nature communications*, 6, p.6923.
- Olson, W.K. et al., 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology*, 313(1), pp.229–237.
- Ösapay, K. & Case, D.A., 1994. Analysis of proton chemical shifts in regular secondary structure of proteins. *Journal of Biomolecular NMR*, 4(2), pp.215–230.
- Padlan, C. et al., 2014. An RNA aptamer possessing a novel monovalent cation-mediated fold inhibits lysozyme catalysis by inhibiting the binding of long natural substrates. *RNA (New York, N.Y.)*, 20(4), pp.447–461.]
- Paleolog, E.M., 2002. Angiogenesis in rheumatoid arthritis. *Arthritis Research {&} Therapy*, 4(3), p.S81.
- Pardi, A. & Nikonowicz, E.P., 1992. Simple procedure for resonance assignment of the sugar protons in carbon-13 labeled RNAs. *Journal of the American Chemical Society*, 114(23), pp.9202–9203.
- Pardi, A., Wagner, G. & Wüthrich, K., 1983. Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *European Journal of Biochemistry*, 137(3), pp.445–454.
- Park, J.E., Keller, G.A. & Ferrara, N., 1993. The vascular endothelial growth factor (VEGF) isoforms: differential deposition into the subepithelial extracellular matrix and bioactivity of extracellular matrix-bound VEGF. *Molecular biology of the cell*, 4(12), pp.1317–26.
- Parker, M.W. et al., 2012. Structural basis for selective vascular endothelial growth factor-A (VEGF-A) binding to neuropilin-1. *Journal of Biological Chemistry*, 287(14), pp.11082–11089.
- Peretz, D. et al., 1992. Glycosylation of vascular endothelial growth factor is not required for its mitogenic activity. *Biochemical and Biophysical Research Communications*, 182(3), pp.1340–1347.
- Pica, A. et al., 2016. Through-bond effects in the ternary complexes of thrombin sandwiched by two DNA aptamers. *Nucleic acids research*, 45(1), pp.461–469.

- Plouët, J. et al., 1997. Extracellular Cleavage of the Vascular Endothelial Growth Factor 189-Amino Acid Form by Urokinase Is Required for Its Mitogenic Effect. , 272(20), pp.13390–13396.
- Potty, A.S.R. et al., 2009. Biophysical characterization of DNA aptamer interactions with vascular endothelial growth factor. *Biopolymers*, 91(2), pp.145–156.
- Redfield, C. & Robertson, J.P., 1991. Assignment of the NMR Spectra of Homologous Proteins BT - Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy. In J. C. Hoch, F. M. Poulsen, & C. Redfield, eds. Boston, MA: Springer US, pp. 303–316. Available at: http://dx.doi.org/10.1007/978-1-4757-9794-7_23.
- Richardson, J.S., 2015. A context-sensitive guide to RNA & DNA base-pair & base-stack geometry. *Computational Crystallography Newsletter*, 5, pp.47–53.
- Roberts, G.C.K., 1993. *NMR of Macromolecules: A Practical Approach* 1st ed., Oxford: Oxford University Press.
- Robertson, M.P. et al., 2005. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biology*, 3(1).
- Robinson, C.J. et al., 2006. VEGF165-binding sites within heparan sulfate encompass two highly sulfated domains and can be liberated by K5 lyase. *Journal of Biological Chemistry*, 281(3), pp.1731–1740.
- Ruckman, J. et al., 1998. 2'-Fluoropyrimidine RNA-based Aptamers to the 165-Amino Acid Form of Vascular Endothelial Growth Factor (VEGF165). *The Journal of Biological Chemistry*, 273(32), pp.20556–20567.
- Rusconi, C.P. et al., 2002. RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature*, 419(6902), pp.90–94.
- Scaldaferri, F. et al., 2009. VEGF-A Links Angiogenesis and Inflammation in Inflammatory Bowel Disease Pathogenesis. *Gastroenterology*, 136(2), p.585–595.e5.
- Schlünzen, F. et al., 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, 413(6858), pp.814–821.
- Schütze, T. et al., 2011. Probing the SELEX process with next-generation sequencing. *PLoS ONE*, 6(12).
- Schwieters, C. D. et al., 2003. The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance*, 160, pp.65–73.

- Schwieters, C.D. & Clore, G.M., 2001. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *Journal of magnetic resonance*, 152, pp.288–302.
- Schwieters, C.D., Kuszewski, J.J. & Marius Clore, G., 2006. Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48(1), pp.47–62.
- Shen, Y. et al., 2009. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4), pp.213–223.
- Shweiki, D. et al., 1992. Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature*, 359(6398), pp.843–845.
- Sigma-Aldrich, 2012. Bicinchoninic Acid Protein Assay Kit. *Bicinchoninic Acid Protein Assay Kit*, pp.1–6.
- Simons, M., Gordon, E. & Claesson-Welsh, L., 2016. Mechanisms and regulation of endothelial VEGF receptor signalling. *Nature Reviews Molecular Cell Biology*, 10, pp.611–25.
- Simorre, J.-P. et al., 1996a. Correlation of the guanosine exchangeable and nonexchangeable base protons in $^{13}\text{C}/^{15}\text{N}$ -labeled RNA with an HNC-TOCSY-CH experiment. *Journal of biomolecular NMR*, 7(2), pp.153–156.
- Simorre, J.-P. et al., 1996b. Triple-Resonance Experiments for Assignment of Adenine Base Resonances in $^{13}\text{C}/^{15}\text{N}$ -Labeled RNA. *Journal of the American Chemical Society*, 118(22), pp.5316–5317.
- Simorre, J.-P. et al., 1995. Triple resonance HNCCCH experiments for correlating exchangeable and nonexchangeable cytidine and uridine base protons in RNA. *Journal of Biomolecular NMR*, 6(4), pp.427–432.
- Smith, P.K. et al., 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1), pp.76–85.
- Soker, S. et al., 1998. Neuropilin-1 Is Expressed by Endothelial and Tumor Cells as an Isoform-Specific Receptor for Vascular Endothelial Growth Factor. *Cell*, 92, pp.735–745.
- Soker, S. et al., 2002. VEGF165 mediates formation of complexes containing VEGFR-2 and neuropilin-1 that enhance VEGF165-receptor binding. *Journal of Cellular Biochemistry*, 85(2), pp.357–368.
- Stauffer, M.E., Skelton, N.J. & Fairbrother, W.J., 2002. Refinement of the solution structure of the heparin-binding domain of vascular endothelial growth factor using residual dipolar couplings. *Journal of Biomolecular NMR*, 23(1), pp.57–61.

- Stewart, D.E., Sarkar, A. & Wampler, J.E., 1990. Occurrence and role of cis peptide bonds in protein structures. *Journal of molecular biology*, 214(1), pp.253–60.
- Tao, J., Chen, L. & Frankel, A.D., 1997. Dissection of the proposed base triple in human immunodeficiency virus TAR RNA indicates the importance of the Hoogsteen interaction. *Biochemistry*, 36(12), pp.3491–3495.
- Tejero, R. et al., 2013. PDBStat: A universal restraint converter and restraint analysis software package for protein NMR. *Journal of Biomolecular NMR*, 56(4), pp.337–351.
- Tesmer, V.M. et al., 2012. Molecular mechanism for inhibition of G protein-coupled receptor kinase 2 by a selective RNA aptamer. *Structure*, 20(8), pp.1300–1309.
- Tjandra, N. et al., 1997. Use of dipolar 1H – 15N and 1H – 13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Structural Biology*, 4(9), pp.732–738.
- Tschopp, J.F. et al., 1987. Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Research*, 15(9), pp.3859–3876.
- Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968), pp.505–510.
- Tugarinov, V. & Kay, L.E., 2005. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem*, 6(9), pp.1567–1577.
- Vempati, P., Popel, A.S. & Mac Gabhann, F., 2014. Extracellular regulation of VEGF: Isoforms, proteolysis, and vascular patterning. *Cytokine and Growth Factor Reviews*, 25(1), pp.1–19.
- Vranken, W.F. et al., 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4), pp.687–696.
- Vuister, G.W. & Bax, A., 1992. Resolution enhancement and spectral editing of uniformly 13C -enriched proteins by homonuclear broadband 13C decoupling. *Journal of Magnetic Resonance (1969)*, 98(2), pp.428–435.
- Wagner, G., 1983. Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. *Quarterly reviews of biophysics*, 16(1), pp.1–57.
- Wang, S. et al., 2008. Control of endothelial cell proliferation and migration by VEGF signaling to histone deacetylase 7. *Proceedings of the National Academy of Sciences of the United States of America*, 105(22), pp.7738–7743.

- White, R. et al., 2001. Generation of species cross-reactive aptamers using “toggle” SELEX. *Molecular therapy : the journal of the American Society of Gene Therapy*, 4(6), pp.567–73.
- Wiechelman, K.J., Braun, R.D. & Fitzpatrick, J.D., 1988. Investigation of the bicinchoninic acid protein assay: Identification of the groups responsible for color formation. *Analytical Biochemistry*, 175(1), pp.231–237.
- Wiesmann, C. et al., 1997. Crystal Structure at 1.7 Å Resolution of VEGF in Complex with Domain 2 of the Flt-1 Receptor. *Cell*, 91(5), pp.695–704.
- Wild, K., Sinning, I. & Cusack, S., 2001. Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. *Science (New York, N.Y.)*, 294(5542), pp.598–601.
- Williamson, M.P., 1990. Secondary-structure dependent chemical shifts in proteins. *Biopolymers*, 29(10–11), pp.1423–1431.
- Wishart, D.S., Sykes, B.D. & Richards, F.M., 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of Molecular Biology*, 222(2), pp.311–333.
- Witmer, A., 2003. Vascular endothelial growth factors and angiogenesis in eye disease. *Progress in Retinal and Eye Research*, 22(1), pp.1–29.
- Wittekind, M. & Mueller, L., 1993. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B*, 101(2), pp.201–205.
- Word, J.M. et al., 1999. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *Journal of molecular biology*, 285(4), pp.1711–33.
- Wuttke, D.S. et al., 1997. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *Journal of molecular biology*, 273(1), pp.183–206.
- Yatime, L. et al., 2015. Structural basis for the targeting of complement anaphylatoxin C5a using a mixed L-RNA/L-DNA aptamer. *Nature communications*, 6, p.6481.
- Zhang, O. et al., 1994. Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *Journal of Biomolecular NMR*, 4(6), pp.845–858.
- Zuiderweg, E.R. & Fesik, S.W., 1989. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry*, 28(6), pp.2387–2391.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), pp.3406–3415.

Zwieb, C., 1992. Recognition of a tetranucleotide loop of signal recognition particle RNA by protein SRP19. *Journal of Biological Chemistry*, 267(22), pp.15650–15656.

Appendix A1 NMR experimental parameters

Table A1.1 NMR experimental parameters for HBD resonance assignments.

Experiment	Spectrometer (MHz)	nt	ni	ni2	sw (Hz)	sw1 (Hz)	sw2 (Hz)	np	d1 (s)	mix (s)
HNCA	600	16	50	42	11990	4526	2100	2048	1	-
HN(CO)CA	800	16	50	40	11261	6027	2950	2048	1.7	-
HNCO	800	16	50	40	11261	3770	2950	2048	1.7	-
HN(CA)CO	600	16	50	42	11990	3770	2100	2048	1.1	-
HNCACB	800	16	50	40	11261	16071	2950	2048	1.7	-
CBCA(CO)NH	600	16	65	50	9001	12070	2600	2048	1	-
HCCH-TOCSY	600	16	85	80	10000	8000	10000	2048	1	-
¹³ C, ¹⁵ N SIM-NOESY-HSQC	800	16	56	40	18382	9599	5981	3404	1.7	0.2
¹⁵ N TOCSY-HSQC	600	16	60	50	8000	8000	1800	2048	1.2	0.08
¹⁵ N NOESY-HSQC	600	16	90	60	8000	8000	1800	2048	1	0.15

Table A1.2 NMR experimental parameters for aptamer resonance assignments.

Experiment	Spectro -meter (MHz)	nt	ni	ni2	sw (Hz)	sw1 (Hz)	sw2 (Hz)	np	d1 (s)	mix (s)
¹ H- ¹ H NOESY (5 °C)	600	96	200	-	14006	14006	-	4096	1	0.2
¹ H- ¹ H NOESY (25 °C)	600	96	200	-	14006	14006	-	4096	2	0.2
Imino-optimized HSQC (10 °C)	600	256	100	-	14006	2400	-	2048	1	-
Imino-optimized HSQC (25 °C)	600	128	100	-	14006	2400	-	2048	1	-
Amino- optimized HSQC (10 °C)	600	224	100	-	14006	2400	-	2048	1	-
Amino- optimized NOESY-HSQC*	900	64	660	-	19841	19841	-	3970	1.7	0.2
Aromatic- optimized NOESY-HSQC*	900	16	120	64	14045	10000	6786	2048	1.7	0.2
Deoxyribose- optimized NOESY-HSQC*	900	16	120	96	14045	10000	9049	2048	1.7	0.2
3D HCCH- COSY	800	8	60	80	11000	7000	11000	3000	1.2	-
HNN-COSY (10 °C)	600	128	120	-	16000	6600	-	4096	1.5	-

*Experiments run by Dr. Geoff Armstrong, formerly of the University of Colorado.

Locations of NMR experiment files

HBD experiments located in:

/home/sabrina/Data/VEGF/HBD111-165/

HNCA

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_ghnca_25C_082814.fid/

HN(CO)CA

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_ghncoca_25C_090414.fid/

HNCO

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_ghnco_25C_091014.fid/

HN(CA)CO

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_ghncaco_25C_091014.fid/

HNCACB

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_ghncacb_25C_090114.fid/

CBCA(CO)NH

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_13C_15N_gcbea_co_nh_032215.fid/

HCCH-TOCSY

Titration_082114/

13C15N-HBD111-165_DNA_aptamer_3D_1H_1H_13C_HCCH-TOCSY_25C_120214.fid/

¹³C, ¹⁵N- SIM-NOESY-HSQC

Titration_082114/knoesy/

13C15N-HBD_DNA_apt_3D_knoesy_inter_ver2_gsa_decreased_sw2_25C_111414.fid/

¹⁵N-TOCSY-HSQC

Titration_052314/1-1_complex/15N-HBD111-

165_DNA_aptamer_3D_1H_1H_15N_gtocsyNhsqc_80ms_25C_060214.fid

¹⁵N-NOESY-HSQC

Titration_052314/1-1_complex/15N-HBD111-

165_DNA_aptamer_3D_1H_1H_15N_gnoesyNhsqc_25C_062514.fid

NOESY experiments located in:

/home/sabrina/Data/VEGF/HBD111-165/Titration/Titration_102413/1-1_complex/

2D ¹H-¹H NOESY (5 °C)

HBD111-165_DNAaptamer_2D_1H_1H_noesy11echo_5C_020514.fid/

2D ¹H-¹H NOESY (25 °C)

HBD111-165_DNAaptamer_2D_1H_1H_noesy11echo_25C_020314.fid/

All other aptamer experiments located in:

/home/sabrina/Data/VEGF/HBD111-165/Titration_042214/1-1_complex/

2D Imino-optimized HSQC (10 °C)

13C-15N_DNA_aptamer_HBD111-165_2D_1H_15N_gNhsqc_imino_10C_032215.fid/

2D Imino-optimized HSQC (25 °C)

13C-15N_DNA_aptamer_HBD111-165_2D_1H_15N_gNhsqc_imino_25C_032515.fid/

3D Amino-optimized HSQC (10 °C)

13C-15N_DNA_aptamer_HBD111-165_2D_1H_15N_gNhsqc_amino_nt256_10C_032515.fid/

2D plane of amino-optimized NOESY-HSQC

NOESY/

15N_NOESYHSQC_2D_Amino_13C15N_DNA_Aptamer_HBD_111-165_020715.fid/

2D plane of imino-optimized NOESY-HSQC

NOESY/

15N_NOESYHSQC_2D_Imino_13C15N_DNA_Aptamer_HBD_111-165_020715.fid/

3D Aromatic-optimized NOESY-HSQC

D2O/900data/NUS_13CNOESYHSQC/

13C_NOESYHSQC_3D_ arom_13C15N_DNA_Aptamer_HBD_111-165_D2O-051815.fid/

3D Deoxyribose-optimized NOESY-HSQC

D2O/900data/NUS_13CNOESYHSQC/

13C_NOESYHSQC_3D_13C15N_DNA_Aptamer_HBD_111-165_D2O-051615.fid/

3D HCCH-COSY

13C-15N_DNA_aptamer_HBD111-165_3D_hcchcosy_25C_052614.fid/

2D HNN-COSY (10 °C)

13C-15N_DNA_aptamer_HBD111-165_2D_1H_15N_HNNcosy_long_25C_032515.fid/

Appendix A2 Backbone connectivities for the HBD bound to the aptamer

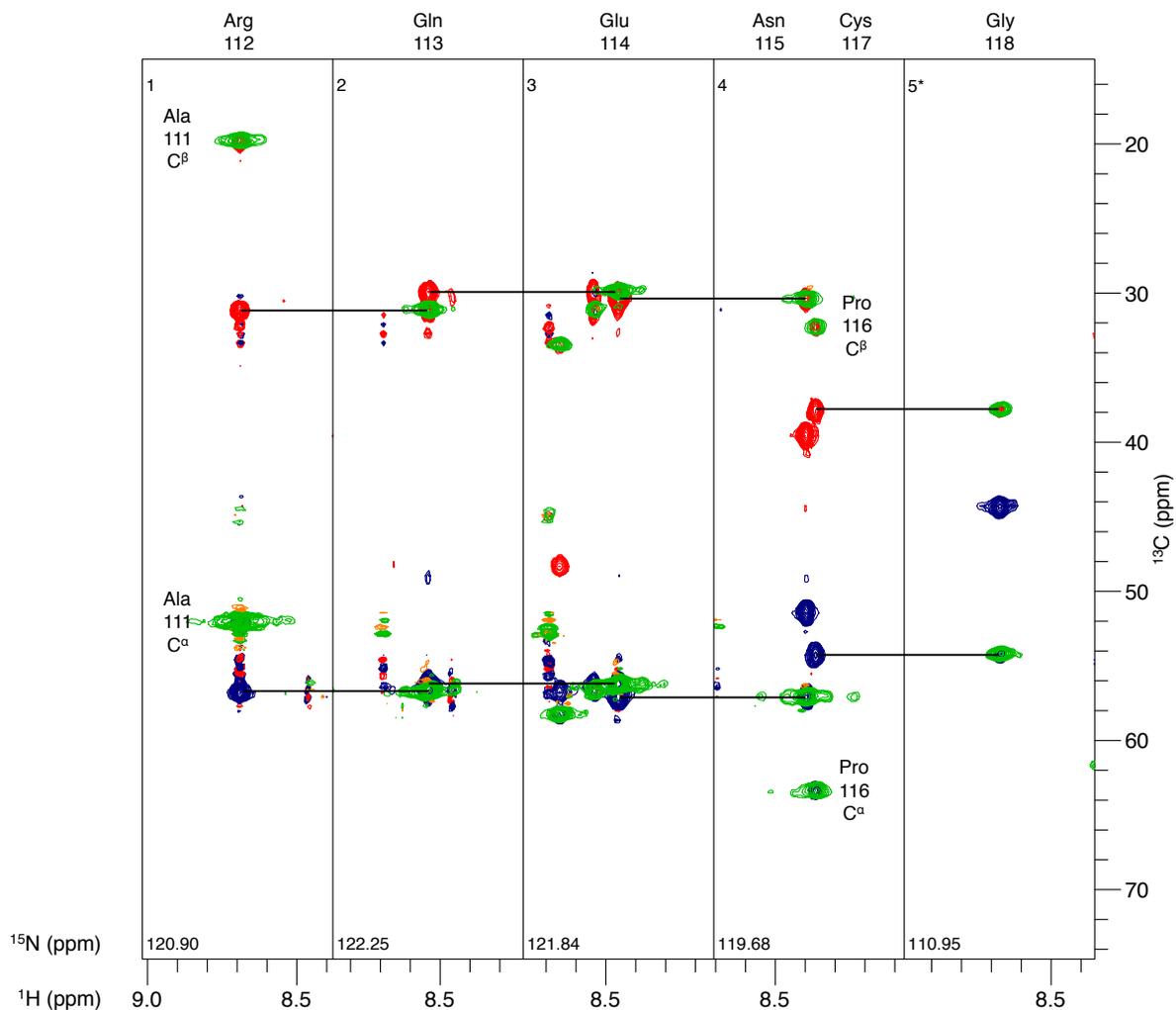


Figure A2.1 Backbone connectivities for Arg112 to Gly118. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. The amide group of Ala111 does not show up in the spectra due to fast exchange of the protons in the N-terminal NH_3^+ group, but the alpha and beta carbons of Ala111 are seen as *i*-1 peaks in Arg112. There is also a break in the connections at Pro116 due to the lack of amide group in proline residues. The alpha and beta carbons of Pro116 are seen as *i*-1 peaks in Cys117.

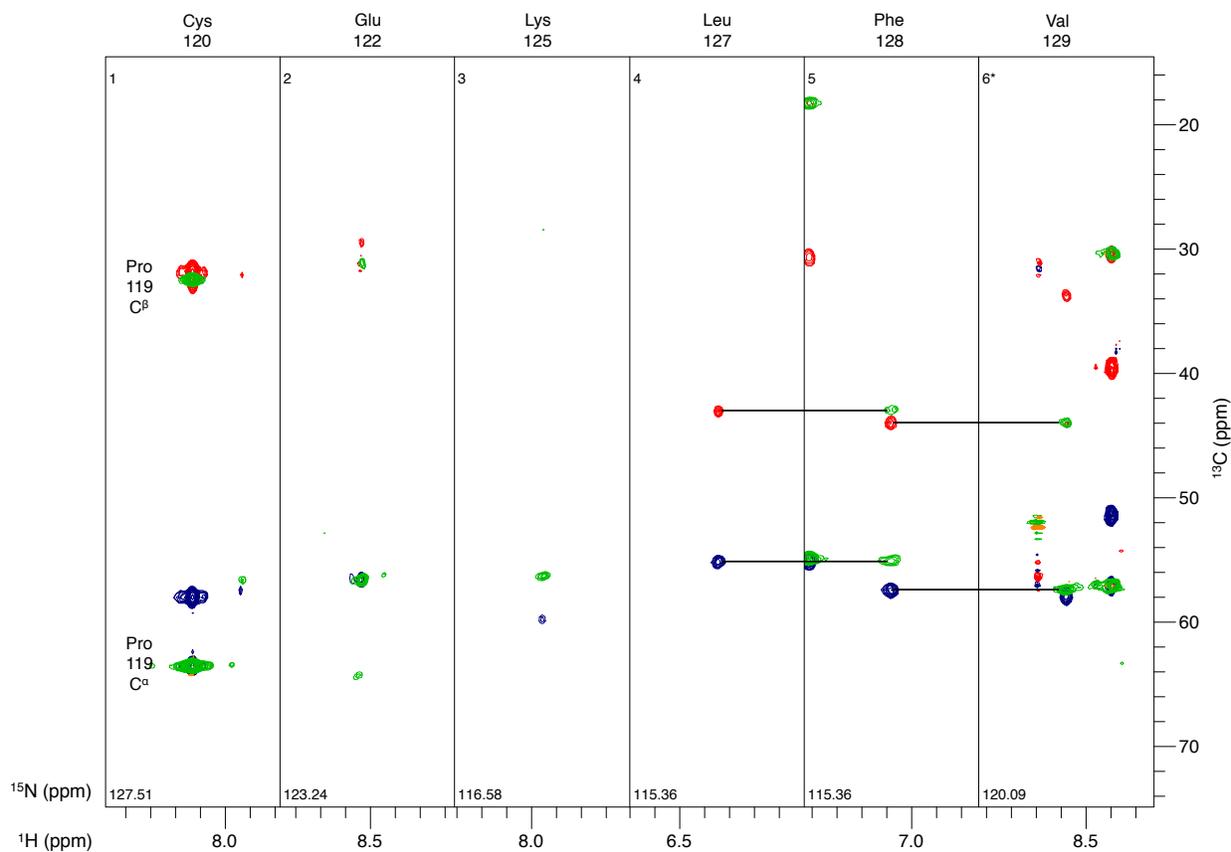


Figure A2.2 HNCACB/CBCA(CO)NH backbone connectivities for Cys120 to Val129. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. The signal/noise was poor for residues in this region of the protein. Ser121, Arg123, and Arg124 did not show up in the HNCACB/CBCA(CO)NH spectra but could be connected using the HNCA/HN(CO)CA. Unlabeled histidine was used as a supplement in the *P. pastoris* medium during protein expression, so His126 was not enriched with ^{13}C and ^{15}N beyond natural abundance levels.

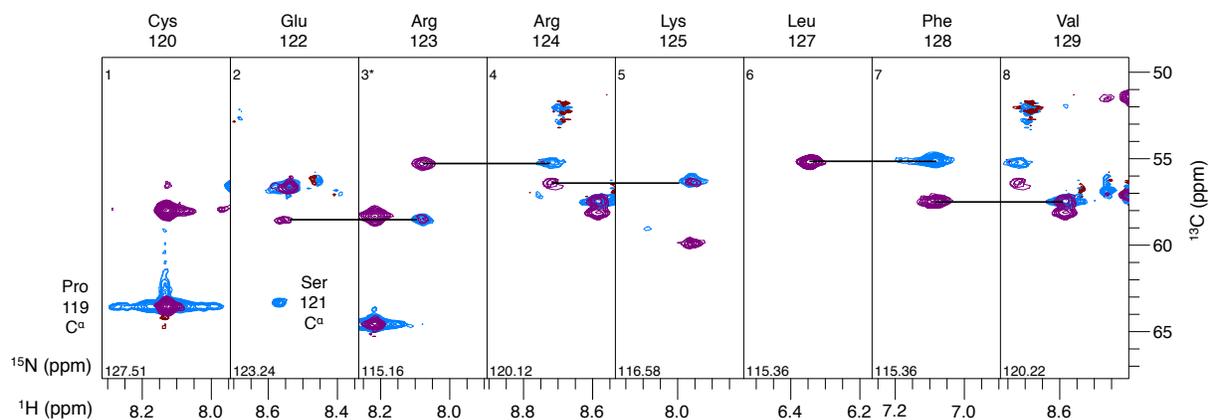


Figure A2.3 HNCA/HN(CO)CA backbone connectivities for Cys120 to Val129. The HNCA (purple) shows the alpha carbons for both the *i* and *i*-1 residues. The HN(CO)CA (blue) shows only the alpha carbon of the *i*-1 residue. The two breaks in the connection occur at Ser121, which has poor signal to noise, and His126, which is not labeled with ^{13}C and ^{15}N .

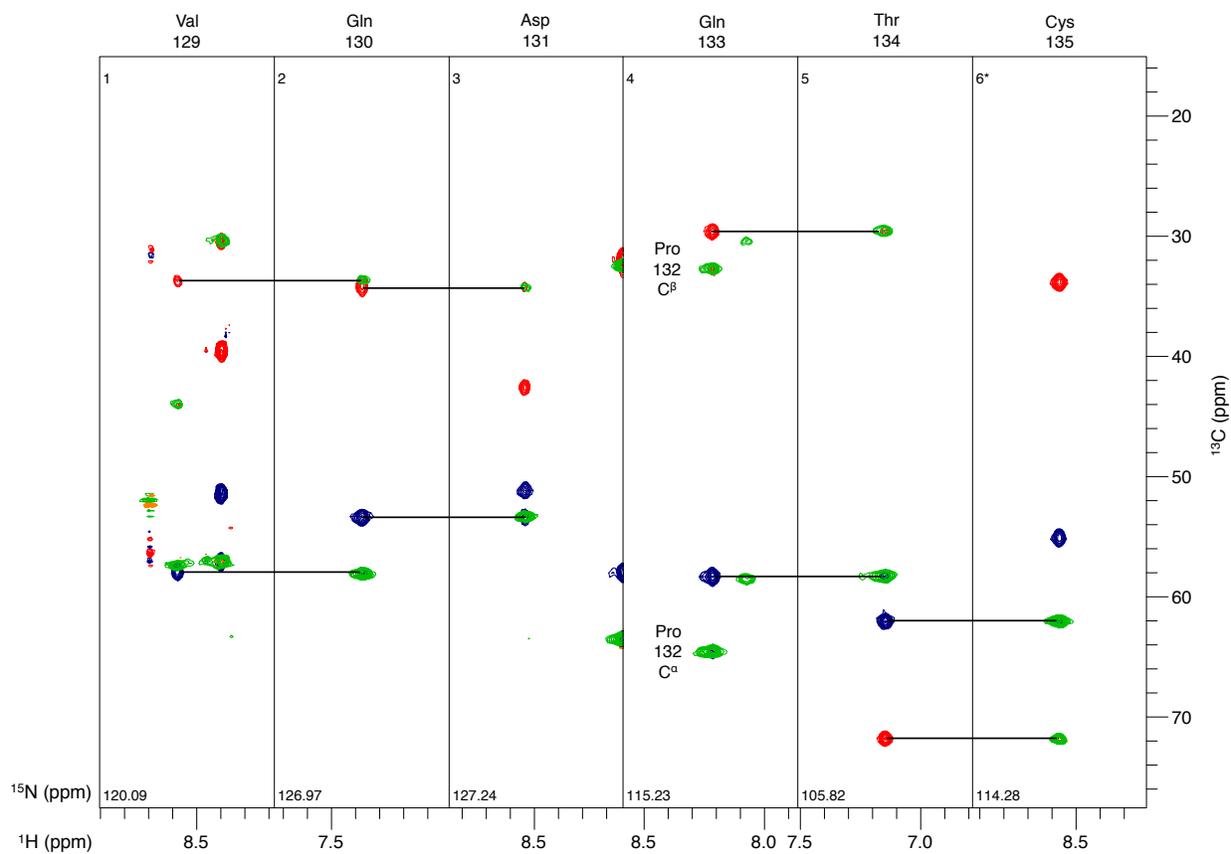


Figure A2.4 Backbone connectivities for Val129 to Cys135. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. There is a break in the connections at Pro132 due to the lack of amide group in proline residues. The alpha and beta carbons of Pro132 are seen as *i*-1 peaks in Gln133.

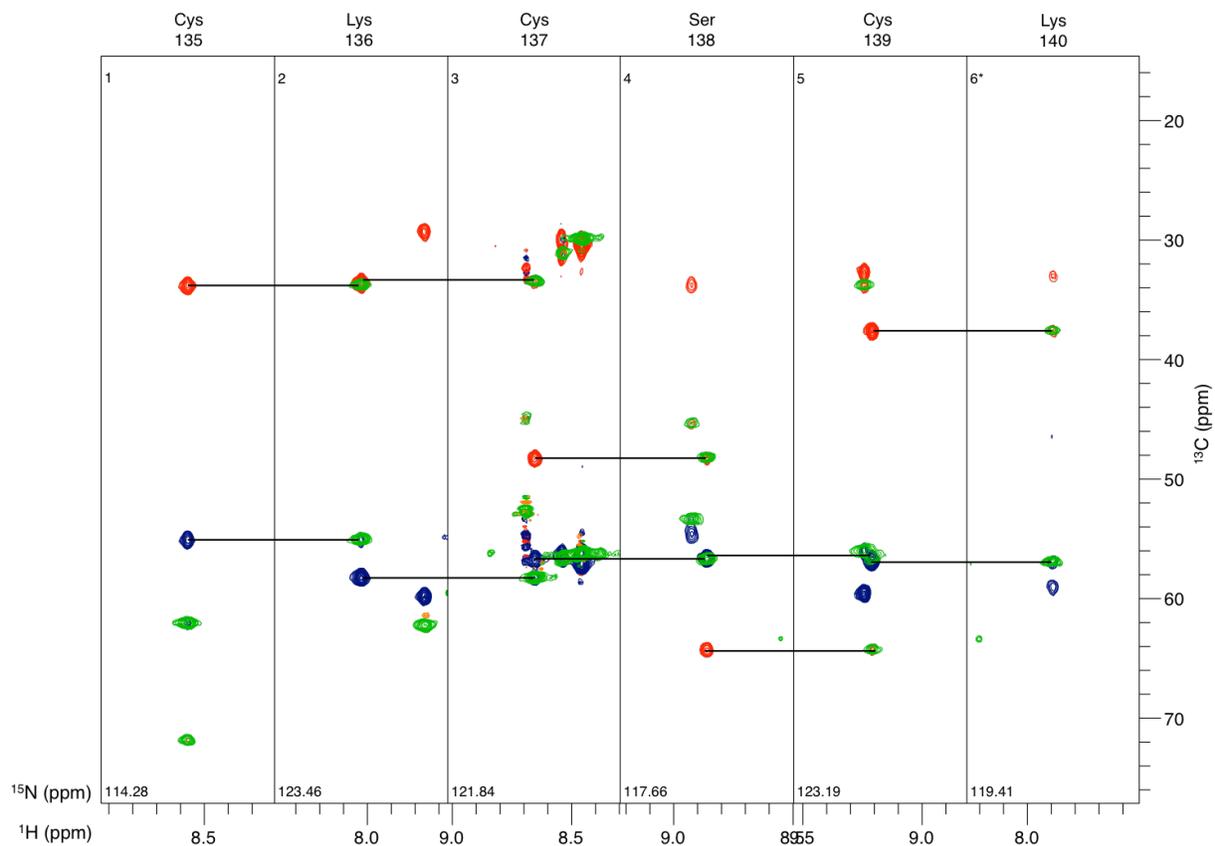


Figure A2.5 Backbone connectivities for Cys135 to Lys140. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue.

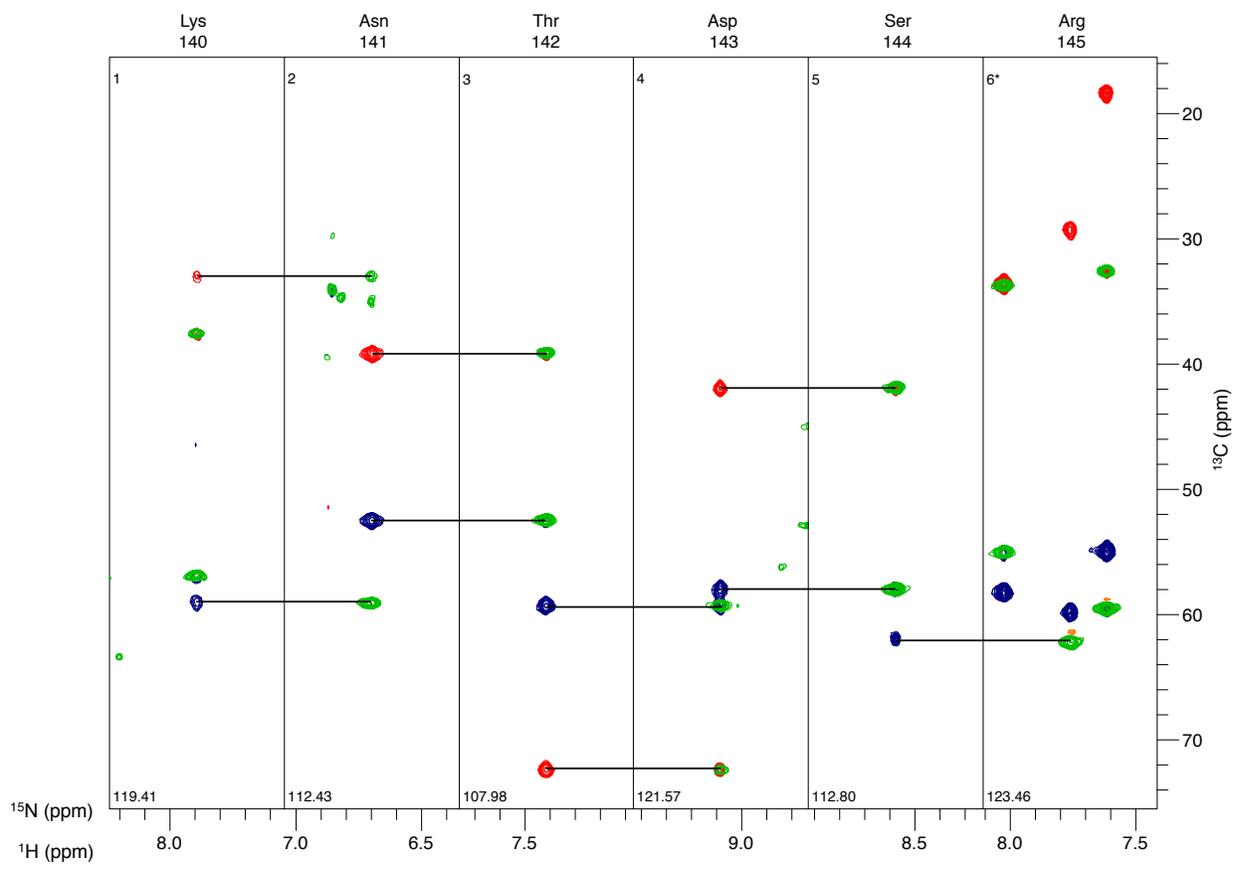


Figure A2.6 Backbone connectivities for Lys140 to Arg145. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. There is no peak corresponding to the beta carbon of Ser144.

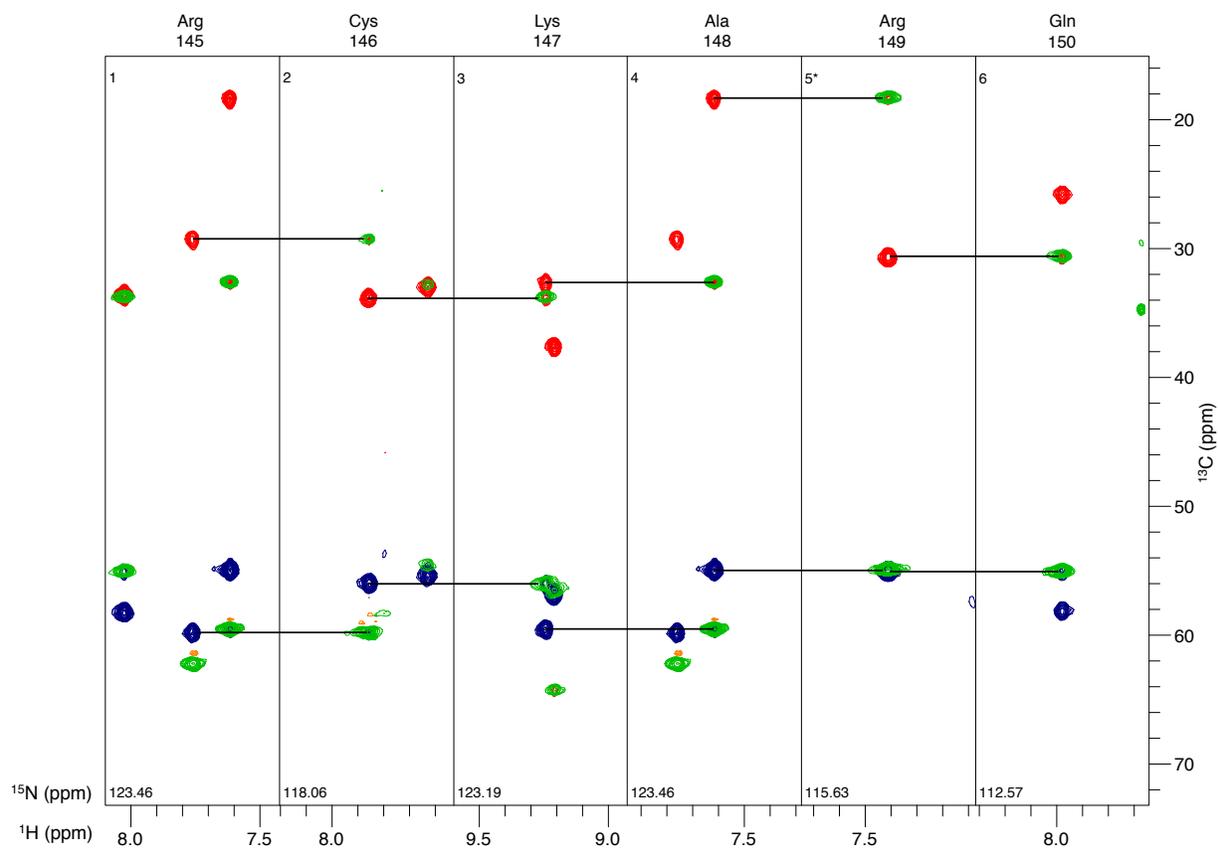


Figure A2.7 Backbone connectivities for Arg145 to Gln150. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue.

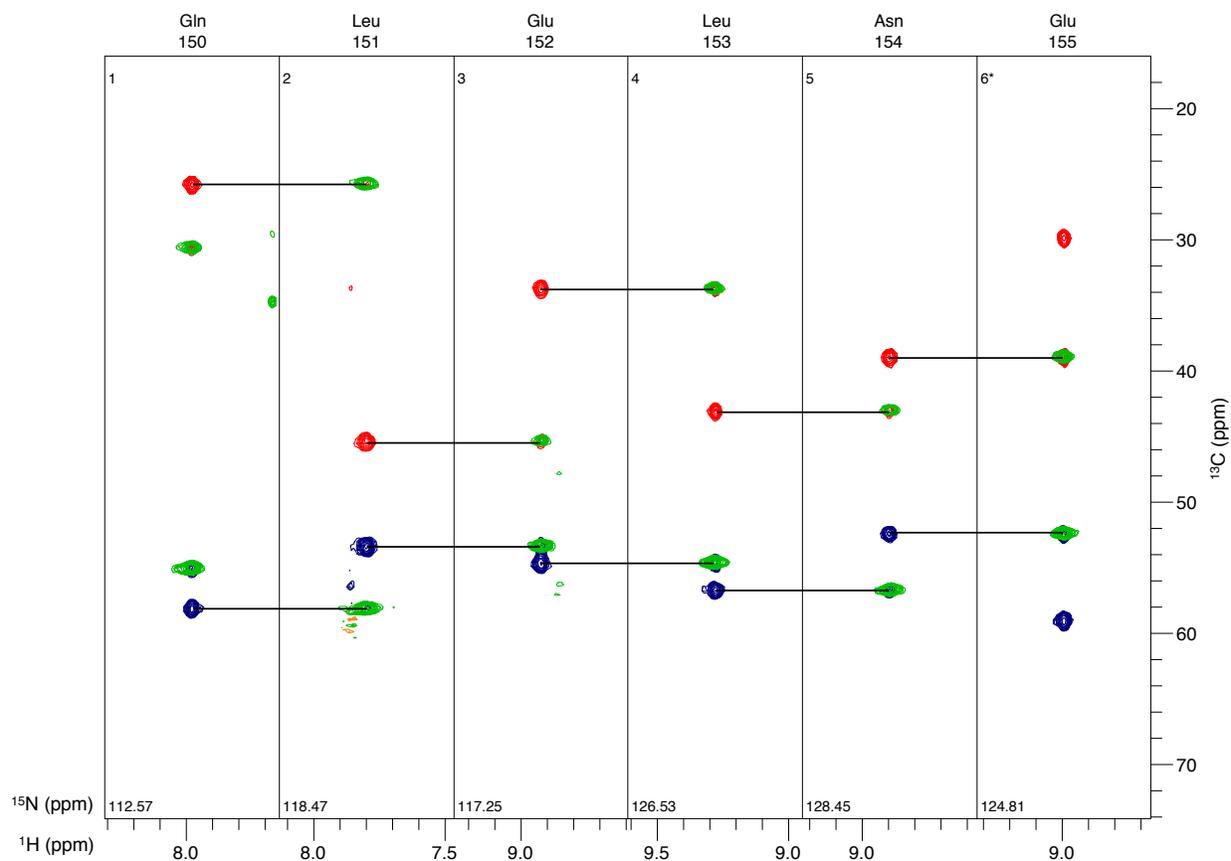


Figure A2.8 Backbone connectivities for Gln150 to Glu155. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue.

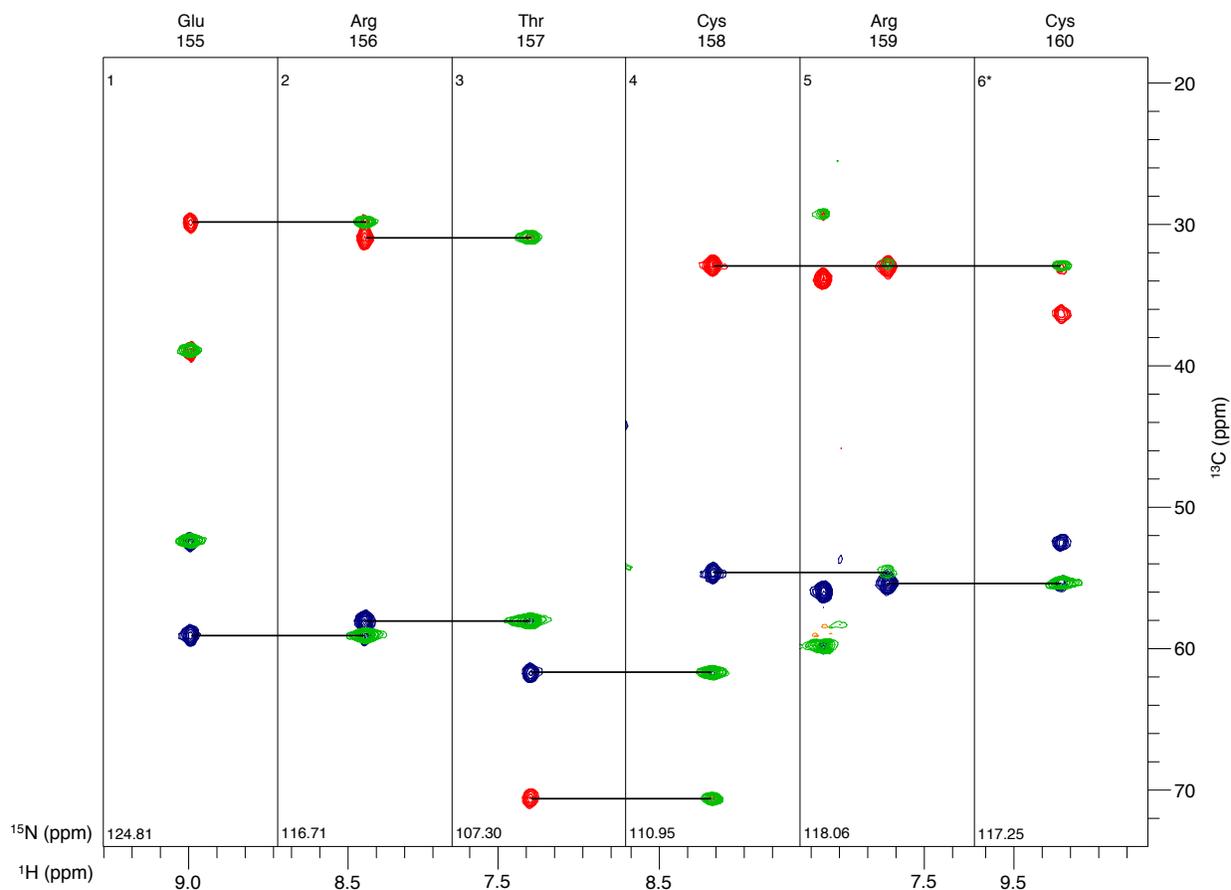


Figure A2.9 Backbone connectivities for Glu155 to Cys160. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the i and $i-1$ residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the $i-1$ residue.

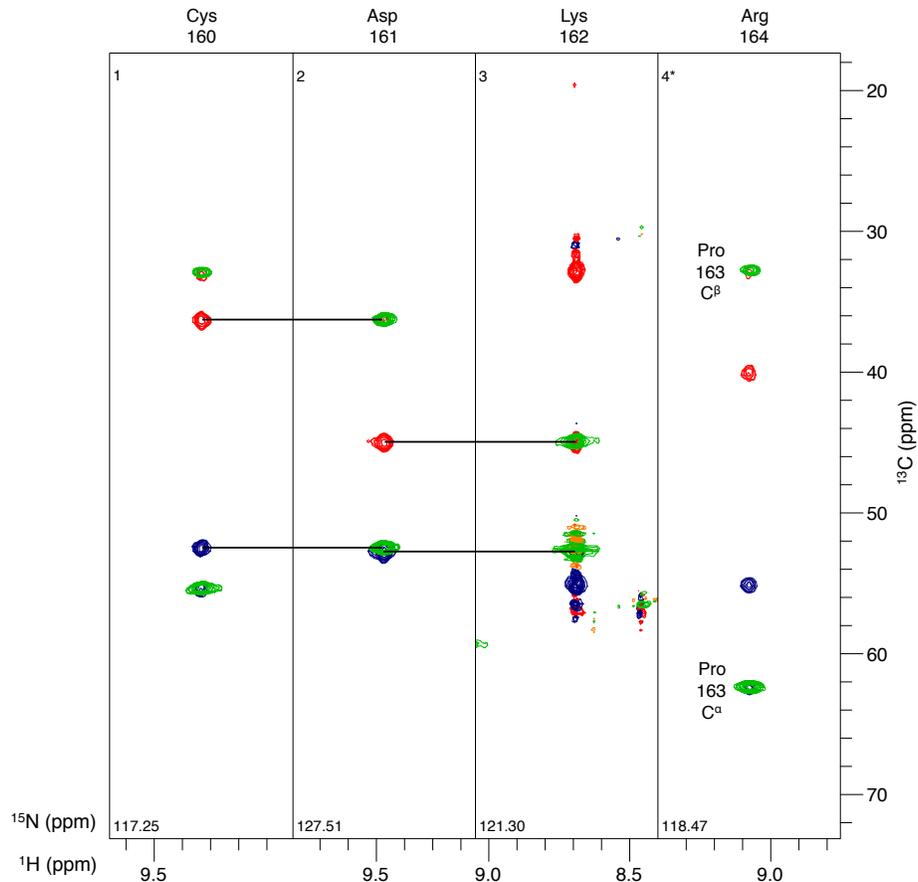


Figure A2.10 Backbone connectivities for Cys160 to Arg164. The HNCACB spectrum (positive contours in blue, negative contours in red) shows the alpha and beta carbons for the *i* and *i*-1 residues. The CBCA(CO)NH spectrum (green) shows only the alpha and beta carbons for the *i*-1 residue. There is a break in the connection at Pro163, but the alpha and beta carbons for Pro163 are seen as the *i*-1 peaks in Arg164. No peaks were identified as belonging to Arg165.

Appendix A3 HBD resonance assignments*

Table A3.1 HBD backbone resonance assignments

Number	Residue	HN (ppm)	N (ppm)	HA (ppm)	CA (ppm)	CO (ppm)
111	Ala			4.12	52.0	173.7
112	Arg	8.69	120.9	4.35	56.8	176.1
113	Gln	8.53	123.2	4.30	56.3	175.8
114	Glu	8.54	122.3	4.21	57.0	175.7
115	Asn	8.40	119.9	4.94	51.4	173.2
116	Pro	-		4.49	63.4	177.2
117	Cys	8.36	119.6	4.58	54.3	176.5
118	Gly	8.67	110.9	3.65, 4.26	44.1	170.7
119	Pro	-		4.36	63.5	176.1
120	Cys	8.09	127.2	4.16	57.6	181.5
121	Ser				56.6	173.4
122	Glu	8.57	123.3	4.12	58.5	177.1
123	Arg	8.08	115.2	4.56	55.3	176.2
124	Arg	8.71	120.2		56.4	175.3
125	Lys	7.96	116.8	3.63	59.4	177.1
126	His					
127	Leu	6.36	115.4	3.32	55.0	176.6
128	Phe	7.08	115.3	4.72	57.1	172.7
129	Val	8.61	120.3	3.60	58.0	170.1
130	Gln	7.39	127.0	5.02	52.9	174.0
131	Asp	8.54	127.3	5.00	50.7	176.1
132	Pro	-		4.35	64.5	177.2
133	Gln	8.25	115.2	4.24	58.0	177.4
134	Thr	7.16	105.8	4.36	61.8	176.0
135	Cys	8.58	114.3	4.38	54.6	175.0
136	Lys	8.04	123.6	4.06	57.9	175.0
137	Cys	8.66	121.9	5.32	56.2	174.4
138	Ser	8.88	117.7	4.95	56.2	173.4
139	Cys	9.23	123.3	4.09	56.4	174.0
140	Lys	7.84	119.5	3.32	58.8	177.0
141	Asn	6.72	112.4	4.48	52.4	173.8

Number	Residue	HN (ppm)	146	HA (ppm)	CA (ppm)	CO (ppm)
142	Thr	7.43	108.0	4.63	58.9	174.3
143	Asp	9.08	121.6	4.23	57.7	178.6
144	Ser	8.60	112.7	4.01	62.0	177.2
145	Arg	7.76	123.6	4.26	59.4	179.6
146	Cys	7.87	118.1	4.42	55.5	178.1
147	Lys	9.27	123.3	4.33	59.0	180.5
148	Ala	7.63	123.6	4.29	54.4	178.4
149	Arg	7.44	115.8	4.60	55.1	174.7
150	Gln	7.99	112.7	3.90	57.7	174.3
151	Leu	7.83	118.7	4.71	53.4	175.0
152	Glu	8.94	117.5	4.71	54.4	175.3
153	Leu	9.29	126.7	4.30	56.3	175.7
154	Asn	8.95	128.5	4.74	52.2	176.0
155	Glu	9.01	124.9	4.00	58.6	176.0
156	Arg	8.46	116.8	4.37	57.6	177.8
157	Thr	7.40	107.3	4.14	61.8	175.3
158	Cys	8.32	111.0	4.46	54.4	173.3
159	Arg	7.63	118.0	4.68	55.2	177.4
160	Cys	9.38	117.4	5.02	52.1	174.3
161	Asp	9.50	127.5	5.00	52.4	175.1
162	Lys	8.70	121.5	4.48	54.7	174.4
163	Pro	-		4.38	62.4	178.5
164	Arg	9.08	118.6	4.07	54.8	174.3
165	Arg	8.16	118.9		51.4	173.2

*Chemical shifts were referenced using trimethylsilylpropanoic acid (TSP) at 0 ppm. Referencing for ¹³C and ¹⁵N was done using the SetReference script.

SetReference

```
#!/usr/local/bin/python
```

```
"""
```

SetReference.py: This program calculates the ppm of the
1H, 13C, & 15N carriers given the position
(in point #) & chemical shift (in ppm) of
a reference signal.

Usage - place this file (source code) and SetReference (executable
script) into a folder in your path. If you aren't sure
what I mean, create a folder in your home directory called 'bin'
copy this file and the script into bin
and add the following line to your .cshrc file
set path=(~username/bin \$path).

Written by Justin Douglas October 20th 2008
University of Colorado, Pardi Lab
justindo@gmail.com

```
"""
```

```
import string  
import sys
```

```
def GetProcpaVar(varname):  
    procpafile=open('procpa','r')  
    lines=procpafile.readlines()  
  
    linecounter = 0  
  
    for i in lines:  
        words=i.split()  
        if len(words) > 0:  
            firstword=words[0]  
            if firstword == varname:  
                important_line = lines[linecounter+1]  
  
            linecounter=linecounter+1  
  
    array=important_line.split()  
    del array[0]
```

```

return array

if __name__=="__main__":

    sfrq_array=GetProcpaVar('sfrq')
    sfrq=string.atof(sfrq_array[0])
    sw_array=GetProcpaVar('sw')
    sw=string.atof(sw_array[0])
    dfrq_array=GetProcpaVar('dfrq')
    dfrq=string.atof(dfrq_array[0])
    if dfrq == 0.0:
        dfrq=1.0
    dfrq2_array=GetProcpaVar('dfrq2')
    dfrq2=string.atof(dfrq2_array[0])
    if dfrq2 == 0.0:
        dfrq2=1.0
    temperature_array=GetProcpaVar('temp')
    temperature=273.15+string.atof(temperature_array[0])

    #From Wishart et al. JBNMR (1995) 6, 135
    c13h1gamma_ratio = 0.251449453
    n15h1gamma_ratio = 0.101329118
    c13temperature_ratio = 1.04e-9
    n15temperature_ratio = 2.74e-10

    line='Please enter the number of complex data points:\n'
    sys.stdout.write(line)
    npoint_str=sys.stdin.readline()
    npoint = string.atoi(npoint_str)
    line='Please enter the reference point:\n'
    sys.stdout.write(line)
    refpoint_str = sys.stdin.readline()
    refpoint = string.atoi(refpoint_str)
    line='Please enter the ppm of the reference peak:\n'
    sys.stdout.write(line)
    refppm_str=sys.stdin.readline()
    refppm=string.atof(refppm_str)

    #absolute frequency of the reference peak
    npoint_from_carrier = npoint/2 + 1 - refpoint #positive number is downfield
    sw_per_point = sw/npoint
    h1reffreq_hz=sfrq*10**6+sw_per_point*npoint_from_carrier

    h1zeroppmfreq_hz = h1reffreq_hz - refppm*sfrq

    h1carrier_ppm = (sfrq*10**6-h1zeroppmfreq_hz)/sfrq

```

```

c13zeroppmfreq_hz = h1zeroppmfreq_hz*(c13h1gamma_ratio + (temperature-
300.0)*c13temperature_ratio)
n15zeroppmfreq_hz = h1zeroppmfreq_hz*(n15h1gamma_ratio + (temperature-
300.0)*n15temperature_ratio)

c13carrier_ppm = (dfrq*10**6-c13zeroppmfreq_hz)/dfrq
n15carrier_ppm = (dfrq2*10**6-n15zeroppmfreq_hz)/dfrq2

outputfilename = "chemicalshift_reference.out"
output=open(outputfilename,'w')

line1='H1 carrier = '+str(h1carrier_ppm)+'\n'
line2='C13 carrier = '+str(c13carrier_ppm)+'\n'
line3='N15 carrier = '+str(n15carrier_ppm)+'\n'

output.write(line1)
sys.stdout.write(line1)
output.write(line2)
sys.stdout.write(line2)
output.write(line3)
sys.stdout.write(line3)

output.close()

```

Table A3.2 HBD side chain resonance assignments*

Number	Residue	HB		CB	HG		CG	HD		CD/ND	HE/HZ		CE/NE/CZ
111	Ala	1.54		19.9	-	-	-	-	-	-	-	-	-
112	Arg	1.62	1.80	31.2			27.3	3.23		43.7	-	-	-
113	Gln	1.98	2.08	30.0	2.37		34.1	-	-	-	6.89	7.57	112.1
114	Glu	1.99	2.34	30.5	2.21		36.4	-	-	-	-	-	-
115	Asn	2.70	2.77	39.5	-	-	-	6.91	7.50	112.1	-	-	-
116	Pro	1.99	2.20	32.3			27.0	3.75	3.70	50.9	-	-	-
117	Cys	2.58	3.03	37.7	-	-	-	-	-	-	-	-	-
118	Gly	-	-	-	-	-	-	-	-	-	-	-	-
119	Pro	2.26	1.90	32.1	1.89	2.09	28.2	3.74	4.05	51.6	-	-	-
120	Cys	1.96	2.26	31.3	-	-	-	-	-	-	-	-	-
121	Ser			64.4	-	-	-	-	-	-	-	-	-
122	Glu			30.4			36.6	-	-	-	-	-	-
123	Arg	1.65	1.85					3.21			-	7.20	93.4
124	Arg	1.18	1.37								-	6.63	97.3
125	Lys			32.4							-		
126	His	-	-	-	-	-	-	-	-	-	-	-	-
127	Leu	0.30	0.81	42.8	-	0.48	26.5	0.35	0.36	23.12, 24.49	-	-	-
128	Phe	2.30	2.62	44.0	-	-	-	-	6.56		-	7.03	
129	Val	-	0.57	33.4	0.83	0.92	16.45, 19.34	-	-	-	-	-	-
130	Gln	1.38	1.66	33.8	1.67	1.82	34.3	-	-	-	-	-	-
131	Asp	2.88	3.00	42.6	-	-	-	-	-	-	-	-	-
132	Pro	1.90	2.24	32.1	1.82	2.07	27.7	3.72	4.02	51.9	-	-	-
133	Gln	2.28	2.18	29.5	2.31	2.45	34.8	-	-	-	6.85	7.70	112.1
134	Thr	-	4.32	71.7	-	1.19	50.7	-	-	-	-	-	-
135	Cys	3.30		33.4	-	-	-	-	-	-	-	-	-
136	Lys	1.85		33.4	1.20	1.45	25.6	1.72	1.82	29.2	1.91	3.04	42.6
137	Cys	2.33	3.26	47.7	-	-	-	-	-	-	-	-	-
138	Ser	3.06	3.82	64.3	-	-	-	-	-	-	-	-	-
139	Cys	2.60	4.09	37.3	-	-	-	-	-	-	-	-	-
140	Lys	1.76	2.01	27.0			25.8				-	-	-
141	Asn	2.63	2.53	39.0	-	-	-	-	-	-	-	-	-
142	Thr	-	4.54	72.5	-	1.42	22.8	-	-	-	-	-	-
143	Asp	2.59	2.69	41.5	-	-	-	-	-	-	-	-	-
144	Ser				-	-	-	-	-	-	-	-	-
145	Arg	1.75	2.01	29.3			27.1	3.34		44.1	-	7.43	95.4
146	Cys	2.55	2.77	33.4	-	-	-	-	-	-	-	-	-
147	Lys	2.03	1.81	32.2	1.68	1.54	25.9	1.76	1.65	29.2	-	3.06	42.1
148	Ala	1.62		18.0	-	-	-	-	-	-	-	-	-

Number	Residue	HB		CB	HG		CG	HD		CD/ND	HE/HZ		CE/NE/CZ
149	Arg	1.77	2.17	30.7			27.1	3.27		44.3	-	7.3 4	93.5
150	Gln	2.30	2.45	25.7			35.2	-	-	-	6.73	7.5 7	112.1
151	Leu	1.42		45.4	-	1.00	26.6	0.76		22.4	-	-	-
152	Glu	1.86	2.12	33.6			35.9	-	-	-	-	-	-
153	Leu	1.80	0.95	42.8	-	0.77	27.2	0.39		24.2	-	-	-
154	Asn	2.68	3.46	38.5	-	-	-	6.99	8.27	115.28 (N)	-	-	-
155	Glu	1.96	2.24	29.5	2.35		36.8	-	-	-	-	-	-
156	Arg	1.63	1.98	30.8	1.15		27.7	3.24		43.4	-	-	-
157	Thr	-	4.34	70.4	-	1.15	21.0	-	-	-	-	-	-
158	Cys	2.95	3.55	32.6	-	-	-	-	-	-	-	-	-
159	Arg	1.54	1.95	32.6			28.6	3.24	3.45	44.2	-	-	-
160	Cys	2.85	2.94	36.0	-	-	-	-	-	-	-	-	-
161	Asp	2.55	2.86	45.0	-	-	-	-	-	-	-	-	-
162	Lys	1.83	1.72	32.4	1.53		24.6				-	3.0 3	42.1
163	Pro	1.77		33.0			27.3				-	-	-
164	Arg	2.22		39.5							-	-	-
165	Arg			40.9							-	-	-

*Chemical shifts were referenced using trimethylsilylpropanoic acid (TSP) at 0 ppm. Referencing for ^{13}C and ^{15}N was done using the SetReference script.

Appendix A4 TALOS+ dihedral angle predictions

Table A4.1 HBD dihedral angles predicted by TALOS+

RESID	RESNAME	PHI	PSI	DPHI	DPSI	DIST	S2	COUNT	CLASS
%4d	%s	%8.3f	%8.3f	%8.3f	%8.3f	%8.3f	%5.3f	%2d	%s
112	R	-98.33	152.42	32.86	31.36	31.95	0.21	10	Dyn
113	Q	-92.80	130.70	71.85	25.12	15.04	0.27	8	Dyn
114	E	-86.75	135.43	58.93	29.58	17.75	0.33	10	Dyn
115	N	-104.25	117.95	19.35	26.21	52.85	0.44	10	Dyn
116	P	-61.75	150.12	5.39	9.72	53.79	0.56	10	Dyn
117	C	-132.63	156.95	49.85	20.09	61.95	0.75	10	Warn
118	G	-155.15	165.37	44.48	30.48	81.27	0.77	10	Warn
119	P	-69.55	138.74	15.75	26.77	91.98	0.80	10	Good
121	S	-61.37	-30.58	5.49	20.53	89.68	0.83	10	Good
122	E	-68.39	-33.98	8.29	16.06	64.20	0.78	10	Good
123	R	-92.68	-3.53	9.33	11.57	32.38	0.76	10	Warn
124	R	-112.87	166.57	16.96	20.53	33.41	0.77	5	Warn
125	K	-59.73	-36.53	11.69	12.30	69.21	0.81	10	Good
126	H	-69.76	-28.47	11.34	17.21	55.44	0.84	10	Good
127	L	-94.53	-14.08	21.26	26.58	53.78	0.87	10	Good
128	F	-142.22	150.78	14.73	16.70	31.59	0.90	10	Good
129	V	-120.09	130.60	13.23	7.14	31.02	0.92	10	Good
130	Q	-121.48	133.24	20.11	8.13	30.38	0.91	10	Good
131	D	-94.83	116.59	25.45	23.26	63.81	0.88	10	Good
132	P	-55.65	-33.22	4.68	4.69	43.07	0.85	10	Good
133	Q	-64.27	-36.80	6.17	16.09	40.82	0.85	10	Good
134	T	-103.06	-11.14	14.30	25.91	45.04	0.85	10	Good
135	C	-124.12	151.79	12.85	10.56	48.85	0.86	10	Good
136	K	-123.59	131.80	31.06	23.69	37.58	0.85	10	Good
137	C	-124.82	149.38	11.36	8.21	23.71	0.86	10	Good
138	S	-116.84	140.43	28.33	16.54	67.30	0.86	10	Good
139	C	-61.16	-21.61	7.70	10.00	52.64	0.86	7	Warn
140	K	-64.16	-37.51	8.38	5.93	42.98	0.86	10	Good
141	N	-109.39	1.25	12.94	20.96	27.88	0.87	10	Good
142	T	-120.16	162.57	32.00	6.23	33.95	0.89	10	Good
143	D	-62.17	-39.55	5.68	3.57	25.05	0.92	10	Good
144	S	-62.42	-42.11	3.80	6.96	18.16	0.92	10	Good
145	R	-69.25	-38.43	7.55	12.52	49.86	0.91	10	Good

RESID	RESNAME	PHI	PSI	DPHI	DPSI	DIST	S2	COUNT	CLASS
146	C	-122.90	151.77	21.11	28.85	55.26	0.90	10	Good
147	K	-63.78	-36.21	8.11	11.00	37.64	0.89	10	Good
148	A	-63.77	-26.11	3.16	8.29	14.84	0.88	10	Good
149	R	-94.11	16.57	7.42	5.72	19.75	0.88	10	Good
150	Q	58.34	39.32	5.03	13.29	21.54	0.89	10	Good
151	L	-115.36	153.01	22.99	13.81	24.39	0.90	10	Good
152	E	-133.46	140.58	21.36	15.89	17.73	0.88	10	Good
153	L	-92.15	125.81	25.09	8.60	24.68	0.85	10	Good
154	N	-91.69	114.96	21.61	25.71	26.58	0.84	10	Good
155	E	-60.30	-24.83	5.64	11.99	39.36	0.84	10	Good
156	R	-77.33	-33.01	12.18	12.68	24.88	0.86	10	Good
157	T	-96.70	-8.38	13.25	23.72	47.35	0.87	10	Good
158	C	-132.37	160.40	22.76	9.47	38.32	0.90	10	Good
159	R	-119.58	136.87	21.97	10.31	36.12	0.90	10	Good
160	C	-132.11	143.89	9.40	12.03	32.19	0.87	10	Good
161	D	-118.14	140.42	13.00	14.56	35.66	0.81	10	Good
162	K	-90.32	118.57	32.96	32.51	74.49	0.76	10	Good
163	P	-64.82	151.14	6.01	8.31	88.40	0.73	10	Good

PHI: predicted torsion angle C(i-1)—N(i)—CA(i)—C(i)

PSI: predicted torsion angle N(i)—CA(i)—C(i)—N(i+1)

DPHI and DPSI: estimated standard deviations of the prediction errors in PHI and PSI

DIST: TALOS database matching score

S2: RCI chemical shift order parameter

COUNT: Number of database triplets used to form the torsion angle predictions

CLASS: classification of the prediction result: “None” for no prediction made; “Strong/Generous” is marked when the prediction is likely to be good because there was a majority consensus in the database matches; “Warn” when there is no consensus in the database matches and the prediction should not be used; “Dyn” when the S2 value indicates that the residue is likely dynamic.

Appendix A5 Aptamer sequential walk

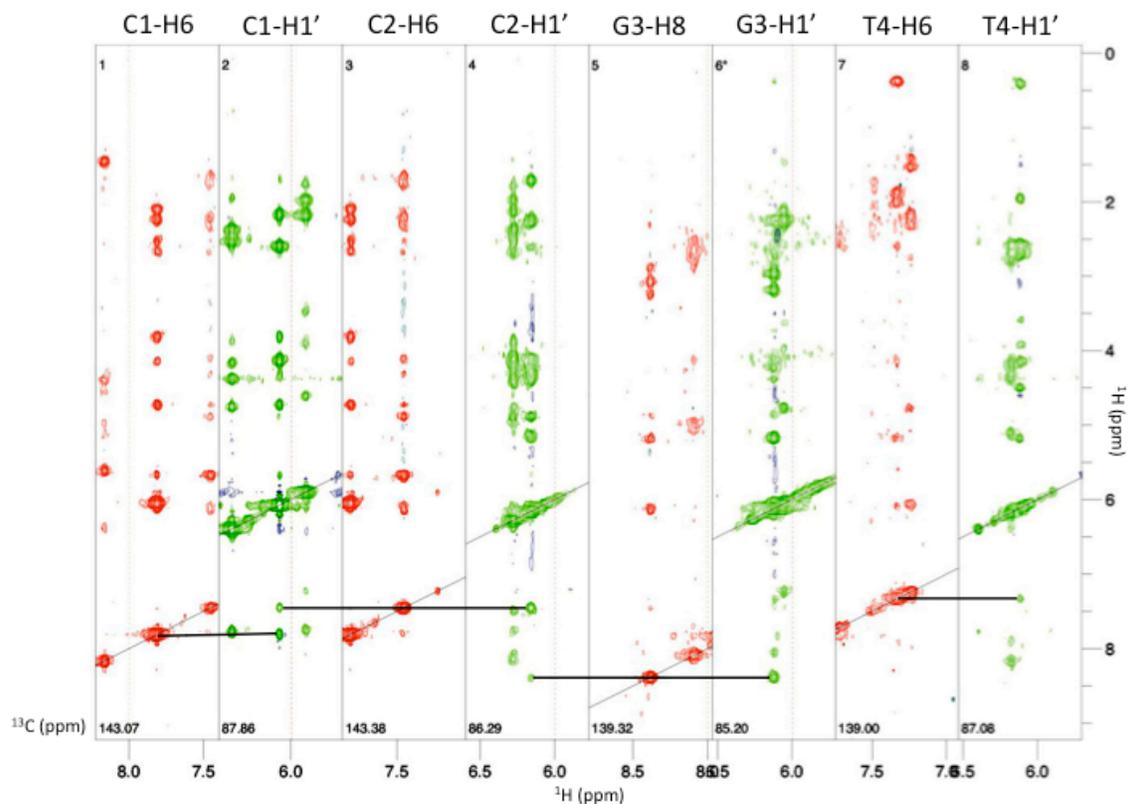


Figure A5.1 Sequential walk from C1 to T4. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1').

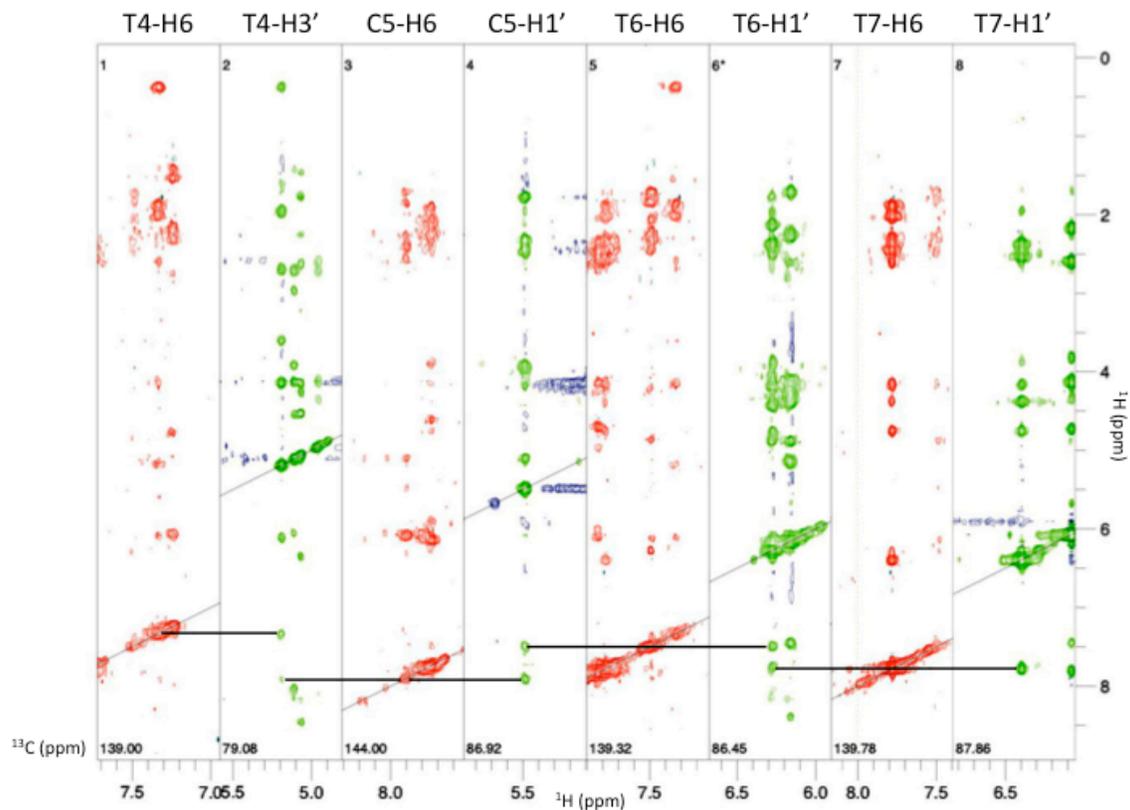


Figure A5.2 Sequential walk from T4 to T7. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1').

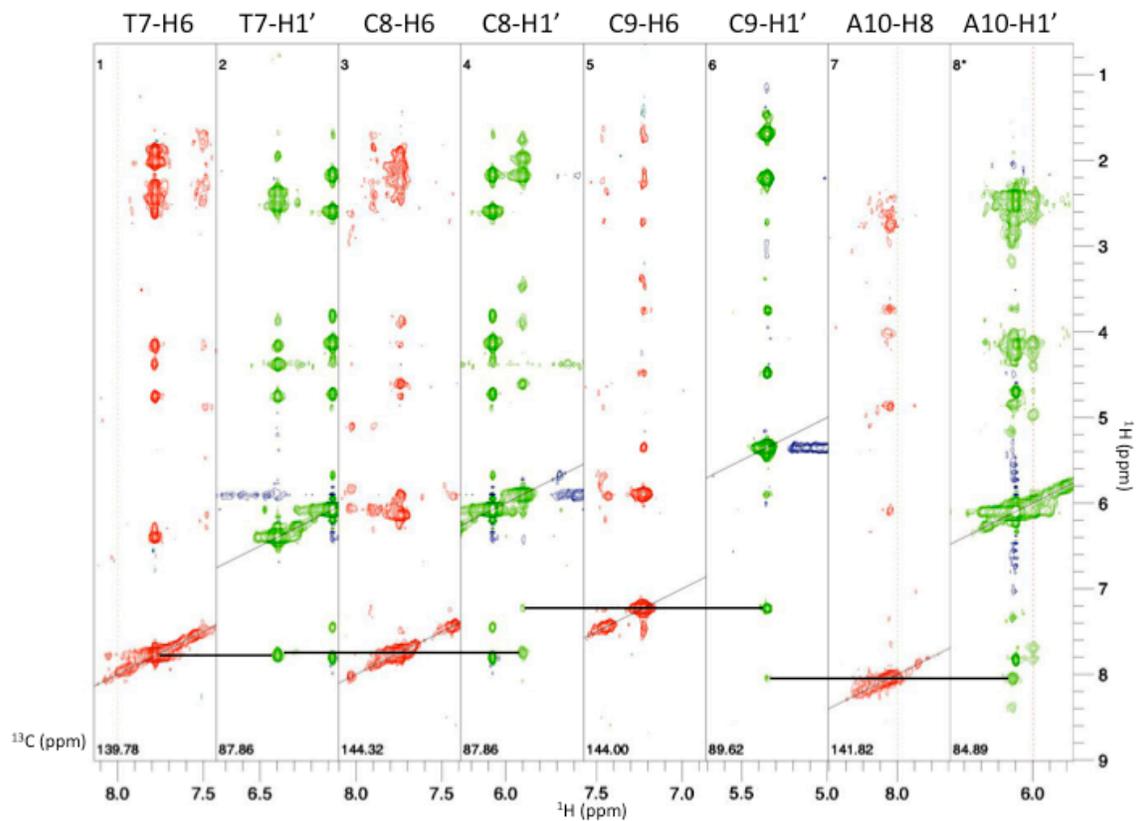


Figure A5.3 Sequential walk from T7 to A10. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1').

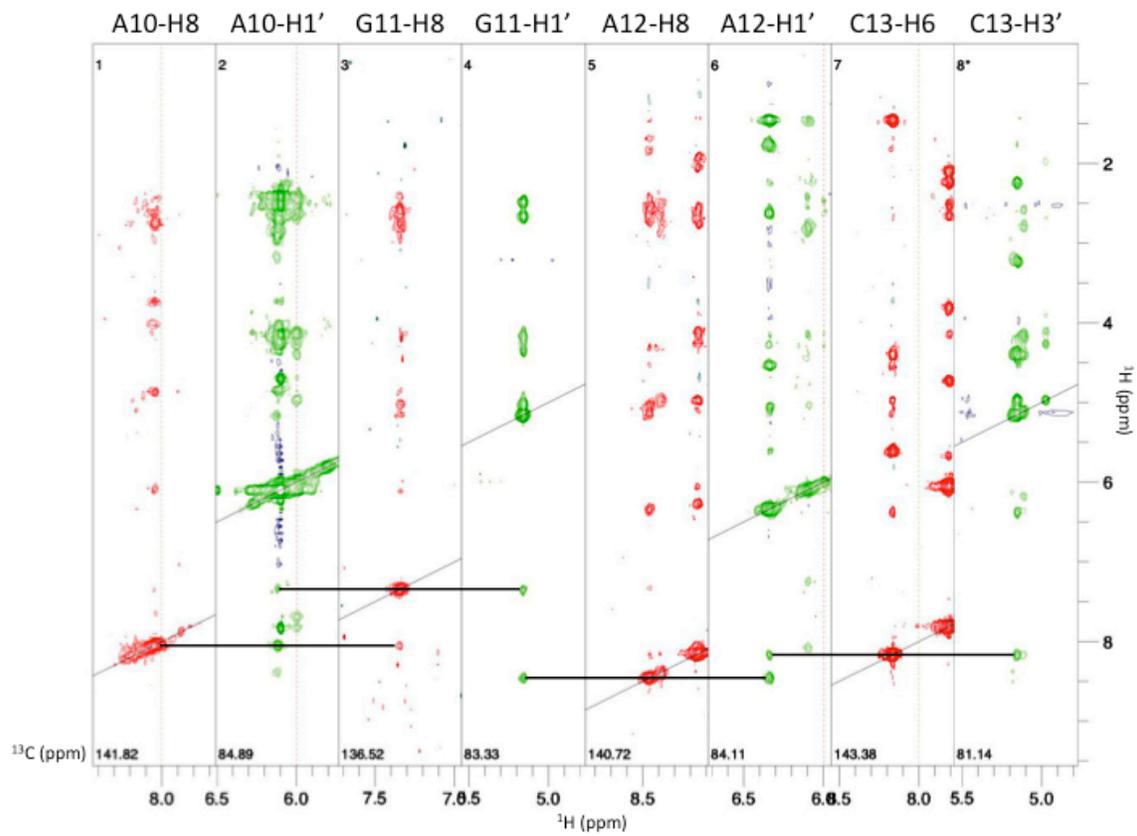


Figure A5.4 Sequential walk from A10 to C13. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1').

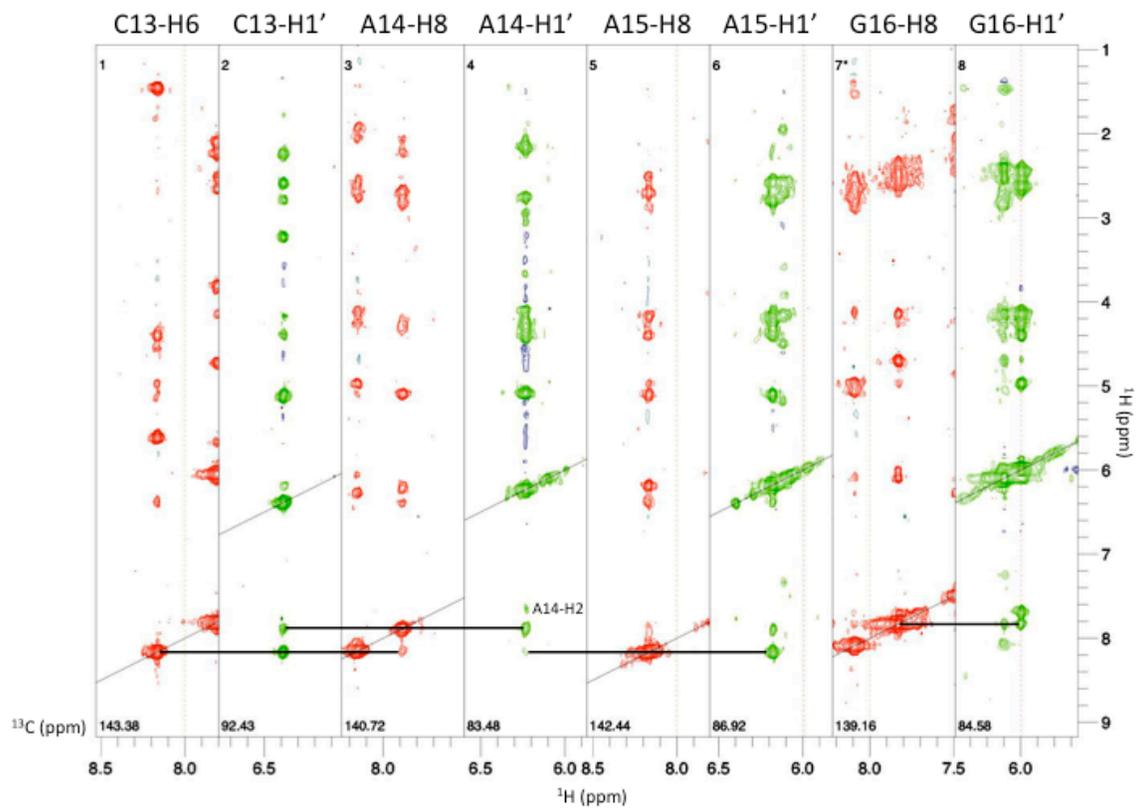


Figure A5.5 Sequential walk from C13 to G16. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1').

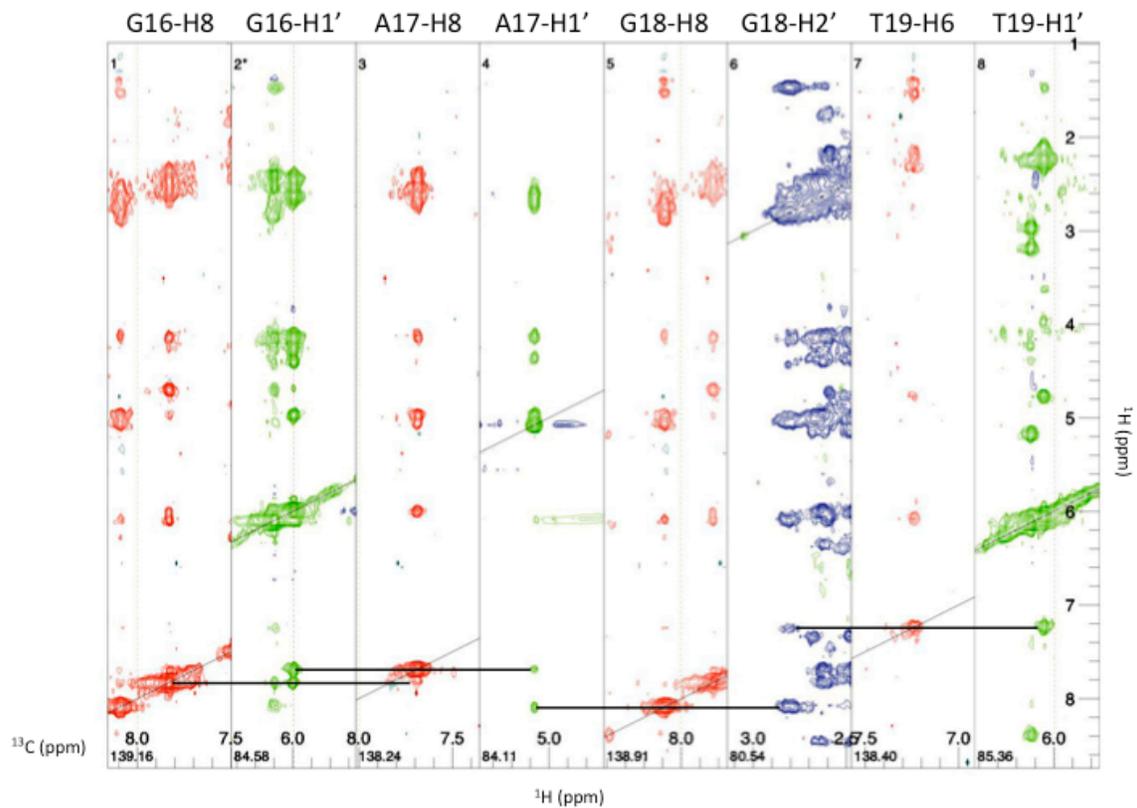


Figure A5.6 Sequential walk from G16 to T19. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1') or blue (H2'/H2'').

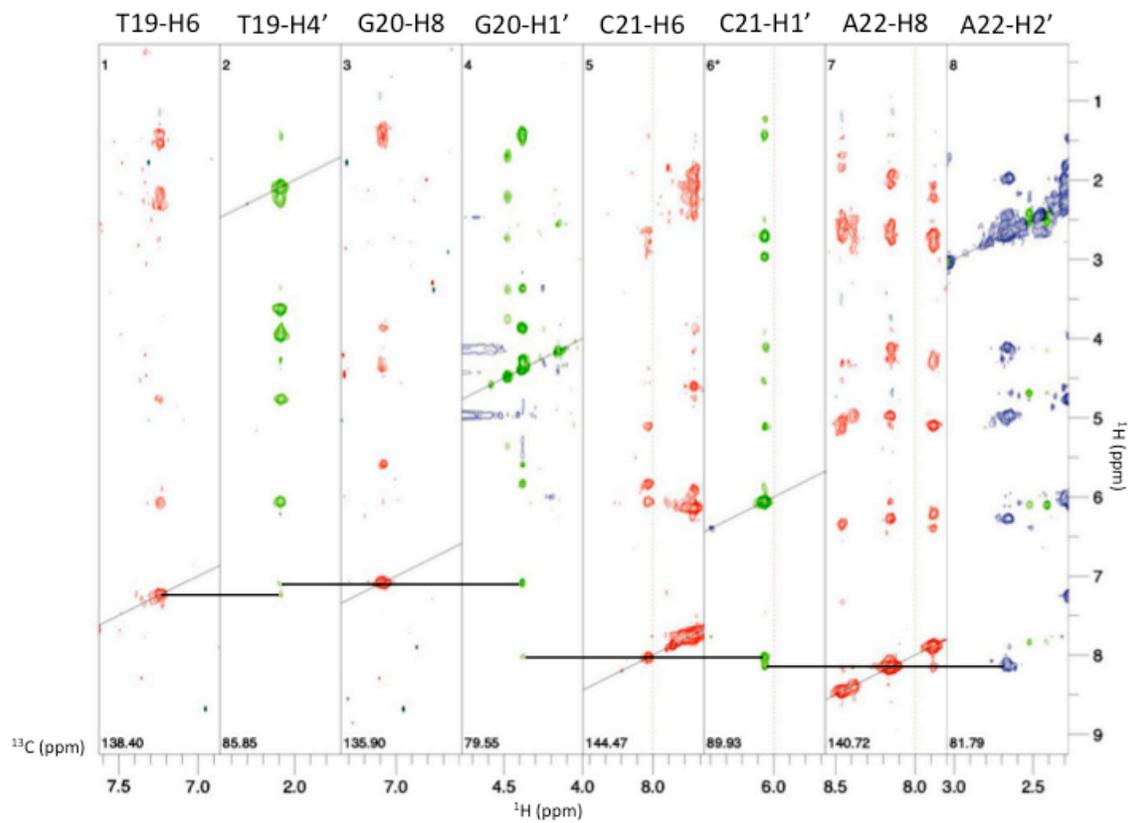


Figure A5.7 Sequential walk from T19 to A22. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1') or blue (H2'/H2'').

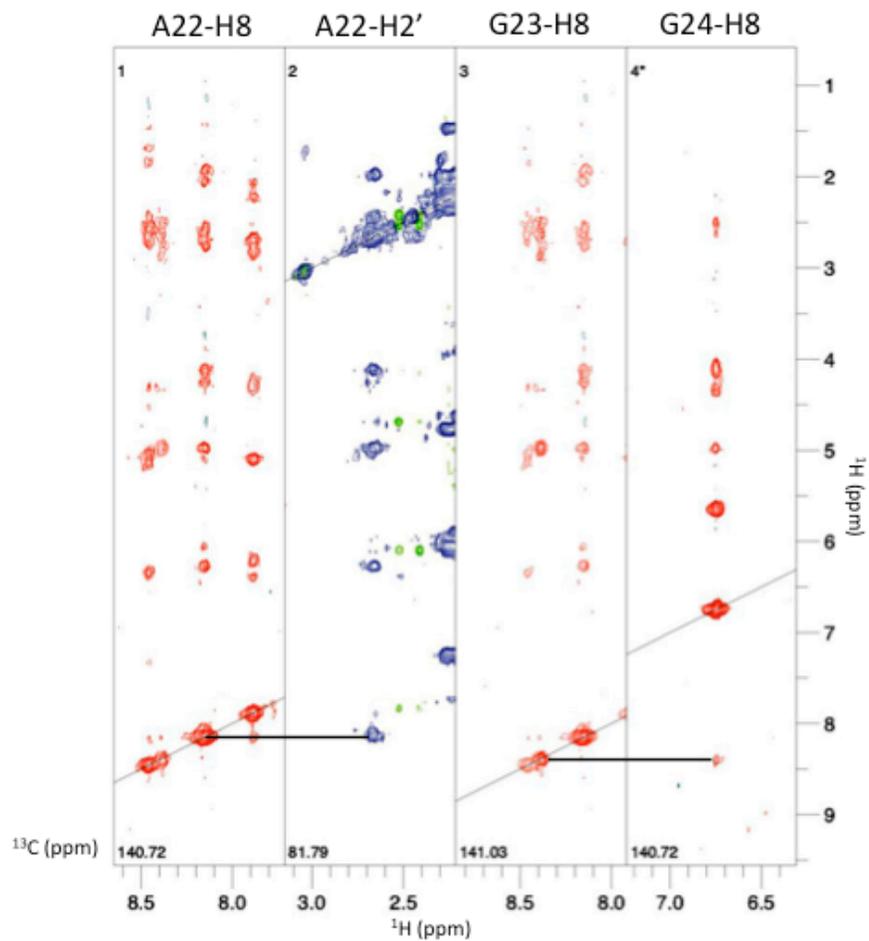


Figure A5.8 Sequential walk from A22 to G24. The aromatic-optimized NOESY-HSQC is shown in red. The deoxyribose-optimized NOESY-HSQC is shown in green (H1') or blue (H2'/H2'').

Appendix A6 Aptamer resonance assignments

Table A6.1 Aptamer aromatic resonance assignments*

Aromatic atoms		N1	H1	C2	H2	N3	H3	C5	H5	C6	H6	C8	H8
1	C		-		-		-	100.2	6.04	143.5	7.82	-	-
2	C		-		-	196.5	-	98.9	5.68	143.8	7.46	-	-
3	G	148.7	12.68		-		-		-		-	139.6	8.39
4	T		-		-	159.4	14.30		-	139.4	7.34	-	-
5	C		-		-	196.7	-	100.1	6.08	144.3	7.91	-	-
6	T		-		-	160.7	14.47		-	139.7	7.49	-	-
7	T		-		-				-	140.1	7.78	-	-
8	C		-		-		-	99.2	6.14	144.7	7.74	-	-
9	C		-		-		-	97.6	5.90	144.5	7.23	-	-
10	A		-	154.4	8.07	233.9	-		-		-	142.1	8.05
11	G	147.2	13.18		-		-		-		-	137.0	7.35
12	A	222.4	-	156.2	8.27		-		-		-	141.0	8.46
13	C		-		-	199.1	-	101.3	5.60	143.7	8.16	-	-
14	A		-	154.6	7.98		-		-		-	142.8	8.17
15	A		-	154.7	7.65				-		-	141.1	7.89
16	G				-		-		-		-	139.5	7.83
17	A		-	155.8	8.41		-		-		-	138.5	7.69
18	G	151.0	12.49		-	236.5	-		-		-	139.4	8.09
19	T		-		-	159.2	13.73		-	139.0	7.25	-	-
20	G				-		-		-		-	136.1	7.09
21	C		-		-		-	97.8	5.84	144.9	8.03	-	-
22	A		-	155.7	8.16		-		-		-	141.1	8.16
23	G	147.7	11.14		-		-		-		-	141.2	8.39
24	G	146.9	13.44		-		-		-		-	141.1	6.75
25	G	147.8	13.01		-		-		-		-		

*Chemical shifts were referenced using trimethylsilylpropanoic acid (TSP) at 0 ppm. Referencing for ^{13}C and ^{15}N was done using the SetReference script.

Table A6.2 Aptamer deoxyribose resonance assignments*

Deoxyribose atoms		H1'	C1'	H2'	H2''	C2'	H3'	C3'	H4'	C4'	H5'	H5''	C5'
1	C	6.07	87.9	2.17	2.59	39.7	4.73	77.7	4.15	88.2	3.81		63.6
2	C	6.16	86.3	1.71	2.25	38.2			4.31	86.5			
3	G	6.12	85.2	2.97	3.18	39.0	5.18	81.6					
4	T	6.12	87.1	1.96	2.68	38.2	5.20	79.1	4.15	86.5	3.60	4.11	66.4
5	C	5.48	86.9	2.32	2.47	38.2	5.10	76.0	3.96	86.5			
6	T	6.27	86.5	2.12	2.38	40.1			4.44	86.8			
7	T	6.40	87.9	2.38	2.52	40.5			4.38	87.7	4.16		68.5
8	C	5.90	87.9	1.98	2.17	41.4	4.61	77.7	3.47	87.4	3.83	3.90	67.1
9	C	5.35	89.6	1.69	2.21	42.5	4.47	79.9	3.75	86.2	2.73	3.38	68.0
10	A	6.12	84.9	2.70	2.87	38.7			4.25	89.0	3.72	4.01	70.3
11	G	5.16	83.3	2.49	2.67	40.1	5.02	77.4	4.36	86.5			
12	A	6.34	84.1				5.07	79.1	4.53	88.3			
13	C	6.39	92.4	2.24	3.22	44.4	5.15	81.1	4.97	89.1			
14	A	6.18	86.9	2.59	2.79	39.1	5.11	80.8					
15	A	6.23	83.5	2.15	2.75	40.8	5.08	76.9	4.29	86.2	3.69	4.28	66.1
16	G	5.99	84.6	2.45	2.62	40.1	4.97	79.6	4.40	87.3			
17	A	5.08	84.1		2.71	39.1	4.96	78.8	4.35	87.3			
18	G	6.09	84.4		2.81	40.2							
19	T	6.05	85.4	2.23	2.26	41.4			2.09	85.9	3.63	3.95	66.8
20	G	5.59	84.7	1.40	1.47	36.5	4.38	79.6	4.26	89.0	3.36	3.86	69.4
21	C	6.06	89.9	2.70	2.96	37.4	5.11	78.9					
22	A			1.98	2.67	41.4	4.97	80.8	4.26	88.0	3.89	4.11	69.9
23	G												
24	G								4.36	88.2			
25	G												

*Chemical shifts were referenced using trimethylsilylpropanoic acid (TSP) at 0 ppm. Referencing for ^{13}C and ^{15}N was done using the SetReference script.

Table A6.3 Aptamer resonance assignments for exocyclic groups*

Exocyclic atoms		amino N	amino H		Methyl C	Methyl H
1	C	97.6	7.00	8.37	-	-
2	C	99.7	7.33	8.95	-	-
3	G				-	-
4	T	-	-	-		1.93
5	C	101.6	7.67	9.16	-	-
6	T	-	-	-		1.77
7	T	-	-	-		1.95
8	C				-	-
9	C				-	-
10	A	80.4	7.03	7.34	-	-
11	G					
12	A	86.9	7.86	8.23	-	-
13	C	95.9	6.73	7.41	-	-
14	A				-	-
15	A				-	-
16	G					
17	A				-	-
18	G					
19	T	-	-	-		1.46
20	G					
21	C					
22	A				-	-
23	G					
24	G					
25	G					

*Chemical shifts were referenced using trimethylsilylpropanoic acid (TSP) at 0 ppm. Referencing for ^{13}C and ^{15}N was done using the SetReference script.

Appendix A7 Xplor-NIH scripts used for structure calculations

A7.1 Script used for simulated annealing

```
#
# Total number of structures.
#
nstructures = 200

#
# Base name for output PDB files.
# ---
# This string must contain the "STRUCTURE" literal to be replaced by the
# structure number in the PDB filename. The (optional) "SCRIPT" literal is
# replaced by the name of this file (or stdin if redirected using <).
#
outfilename = 'SCRIPT_STRUCTURE.pdb'

#
# Set random seed.
#
import protocol
protocol.initRandomSeed(24021) # by specific seed number

#
# Read PSF file(s).
#
protocol.initStruct(['DNA.psf', 'HBD.psf'])

#
# Load paramaters.
# ---
# Read covalent and nonbonded parameters from parameter file(s).
# Note that only covalent parameters for bond lengths and angles, and impropers
# dihedrals are used in this script. The torsion angle parameters, if any, are
# omitted because they are provided by a statistical potential below.
#
protocol.initParams(['nucleic-3.1.par', 'protein.par'])

#
# Generate extended conformation with satisfied covalent geometry.
```

```

#
protocol.genExtendedStructure()

#read input structure
#import glob
#infilename = glob.glob('complex.pdb')[0]
#protocol.initCoords(infilename)

#Add disulfide bonds
#protocol.addDisulfideBond('(resid 117 and name SG)','(resid 135 and name SG)')
#protocol.addDisulfideBond('(resid 120 and name SG)','(resid 137 and name SG)')
#protocol.addDisulfideBond('(resid 139 and name SG)','(resid 158 and name SG)')
#protocol.addDisulfideBond('(resid 146 and name SG)','(resid 160 and name SG)')

#
# Create a PotList() to contain the energy terms that will be active
# during structure calculations.
#
from potList import PotList
potList = PotList()

#
# Lists highTempParams and rampedParams will hold simulationTools.StaticRamp and
# simulationTools.MultRamp objects to handle parameter changes between the high
# temperature and annealing stage, and within the annealing stage (e.g., ramped
# force constants).
#
from simulationTools import StaticRamp, MultRamp, InitialParams, AnnealIVM

highTempParams = []
rampedParams = []

# Below, the entire setup of each energy term (including their treatment during
# the high temperature and annealing stages) is performed in self-contained
# sections, so that removal of a term or addition of a new one can be done
# simply by commenting out or adding the corresponding section, respctively.

#
# Set up distance restraint potential (e.g., from NOEs).
#
import noePotTools
noe = PotList('noe')

```

```

for (name, scale, table) in [('all', 1, 'all_noe_no_bp.tbl'),
    # add entries for more restraint tables here
    ]:
    pot = noePotTools.create_NOEPot(name, table)
    # pot.setPotType("soft") # if you suspect there are bad NOEs
    pot.setScale(scale)
    noe.append(pot)
potList.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale( VALUE )"))

#
# Set up torsion angle restraint potential (e.g., from J-couplings).
#
import xplorPot
dihedralTables = ['DNA_dih.tbl', 'HBD_111-165_dih.tbl']
protocol.initDihedrals(dihedralTables)
potList.append(xplorPot.XplorPot('CDIH'))
highTempParams.append(StaticRamp("potList['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("potList['CDIH'].setScale(200)"))

#
# Set up potential for base-pair planarity restraints.
#
#protocol.initPlanarity('ssDNA_plane.tbl')
#potList.append(xplorPot.XplorPot('PLAN'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up statistical torsion angle potential (torsionDB).
#
import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name='torsiondb')
potList.append(torsiondb)
rampedParams.append(MultRamp(0.5, 4, "torsiondb.setScale(VALUE)"))

#
# setup parameters for atom-atom repulsive term. (van der Waals-like term)
#
from repelPotTools import create_RepelPot,initRepel
repel = create_RepelPot('repel')
potList.append(repel)
rampedParams.append( StaticRamp("initRepel(repel,useI4=False") ) )

```

```

rampedParams.append( MultRamp(.004,4, "repel.setScale( VALUE)") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("initRepel(repel,
    use14=True,
    scale=0.004,
    repel=1.2,
    moveTol=45,
    interactingAtoms="name C1"
)")) )

# Selected 1-4 interactions.
import torsionDBPotTools
repel14 = torsionDBPotTools.create_Terminal14Pot('repel14')
potList.append(repel14)
highTempParams.append(StaticRamp("repel14.setScale(0)")
rampedParams.append(MultRamp(0.004, 4, "repel14.setScale(VALUE)"))

#
# Set up bond length potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('BOND'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up bond angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('ANGL'))
rampedParams.append(MultRamp(0.4, 1.0, "potList['ANGL'].setScale(VALUE)"))

#
# Set up improper dihedral angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potList.append(xplorPot.XplorPot('IMPR'))
rampedParams.append(MultRamp(0.1, 1.0, "potList['IMPR'].setScale(VALUE)"))

#
# Done with energy terms.

#
# Give atoms uniform weights, except for anisotropy axes (if any).

```

```

#
protocol.massSetup()

#
# Set up IVM object(s).
#

# IVM object for torsion-angle dynamics/minimization.
import ivm
dyn = ivm.IVM()

# Make ligand (resid 34) rigid.
#dyn.group("resid 34 and (name n1 or name c1 or name c2 or name o2
#   or name n3 or name c3 or name c4 or name n9 or name c8
#   or name n7 or name c5 or name c6 or name o6 or name h7)")

# Argument flexRiboseRing below is a string that selects residues whose ribose
# rings will have all endocyclic angles flexible. In general, all ribose rings
# should be selected. In this example, the non-RNA ligand (residue 34) has to
# be excluded.
protocol.torsionTopology(dyn, flexRiboseRing='resid 1:25')

## Optional IVM object for final Cartesian minimization.
##minc = ivm.IVM()
##protocol.cartesianTopology(minc)

#
# Temperature set up.
#
temp_ini = 3500.0 # initial temperature
temp_fin = 25.0  # final temperature

def calcOneStructure(loopInfo):
    """Calculate a structure.

    """
    # Generate initial structure by randomizing torsion angles.
    import monteCarlo
    monteCarlo.randomizeTorsions(dyn)

    # Randomize position of ligand.
    # (In the case where there is no ligand, comment out the next two lines.)
    #import atomAction

```

```

#atomAction.randomizeDomainPos('resid 34', deltaPos=45)

# Set torsion angles from restraints.
# (They start satisfied, allowing the shortening of high temp dynamics.)
import torsionTools
torsionTools.setTorsionsFromTable(dihedralTables)

# Fix up covalent geometry.
# (The torsion restraints may include ring torsions and distort geometry.)
while True:
    try:
        protocol.fixupCovalentGeom(maxIters=100, useVDW=1)
        break
    except protocol.CovalentViolation:
        pass

#
# High Temperature Dynamics Stage.
#

# Initialize parameters for high temperature dynamics.
InitialParams(rampedParams)
InitialParams(highTempParams) # purposely overrides some
                               # setups in rampedParams

# Set up IVM object and run.
protocol.initDynamics(dyn,
                    potList=potList,
                    bathTemp=temp_ini,
                    initVelocities=1,
                    finalTime=15, # run for finalTime or
                    numSteps=15001, # numSteps * 0.001, whichever is less
                    printInterval=100)

dyn.setETolerance(temp_ini/100) # used to find stepsize (default: temp/1000)

dyn.run()

#
# Simulated Annealing Stage.
#

# Initialize parameters for annealing.
InitialParams(rampedParams)

# Set up IVM object for annealing.

```

```

protocol.initDynamics(dyn,
                      potList=potList,
                      finalTime=0.7, # run for finalTime or
                      numSteps=701, # numSteps * 0.001, whichever is less
                      printInterval=100)

# Set up cooling loop and run.
AnnealIVM(initTemp=temp_ini,
           finalTemp=temp_fin,
           tempStep=7.0,
           ivm=dyn,
           rampedParams=rampedParams).run()

#
# Torsion angle minimization.
#
protocol.initMinimize(dyn,
                      potList=potList,
                      printInterval=50)
dyn.run()

## #
## # Cartesian minimization (optional).
## #
## protocol.initMinimize(minc,
##                       potList=potList,
##                       dEPred=10)
## minc.run()

from simulationTools import StructureLoop
StructureLoop(numStructures=nstructures,
              pdbTemplate=outfilename,
              doWriteStructures=True,
              structLoopAction=calcOneStructure,
              # Arguments for generating structure statistics:
              genViolationStats=True,
              averageSortPots=[potList['noe'], # terms for structure sorting
                               potList['CDIH']],
              averageTopFraction=0.05, # top fraction of structs. to report on
              averagePotList=potList, # terms analyzed
              averageFitSel='not (name H* or rename ANI)', # selection to fit
              ).run()
# to average struct.
# and report precision

```

A7.2 Script used for refinement

```
#
# Total number of structures.
#
nstructures = 200

#
# Base name for output PDB files.
# ---
# This string must contain the "STRUCTURE" literal to be replaced by the
# structure number in the PDB filename. The (optional) "SCRIPT" literal is
# replaced by the name of this file (or stdin if redirected using <).
#
outfile = 'SCRIPT_STRUCTURE.pdb'

#
# Set random seed.
#
import protocol
protocol.initRandomSeed(40888) # by specific seed number

#
# Read PSF file(s).
#
protocol.initStruct(['HBD.psf', 'DNA.psf'])

#
# Load parameters.
# ---
# Read covalent and nonbonded parameters from parameter file(s).
# Note that only covalent parameters for bond lengths and angles, and impropers
# dihedrals are used in this script. The torsion angle parameters, if any, are
# omitted because they are provided by a statistical potential below.
#
protocol.initParams(['nucleic-3.1.par', 'protein.par'])

#
# Read input structure.
#
import glob
infile = glob.glob('*.best')[0] # name of top-ranked struct. from fold.py
```

```

infilename = glob.glob('sa1_52.pdb')[0] # name of top-ranked structure from sa_complex.py
protocol.initCoords(infilename)

#
# Create a potList.PotList() to contain the energy terms that will be active
# during structure calculations.
#
import potList
potlist = potList.PotList()

#Add Disulfide bonds
#10/31/16 SH
protocol.addDisulfideBond('(resid 117 and name SG)','(resid 135 and name SG)')
protocol.addDisulfideBond('(resid 120 and name SG)','(resid 137 and name SG)')
protocol.addDisulfideBond('(resid 139 and name SG)','(resid 158 and name SG)')
protocol.addDisulfideBond('(resid 146 and name SG)','(resid 160 and name SG)')

#
# Lists highTempParams and rampedParams will hold simulationTools.StaticRamp and
# simulationTools.MultRamp objects to handle parameter changes between the high
# temperature and annealing stage, and within the annealing stage (e.g., ramped
# force constants).
#
from simulationTools import StaticRamp, MultRamp, InitialParams, AnnealIVM

highTempParams = []
rampedParams = []

# Below, the entire setup of each energy term (including their treatment during
# the high temperature and annealing stages) is performed in self-contained
# sections, so that removal of a term or addition of a new one can be done
# simply by commenting out or adding the corresponding section, respectively.

#
# Set up RDC potential.
#

# List with RDC data.
# ---
# Each entry, a tuple, must contain strings for (in this order): arbitrary name
# for the alignment medium, arbitrary name for the experiment (e.g., "CH" for
# C-H RDCs), and the path of the corresponding restraint table.
#     medium exp. table

```

```

#rdcData = [('pf1', 'CH', 'dip_all.tbl'),
#           # add additional entries here
#           ]

# Orientation tensor(s).
#import varTensorTools
#tensors = {}
#for item in rdcData:
#  medium = item[0]
#  tensors[medium] = varTensorTools.create_VarTensor(medium)

# RDCs per se.
#import rdcPotTools
##rdcPotTools.correctGyromagneticSigns() # change sign of NH RDCs after scaling
#rdcs = potList.PotList('rdc')
#for (medium, exp, table) in rdcData:
#  name = '%s_%s' % (exp, medium)
#  rdc = rdcPotTools.create_RDCPot(name, table, tensors[medium])
#  rdcPotTools.scale_toCH(rdc)
#  rdcs.append(rdc)
#potlist.append(rdcs)
#rampedParams.append(MultRamp(0.05,1.0, "rdcs.setScale(VALUE)"))

# Calculate initial tensor orientation and setup tensor calculation during
# simulated annealing.
#for tensor in tensors.keys():
#  highTempParams.append(StaticRamp(
#    "varTensorTools.calcTensor(tensors['%s'])" % medium))

#
# Set up distance restraint potential (e.g., from NOEs).
#
import noePotTools
noe = potList.PotList('noe')
for (name, scale, table) in [('all', 1, 'all_noe.tbl'),
# add entries for more restraint tables here
#
#]:
#  pot = noePotTools.create_NOEPot(name, table)
#  pot.setPotType("soft") # if you suspect there are bad NOEs
#  pot.setScale(scale)
#  noe.append(pot)
potlist.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale(VALUE)"))

```

```

#hbda- distance/angle bb hbond terms
#
import xplorPot
protocol.updatePseudoAtoms()

#hbdb - knowledge-based backbone hydrogen bond term
#
protocol.initHBDB()
potlist.append( xplorPot.XplorPot('HBDB') )

#
# Set up torsion angle restraint potential (e.g., from J-couplings).
#
dihedralTables = ['DNA_dih.tbl', 'HBD_111-165_dih.tbl']
protocol.initDihedrals(dihedralTables)
potlist.append(xplorPot.XplorPot('CDIH'))
highTempParams.append(StaticRamp("potlist['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("potlist['CDIH'].setScale(200)"))

#
# Set up potential for base-pair planarity restraints.
#
protocol.initPlanarity('ssDNA_plane.tbl')
potlist.append(xplorPot.XplorPot('PLAN'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up statistical torsion angle potential (torsionDB).
#
import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name='torsiondb')
potlist.append(torsiondb)
rampedParams.append(MultRamp(0.5, 4, "torsiondb.setScale(VALUE)"))

#
# Setup interatomic repulsion potential (van der Waals-like term).
#
potlist.append(xplorPot.XplorPot('VDW'))
# C1'-only interactions with large atomic radius to improve sampling.
highTempParams.append(StaticRamp("""protocol.initNBond(cutnb=100,
                                     rcon=0.004,
                                     tolerance=45,

```

```

        repel=1.2,
        selStr="name C1'")'''''))
# All interactions active with more realistic atomic radii.
rampedParams.append(StaticRamp("protocol.initNBond(nbxmod=4)")
rampedParams.append(StaticRamp(
    "xplor.command('param nbonds repel 0.9 end end'"))
rampedParams.append(MultRamp(0.004, 4.0,
    "xplor.command('param nbonds rcon VALUE end end'"))

#
# Set up bond length potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potlist.append(xplorPot.XplorPot('BOND'))
# (The setup of this term remains unchanged throughout; no need to involve
# highTempParams and/or rampedParams.)

#
# Set up bond angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potlist.append(xplorPot.XplorPot('ANGL'))
rampedParams.append(MultRamp(0.4, 1.0, "potlist['ANGL'].setScale(VALUE)")

#
# Set up improper dihedral angle potential.
# (Needed even if no Cartesian minimization is used, for "broken" rings.)
#
potlist.append(xplorPot.XplorPot('IMPR'))
rampedParams.append(MultRamp(0.1, 1.0, "potlist['IMPR'].setScale(VALUE)")

#
# Set up statistical base-base positional potential.
# ---
# The selection argument in initOrie below should select all residues to be
# affected by the potential.
# In this particular example, the small-molecule ligand (residue 34) is omitted,
# along with residue 27, known from preliminary calculations to stick out of the
# structure. In general, however, all RNA residues should be selected (e.g.,
# for an isolated RNA molecule use selection='all').
# Reference: Clore, GM & Kuszewski, J, (2003) J. Am. Chem. Soc. 125:1518-1525.
#
#protocol.initOrie(system='rna', selection='resid 1:26 or resid 28:33')

```

```

protocol.initOrie(system='rna', selection='resid 1:25')
potlist.append(xplorPot.XplorPot('ORIE'))
rampedParams.append(MultRamp(0.002,0.3,"xplor.command('orie scale VALUE end')"))

#
# Done with energy terms.

#
# Give atoms uniform weights, except for anisotropy axes (if any).
#
protocol.massSetup()

#
# Set up IVM object(s).
#

# IVM object for torsion-angle dynamics/minimization.
import ivm
dyn = ivm.IVM()

# Make ligand (resid 34) rigid.
#dyn.group("resid 34 and (name n1 or name c1 or name c2 or name o2
#   or name n3 or name c3 or name c4 or name n9 or name c8
#   or name n7 or name c5 or name c6 or name o6 or name h7)")
#
# Orientation tensor setup.
# ---
# If RDCs are not included in the calculations by commenting out their setup
# section above, this set up has to be commented out also.
#for tensor in tensors.values():
#   tensor.setFreedom("fixDa, fixRh") # fix tensor Rh, Da, vary orientation
#   tensor.setFreedom("varyDa, varyRh") # vary tensor Rh, Da, vary orientation

# Argument flexRiboseRing below is a string that selects residues whose ribose
# rings will have all endocyclic angles flexible. In general, all ribose rings
# should be selected. In this example, the non-RNA ligand (residue 34) has to
# be excluded.
protocol.torsionTopology(dyn, flexRiboseRing='resid 1:25')

## Optional IVM object for final Cartesian minimization.
##minc = ivm.IVM()
##
##for tensor in tensors.values():
##   tensor.setFreedom("varyDa, varyRh") # allow all tensor parameters float here

```

```

##
##protocol.cartesianTopology(minc)

#
# Temperature set up.
#
temp_ini = 1000.0 # initial temperature
temp_fin = 25.0 # final temperature

def calcOneStructure(loopInfo):
    """Calculate a structure.

    """
    #
    # High Temperature Dynamics Stage.
    #

    # Initialize parameters for high temperature dynamics.
    InitialParams(rampedParams)
    InitialParams(highTempParams) # purposely overrides some
        # setups in rampedParams

    # Set up IVM object and run.
    protocol.initDynamics(dyn,
        potList=potlist,
        bathTemp=temp_ini,
        initVelocities=1,
        finalTime=15, # run for finalTime or
        numSteps=15001, # numSteps * 0.001, whichever is less
        printInterval=100)

    dyn.setETolerance(temp_ini/100) # used to find stepsize (default: temp/1000)

    dyn.run()

    #
    # Simulated Annealing Stage.
    #

    # Initialize parameters for annealing.
    InitialParams(rampedParams)

    # Set up IVM object for annealing.

```

```

protocol.initDynamics(dyn,
                    potList=potlist,
                    finalTime=0.70, # run for finalTime or
                    numSteps=701, # numSteps * 0.001, whichever is less
                    printInterval=100)

# Set up cooling loop and run.
AnnealIVM(initTemp=temp_ini,
          finalTemp=temp_fin,
          tempStep=7.0,
          ivm=dyn,
          rampedParams=rampedParams).run()

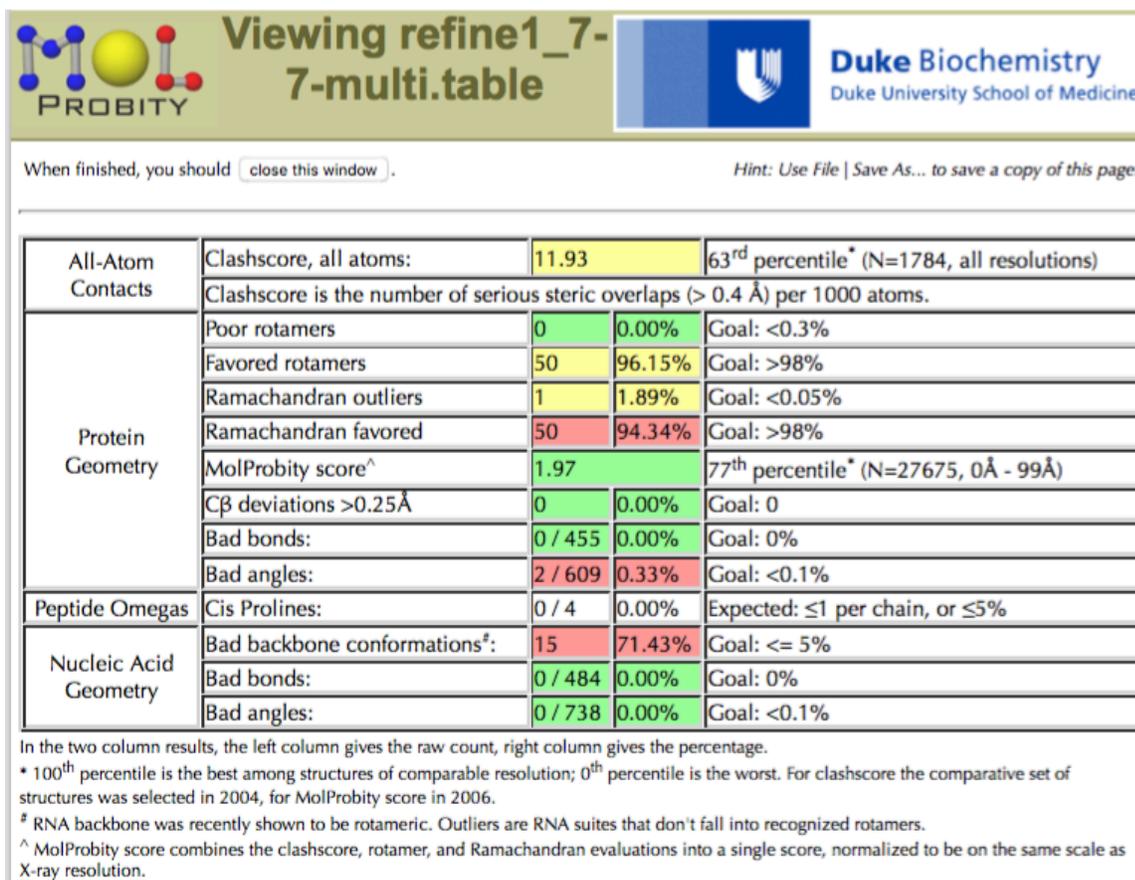
#
# Torsion angle minimization.
#
protocol.initMinimize(dyn,
                    potList=potlist,
                    printInterval=50)
dyn.run()

## #
## # Cartesian minimization (optional).
## #
## protocol.initMinimize(minc,
##                       potList=potlist,
##                       dEPred=10)
## minc.run()

from simulationTools import StructureLoop
StructureLoop(numStructures=nstructures,
             pdbFilesIn=infilename,
             pdbTemplate=outfilename,
             doWriteStructures=True,
             structLoopAction=calcOneStructure,
             # Arguments for generating structure statistics:
             genViolationStats=True,
             averageSortPots=[potlist['noe'], # terms for structure sorting
                             potlist['CDIH']],
#                             potlist['rdc']],
             averageTopFraction=0.05, # top fraction of structs. to report on
             averagePotList=potlist, # terms analyzed
             averageFitSel='not (name H* or rename ANI)',
             ).run()

```

Appendix A8 MolProbity results from all structures in the ensemble





Viewing refine1_29-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should .

Hint: Use File | Save As... to save a copy of this page.

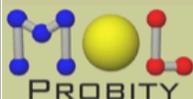
All-Atom Contacts	Clashscore, all atoms:	8.35		80 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	1.92%	Goal: <0.3%
	Favored rotamers	50	96.15%	Goal: >98%
	Ramachandran outliers	1	1.89%	Goal: <0.05%
	Ramachandran favored	52	98.11%	Goal: >98%
	MolProbity score [^]	1.67		90 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [‡] :	15	71.43%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

[‡] RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_53-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

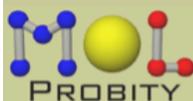
All-Atom Contacts	Clashscore, all atoms:	13.13		57 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	3.85%	Goal: <0.3%
	Favored rotamers	46	88.46%	Goal: >98%
	Ramachandran outliers	1	1.89%	Goal: <0.05%
	Ramachandran favored	50	94.34%	Goal: >98%
	MolProbity score [^]	2.46		49 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	16	76.19%	Goal: ≤5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_54-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should .

Hint: Use File | Save As... to save a copy of this page.

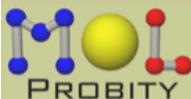
All-Atom Contacts	Clashscore, all atoms:	8.35		80 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	49	94.23%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	49	92.45%	Goal: >98%
	MolProbity score [^]	1.92		80 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	16	76.19%	Goal: ≤= 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_57-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

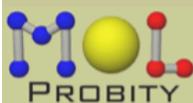
All-Atom Contacts	Clashscore, all atoms:	4.77		94 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	3.85%	Goal: <0.3%
	Favored rotamers	48	92.31%	Goal: >98%
	Ramachandran outliers	1	1.89%	Goal: <0.05%
	Ramachandran favored	50	94.34%	Goal: >98%
	MolProbity score [^]	2.08		72 nd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	15	71.43%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

[#] RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_64-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

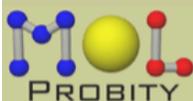
All-Atom Contacts	Clashscore, all atoms:	11.34		65 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	5.77%	Goal: <0.3%
	Favored rotamers	48	92.31%	Goal: >98%
	Ramachandran outliers	1	1.89%	Goal: <0.05%
	Ramachandran favored	50	94.34%	Goal: >98%
	MolProbity score [^]	2.53		45 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	15	71.43%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_106-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

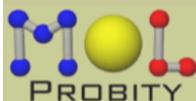
All-Atom Contacts	Clashscore, all atoms:	10.74		68 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	3.85%	Goal: <0.3%
	Favored rotamers	47	90.38%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	51	96.23%	Goal: >98%
	MolProbity score [^]	2.25		62 nd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	15	71.43%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

[#] RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_126-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

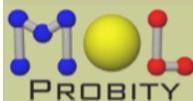
All-Atom Contacts	Clashscore, all atoms:	7.16		86 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	1.92%	Goal: <0.3%
	Favored rotamers	51	98.08%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	50	94.34%	Goal: >98%
	MolProbity score [^]	1.99		76 th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	16	76.19%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_140-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

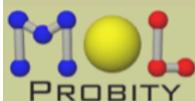
All-Atom Contacts	Clashscore, all atoms:	7.16		86 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	1.92%	Goal: <0.3%
	Favored rotamers	48	92.31%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	51	96.23%	Goal: >98%
	MolProbity score [^]	1.86		82 nd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
	Bad angles:	2 / 609	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	15	71.43%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

[#] RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.



Viewing refine1_147-2- multi.table



Duke Biochemistry
Duke University School of Medicine

When finished, you should [close this window](#) .

Hint: Use File | Save As... to save a copy of this page.

All-Atom Contacts	Clashscore, all atoms:	6.56		89 th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	5.77%	Goal: <0.3%
	Favored rotamers	49	94.23%	Goal: >98%
	Ramachandran outliers	1	1.89%	Goal: <0.05%
	Ramachandran favored	49	92.45%	Goal: >98%
	MolProbity score [^]	2.41		52 nd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 455	0.00%	Goal: 0%
Bad angles:	2 / 609	0.33%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 4	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad backbone conformations [#] :	17	80.95%	Goal: ≤ 5%
	Bad bonds:	0 / 484	0.00%	Goal: 0%
	Bad angles:	0 / 738	0.00%	Goal: <0.1%

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

RNA backbone was recently shown to be rotameric. Outliers are RNA suites that don't fall into recognized rotamers.

[^] MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

Appendix A9 Experiments performed with HBD₁₀₅₋₁₆₅

A9.1 Introduction

Previous work in our lab has shown that there is a large difference between the K_D values of the complexes that form with either VEGF₁₆₅ or HBD₁₁₁₋₁₆₅ and Macugen, which have K_D values of 50 pM and 12 nM respectively. The observation of these differences in the K_D of the HBD₁₁₁₋₁₆₅-Macugen complex compared to that of the complex with full-length VEGF₁₆₅ resulted in the hypothesis that other residues of VEGF₁₆₅ were contributing to the interaction. This led to the design of a new HBD construct consisting of residues 105 to 165 (Figure A4.1). Based on the available structures of the RBD and HBD, these additional amino acids reside within what may be a linker between the two domains. Therefore, it was hypothesized that the four additional basic amino acids would form electrostatic interactions with the aptamers. We used isothermal titration calorimetry (ITC) to compare the two constructs.

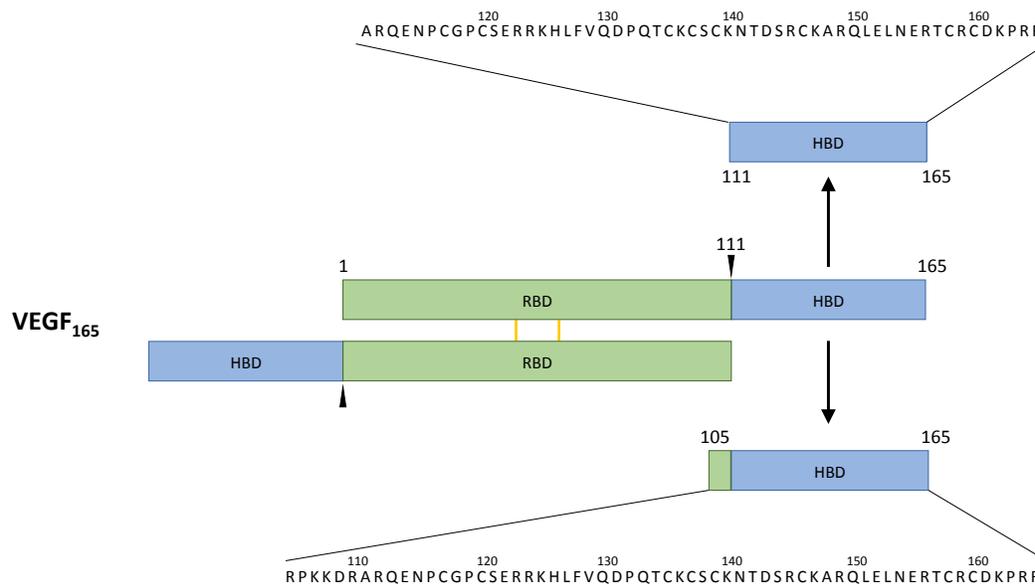


Figure A9.1. HBD₁₀₅₋₁₆₅ construct design.

A9.2 Materials and methods

Samples of HBD₁₀₅₋₁₆₅ and HBD₁₁₁₋₁₆₅ were prepared as described in the text. For ITC, both HBD constructs were concentrated to 50 μM in 10 mM sodium phosphate, pH 7.4, 100 mM NaCl, and 3 mM KCl. The lyophilized aptamer (IDT) was resuspended at a concentration of 5 μM in 10 mM sodium phosphate, pH 7.4, 100 mM NaCl, and 3 mM KCl. ITC was performed at 25 °C using an iTC₂₀₀ Microcalorimeter. The titration included forty-two 1.6 μL injections of HBD into the aptamer with three minutes of equilibration time between injections.

The NMR samples of HBD₁₀₅₋₁₆₅ were in a different buffer than described in the text: 10 mM Tris pH 7.5, 1 mM EDTA, and 10% D₂O/90% H₂O instead of sodium phosphate. A sample of HBD₁₀₅₋₁₆₅ was prepared using the fractions that eluted in the first peak from the heparin affinity column. A 2D ¹H-¹⁵N HSQC was collected on each sample.

A9.3 Results and discussion

A9.3.1 ITC to compare the two HBD constructs

For HBD₁₀₅₋₁₆₅ binding to the DNA aptamer, the data measured $n = 1.14 \pm 0.05$, $K_D = 15 \pm 5$ nM, $\Delta H^\circ = -26.8 \pm 0.8$ kcal/mol, and calculated $\Delta S^\circ = -54 \pm 3$ cal/mol/deg at 25°C (Figure A9.1). For HBD₁₁₁₋₁₆₅ binding to the DNA aptamer, the data measured $n = 1.04 \pm 0.03$, $K_D = 39 \pm 3$ nM, $\Delta H^\circ = -27.4 \pm 0.3$ kcal/mol, and calculated $\Delta S^\circ = -58 \pm 1$ cal/mol/deg at 25°C (Figure A9.2). The ITC data showed that complex formation with the DNA aptamer is enthalpically driven. These experiments did not show a dramatic difference in binding between the two constructs, so we moved forward with HBD₁₁₁₋₁₆₅, which expressed better in the *Pichia pastoris* system.

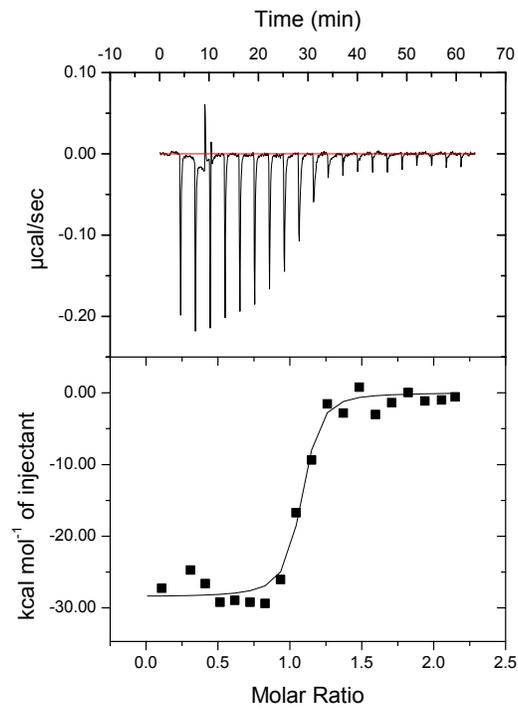


Figure A9.2 Representative ITC experiment of HBD₁₀₅₋₁₆₅ binding to the aptamer

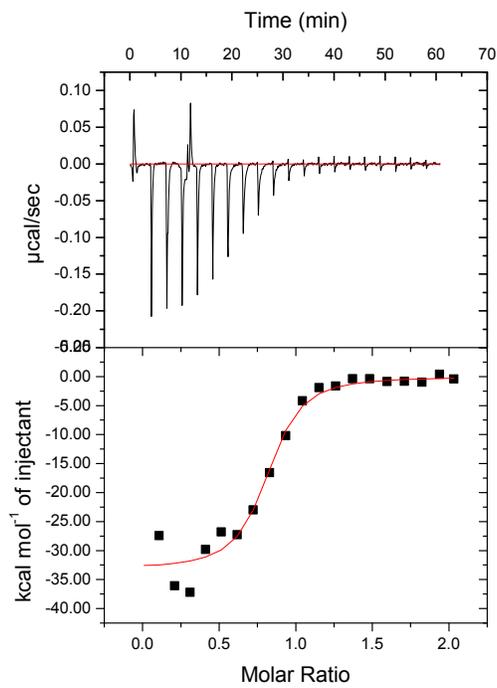


Figure A9.3 Representative ITC experiment of HBD₁₁₁₋₁₆₅ binding to the aptamer

A9.3.2 NMR to determine the structural integrity of HBD₁₀₅₋₁₆₅

The ¹H-¹⁵N HSQC of the HBD sample from the first elution from the heparin affinity column showed that the protein was mostly unstructured (Figures A9.3 and A9.4). Since the properly folded HBD contains four disulfide bonds, it was possible that this misfolded HBD had the incorrect disulfide bonds. To test this hypothesis, β-mercaptoethanol was added to a final concentration of 1 mM. The sample was incubated at room temperature for an hour before a 16-hour HSQC was collected (Figure A9.5). The NMR spectrum showed that the addition of β-mercaptoethanol allowed the HBD to refold into the proper conformation.

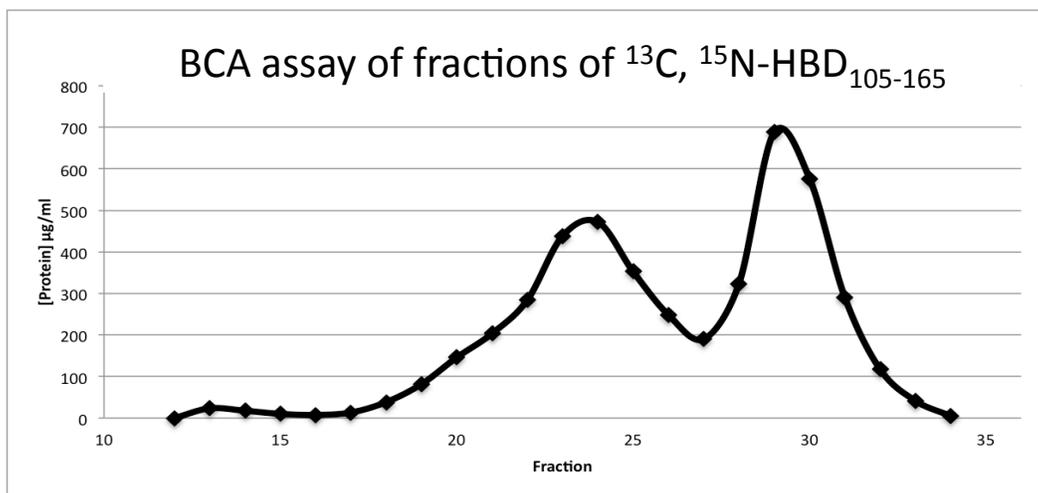


Figure A9.4 Representative BCA assay of the HBD₁₀₅₋₁₆₅ heparin affinity column fractions.

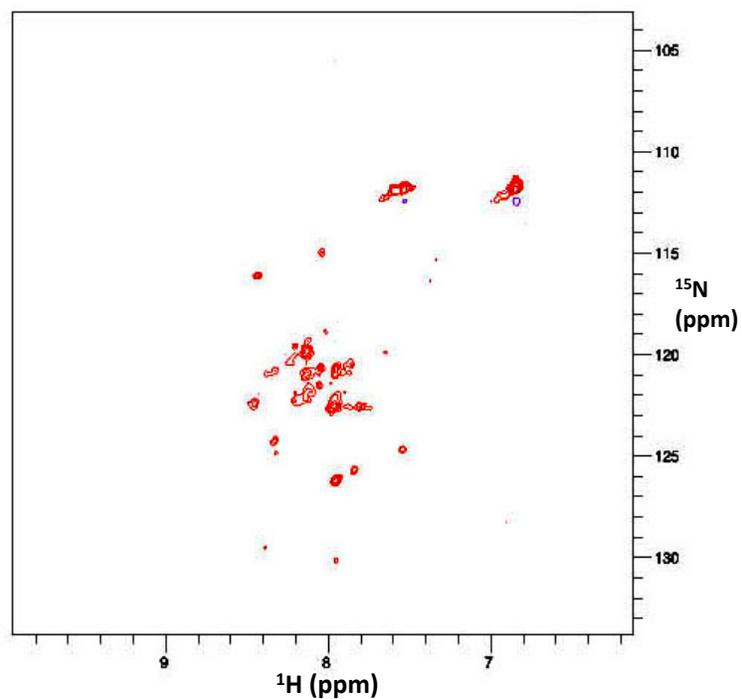


Figure A9.5 ^1H - ^{15}N HSQC of the misfolded HBD sample.

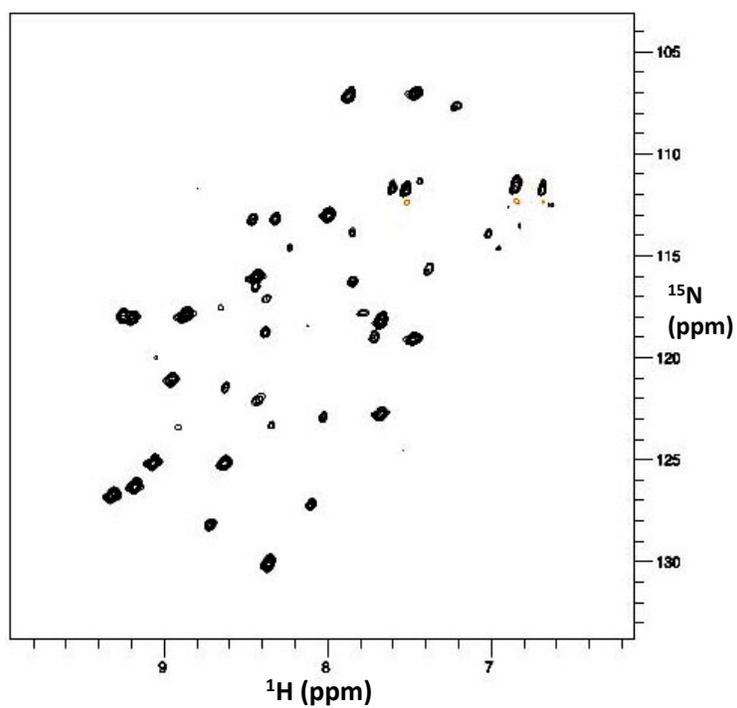


Figure A9.6 ^1H - ^{15}N HSQC of the previously misfolded HBD sample after addition of 1 mM β -mercaptoethanol.