RICE UNIVERSITY

Human UP1 as a model for understanding purine recognition in the
family of proteins containing the RNA Recognition Motif (RRM)

by

Jeffrey C. Myers

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Yousif Shamoo, Assistant Professor, Thesis Advisor
Biochemistry and Cell Biology

John S. Olson, Professor,
Biochemistry and Cell Biology

Kathleen S. Matthews, Professor,
Biochemistry and Cell Biology

James McNew, Assistant Professor,
Biochemistry and Cell Biology

Robert Raphael, Assistant Professor,
Bioengineering

November 10, 2004

UMI Number: 3168109

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3168109

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA Recognition Motif (RRM)

By

Jeffrey C. Myers

Heterogeneous ribonucleoprotein A1 (hnRNP A1) is a prototype for the family of eukaryotic RNA processing proteins containing the common RNA recognition motif (RRM). The region consisting of residues 1-195 of hnRNP A1 is referred to as UP1. This region has two RRM and has a high affinity for both single-stranded RNA and the human telomeric repeat sequence d(TTAGGG)_n. We have used UP1's novel DNA binding to investigate how RRM bind nucleic acid bases through their highly conserved RNP consensus sequences. Nine complexes of UP1 bound to modified telomeric repeats were investigated using equilibrium fluorescence binding and X-ray crystallography. In two of the complexes, alteration of a guanine to either 2-aminopurine or nebularine resulted in an increase in K_d from 70 nM to 160 and 280 nM, respectively. The loss of orienting interactions between UP1 and the substituted base allows it to flip between *syn* and *anti* conformations. Substitution of the same base with 7-deaza-guanine preserves the O6/N1 contacts but still increases the K_d to 250 nM, a result suggesting that it is not simply the loss of affinity that gives rise to base mobility but also the stereochemistry of

the specific contact to O6. Although these studies provide details of UP1 interactions to nucleic acids, three general observations about RRM s are also evident: 1) as suggested by informatic studies, main chain to base hydrogen bonding make up an important aspect of ligand recognition; 2) steric clashes generated by modification of a hydrogen bond donor-acceptor pair to a donor-donor pair are poorly tolerated; and 3) a conserved lysine position proximal to RNP-2 (K₁₀₆-IFVGGI) orients the purine to allow stereochemical discrimination between adenine and guanine based on the 6-position. This single interaction is well-conserved in known RRM structures and appears to be a broad indicator for purine preference in the larger family of RRM proteins.

TABLE OF CONTENTS

ABSTRACT	ii
CHAPTER 1: INTRODUCTION	1
1.1 Heterogeneous Ribonucleoproteins (hnRNPs)	1
1.2 hnRNP A1 Structure and Function	3
1.3 M9 Domain and Nuclear Import	5
1.4 M9 Domain and Nuclear Export	7
1.5 RGG Domains of hnRNP A1 and A2	9
1.6 RNA Recognition Motif (RRM)	10
1.7 Features of RNA Recognition	12
1.8 UP1/hnRNP A1 and Telomere Biogenesis	14
1.9 hnRNP A1 binding Studies Using Systematic Evolution of Ligands by Exponential enrichments (SELEX)	16
1.10 UP1 Destabilizes G-tetrads	18
1.11 UP1 Crystal Structure	19
CHAPTER 2: MATERIALS AND METHODS	23
2.1 UP1 Growth	23
2.2 UP1 Purification	24
2.3 Oligonucleotides	26
2.4 Crystallization of UP1:hTR Complexes	27
2.5 Data Collection	28

2.6 Building heterocompounds for modified bases used	
in crystallographic refinement	29
2.7 Refinement	31
2.8 Equilibrium Fluorescence Titrations	34
2.9 Thermal Denaturation of Quadruplex Nucleic Acids	36
2.10 Static Light Scattering	37

CHAPTER 3. CRYSTALLIZATION OF UP1 BOUND

TO OLIGONUCLEOTIDES USED IN THIS STUDY	40
3.1 Oligonucleotides Used in this Study	40
3.2 Crystallization of UP1 to Modified hTRs	40
3.3 Complexes that Crystallize but do not Diffract	47
3.4 Complexes that Crystallize and do Diffract	47

CHAPTER 4 UP1 AND NUCLEIC ACIDS IN SOLUTION

4.1 Introduction	49
4.2 Light Scattering Data Shows Up1 Binds as a Dimer	
to Nucleic Acids in Solution	50
4.3 UP1 Destabilizes G-tetrads	53
4.4 Thermal Denaturation of G-tetrads	55
4.5 Circular dichroism (CD) of modified oligonucleotides	56
4.6 Structure-based Incorporation of a Fluorescent Guanine	
Analog as a Tool for Studying UP1:hTR Binding	59

4.7 Use of 6-methyl-8-(2-deoxy- β -ribofuranosyl)isoxanthopteridine (6-MI) in hTRs as a Reporter Molecule for UP1:hTR Interactions _____	61
4.8 6-MI Gain-of-Signal Depends Strongly upon Change of Environment During Complex Formation _____	65
4.9 Fluorescence Binding Studies of UP1 to TR2-6F as a Function of Salt Concentration and pH _____	66
4.10 Competition Binding Studies _____	74
4.11 Competition Binding Experiments of UP1:TR2-6F and hTRs that Contain Modified Bases _____	74
4.12 Conclusions from Competition Binding Experiments _____	80

CHAPTER 5 ANALYSIS OF Ade-9 BINDING SURFACE USING 7-DEAZA-

ADENINE AND NEBULARINE _____	81
5.1 Ade-9 Binding Pocket _____	81
5.2 Substitution of Ade-9 with Nebularine _____	82
5.3 Substitution of Ade-9 with 7-deaza-adenine _____	82
5.4 Substitutions that Failed to Produce Crystals _____	85
5.4a Substitution of Ade-9 with Inosine and 4-methylindole _____	85
5.4b Substitution of Ade-9 with Guanine and Thymine _____	86
5.5 Conclusions _____	87

CHAPTER 6. ANALYSIS OF Gua-10 BINDING SURFACE USING

7-DEAZA-GUANINE, INOSINE, 2-AMINOPURINE, AND NEBULARINE	92
6.1 Gua-10 Binding Pocket	92
6.2 Substitution of Gua-10 with 7-deaza-guanine	93
6.3 Substitution of Gua-10 with 2-aminopurine and Nebularine	94
6.4 Substitution of Gua-10 with Inosine	96
6.5 Gua10 Binding PocketConclusions	98
6.6 Comparison of the Gua10 and Gua4 Binding Pockets	101

CHAPTER 7. ANALYSIS OF Gua-11 BINDING SURFACE USING

INOSINE AND 2-AMINOPURINE	109
7.1 Gua-11 Binding Pocket	109
7.2 Substitution of Gua-11 with 2-aminopurine	109
7.3 Substitution of Gua-11 with Inosine and 6-MI	111
7.4 Conclusions	113

CHAPTER 8. SUMMARY AND CONCLUSIONS

8.1 Comparison of UP1:TR2 Complex with other	
RRM:nucleic acid Complexes	117
8.2 RRM Hydrophobic Core	121
8.3 UP1 Consensus Binding Sequence	127
8.3a Characterizing the Pyrimidine Binding Pocket	129
8.3b UP1 Adenine Binding Pocket	131

8.3c UP1 Guanine Binding Pockets	132
----------------------------------	-----

REFERENCES	137
------------	-----

LIST OF TABLES

Table 3.1 Published crystallization protocols for UP1 and UP1:TR2 complexes	42
Table 3.2 Summary of oligonucleotides used in crystallization studies with UP1	44
Table 3.3 Crystal data summary of UP1 bound to modified oligonucleotides	48
Table 4.1 Dissociation Constants measured for UP1 binding to TR2-6F at increasing concentrations of NaCl	69
Table 4.2 Dissociation Constants measured for UP1 binding to TR2-6F as a function of pH	69
Table 4.3 Dissociation constants for UP1 binding to oligonucleotides from competition assays	76
Table 6.1 Dissociation constants for UP1 binding to oligonucleotides from competition assays	94
Table 8.1 Summary of base-stacking and hydrogen bonding interactions from the available RRM:nucleic acid complexes	121

LIST OF FIGURES

Figure 1.1 Base specific hydrogen bond donors and acceptors _____	11
Figure 1.2 Protein:RNA binding strategies identified by ENTANGLE _____	12
Figure 1.3 SELEX results for hnRNP A1 _____	17
Figure 1.4 G-tetrad stabilized with a monovalent K^+ _____	18
Figure 1.5 Crystal structure of UP1 bound to the human telomeric repeat (hTR) _____	22
Figure 3.1 Summary of modified bases used in this study _____	41
Figure 3.2 Examples of UP1 crystals _____	45
Figure 4.1 Light scattering studies of UP1 complexed to d(TTAGGG) ₂ _____	52
Figure 4.2 G-tetrad stabilized with a monovalent K^+ _____	53
Figure 4.3 Thermal denaturation of G-tetrad DNA by UP1 _____	56
Figure 4.4 Circular dichroism spectra of oligonucleotides _____	58
Figure 4.5 Comparison of Gua to the fluorescent Gua analog 6MI _____	62
Figure 4.6 RIBBONS diagram of UP1 bound to the hTR sequence _____	64
Figure 4.7 Fluorescence titration of UP1 with TR2-6F and TR2-11F _____	66
Figure 4.8 Titrations of UP1 with TR2-6F at different concentrations of NaCl _____	67
Figure 4.9 Log K_a versus log [NaCl] _____	68
Figure 4.10 Fluorescence titrations of UP1 with TR2-6F at pH 6.0 _____	70
Figure 4.11 Fluorescence titrations of UP1 with TR2-6F at pH 6.5 _____	71
Figure 4.12 Fluorescence titrations of UP1 with TR2-6F at pH 7.0 _____	72
Figure 4.13 Fluorescence titrations of UP1 with TR2-6F at pH 7.5 _____	73
Figure 4.14 Competition binding experiments with Ade9 substitutions _____	77

Figure 4.15 Competition binding experiments with Gua10 substitutions	78
Figure 4.16 Competition binding experiments with Gua11 substitutions	79
Figure 5.1 Guide to the bases used in the crystallography studies	83
Figure 5.2 Structures of UP1 complexes with substitutions at Adenine-9	84
Figure 5.3 Bases used to replace Ade9 that failed to produce crystals	89
Figure 5.4 Ade9 binding pocket of UP1	90
Figure 5.5 Superposition of Thy in the Ade9 binding pocket	91
Figure 6.1 Examples of modified bases used to replace Gua10	93
Figure 6.2 Structures of UP1 with substitutions of Guanine-10	97
Figure 6.3 Superposition of Ade in the Gua10 pocket with proposed L106Q mutation	104
Figure 6.4 Superposition of Ade in the Gua10 binding pocket with a proposed E135L mutation	105
Figure 6.5 Superposition of Thy in the Gua10 binding pocket	106
Figure 6.6 Superposition of Ade in the Gua4 binding pocket of UP1	107
Figure 6.7 Superposition of Thy in the Gua4 binding pocket of UP1	108
Figure 7.1 Modified bases used in the Gua11 binding pocket of UP1	110
Figure 7.2 Structures of UP1 with substitutions at the Gua11 binding pocket	112
Figure 7.3 Superposition of Ade in the Gua11 binding pocket of UP1	115
Figure 7.4 Superposition of Thy in the Gua11 binding pocket of UP1	116
Figure 8.1 Comparison of available RRM:nucleic acid complexes	119
Figure 8.2 A sequence alignment of the RRM proteins	120
Figure 8.3 The hydrophobic clefts of UP1	122

Figure 8.4 The hydrophobic clefts of Sxl	123
Figure 8.5 The hydrophobic clefts of Hud	123
Figure 8.6 The hydrophobic clefts of PABP	125
Figure 8.7 The hydrophobic cleft of U1A	125
Figure 8.8 The hydrophobic clefts of Rbd12	126
Figure 8.9 Summary of stacking and hydrogen bonds between UP1 and TR2	128
Figure 8.10 Proposed effects of L15Q and D42N mutations in the Gua5 site	136

CHAPTER 1. INTRODUCTION

1.1 Heterogeneous Ribonucleoproteins (hnRNPs)

The eukaryotic nucleus contains an abundant family of RNA binding proteins called heterogeneous ribonucleoproteins or hnRNPs. Although hnRNPs were among the first eukaryotic RNA binding proteins isolated, their role *in vivo* is complex and not fully delineated. HnRNPs are tightly associated with polymerase-II (pol-II) transcripts, and these proteins have been shown to be involved in diverse cellular functions both inside the nucleus and the cytoplasm. These functions include involvement in pre-mRNA splicing, mRNA transport, recombination, polyadenylation, translational silencing, mRNA stability, heat-shock response, and oncogenesis, and telomere biogenesis (Calvio *et al.*, 1995; Dreyfuss *et al.*, 1993; Gattoni *et al.*, 1996; LaBranche *et al.*, 1998; Lou, 1999; Ostareck *et al.*, 1997; Shahied-Milam *et al.*, 1998; Shih and Claffey, 1999). Numerous isoforms of hnRNPs are generated by alternative pre-mRNA splicing and post-translational modifications (Krecic and Swanson, 1999), and common features of these proteins include the presence of highly conserved RNA recognition motifs (RRM) and domains rich in arginines and glycines known as RGG domains, which are generally considered to be non-sequence specific RNA binding motifs. Additional domains are involved in protein-protein interactions, and together these elements package heterogeneous nuclear RNA into hnRNP particles.

Of the roughly 20 hnRNPs identified, early studies of hnRNPs focused on six “core” proteins that migrate as three doublets on SDS-PAGE and were named hnRNP A1, A2, B1, B2, C1, and C2 (Beyer *et al.*, 1977). Sedimentation analysis of nuclear

extracts show that these proteins sediment at 40 S in conjunction with pre-mRNA (Buvoli *et al.*, 1992) and form (C1)₃C2, (A2)₃B1, and (A1)₃B2 tetramers that package RNA into 20-25 nm particles (McAfee *et al.*, 1997b). It has been suggested that the (C1)₃C2 tetramers first associate with a 700-nucleotide hnRNA fragment forming a triangular shaped 19 S assembly to which three (A2)₃B1 complexes can bind. This complex (35 S) then matures to a 40 S complex by binding with (A1)₃B2 (Huang *et al.*, 1994). A1, A2, B1, and B2 hnRNPs each contain two RRM and an RGG box, whereas hnRNP C1 and C2 are single RRM proteins without an RGG box.

The packaging of nuclear RNA into ribonucleosomes by hnRNP proteins is analogous to the packaging of DNA by histones. These structures regulate the exposure of nucleic acid to nucleases and in ribonucleosomes and aid in RNA processivity by selectively exposing pre-mRNA to factors involved in splicing, polyadenylation and transport. In the case of splicing, the condensation of pre-mRNA by hnRNP complexes may be required to conserve its single stranded form and for selective exposure of splice sites (McAfee *et al.*, 1997a). Thus, long intronic regions can be removed by positioning distal splice sites close together so that splicing factors, such as snRNP complexes, can physically contact both ends of the splice site. Although cooperative interaction of entire hnRNP and snRNP complexes has not been observed, immunoprecipitation studies of A1 and U1, U2, U4, and U5 snRNPs have shown that A1 binds selectively to U2 and U4 which assemble at the 5' splice site (Buvoli *et al.*, 1992), and hnRNP C proteins have been shown to bind U1, U2, and U6 snRNAs (Shahied-Milam *et al.*, 1998). These studies, along with those that have shown A1 interaction with snRNPs at 3' splice sites (Barabino *et al.*, 1990; Guthrie, 1991; Minvielle-Sebastia, 1998; Ruby and Abelson,

1988; Seraphin and Rosbash, 1989), provide strong evidence for concerted interaction between hnRNPs and snRNPs during pre-mRNA splicing and processing.

One of the most interesting hnRNP complexes is the 75 S Balbiani ring (BR) pre-mRNP particle that carries the transcript for high molecular weight salivary proteins in *Chironomus tentans* (Krecic and Swanson, 1999). During pre-mRNA transcription, an hnRNP protein, hrp36, associates with the entire length of the BR pre-mRNA during transcription forming a 7 nm fiber which coils into a 19 nm fiber. As transcription continues, the protein:DNA complex folds into a ribbon-like 26 nm fiber and continues to bend and fold until maturing into a ring-shaped BR RNP. Nuclear export requires unwinding of the ring structure for passage through the nuclear pore complex, and a protein with high homology to human hnRNP A1 is associated with the BR RNA during nuclear export (Nakielnny and Dreyfuss, 1999). Electron microscopy and electron tomography has been used to study large nuclear RNP (lnRNP) particles, and these studies have shown that lnRNP particles contain pre-mRNA processing factors. The BR hnRNP complex demonstrates that hnRNPs can form dynamic structures that are intimately involved with pre-mRNA.

1.2 HnRNP A1 Structure and Function

HnRNP A1 is one member of the six “core” hnRNP proteins and is a 319 amino acid protein (34 kD) made up of four domains (Cobianchi *et al.*, 1986; Merrill *et al.*, 1987). The N-terminal two-thirds of the protein contains two canonical RNA recognition motifs (RRMs) ~90 amino acids in length that have been shown to bind tightly to ssRNA (Nadler *et al.*, 1991; Shamoo *et al.*, 1994). The C-terminal region of hnRNP A1 is

comprised largely of small, highly flexible residues (40% glycine) that make several RGG motifs which contribute to hnRNP A1 cooperativity and affinity for ssRNA (Cartegni *et al.*, 1996; Nadler *et al.*, 1991). A 38 residue M9 nuclear localization/nuclear import signal is found C-terminal to the RGG motifs, as are several potential phosphorylation (Cobianchi *et al.*, 1993) and methylation sites that mediate hnRNP A1 localization and affinity for ssRNA (Kim *et al.*, 1997). Partial proteolysis studies of hnRNP A1 showed that the C-terminal 124 residues could be removed readily to produce UP1. UP1 was originally discovered in calf thymus extracts and was thought to be a DNA helicase, thus it was given the name unwinding protein 1 (UP1). It was later discovered that UP1 is a proteolytic fragment of hnRNP A1, and, although UP1 has not been shown to have helicase activity, the name UP1 has remained intact. UP1 is made up of residues 1-195 and includes both RRM s as well as a short basic region just C-terminal of RRM2 that is important to nucleic acid binding (Merrill *et al.*, 1988). Unlike most RRM containing proteins, UP1 is able to bind with high affinity to either ssDNA or ssRNA (Abdul-Manan and Williams, 1996; Nadler *et al.*, 1991; Shamoo *et al.*, 1994). The ability to bind DNA in a sequence-specific fashion through two RRM s appears to be unique to hnRNP A1/UP1 and presents an opportunity to study nucleic acid recognition via modified DNA bases.

Human hnRNP A1 has been studied extensively and appears to be involved in numerous cellular activities of the hnRNP complex, including retention of intron-containing pre-mRNA within the nucleus (Reed and Hurt, 2002) as well as already established roles in alternative splicing (Blanchette and Chabot, 1997; Blanchette and Chabot, 1999; Burd and Dreyfuss, 1994; Dreyfuss *et al.*, 2002; Hutchison *et al.*, 2002; Marchand *et al.*, 2002;

Mayeda and Krainer, 1992; Simard and Chabot, 2000; Zhu *et al.*, 2001) and RNA transport (Gallouzi and Steitz, 2001; Izaurralde *et al.*, 1997; Pollard *et al.*, 1996; Visa *et al.*, 1996). A1 is an abundant protein that has been shown to shuttle from the nucleus to the cytoplasm at a rate of about 120,000 molecules per minute and then localize back to the nucleus at such a rate that the steady state localization is in the nucleus (Michael *et al.*, 1995). In addition, hnRNP A1 has been demonstrated to bind purine-rich DNAs such as the human telomeric repeat (hTR) ssDNA, d(TTAGGG)_n (Dallaire *et al.*, 2000; Fiset and Chabot, 2001; Ishikawa *et al.*, 1993; LaBranche *et al.*, 1998), and the mouse minisatellite (MN) repeat DNA, d(GGCAG)_n (Fukuda *et al.*, 2002). The observation that UP1 binds with high affinity to these repeat sequences has suggested an *in vivo* role for hnRNP A1 and its proteolytic derivative UP1 in chromosome maintenance (Dallaire *et al.*, 2000; Fiset and Chabot, 2001).

1.3 M9 Domain and Nuclear Import

Nuclear import of hnRNP A1 has been shown to be mediated through factors that interact with the M9 domain, such as 90 kD karyopherin-β2 (also referred to as transportin in the literature) and 101 kD MIP (Fridell *et al.*, 1997; Siomi *et al.*, 1997). Although M9 bears no similarity to the "classical" basic nuclear localization signal (NLS), karyopherin-β2 shares 24% identity with human karyopherin-β which is involved with nuclear import of proteins bearing the basic NLS (Nakielnny and Dreyfuss, 1996). Proteins that contain this sequence are bound by karyopherin-α at the NLS, and karyopherin-α acts as an adapter protein for docking with karyopherin-β. The NLS-karyopherin-α/β complex then interacts with the nuclear pore complex (NPC), and the

protein is imported into the nucleoplasm in a Ran:GTP dependent fashion. Inhibitors and competitors of karyopherin- α and karyopherin- β that affect nuclear localization of proteins that contain a classical NLS have no effect on M9 mediated nuclear import (Pollard *et al.*, 1996).

Unlike karyopherin- β 1 which requires an adapter, karyopherin- α , for nuclear localization, karyopherin- β 2 binds directly to the glycine rich M9 domain of hnRNP A1, and the complex is imported into the cell nucleus in a process that appears to be mediated by Ran and GTP hydrolysis, similar to the karyopherin- α/β -NLS nuclear import (Bonifaci *et al.*, 1997). In the case of hnRNP A1, dissociation of karyopherin- β 2 and A1 has been shown to occur in the presence of Ran:GTP *in vitro* (Siomi *et al.*, 1997), and it is likely that this dissociation occurs in the nucleoplasm *in vivo* so that karyopherin- β 2:RanGTP can return to the cytoplasm and A1 can associate with pre-mRNAs and other hnRNP proteins to aid in pre-mRNA processing (Caceres *et al.*, 1994; Choi *et al.*, 1986; Mayeda and Krainer, 1992).

Although the classical NLS and M9 domains bear little resemblance to each other, the nuclear import pathways of the two appear to be evolutionarily related. Both β 1 and β 2 have been shown to compete for immobilized nucleoporin protein Nup98 in solution. Increased levels of karyopherin- α/β 1 inhibits import of M9 domains, whereas increased levels of karyopherin- β 2 inhibit NLS import *in vivo* (Bonifaci *et al.*, 1997). Additionally, karyopherin- β 2-mediated nuclear localization is inhibited by RanQ64L, a Ran mutant that is unable to hydrolyze GTP, and this mutant also inhibits import of classical NLS containing proteins (Melchior *et al.*, 1995; Nakielnny and Dreyfuss, 1996). Nuclear import of hnRNP A1 is also inhibited in the presence of guanylyl imidodiphosphate, a

nonhydrolyzable GTP analog (Bonifaci *et al.*, 1997). The crystal structure of the karyopherin- β 2-Ran:GppNHp complex, where GppNHp is a nonhydrolyzable GTP analog, shows that Ran interacts with one of two arches on karyopherin- β 2 formed by 18 tandem repeating sequences (identified as HEAT repeats) (Andrade and Bork, 1995). Deletion mutations have mapped hnRNP A1 binding to the other arch (Chook and Blobel, 1999; Fridell *et al.*, 1997). The crystal structures of Ran:GDP and Ran:GppNHp show that Ran undergoes extensive structural rearrangements depending on the phosphorylation state of the ligand, and the region involved in binding karyopherin- β 2 to A1 is affected so that A1 dissociates from the complex occurs in the presence of RanGTP (Chook and Blobel, 1999; Stewart *et al.*, 1998). Thus, the biochemical studies of hnRNP A1, karyopherin- β 2, and Ran:GTP and the crystal structures of the karyopherin- β 2-Ran:GppNHp complex and Ran:GDP are consistent with a model in which karyopherin- β 2 binds hnRNP A1 through the M9 domain, the complex is transported into the nucleoplasm, and A1 dissociates from the complex in a Ran:GTP dependent fashion.

1.4 M9 Domain and Nuclear Export

The mechanisms of mRNA export by hnRNP A1 are less understood than those of A1 import, but the M9 domain appears to be involved and has been shown to be necessary and sufficient for export of proteins normally confined to the nucleus (Michael *et al.*, 1995). Nuclear export of hnRNP A1 is mediated by pol-II activity, and A1 is found to accumulate in the cytoplasm in cells treated with pol-II inhibitors (Pinol-Roma and Dreyfuss, 1992). Arginine methylation has also been shown to modulate nuclear export of hnRNP proteins (Nakielnny and Dreyfuss, 1999). Methylation of a domain

normally associated with RNA binding, the RGG domain, has been shown to regulate subcellular localization of a similar protein, hnRNP A2 (Nichols *et al.*, 2000). In order for A1 to exit the nucleus bound with mRNA, the hnRNP complex must dissociate. Non-shuttling hnRNP proteins contain nuclear retention signals (NRS) that override hnRNP nuclear export signals (Nakielnny and Dreyfuss, 1996), and these proteins do not dissociate from the hnRNP complex prior to mRNA maturation. Additionally, it has been shown that introns prevent RNAs from exiting the nucleus through a saturable retention mechanism (Nakielnny and Dreyfuss, 1997). It is therefore possible that intron removal and dissociation of NRS-containing hnRNPs from the hnRNP complex are interrelated.

Both *in vitro* and *in vivo* studies of a spliced mRNA have shown that, following intron removal, factors are recruited to form an exon-exon junction complex (EJC) (Kim and Dreyfus, 2001; LeHir *et al.*, 2000). Six components of the EJC were reviewed by Dreyfuss *et al.* (Dreyfuss *et al.*, 2002), and a role in mediating mRNA export by some or all of the proteins in this complex is implied. Splicing increases export efficiency of some mRNAs over that of mRNAs transcribed by intronless complementary DNAs, whereas EJCs do not associate (Kataoka, 2000). One protein that contributes to the EJC that may link RNA splicing to RNA export in vertebrates is Aly/REF which has also been shown to interact with a protein known as TAP, a mRNA binding protein that has been suggested to interact in conjunction with hnRNP A1 in mRNA export. In addition to TAP, several other mRNA binding proteins have been implicated in similar functions. These include Gle2p and Dbp5p, all of which have been shown to bind mRNA, interact at the NPC, and shuttle between the nucleus and cytoplasm (Nakielnny and Dreyfuss, 1999).

Dbp5p also has RNA helicase activity (Tseng *et al.*, 1998), and the unfolding of mRNPs might be necessary for passage through the NPC. Balbiani ring transcripts are transported through the NPC by an hnRNP protein similar to A1, and electron microscopy shows that significant unfolding of the mRNP occurs after dissociation of non-shuttling hnRNP proteins prior to nuclear export (Daneholt, 1997). Although many elements of hnRNP A1 and mRNA nuclear export have been identified, the details that regulate transport are not fully understood.

1.5 RGG Domains of hnRNP A1 and A2

HnRNP A1 and A2 have been studied extensively, and they share ~80% identity in the N-terminal RRM region and ~60% identity in the glycine-rich C-terminal region (Nichols *et al.*, 2000). Both proteins have RGG domains, which are motifs rich in arginine and glycine that have been shown to bind indiscriminately to nucleic acid. Glycine repeats may give the domain flexibility, whereas the arginine guanidino nitrogens provide affinity for the RNA phosphate backbone. Interestingly, the arginines in the RGG domains of hnRNP A proteins are targets of an uncommon asymmetric dimethylation to the guanidino nitrogens to produce N^G , N^G -dimethylarginine (Nichols *et al.*, 2000), and ~65% of the arginines in hnRNPs are methylated in the cell nucleus (Boffa *et al.*, 1977; Nichols *et al.*, 2000). If the RGG domain plays a substantial role in RNA binding, it might be expected that arginine methylation would decrease binding, as found for A1 (Kim and Dreyfus, 2001). Both A1 and A2 are targets for the nuclear protein/histone-specific arginine methyltransferase PRMT1, and studies of hnRNP A2 have shown that the methylation state of the RGG domain regulates its cellular

localization (Nichols *et al.*, 2000). Inhibition of methyltransferase resulted in cytoplasmic accumulation of endogenous hnRNP A2 *in vivo* and deletion of the RGG domain in an hnRNP A2 mutant showed a four-fold greater accumulation in the cytoplasm than wild-type protein.

It remains to be seen if the methylation state of the A1 RGG domain has an effect on subcellular localization as it does in A2, and it is still not clear how the arginine methylation state regulates localization in the context of other nuclear localization domains involved in nuclear export. The M9 domain appears to be involved in both nuclear import and nuclear export of A1. The RGG domain may help regulate localization depending on the interaction of A1 with RNA, and the binding of A1 to RNA may be regulated by RGG methylation. Thus, while A1 and A2 have a high degree of homology and have similar domains, regulation of cellular localization is still not fully understood.

1.6 RNA Recognition Motif (RRM)

The RNA recognition motif (RRM), also referred to as the RNA Binding Domain (RBD), is comprised of ~90 amino acids and is made up of a four-stranded β -sheet and two α -helices (Birney E., 1993). RRM-containing proteins are typically involved in all aspects of RNA processing, including splicing, alternative splice site selection, transport, and turnover (Faustino and Cooper, 2003; Lorsch, 2002; Nasim *et al.*, 2002). Many proteins, such as nucleolin and poly-A binding proteins, have multiple RRMs. The InterPro Database at EMBL currently lists 455 human proteins containing putative RRMs, making it the most common RNA binding motif (Mulder *et al.*, 2003).

Understanding the atomic basis for RRM specificity is an important step toward piecing together the puzzle of how numerous RRM containing proteins compete for similar RNA sequences during alternative splice site selection and RNA processing. In addition, a fuller understanding of RRM-RNA interactions will provide the basis for engineering specificity in these motifs and permit their development as tools for investigating patterns of splicing and gene expression (Hall, 2002).

Figure 1.1. Potential base hydrogen bond donors and acceptors. Arrows pointing toward positions designate potential hydrogen bond acceptors and arrows pointing away from positions designate potential hydrogen bond donors.

1.7 Features of RNA Recognition

Proteins that bind nucleic acids have a wide range of mechanisms available to facilitate binding. Non-sequence specific protein:nucleic acid interactions can be made to the negatively charged phosphate backbone, whereas sequence-specific binding in double stranded RNAs and DNAs can occur through hydrogen bonds to the major and minor grooves. Proteins that bind single-stranded nucleic acids and exposed bases of hairpin loops have greater access to bases and can utilize binding pockets that derive base specificity through steric exclusions and Watson-Crick like hydrogen bonding arrangements (Figure 1.1). The stacking of aromatic side chains with bases provides stability in binding pockets, and specificity for RNA over DNA can be driven by hydrogen bonds made to the ribose 2'-OH of RNA.

Identifying general patterns of protein:nucleic acid interactions is a first step toward understanding more specific systems such as RRM:RNA interactions. The number

Figure 1.2. Summary of common binding strategies used in protein:RNA binding found by ENTANGLE (Allers and Shamoo, 2001). Coordinates from 42 structures were used and no conserved binding domains represented more than 15% of the total interactions. Note the extensive usage of main chain amides and carbonyls.

of protein:RNA complexes in the protein database (PDB) continues grow every year. Computer programs such as ENTANGLE (Allers and Shamoo, 2001) can be used to study elements of sequence specific binding by downloading the coordinates of these complexes from the PDB and identifying hydrogen bonds, electrostatic interactions, van der Waals attractions, hydrophobic interactions, and stacking interactions based on distances and bond angles. One study compared protein:RNA interactions of 42 protein:RNA complexes where no specific binding domain (such as an RRM) represented more than 15% of the total interactions, and the results are summarized in Figure 1.2 (Allers and Shamoo, 2001). Main chain carbonyl and amides were shown to contribute to the majority of hydrogen bond donors and acceptors, consistent with recent studies where main chain interactions account for about one-third of base-specific hydrogen bonding interactions (Allers and Shamoo, 2001; Hoffman *et al.*, 2004; Treger, 2001).

Main chain atoms interact with bases by forming a close fit, which may aid in building a nucleotide recognition pocket based both on hydrogen bonding and the shape of the base. The RRM also has a pair of consensus sequences, RNP2 (L-F/Y-V/I-G-N/D-L) and RNP1 (G-F-G-F-V/I-polar-F), whose stacking interactions between aromatic amino acids and RNA bases are highly conserved. These sequences are also essential to building a conserved nucleotide recognition surface that puts the base edge in position to make specific interactions to the protein.

In addition to main chain recognition in RRM, protein side chains play a broad role in base-specific hydrogen bonding interactions. Molecular dynamics simulations have suggested that certain highly ionized bidentate interactions, such as lysine or arginine to the N7/O6 position of guanine, are likely to be much stronger than non-

ionized interactions, such as asparagine to the N7/O6 of guanine (Cheng *et al.*, 2003; Cheng, 2004). Studies have shown that ionized hydrogen bonding interactions from lysine and arginine to the guanine N7/O6 are statistically over-represented (Allers and Shamoo, 2001; Treger, 2001). The role of these highly specific interactions in positioning and binding nucleic acids is important to understanding RNA recognition. Experimental investigation into these interactions provides both validation and further insight for future studies into the mechanism of ligand recognition for this important class of proteins.

1.8 UP1/hnRNP A1 and Telomere Biogenesis

Perhaps the most intriguing characteristic of hnRNP A1 is its role in telomere biogenesis (Fiset and Chabot, 2001; LaBranche *et al.*, 1998). Chromosomal ends in humans vary in length from 5,000 to 15,000 nucleotides of a repeating TTAGGG motif, and mortal cell populations in humans typically lose 65bp of telomeric length for each cell generation (Counter *et al.*, 1992; Lingner and Cech, 1998). Telomeres shorten as a result of asymmetric replication during each cell cycle, and chromosomes become less stable, leading to degradation and illegitimate recombination (Collins, 2000; Counter *et al.*, 1994).

Telomere length is stabilized in germ cells and many tumor cells by the action of telomerase. Telomerase is a ribonucleoprotein with reverse transcriptase activity that recognizes and binds telomeric 3' overhangs and uses its RNA to produce a complementary TTAGGG strand to extend the moiety (Weilbaecher and Lundblad, 1999). The overhang is primed, and DNA polymerase- α synthesizes the telomeric repeat so that the original length is conserved. Telomerase does not bind the dsDNA blunt ends

formed by leading strand polymerase activity (Lingner and Cech, 1998), but a 5'-3' exonuclease has been proposed that would generate overhangs by trimming the newly synthesized CA rich strand (Wellinger *et al.*, 1996). This step may be necessary for chromosomal end protection, as overhangs in mammals have been shown to form stable “T-loop” structures that result when the 3' overhang loops back and associates with duplex telomeric DNA (Evans *et al.*, 1999). Telomerase is generally active only in mammalian germ cells, and reactivation of this enzyme in somatic cells can induce metastatic tumors (Bryan and Cech, 1999).

Proteins in the hnRNP family have been shown to modulate telomeric activities (Eversole and Maizels, 2000). A1 and UP1 have been shown to protect telomeric sequences from nuclease activity and restore telomeric length *in vivo* (Dallaire *et al.*, 2000; LaBranche *et al.*, 1998). LaBranche *et al.* (1998) performed a series of *in vivo* and *in vitro* experiments that showed UP1 could restore telomeric length in cells deficient of A1 transcripts. Retroviral insertions into mouse erythroleukaemic cell lines produced a phenotype in which A1 was not produced to any detectable level. Controls maintained telomeric extensions, whereas A1 deficient strains showed significant shortening. Restoration of A1 expression in A1 deficient cell lines restored telomeric length to normal levels after 46 cycles.

Since A1 has been shown to be a factor in mRNA maturation, it was necessary to show that telomere restoration was the direct result of A1 activity at the telomeres and not the result of A1 directing alternative splice site selection of a pre-mRNA whose product controls telomere length. In order to rule out this possibility, UP1 was expressed to test whether telomere length could be restored in A1 deficient cells. UP1 is the amino

portion of A1 that encompasses the two RRM motifs. UP1 does not have the ability to affect splice site selection, although it does have sequence specific affinity for RNA and telomeric ssDNA (Abdul-Manan and Williams, 1996; LaBranche *et al.*, 1998; Myers and Shamoo, 2004). Expression of UP1 in A1 deficient cell lines resulted in telomere restoration in 86 cycles. The suggestion that UP1 is present in cells *in vivo* is controversial, since proteolysis of A1 may occur in cell lysates following cell fractionation (Pandolfo *et al.*, 1985; Riva *et al.*, 1986). However, studies comparing telomerase activity between A1 and UP1 showed that only UP1 could recover telomerase activity from a cell lysate (LaBranche *et al.*, 1998). This result suggests that A1 may undergo proteolysis to form UP1 *in vivo*, and UP1 is a factor in regulating telomerase activities.

1.9 hnRNP A1 binding Studies Using Systematic Evolution of Ligands by Exponential Enrichments (SELEX)

High affinity RNA binding targets for hnRNP A1 were examined using SELEX (Systematic Evolution of Ligands by Exponential enrichments) (Tuerk, 1997), and the "winner" sequence to which A1 bound with high affinity strongly resembles 5' and 3' splice site consensus sequences (Figure 1.3) (Burd and Dreyfuss, 1994). Considering that A1 binds promiscuously to pre-mRNA, it might seem unlikely that the protein would show sequence-specific specificity for RNA. The two conserved RRMs that make UP1 and the amino terminus of A1 are likely candidates for specificity, but if the RGG domain in A1 is involved in binding, A1 would likely bind indiscriminately.

The results of the SELEX experiments are consistent with studies in which A1 has been shown to affect alternative RNA splicing *in vivo* through its interaction with splicing regulators and pre-mRNA (Simard and Chabot, 2000). Essential splicing factor ASF/SF2 of the SR protein family compete for 5' and 3' splice sites where an excess of A1 forces ASF/SF2 to select distal rather than proximal sites (Bai *et al.*, 1999; Caceres *et al.*, 1994; Hanamura *et al.*, 1998; Mayeda and Krainer, 1992; Pollard *et al.*, 2000). The SELEX winner sequence strongly resembles both the 5' and 3' consensus splice site sequences. The winner sequence also resembles the ssDNA human telomeric repeat (hTR) and mouse minisatellite (MN) repeat which UP1 has been shown to bind (Fiset and Chabot, 2001; Fukuda *et al.*, 2002). Thus, the RRM of A1 and UP1 dictate sequence specific binding affinity. Although the RRM is typically thought of as an RNA binding surface, the RRM of A1/UP1 do not discriminate between RNA and ssDNA.

"winner" sequence	U A G G G A/U
5' splice site consensus sequence	C/A A G G U A/G A G U
3' splice site consensus sequence	C/U A G G/A
human telomeric repeat sequence	T A G G G T
mouse minisatellite repeat	C A G G G

Figure 1.3. Comparison of "winner" sequence from SELEX binding experiment performed on hnRNP A1 to physiologically relevant close matches of this sequence.

1.10 UP1 Destabilizes G-tetrads

A prominent feature of the SELEX "winner sequence", the hTR, and MN repeats is the fact that they are all rich in guanine. Guanine-rich polynucleotides can form stable G-tetrad structures under a wide range of pH and salt conditions (Castrignano *et al.*, 2002; Showalter and Hall, 2004). G-tetrads are planar arrangements of four guanines such that each base makes hydrogen bonds to neighboring guanines through N1, N2, O6, and N7 positions (Figure 1.4). Tetrads can be made from interstrand or intrastrand bases

Figure 1.4. G-tetrad stabilized with a monovalent K^+ in the center. Stable G-tetrads can also be made with Na^+ and multiple tetrads can stack to form G-quadruplex structures. Quadruplexes can be made from either interstrand or intrastrand bases. Dashed lines indicate hydrogen bonds.

arranged parallel or anti-parallel, and multiple tetrads can stack to form quadruplexes. The center of the G-tetrad contains a monovalent cation such as Na^+ or K^+ , and the stability of the structure is dependent upon the type and concentration of the cation (Shafer and Smirnov, 2001). In some cases, Ade in the hTR sequence also forms an A-tetrad. NMR data have shown that Ade can form tetrads in either the *anti* or *syn* conformation depending on the number of Thy separating the d(AGGG) repeats (Gavathiotis and Searle, 2003). However, the hTR oligonucleotide used in the crystal studies with UP1 was d(TTAGGG)₂ (Ding *et al.*, 1999), and the crystal structure of a similar hTR, d(TAGGGT)₂, showed that Ade does not participate in tetrad formation (Parkinson *et al.*, 2002).

Fukuda *et al.* (2002) have shown that UP1 is able to unfold unimolecular quadruplex DNA derived from mouse MN repeats from the sequence d(GGCAG)_n *in vitro*, and other melting studies confirm these findings with this sequence as well as with the hTR repeat (Myers *et al.*, 2003). UP1 can destabilize quadruplex structures by binding to available single-stranded nucleic acid and thereby shift the equilibria. Single-stranded DNA binding proteins (SSBs) use a similar mechanism for removing secondary structures during DNA replication, and UP1 may function in an analogous manner to prevent G-tetrad stabilized quadruplexes. Presumably, the removal of such quadruplex structures would aid in telomerase processivity and fidelity much as SSBs aid in DNA replication.

1.11 UP1 Crystal Structure

The structure of UP1 has been determined alone (Shamoo *et al.*, 1997; Xu *et al.*, 1997) and with the short telomeric DNA sequence d(TTAGGG)₂ (Ding *et al.*, 1999). To date, the full length hnRNP A1 protein has not been crystallized, and the flexible glycine rich portions C-terminal to the RRM would likely present difficulties in obtaining crystals. In the UP1:hTR co-crystal, the arrangement of protomers showed that the nucleic acid binding groove of UP1 is made by two copies of UP1 that bind ssDNA such that the two DNA strands bind in an anti-parallel fashion (Figure 1.5). The ssDNA in the crystal structure of Ding *et al.* (1999) has an extended conformation (average interatomic phosphate distance of 6.6 Å), with each RRM making a network of stacking, hydrogen bonding, hydrophobic, van der Waals, and electrostatic interactions to about 4 nucleotides. This arrangement is in excellent agreement with prior estimates of DNA binding site size based on polynucleotide binding studies (Shamoo *et al.*, 1994). Thirty-two amino acids contribute 41 total interactions, 19 of which are hydrogen bonds. Of the 19 hydrogen bond interactions, slightly more than half are contributed by amino acid side chains (11), and the rest are from the main chain (8). The majority of these hydrogen bonds (15) interact with the bases of the TR2 sequences, and only 4 are made to the phosphate backbone. Contacts to the 2'-OH of an RNA analog of the TR2 sequence might be expected, as is found in RRM:RNA complexes, and the extent of binding at this position is expected to represent an important element in the differentiation of ssRNA and ssDNA.

A comparison of the UP1:hTR structure with other RRM:nucleic acid complexes reveals that slightly more than half of the hydrogen bonding and stacking interactions are made from conserved residues. The majority of stacking interactions are highly

conserved, and most are made from aromatic residues in the RRM $\beta 1$ and $\beta 3$ regions. When looking at hydrogen bonds alone, the majority of residues that interact with nucleic acid are not conserved. In all cases, the non-structured region C-terminal to the RRM $\beta 4$ makes more hydrogen bonds to the nucleic acid substrate than any other structural domain.

The anti-parallel binding of two copies of substrate to UP1 is unique among the six RRM:RNA complexes solved to date since other multiple RRM proteins bind their substrate across RRMs of the same copy of protein. Although different, the manner in which UP1 binds the hTR in the crystal structure is consistent with an *in vivo* role of bringing distal splice sites together. The splice sites flanking an intronic region of pre-mRNA would run anti-parallel to each other if the intron were looped out, and multiple copies of hnRNP A1 could bind across the strands at the splice site and provide stability for splicing factors. The same might be true with regard to telomeres, where the 3' overhang has been shown to loop back into the duplex DNA forming stable structures called t-loops (Ford *et al.*, 2002). The overhang may be protected by A1/UP1 by binding to hTRs across the loop that would be crossing in an anti-parallel fashion.

Figure 1.5 Structure of UP1 bound to the human telomeric repeat (hTR), d(TTAGGG)₂. Two copies of UP1 form a crystallographic dimer (one copy is red and the second is blue) to which the DNA binds in an antiparallel manner to RRM1 (residues 1-92) of the first copy of UP1 and to RRM2 (residues 93-195) of the second copy of UP1.

CHAPTER 2. MATERIALS AND METHODS

2.1 UP1 Growth

A clone of UP1 was provided by Y. Shamoo and was originally constructed using the T7 expression vector p7A7W and the 1-196 fragment of hnRNP A1. This clone, pYS-45 and pYS-46, was transformed into the *E. coli* BL21(DE3) cell line and glycerol stocks were made. Glycerol stocks made with transformed pYS-45 clones were often contaminated with phage, and cells tended to lyse following mid-log induction. UP1 grown from glycerol stocks of pYS-46 occasionally produced the same result. As a consequence, LB plates with 0.1 mg/ml carbenicillin were streaked with pYS-46 glycerol stock in order to produce colonies for a mini-prep extraction of the plasmid. Four colonies were selected and minipreps made from each using the Promega Wizard Kit and kit protocol. The samples were labeled 1-4, dated 01-18-01, and stored at -80°C . All subsequent growths were performed using freshly transformed cells from miniprep 1, JCM, 01-18-01.

Early strategies for expression and purification of UP1 followed the procedures outlined by Shamoo *et al.* (1994), although some changes were made. For small-scale growths, a colony grown on 0.1 mg/ml carbenicillin LB plates from freshly transformed cells was selected and grown overnight in 50 ml of 2XYT with either 1 mg/ml carbenicillin or 1 mg/ml ampicillin at 37°C . One ml of the overnight growth was used to inoculate 1 L of LB with 0.1 mg/ml ampicillin in 4 L baffle flasks. A typical growth yielded ~ 30 g of cells per L of media. A 15 L fermenter growth was also performed using Terrific Broth (TB). A culture (300 ml) of a 1 L overnight at midlog growth was

added to the fermenter and allowed to grow to $OD_{595} = 0.5$ to 0.7 before induction with 0.65 mM IPTG. Cells were allowed to grow for 9 hours after induction, and growth tended to slow after about 8 hours, as pH control could not be maintained by the KH_2PO_4 and K_2HPO_4 buffer. The pH started at 7.4 and ended at 8.3 with an optical density of 2.25 . Antifoam was added but foaming still occurred, and the fermenter shut down on several occasions after induction as a result. Cells were pelleted in a 6 L bucket centrifuge for 30 minutes at $5,000$ rpm and 4 °C and frozen at -80 °C in 6 aliquots of about 65 g each, with a total cell yield of 394 g. The frozen cell paste was stored in weigh boats dated 06-21-01.

2.2 UP1 Purification

Cells were thawed on ice and disrupted by sonication in 100 to 200 ml of 10 mM Tris-HCl, pH 8.0 , 50 mM NaCl, 0.1 mM EDTA, 0.1 mM PMSF, 1 mM DTT, and 10% v/v glycerol (buffer A). Cells were sonicated with 4 bursts for 1 minute at 50% duty cycle at power level 6. The lysate was pelleted in an SS-34 rotor at 4 °C and $12,500$ rpm ($18,675$ RCF).

A DE52 (DEAE) anion exchange column was prepared by dissolving 35 g of resin in 250 ml of 0.1 M Tris-HCl, pH 8.5 , 1 M NaCl and resuspending for 3 washes. The resin was then washed and poured into a column in a buffer of 10 mM Tris-HCl, pH 8.0 , 50 mM NaCl, and 0.1 mM EDTA. After the column material settled, it was washed with 500 ml of buffer A. A ssDNA cellulose column was also prepared by dissolving 0.6 g of calf thymus DNA in 500 ml of 10 mM K_3PO_4 at pH 7.8 overnight at 4 °C. The dissolved DNA was boiled for 20 min in 10 ml pyrex test tubes and quickly cooled in a

slurry of NaCl and ice. The chilled solution was pooled in a beaker and the liquid exchanged with 1 M Tris-HCl, pH 7.4. Cellulose was then mixed with the solution (100 g of CFII) and spread out on a pan to dry for two or three days. The dried solution was then lyophilized for an additional two days and then stored at -80°C . The ssDNA cellulose column was prepared by washing in buffer A until 1 ml of wash had an absorbance at 280 nm < 0.1 . The DE52 column and ssDNA cellulose column were connected in tandem so that UP1 flowed through the DE52 resin (whereas other *E. coli* proteins bound to it), and UP1 was bound to the ssDNA cellulose resin. After washing in buffer A until the absorbance at 280 nm was < 0.1 , the DE52 column was removed and UP1 eluted from the ssDNA cellulose column with a step gradient using buffer A + 1.0 M NaCl. The protein was then dialyzed against 1 L of 20 mM MES, pH 6.0, 35 mM NaCl, 1 mM EDTA, 1 mM DTT, and 10% glycerol (v/v) (Buffer B) at 4°C overnight. A cloudy precipitate was always found in the dialysis bag the following day with pH and NaCl concentration reduction, and the precipitate was pelleted at 12,500 rpm in an SS-34 rotor (18,675 RCF) for 60 minutes. The supernatant was removed and passed through a Sterivex-GP™ 0.22 μm syringe filter prior to being loaded on a Pharmacia™ FPLC. At this point the lysate was usually 100 to 200 ml in volume, and the ratio of absorbance at 280 nm to 260 nm was normally ≈ 1.5 , indicating that DNA was sufficiently removed from the UP1 fraction. The total amount of UP1 was determined by measuring absorbance at 280 nm in equation (1):

$$A = \epsilon cl \quad (1)$$

where A is absorbance at 280 nm, ϵ is the extinction coefficient ($11,050 \text{ Lcm}^{-1}\text{mole}^{-1}$) at 280 nm, c is the concentration, and l is the pathlength of the cuvette (1 cm). The molar

amount of UP1 was determined by multiplying the concentration by the total volume of sample.

UP1 was then bound to a 5 ml Pharmacia HiTrapTM Heparin affinity column and eluted using a linear gradient from 0.035 M to 1.0 M NaCl over 15 column volumes. UP1 elutes early from the heparin column, and elution was monitored by absorbance at 280 nm. Peaks were pooled and the protein dialyzed into buffer B overnight. Samples from each peak and flowthrough were loaded onto a 12% poly acrylamide gel (SDS-PAGE) to identify UP1. The UP1 elution peak was again passed through a 0.22 μ m syringe filter prior to being loaded on a Pharmacia HiTrapTM Sepharose SPTM FF cation exchange column. UP1 was eluted from the Sepharose column in the same manner as with the Heparin column and was sufficiently pure at this point for crystallography. Overloaded 12% PAGE generally showed only trace amounts of other proteins. Purified UP1 was dialyzed against 20 mM MES, pH 6.0, 300 mM NaCl, 1 mM EDTA, and 0.1% β -mercaptoethanol (v/v), and the total amount of UP1 was again determined spectroscopically. The volume needed for a concentration of $\approx$ 40 mg/ml was determined, and UP1 was concentrated to this volume using a Vivascience Vivaspin 20, 10,000 MWCO concentrator. Concentrated UP1 was divided into 50 μ l aliquots in 1.5 ml Eppendorf tubes and flash frozen in liquid nitrogen and stored at -80°C .

2.3 Oligonucleotides

All oligonucleotides were obtained commercially (IDT, MWG Biotech Inc., TriLink Biotechnologies for 6MI) and purified with a Shimadzu HPLC (LC-10AT VP pumps and SPD-10A VP UV-VIS detector) using a Vydac C4 reverse phase column.

Dried oligonucleotides were hydrated in 0.8 ml of 25 mM ammonium acetate, pH 6.0, and loaded onto the column. A 30 minute gradient from 25 mM ammonium acetate to a buffer of 50% acetonitrile, 50% 25 mM ammonium acetate was used to elute oligonucleotides from the C4 column at 2 ml/minute flow rate and allowed good separation of the proper length species. The purified peaks were collected and frozen on dry ice and then lyophilized for one or two days. The dried oligonucleotides were then suspended in 0.5 ml of ddH₂O, lyophilized until dry (2 cycles), and finally hydrated in TE buffer, pH 8.0 (10 mM Tris-HCl, 1mM EDTA) to a volume that yielded $\approx$ 5.0 mM oligonucleotide.

2.4 Crystallization of UP1:hTR Complexes

Early crystallization trials were conducted with non-complexed UP1 using the previously published conditions of Shamoo *et al.* (1997) and Jokhan *et al.* (1997), but these trials were unsuccessful. After purification of the TR2 telomeric repeat oligonucleotide (TTAGGG)₂ used by Ding *et al.* (1999), a crystallization trial was prepared using buffers from a previous attempt to crystallize the non-complexed form using a mother liquor of 50 mM Tris-HCl, pH 8.1, 50 mM NaCl, and 28-32% PEG 1500. UP1 and TR2 were mixed in a 1:1.2 molar ratio and incubated on ice for 1.5 hours. The hanging drop method was used, and the drops consisted of 2 μ l of UP1:TR2 and 2 μ l of mother liquor from 27-32% PEG 1550. Crystals were seen 2 days later and reached full size in 4 days. The largest crystal was grown in 27% PEG 1550. From the conditions above, UP1:TR2 crystallized in space group P4₃2₁2 with unit cell dimensions of

$a=b=51.09\text{\AA}$, $c=171.95\text{\AA}$. Interestingly, the space group and unit cell dimensions are consistent with Ding *et al.* (1999), although the crystallization mother liquor is different.

As a general rule, crystals of UP1 complexed with hTRs that contained modified bases were grown from fairly low initial protein concentrations of around 12 mg/ml. The protein:DNA molar ratio was unique for each crystal but was generally higher for modified complexes (1:1.5-1:1.7) than for the native UP1:TR2 complex (1:1-1:1.2). The initial ratio of complex:precipitant in the drop was a crystallization variable, but most of the modified oligonucleotides required a ratio of at least 1:2.5. Some do not crystallize at all in ratios that approach 1:1, where UP1:TR2 crystals grow. Crystals took 3-5 days to grow, and roughly 50% of the complexes produced diffraction quality crystals.

2.5 Data Collection

Diffraction data for UP1:TR2-6F, UP1:TR2-11F, UP1:G(10)NEB, UP1:G(11)INO, UP1:G(10)2AP, UP1:TR18, and UP1:TR2-2C were collected on a Rigaku R-Axis IV⁺⁺ image plate detector. Data from UP1:A(9)7DA, UP1:A(9)NEB, and UP1:G(10)7DG were collected on a Rigaku JupiterTM CCD detector. The home source X-ray generator was an MSC RU-H3R with a rotating copper anode, and power was set at 50 kV and 100 mA. The beam was focused with Osmic MirrorsTM and collimated to 0.3 μm . Crystals were flash-frozen in a cryostream at -170°C , and data were collected under cryogenic conditions. Data for UP1-G(10)Ino and UP1-G(11)2AP were collected at the CHESS A1 beam line on an ADSC Quantum IV CCD Detector using a beam wavelength of $0.9764\text{\AA}$. Source to detector distance and angle were varied from 120-150 cm and $20\text{-}25^\circ$, respectively, for data collected on the JupiterTM CCD detector, but

fixed at 150 cm and 0° for data collected on the R-Axis IV++™. Detector distance was 180 cm for data collected at CHESS, with 10 sec exposures and 0.5° oscillations. Home source exposures were generally collected over at least 90° of crystal rotation (a minimum of 45° is required in the P4₃2₁2 space group) with 0.5° oscillations, and exposure times for each oscillation were 180-240 sec for the Jupiter™ CCD and 300-600 sec for the R-Axis IV++™. The UP1:TR2-11F crystal diffracted to 2.6 Å, but all other crystals diffracted to at least 2.1 Å with better than 89% completeness. All models fit the data better than $R = 0.25$ and $R_{\text{free}} = 0.29$.

2.6 Building heterocompounds for modified bases used in crystallographic refinement

Each modified base used in this study required the creation of parameter files for use in crystallographic refinement. The initial structures were built as pdb files using a graphical JAVA interface in XPLO2D (Kleywegt, 1995). Parameter, topology, and pdb files were then generated by XPLO2D, and these were used in the refinement process. The modified base pdb files were inserted in the proper position of the search model, and some modifications to the nucleic acid parameter, topology, and link files were required due to the unique alphabetical identifier assigned to each atom of the modified bases.

The program CNS (Brunger *et al.*, 1998) requires a molecular torsion file (.mtf) for refinement, and these were generated using the “generate” script in CNS. This process will default to placing 2'-OH groups on nucleic acid pdb's unless the protein chain and nucleic acid chain of the search model are separated into two different pdb files. Additionally, the command “END” must be entered after the last entry of both pdb

files, and “Rename Segid” should be selected as “true” in the “generate” input file (generate.inp). The structures described herein use the letter “A” to identify the protein and “B” for the nucleic acid. The command “SEGID B” must be entered in the “RNA to DNA conversion” box of the “generate” input file to prevent the addition of 2’-OH groups to the output pdb file. It is also important to verify that the chain and segment identification of all atoms within a pdb input file are identical. In one case where a modified base pdb file was inserted into the nucleic acid search model, the modified base used the segment identification “DNA”, whereas all other atoms in the search model used the letter “B”. This resulted in the truncation of the first atom of the modified base and that of the base that followed it.

The “generate” input file also requires parameter, topology, and link files and XPLO2D (Kleywegt, 1995) was used to generate these files for the modified bases used in this study. The parameter file specifies various constants for conformational and nonbonded energy terms. The topology file is required for identifying atoms, and the pdb nomenclature differs from that used in the parameter, topology, and final mtf file. The topology file is also needed for describing connectivity and bond angles. The link file was modified to indicate that bases bearing the names of those used in this study should be connected with standard biological bases in the 5’ to 3’ fashion.

The output .pdb and .mtf files from “generate” were used as inputs for running a simulated annealing process call “anneal”. “Anneal” was used as the first step after “generate” to identify mismatches in the naming of atoms in the modified bases in the topology and parameter files. The process would run until failing to find bond parameters for modified base atoms with a “name” that did not match a “type” between

the two files. The parameter files were corrected to match the topology files, and the corrections were mostly confined to the ribose sugar and phosphate backbone. After correcting the bond parameters, another cycle of “anneal” would stall where angle parameters were absent, again the result of failing to match atom types between the topology and parameter files. The process of correcting the parameter files was continued for dihedral, torsion, and nonbonded parameters until “anneal” was able to run without stalling.

After correcting the above files to accommodate modified bases, a “.dat” file for use in the program O was modified so that modified bases could be displayed linked to their 5’ and 3’ neighbors. The file all.dat was modified by naming the heterocompounds and identifying all bonds within them in the same format as biological bases. The new .dat file was stored in the same directory as that where O was being used and was loaded when prompted when starting O.

2.7 Refinement

Scaled and averaged reflection data from CRYSTAL CLEAR (Pflugrath, 1999) was converted from d*TREK format (.ref file) to SCALE PACK format (.hkl file) using a script called “a.out” (available in the “jmyers” home directory) so that CNS could be used in refinement. The header information from the .ref files was removed to allow the conversion. Five percent of the reflection data was randomly selected as the “test set” from the .hkl file and designated as “test = 1” in the generate.cv script in CNS. The test set was used to determine R_{free} during refinement, and these data remained untouched

throughout the process to prevent model bias. Reflection files with a test set had the suffix “.cv”, and the same file was used throughout refinement.

The initial search model for molecular replacement was 2UP1 from Ding *et al.* (1999) and did not contain modified bases. Cross rotation searches were performed with resolution limits set at 20-4 Å. Observed data cutoff criteria were left at CNS defaults of $\text{obs}/\sigma < 0.0$ and $\text{obs} > 1000$. The “fast direct” rotation function was selected, and no user-defined limits were placed on searching the asymmetric unit. Parameters for conducting the “fast direct” search and “cluster analysis” were left at the default CNS values.

Translation searches using data from the cross rotation were performed over a resolution range of 20.0-3.0 Å. Trials using higher resolution ranges of 20.0-2.5 Å or 20.0-2.0 Å were time consuming and rarely led to a better fit model. Data cutoff criteria and outlier cutoffs were left at 0.0 and 1,000 respectively, and all other CNS defaults were left unchanged.

Rigid body refinement was performed over the full resolution range of the data, generally from 20.0-2.0 Å. The default obs/σ criteria was kept at 0.0, and the outlier cutoff was generally left at 10,000, although a lower value of 1,000 would have been more appropriate. The number of minimization steps was typically left at the default of 20, and higher numbers generally did not give a significantly better fit. Only one cycle was performed for each structure, as multiple cycles had little effect on model statistics.

B-factor minimization was performed in two different ways with the “bindividual” and “bgroup” commands. The bindividual refinement is restrained, and the default restraint parameters were never changed. The bgroup refinement separates each

residue into two groups: one for main chain atoms, another for side chain atoms in proteins, a group for the sugar-phosphate backbone, and another for the bases in nucleic acids. An unrestrained B-factor refinement was performed on both groups, and it was important to insert the appropriate modified base residue name in the “select atoms in group” box of the input file. Refinement for both was performed over the entire resolution range of the data with default values for the various parameters. Generally the unrestrained bgroup refinement provided greater reductions in R_{free} , but the refinement that gave the greatest reduction was used in the next stage of refinement.

The best models following B-factor refinement were entered into a cycle of simulated annealing prior to manual refinement in O. Many of the defaults CNS chooses for the input file were changed in order to achieve better R_{free} . The initial number of steps of minimization was set to 200, the “type of molecular dynamics” was selected as torsion, the number of minimization steps to regularize was set at 100, and a slow cool annealing schedule was used. The starting temperature was changed to 4,000 K from the 2,500 K default, and, although temperatures as high as 10,000 K were tried, 4,000 K tended to produce the best results. The rate of cooling was changed from 50 K per step to 25. Smaller steps were tried, but 25 proved to be as effective and less time consuming. The number of molecular dynamics steps was 1000, and only one cycle of simulated annealing was selected for refinement prior to manual rebuild. As many as 12 trials with different initial velocities were run on the final model, and the best R_{free} was selected from among them. The maximum unbranched chain length and maximum number of distinct bodies had to be greater than the defaults of 50 and 5 respectively, and normally a value of 150 was entered for chain length and 50 entered for number of distinct bodies to

prevent the annealing process from generating an error and stalling. A typical error for these values being too small could be identified by the warning "too many trees" in the output file.

2.8 Equilibrium Fluorescence Titrations

Equilibrium binding experiments were performed using an SLM 8100 spectrometer (SLM Instruments). When performing "forward titrations", UP1 was titrated into a cuvette containing oligonucleotides incorporating 6MI and measuring the gain of signal as a function of complex formation (Lohman and Bujalowski, 1991). An excitation wavelength of 340 nm and emission of 430 nm was used at 25 °C with a 1 cm path length cuvette that contained a magnetic stir disk (Hawkins *et al.*, 1997). Ten measurements were taken 450 sec after each addition of UP1 and the average taken as one data point. A control of oligonucleotide alone was also used to correct for changes in the detector baseline and photobleaching. For titrations with varying NaCl concentrations, TR2-6F concentration was 250 nM in 20 mM MES, pH 6.0, and 4 nm band-passes were used for both the excitation and emission monochromators without polarizers. Titrations with variable pH used 10 nM TR2-6F in 20 mM MES (pH 6.0, 6.5) or 20 mM HEPES (pH 7.0, 7.5) in 300 mM NaCl, and 8 nm band-passes were used for both monochromators without polarizing filters. Titration curves were fitted using Dynafit (BioKin Ltd., Pullman, WA).

An early competition binding experiment where TR2 competed for TR2-6F was accomplished with an equimolar amount of UP1 and TR2-6F of 0.5 μ M in 100 mM NaCl and 20mM MES pH 6.0. The UP1:TR2-6F mixture was incubated on ice for 1 hour and allowed to warm to 25 °C in the spectrometer for 15 min prior to starting the experiment.

A sample of TR2 was slowly titrated into the cuvette ("reverse titration"), and the loss of signal was measured. Excitation and emission wavelengths were 340 nm and 430 nm, respectively, and 4 nm band-passes were used without polarizing filters.

Competition binding experiments using modified oligonucleotides were performed using an SLM 8100 spectrometer (SLM Instruments) with polarizing filters (0° excitation and 90° emission). UP1 and TR2-6F were mixed at equimolar concentrations of 150 nM in 20 mM HEPES, pH 7.4, 300 mM NaCl, and incubated on ice for at least one hour prior to the experiment, and allowed to warm to 25 °C for 15 min in a 1 cm path length stirring quartz cuvette. Competitor oligonucleotides were diluted from stock concentrations of 5 mM to 150 μM and brought to a final volume of 400 μl in 20 mM HEPES, pH 7.4, 300 mM NaCl with the 150 nM UP1:TR2-6F mixture so that the total concentration of UP1 and TR2-6F did not change during the titrations. The competitor oligonucleotides were titrated into the cuvette to compete with the fluorescent oligonucleotide from UP1 and the loss of signal recorded (Myers *et al.*, 2003). Excitation and emission wavelengths were 340 nm and 430 nm, respectively, and all four emission and excitation band-passes were set to 8 nm. Measurements were taken using a 10 second time base with five measurements for each titration and 450 seconds between titrations. The average of the five measurements taken at each titration point was used to fit the Hill equation

$$Y = \min + ((\min - \max) / (1 + 10^{((\log X - \log IC_{50}) \text{Hillslope})})) \quad (2)$$

where “min” is the signal at the end of the experiment after a “dark” oligonucleotide has out competed the fluorescent ligand (nonspecific binding), “max” is the signal at the

beginning of the experiment in the absence of competitor (specific binding), IC_{50} is the point at which half the binding sites are occupied by the fluorophore and half are occupied by the competing “dark” oligonucleotide, and the “Hillslope” is the slope of the sigmoidal dose response at the IC_{50} point. The relative K_d of the competitor oligonucleotide to that of the fluorophore was determined by identifying the IC_{50} point from the above fitted equation using the equation of Cheng and Prusoff (1973)

$$K_i = (IC_{50}) / (1 + ([TR2-6F] / K_d)) \quad (3)$$

where $[TR2-6F]$ is the concentration of free fluorescent oligonucleotide and K_d is that of the UP1:TR2-6F complex ($K_d = 130$ nM) that was determined using a “forward titration” (Lohman and Bujalowski, 1991; Myers *et al.*, 2003). In order to obtain a sufficient fluorescent signal for the experiment, equal concentrations of UP1 and fluorescent ligand were used at the beginning of the experiment. Thus, all binding constants derived from competition binding assays should be considered relative to the K_d of the protein and not true K_d 's since the total amount of TR2-6F did not approximate the free concentration.

2.9 Thermal Denaturation of Quadruplex Nucleic Acids

Melting assays were performed in either 150 mM KCl, 10 mM potassium cacodylate, pH 7.0, or 100 mM KCl, 10 mM potassium cacodylate, pH 6.5. Experiments were carried out with 3 μ M d(TAGGT)₄, 3 μ M mouse MN d(GGCAG)₄, and 6 μ M UP1. The melting assays were run on a Perkin-Elmer Lambda 20 UV/VIS Spectrometer that included a PTP6 Peltier System. All melting data were collected at 295 nm as a function of temperature over the range of 5 °C to 85 °C, with a fixed heating rate of 1.0 °C/min. Prior to melting, the DNA was incubated at 85 °C for 5 minutes and then allowed to

anneal to room temperature. For the protein:DNA complexes, the annealed DNA was chilled on ice before UP1 was added. For data collected at temperatures less than 20 °C the chamber of the spectrometer was sparged with nitrogen to reduce humidity.

2.10 Static Light Scattering

A miniDAWN three-angle light-scattering photometer (Wyatt Technology, Santa Barbara, CA) was used to determine the molecular weight of individual proteins and complexes. Samples of 150-300 µg were injected onto a Shodex KW803 gel filtration column at a flow rate of 0.5 ml/min and their elution monitored by a Shimadzu SPD-10Avp UV-VIS and Waters R401 refractive index detectors. All measurements were made in 0.1 M KCl, 10 mM potassium cacodylate, pH 6.5, at 20 °C.

The miniDAWN has three photodiode detectors (45°, 90°, and 135°), and the 90° detector was calibrated before beginning the experiment. Toluene was used for this procedure, as suggested by the manufacturer due to its high, well characterized Rayleigh ratio. Voltages were measured with the laser on and off, and a calibration constant was computed with the ASTRA software. The 45° and 135° detectors were normalized against the calibrated 90° detector in order to compensate for different detector sensitivities and scattering volumes. Bovine serum albumin (BSA) was used in this procedure because it is an isotropic light scatterer, and the normalization was carried out in the same buffer used for the molecular mass experiments. The buffer solution was from the same stock from this point forward since small variations between stocks have an impact on the solvent refractive index and render the normalization coefficients

inaccurate. All solutions injected into the flow cell were filtered with 0.02 μm filters and degassed.

The scattering intensity of a macromolecule is directly proportional to the molecular weight of the molecule, its concentration, and the square of the refractive index of the macromolecule in solution. The light scattering equation ASTRA uses to determine the molar mass of a particle is

$$K^* c / R_\theta = 1/M_w P_\theta + 2A_2 c \quad (4)$$

Where c is the concentration of the scattering molecule, R_θ is the excess Rayleigh ratio of light scattered by the molecule versus the intensity of that scattered by solvent ($R_\theta = (I_\theta - I_{\theta, \text{solvent}}) r^2 / I_0 V$), M_w is the molar mass of the molecule, P_θ is the theoretical scattering function and is unity at zero angle, A_2 is the second virial coefficient and

$$K^* = 4\pi^2 (dn/dc)^2 n_0^2 N_A^{-1} \lambda_0^4 \quad (5)$$

Where dn/dc is the refractive index increment which varies with solute concentration, n_0 is the solvent refractive index, N_A is Avogadro's number, and λ_0 is the vacuum wavelength of the incident light (690 nm).

The excess Rayleigh ratio (R_θ) is measured by the miniDAWN, whereas concentrations and refractive indices of buffer alone and with protein at different concentrations must be determined by the user. A refractometer was used downstream of the miniDAWN to measure refractive index, and a UV-VIS detector (280 nm) was used upstream of the miniDAWN to measure the amount of protein or protein:nucleic acid complex flowing off the size exclusion column, and 100% recovery was assumed. Dead

volumes were measured and used to correct for peak shifts that occur as a result of detectors being connected in series.

Molar mass was determined using a Zimm plot of K^*c/R_θ vs. $\sin^2(\theta/2)$ (Zimm). A dn/dc value of 0.184 was used for BSA and single proteins (Sun and Shamoo, 2003), and dn/dc for UP1:TR2 complexes was determined experimentally from UV-VIS and refractometer data. UV calibration constants were calculated from the known extinction coefficients of the components and confirmed empirically using their absorbance at 280 nm. A line was drawn to connect the data points from each photodiode and extrapolated to zero scattering angle where $P_\theta = 1$. Since the concentration of protein or protein:nucleic acid complex was low enough to allow the A_2 term to drop out, the y-intercept of the extrapolated line is equal to $1/M$.

CHAPTER 3. CRYSTALLIZATION OF UP1 BOUND TO OLIGONUCLEOTIDES USED IN THIS STUDY

3.1 Oligonucleotides Used in this Study

UP1 was bound to oligonucleotides that resemble the human telomeric repeat (hTR) sequence, d(TTAGGG)_n, except for the substitution of modified bases. Modified bases were used to eliminate hydrogen bonds at the protein:nucleic acid interface in order to identify elements that give UP1 base-specific affinity for the hTR and similar sequences. The modified bases used in this study are shown in Figure 3.1, and their ability to donate and accept hydrogen bonds is shown by arrows at the potential donor/acceptor sites.

3.2 Crystallization of UP1 to Modified hTRs

Initial crystallization trials focused on crystallization of the protein alone to show that the newly developed purification protocol was capable of producing UP1 to crystallographic standards of purity. After setting up a 24-well hanging drop tray with UP1 in PEG conditions appropriate for crystallizing UP1 in the absence of substrate (Shamoo *et al.*, 1997; Xu *et al.*, 1997), excess UP1 was mixed with an equal molar concentration of TR2 oligonucleotide as an afterthought and several drops were set up in the protein-only PEG conditions (Table 3.1). After 48 hours, it was observed that UP1 did not crystallize in the absence of TR2 and the UP1:TR2 crystals that grew from the PEG conditions diffracted to 2.0 Å. Although most modified oligonucleotides

FIGURE 3.1 Summary of bases used in this study. Arrows show hydrogen bond donor (arrows point away from the base) and acceptor groups (arrows point toward base).

crystallized in the published ammonium phosphate conditions for UP1:TR2 (Ding *et al.*, 1999), some only crystallized in the PEG conditions.

UP1 was able to form crystals when complexed with 18 out of 23 oligonucleotides used in this study. Of these, 12 produced diffraction quality crystals (Table 3.2). Interestingly, five sequences that did not crystallize did not contain modified bases but instead had natural base substitutions in the hTR. The inability of these sequences to crystallize may be reflective of the disruptive nature of placing improper

bases in the UP1 binding site. The majority of sequences that produced diffraction quality crystals had substitutions that affected the RRM2 binding site of UP1, prompting us to focus on the effects of disrupting interactions at RRM2. It is somewhat surprising

Table 3.1 Comparison of published crystallization protocols for UP1 and UP1:TR2 complexes

Shamoo (protein only)

Protein solution: 18 mg/ml UP1, 30 mM Tris (pH 8.1), 50 mM NaCl, 14%-16% PEG 1500

Precipitant: 50 mM Tris (pH 8.1), 50 mM NaCl, 28-32% PEG 1500

Result: Crystals in 5-7 days, space group $P2_1$, unit cell dimensions $a = 38.1 \text{ \AA}$, $b = 44.0 \text{ \AA}$, $c = 56.1 \text{ \AA}$, $\beta = 94.7^\circ$. Resolution to $1.75 \text{ \AA}$.

Xu (protein only)

Protein solution: >15 mg/ml UP1, 20 mM MES (pH 6.0), 300 mM NaCl, 1 mM EDTA, 0.1% BME

Precipitant: 100 mM Tris (pH 8.5), 25% PEG 4000, 10-15% glycerol or 20% 2-methyl-2,4-pentanediol

Result: Crystals in hours, space group $P2_1$, unit cell dimensions $a = 37.94 \text{ \AA}$, $b = 43.98 \text{ \AA}$, $c = 55.64 \text{ \AA}$, $\beta = 93.9^\circ$. Resolution to $2.0 \text{ \AA}$.

Ding (protein/DNA complex)

Protein/DNA solution: ~12 mg/ml UP1 in 1:1 molar ratio of 12mer human telomeric repeat ssDNA sequence (TTAGGGTTAGGG), 20 mM MES (pH 6.0), 300 mM NaCl, 1 mM EDTA, 0.1% BME

Precipitant: 100 mM Tris (pH 8.5), 2.0 M $(\text{NH}_4)_2\text{HPO}_4$, 15% glycerol

Result: Crystals 3 days, space group $P4_32_12$, unit cell dimensions $a = b = 51.20 \text{ \AA}$, $c = 171.09 \text{ \AA}$. Resolution to $2.1 \text{ \AA}$.

Myers (protein/DNA complex)

Protein/DNA solution: ~22 mg/ml UP1 in 1:1.2 molar ratio of 12mer human telomeric repeat ssDNA sequence, 20 mM MES (pH 6.0), 300 mM NaCl, 1 mM EDTA, 0.1% BME

Precipitant: either 50 mM Tris (pH 8.1), 50 mM NaCl, 28-32% PEG 1550, or 100 mM Tris (pH 8.5), 1.8-2.2 M $(\text{NH}_4)_2\text{HPO}_4$, 15% glycerol

Result: Crystals 3 days, space group $P4_32_12$, unit cell dimensions $a = b = 51.09 \text{ \AA}$, $c = 171.95 \text{ \AA}$. Resolution to $2.0 \text{ \AA}$ (home source, R-axis IV with osmic mirrors).

that modified bases that affected interactions to RRM1 were not tolerated crystallographically considering the high degree of homology between the two RRMs. Binding data were collected on those sequences that produced high quality crystals.

UP1 tended to crystallize readily in the presence of the native hTR in either PEG 1550 or ammonium phosphate over a fairly wide range of pH and salt conditions; however, crystallizing UP1 to oligonucleotides that contained modified bases required greater care. These complexes tended to crystallize in only one of two conditions that UP1 and native hTR have been shown to crystallize (PEG or ammonium phosphate), and the ratios of oligonucleotide to UP1 were higher than those required for wild-type hTRs. Higher ratios of precipitant were also required for complexes that contained modified bases as opposed to the native hTR.

Crystal seeding was important in obtaining crystals of UP1 bound to modified oligonucleotides. Only one complex, UP1:A(9)Neb, crystallized without seeding, but these crystals had soft edges (Figure 3.2h) and diffracted with high mosaicity and low resolution. This result is interesting in that it suggests that seeding can provide nucleation events superior to those that occur naturally in this system. The method of seeding also played a factor in obtaining quality crystals. The best crystals grew from streak seeding with a cat whisker that had lightly touched a UP1:TR2 crystal and was quickly brushed through a drop that had equilibrated for 36 to 48 hours. The surface of the cover slip where the whisker had been brushed was often visible after 24 hours since crystals readily form along the “seed trail”; however, some of the better-diffracting crystals formed at the surface of the drop.

Table 3.2 Summary of oligonucleotides used in crystallization studies with UP1.

Human Telomeric Repeat Sequence:

TR2 d(TTA GGG TTA GGG)

Alternative Sequences:

A. Do not crystallize

T(2)Cyt	d(TCA GGG TCA GGG)
A(3)Thy	d(TTT GGG TTA GGG)
T(8)Cyt	d(TTA GGG TCA GGG)
A(9)Thy	d(TTA GGG TTT GGG)
A(9)Gua	d(TTA GGG TTG GGG)

B. Crystallize but do not diffract

G(4)2,6DP	d(TTA 2,6-diaminopurine GG TTA GGG)
G(4)7DA	d(TTA 7-deaza-guanine GG TTA GGG)
G(5)2AP	d(TTA G 2-aminopurine G TTA GGG)
G(5)7DG	d(TTA G 7-deaza-guanine G TTA GGG)
A(9)4MI	d(TTA GGG TT 4-methylindole GGG)
A(9)Ino	d(TTA GGG TT inosine GGG)

C. Crystallize and diffract

T(2)Cyt	d(TCA GGG TTA GGG)
T(2)5MC	d(T 5-methyl-cytosine A GGG TTA GGG)
G(6)6MI	d(TTA GG 6MI TTA GGG)
A(9)7DA	d(TTA GGG TT 7-deaza-adenine GGG)
A(9)Neb	d(TTA GGG TT nebularine GGG)
G(10)2AP	d(TTA GGG TTA 2-aminopurine GG)
G(10)7DG	d(TTA GGG TTA 7-deaza-guanine GG)
G(10)Ino	d(TTA GGG TTA inosine GG)
G(10)Neb	d(TTA GGG TTA nebularine GG)
G(11)2AP	d(TTA GGG TTA G 2-aminopurine G)
G(11)6MI	d(TTA GGG TTA G 6MI G)
G(11)Ino	d(TTA GGG TTA G inosine G)

Figure 3.2 Examples of UP1 crystals. A, UP1 in the absence of substrate grown in 30% PEG 1550. B, UP1:TR2 crystals grown in 2.0 M $(\text{NH}_4)_2\text{HPO}_4$ (1:1.7 molar ratio UP1:TR2, 2 μl drops (1:1)). C, UP1:TR2 crystals grown in 30% PEG 1550, 1% βog (1:1.7 molar ratio UP1:TR2, 2 μl drops (1:1)). D, UP1:G(4)2,6DP crystals grown in 28-31% PEG 1550 (UP1= 21.6 mg/ml, 1:1 molar ratio UP1:G(4)2,6DP, 1:2 complex:precip, seeded). E, UP1:A(9)7DA crystals grown in 29% PEG 1550 (UP1= 12 mg/ml, 1:1.5-1.7 molar ratio UP1:A(9)7DA, 5.25 μl drops (1:2.5 complex:precip) seeded). F, UP1:A(9)Neb crystals grown in 2.0 M $(\text{NH}_4)_2\text{HPO}_4$ (UP1= 12 mg/ml, 1:1.5-1.7 molar ratio UP1:A(9)Neb, 5.25 μl drops (1:2.5 complex:precip) seeded). G, UP1:A(9)Neb crystals grown in 2.1 M $(\text{NH}_4)_2\text{HPO}_4$ (UP1= 12 mg/ml, 1:1.5-1.7 molar ratio UP1:A(9)Neb, 5.25 μl drops (1:2.5 complex:precip) seeded). H, UP1:A(9)Neb crystals in 2.3 M $(\text{NH}_4)_2\text{HPO}_4$ (UP1= 12 mg/ml, 1:1.5-1.7 molar ratio UP1:A(9)Neb, 5.25 μl drops (1:2.5 complex:precip) not seeded). I, UP1:G(5)2AP crystals grown in crystals grown in 28-31% PEG 1550, not seeded.. J, UP1:G(4)7DG crystals grown in 28-31% PEG 1550, not seeded. K, UP1:G(5)2AP crystals grown in 28-31% PEG 1550, not seeded. L, UP1:G(5)7DG grown in 28-31% PEG 1550, seeded.

Crystals that grew from PEG conditions were often difficult to harvest since PEG tends to form a film on the surface of the drop, but crystals harvested from the surface of ammonium phosphate drops often diffracted quite well, especially if only a small surface of the crystal was in contact with the surface of the drop. Crystals that grew with a large amount of surface area at the drop surface tended to diffract poorly. High quality crystals were harvested from the surface of the cover slip, but the stress of breaking them loose seemed have an effect on peak shape in the diffraction pattern.

Regardless of the attempts to dislodge “PEG film” from crystals grown in PEG conditions, the presence of this film tended to cause ice to form on the crystal in the cryo-stream, and ice rings were often a factor in data collection. Re-annealing the crystal by blocking the cold stream for 10 to 15 seconds and then repeating the process of flash freezing sometimes reduced or eliminated ice rings. Glycerol was also tried as a cryoprotectant, and placing crystals in a solution of mother liquor with 15 to 20% glycerol for 15 to 30 seconds often eliminated icing. However, the best diffraction data were collected in the absence of cryoprotectant in most cases.

Crystals grown in ammonium phosphate conditions did not have the problem of a thick film at the drop surface, and these conditions also contained 15% glycerol in the precipitant. Crystals grown from high ratios of precipitant to complex were generally flash frozen in the cryo-stream directly from the drop, and icing did not occur in most cases. As with the PEG conditions, re-annealing often reduced or eliminated the formation of ice.

3.3 Complexes that Crystallize but do not Diffract

Six UP1:oligonucleotide complexes crystallized but did not diffract X-rays, and the crystal morphologies of four examples were unlike those of complexes that did diffract. UP1:A(9)4MI and UP1:A(9)Ino produced the familiar diamond shaped crystals seen in the UP1:hTR complex, but we were unable to grow them large enough for data collection. In the absence of substrate, UP1 forms rod-like crystals (Figure 3.2a) in PEG 1550, and the crystal morphologies of UP1:G(4)7DG and UP1:G(5)2AP have the appearance of "bundled rods" (Figure 3.2 j, k). It is not clear if UP1 is bound to substrate in these examples, and attempts were made to gather unit cell information. However, these crystals did not diffract well enough to obtain these data. UP1:G(5)7DG formed spherulites (disordered crystalline masses), and UP1:G(5)2AP occasionally formed similar structures (Figure 3.2 l, i). The crystals of UP1:G(4)2,6DP resembled six-pointed stars (Figure 3.2 d) and did not diffract well enough to obtain unit cell information.

3.4 Complexes that Crystallize and Diffract

With the exception of the UP1:T(2)Cyt and UP1:T(2)5MC crystals, the UP1 complexes that crystallized and diffracted to 2.6 Å or better have been fully refined and their coordinates deposited in the protein data bank (PDB) (Table 3.3). All complexes crystallized in space group $p4_32_12$ and had unit cell dimensions of roughly $a=b=51$ Å and $c=172$ Å. All structures were at least 89% complete over their resolution ranges and had R_{merge} less than 10%. All R_{free} were less than 29%, and R_{calc} were 24% or better. The overall RMSD of C α of the modified complexes compared to those of the wild-type structure were less than 1.0 Å.

Table 3.3 Summary of crystal data of UP1 bound to modified oligonucleotides.

	G(10)Neb	G(10)2AP	G(10)Ino	G(10)7deazaG	A(9)Neb
Resolution range (Å)	20-2.1	20-1.9	20-1.8	20-2.0	20-2.0
R _{merge} (%)	8.5	5.6	3.2	8.1	6.2
Completeness over range (%)	95	99	97	96	91
I/σ average	9.6	7.8	46.5	7.9	9.9
I/σ highest resolution bin (0.1 Å)	3.9	1.8	15.3	3.4	3.3
R _{calc} /R _{free}	21.3/25.8	23.8/28.4	20.7/25.8	24.1/27.3	23.3/27.4
Estimated Coordinate error (Å)	0.25	0.27	0.21	0.31	0.28
Luzzatti plot (5Å cutoff) [†]					
Overall r.m.s. deviation (Cα) to wild-type structure (Å)	0.5	0.5	0.6	0.6	0.4
PDB	1U1N	1U1P	1U1Q	1U1M	1U1L
	A(9)7deaza	G(11)2AP	G(11)Ino	G(11)6MI	G(6)6MI
Resolution range (Å)	20-2.0	20-1.8	20-2.0	20-2.1	20-2.6
R _{merge} (%)	7.2	5.1	7.7	7.5	8.2
Completeness over range (%)	89	95	96	96	95
I/σ average	9.2	13.3	12.3	9.4	10.2
I/σ highest resolution bin (0.1 Å)	2.5	3.3	5.1	2.4	4.4
R _{calc} /R _{free}	23.4/26.8	24.0/26.5	23.1/27.3	23.0/26.0	24.0/27.0
Estimated Coordinate error (Å)	0.32	0.26	0.28	0.23	0.30
Luzzatti plot (5Å cutoff) [†]					
Overall r.m.s. deviation (Cα) to wild-type structure (Å)	0.7	0.5	0.5	0.5	0.5
PDB	1U1K	1U1R	1U1O	1PO6	1PGZ

CHAPTER 4 UP1 AND NUCLEIC ACIDS IN SOLUTION

4.1 Introduction

Our goal in this project was to gain a better understanding of the elements that dictate binding specificity of RNA binding proteins that contain a highly conserved RNA binding domain known as the RNA recognition motif (RRM). Using the human protein UP1 as a model, we systematically eliminated individual hydrogen bonds at the protein:nucleic acid interface by substituting modified bases into oligonucleotides that are based on the human telomeric repeat (hTR) sequence, $d(\text{TTAGGG})_n$, which UP1 has been shown to bind with high affinity (Ishikawa et al., 1993; LaBranche et al., 1998; Myers et al., 2003; Myers and Shamoo, 2004). The rationale was that we could identify important elements of binding specificity by measuring the contribution of individual hydrogen bonds to overall binding, and the crystal structures of UP1 bound to these modified nucleotides would help us to better understand the dynamics of RRM binding pockets. The crystal structure of UP1 bound to the hTR shows that UP1 binds as a dimer where two hTR strands contact RRM1 and RRM2 of different copies of protein in an anti-parallel fashion (Figure 1.5) (Ding *et al.*, 1999). UP1 and hnRNP A1 have been shown to bind distal splice sites in pre-mRNA, and the manner in which UP1 binds the hTR in the crystal structure is consistent with such a role where two copies of UP1 bind two strands of DNA arranged in anti-parallel fashion. Distal splice sites might be brought together by an arrangement that allows an intron to loop out, with multiple copies of hnRNP A1 or UP1 holding the splice site together. A similar mode of binding may occur at the end of telomeres where 3' overhangs have been shown to form a loop

structure known as a t-loop (Ford *et al.*, 2002). The crystal structure of the UP1:hTR complex is consistent with a model of A1/UP1 interactions *in vivo*, but in order to correlate our biochemical studies with our crystal structures, it was important to show that UP1 binds nucleic acid as a dimer in solution.

4.2 Light Scattering Data Shows UP1 Binds as a Dimer to Nucleic Acids in Solution

Light scattering was used to show that the UP1:hTR crystal structures are consistent with the way UP1 behaves in solution. Particles in solution scatter light as a function of their mass, concentration, and refractive index (Chapter 2), and the molecular mass of UP1 and UP1 complexed to hTR DNAs was determined by light scattering under conditions similar to those used in the fluorescence, thermal denaturation, and X-ray crystallography studies. UP1 alone was found to be monomeric with a mass of 23.5 kD, which was within 5% of the calculated mass, 22.1 kD, and in good agreement with the accuracy obtained from our known protein standard BSA (Figure 4.1). The UP1:d(TTAGGG)₂ complex had a mass of 41.9 kD and was 19.1% lower than the expected dimer mass assuming two copies of UP1 and of the DNA. A similar underestimate was observed for the mass of DNA polymerase:primer-template complexes (Sun and Shamoo, 2003). No other chromatographic peaks at 280 or 260 nm were observed in the elution profile, suggesting that the low estimate is the result of an error in the calculation of dn/dc and the UV calibration constant for the protein-nucleic acid complex. Accurate UV constants and dn/dc values for the UP1:TR2 complex were more difficult to derive given the high UV absorbance of TR2 and its weak scattering and small dn/dc. The total contribution of TR2 in the UV calibration constant was 85%. Our

binding data indicates that the TR2 oligonucleotide undergoes significant conformational changes when binding UP1, and our CD experiments show that TR2 forms G-tetrads in solution. The disruption of base stacking interactions and G-tetrads that occurs in order for the UP1:TR2 complex to form affects the UV calibration constant in a non-quantifiable way. Since TR2 contributes to the major portion of total absorbance in the complex, this led to significant errors in molecular weight determination.

Figure 4.1 shows the weight-averaged molecular mass across the chromatographic peaks for the bovine serum albumin (BSA) standard, UP1 alone, and UP1 bound to d(TTAGGG)₂. UV absorbance was measured as material flowed from the gel filtration column before flowing through the miniDAWN flowcell and refractometer. The lower molecular weight peaks are somewhat broader than those of the higher molecular weight BSA standard, but all are within the separation range of the Shodex KW803 gel filtration column. The weight-averaged molecular mass of each species as determined by the miniDAWN and ASTRA software is shown as data points for each peak. The BSA standard was also used to normalize the three high-angle detectors because BSA scatters light isotropically, as seen in the low error of the light scattering data. The UP1:TR2 complex and UP1 alone were somewhat more polydisperse at the solution conditions used in these experiments (0.1 M KCl, 10 mM potassium cacodylate, pH 6.5), and higher scattering error is apparent from the data.

Figure 4.1 Light scattering studies of UP1 complexed to d(TTAGGG)₂. Molecular mass is indicated for 9 μ M BSA (calculated mass = 66.4 kD) ($\square$); 9 μ M UP1 (calculated mass = 22.1 kD) ($\circ$); 9 μ M UP1 + 9 μ M d(TTAGGG)₂ (calculated monomer mass = 25.9 kD, dimer = 51.8 kD) ($\triangle$) in 100 mM KCl, 10 mM potassium cacodylate (pH 6.5) at 20°C. The chromatographic peaks are absorbance at 280 nm from the elution of the gel filtration column, and the points are the weight-averaged molecular masses across the peak for each species tested.

4.3 UP1 Destabilizes G-tetrads

Telomere length is stabilized in proliferating tumour cells (Neidle and Parkinson, 2003), and down regulating A1 production in a mouse erythroleukemic cell line leads to shortening of telomeres (LaBranche *et al.*, 1998). Direct involvement of hnRNP A1/UP1 with hTR and telomerase has been suggested (Fiset and Chabot, 2001), although our data would also support a more indirect role for hnRNP A1/UP1 in telomeric maintenance by destabilization of quadruplex structures. Quadruplexes arise from G-rich polynucleotides where the inter or intra strand associations of guanines form stable planar arrangements known as G-tetrads (Figure 4.2). These can stack with one another to form a quadruplex.

Figure 4.2 G-tetrad stabilized with a monovalent K^+ in the center. Stable G-tetrads can also be made with Na^+ and multiple tetrads can stack to form G-quadruplex structures. Quadruplexes can be made from either interstrand or intrastrand bases.

Presumably, the removal of such quadruplex structures would aid in telomerase processivity and fidelity.

Fukuda *et al.* (2002) have shown that UP1 is able to unfold unimolecular quadruplex DNA derived from mouse MN repeats from the sequence d(GGCAG)_n *in vitro*, and our own melting studies confirm these findings with this sequence as well as with the hTR repeat. hnRNP A1 and UP1 can destabilize quadruplex structures by binding to available single-stranded nucleic acid and thereby shift the equilibria toward an extended bound form of G-rich polynucleotides. Single-stranded DNA binding proteins (SSBs) use a similar mechanism for removing secondary structures during DNA replication. UP1 may function in an analogous manner to prevent G-tetrad stabilized quadruplexes.

In order to compare the affinities of UP1 to the hTR and to hTRs containing modified bases, it was necessary to determine if the oligonucleotides that contained modified bases were capable of forming G-tetrads. Although UP1 is a potent destabilizer of these structures, their formation and stability present a potential challenge when comparing dissociation constants from modified oligonucleotides, since each has varying degrees of quadruplex stability. The fluorescent hTR sequence used the pteridine ring compound 6-methyl-8-(2-deoxy- β -ribofuranosyl)isoxanthopterin (6MI) in place of Gua at position G6, where no contacts with RRM1 or RRM2 are made. 6MI was chosen as a Gua analog since it has been shown to form Gua-like Watson-Crick hydrogen bonds in duplex DNA; however, the pteridine was likely to induce steric clashes and prevent tetrad formation. The inability of modified hTR sequences to form G-tetrads would

likely result in overestimating the K_d since destabilizing these structures by UP1 would not be required.

4.4 Thermal Denaturation of G-tetrads

Destabilization of nucleic acid structures by proteins can be readily monitored as a reduction of the nucleic acid melting temperature (T_m) (Munroe and Dong, 1992). Both hnRNP A1 and UP1 are able to reduce the T_m of the poly r(A-U) and d(A-T) by having a much greater affinity for single-stranded nucleic acid compared to double-stranded. Fukuda *et al.* (2002) have shown that UP1 destabilizes the mouse MN repeat and variations of that sequence. We have also found that UP1 is able to strongly destabilize the G-tetrad stabilized quadruplex structures found in both human telomeric d(TAGGGT)₄ and mouse MN d(GGCAG)₅. Both the hTR and MN DNA quadruplexes have characteristic hypochromic shifts at 295 nm, and we have used the depression of this signature signal to monitor UP1 binding and subsequent destabilization of the quadruplex structures (Katahira *et al.*, 1999; Phan and Mergny, 2002). As shown in Figure 4.3, UP1 reduces the T_m of d(TAGGGT)₄ from 67.0 to 36.1 °C at 100 mM KCl, pH 7.0 and that of d(GGCAG)₅ from 56.0 to 17.8 °C at 150 mM KCl, pH 7.0. We were unable to measure a T_m for the interaction of UP1 to d(GGCAG)₅ at the more physiological 150 mM KCl concentration since melting was occurring prior to the start of the experiment at 5 °C. The melting data are consistent with recent circular dichroism studies that indicate UP1 is able to induce changes in the ellipticity of these repeat sequences (Fukuda *et al.*, 2002). Similar studies performed with d(TTAGGG)₂ and r(TTAGGG)₂ also showed a reduced T_m , but these could not be quantitated since the

melting began prior to the starting temperature (5 °C). Substitution of the guanine analog 6MI for either the sixth or eleventh guanine in d(TTAGGG)₂ showed that 6MI did not support G-tetrad formation.

Figure 4.3 Thermal denaturation of G-tetrad DNA by UP1. Normalized melting profiles at 295 nm of hTR4 = d(TAGGGT)₄ and mouse MN4 = (GGCAG)₅ alone, and in the presence of UP1. Destabilization of G-tetrad stabilized quadruplexes is determined by the change in the T_m (T_m = 50% of the hypochromic shift). (A) 3 μ M hTR4 (black); 3 μ M hTR4+6 μ M UP1 (gray) in 150 mM KCl, 10 mM potassium cacodylate pH 7.0. (B) 3 μ M mouse MN4 (black); 3 μ M mouse MN4+6 μ M UP1 (gray) in 100 mM KCl, 10 mM potassium cacodylate, pH 6.5.

4.5 Circular dichroism (CD) of modified oligonucleotides

Since we wished to compare the relative affinities of our modified oligonucleotides, we used circular dichroism to quantitate the ability of each oligonucleotide to form stable G-quadruplexes in the concentration range of our fluorescence studies (Balagurumoorthy and S.K., 1994). The telomeric G-quadruplex formed by d(TTAGGG)_n has a characteristic spectrum (Figure 4.4) and is readily distinguished from single-stranded DNA with characteristic positive peaks at 295 and

250 nm, and a small negative peak at 266 nm (Balagurumoorthy and S.K., 1994). Heat denaturation of the G-quadruplex at 45 °C (Figure 4.4(a)) induces changes in the CD spectra that reduce the positive peak at 295 nm to near baseline and shifts the peak at 250 nm to around 260 nm. As shown in Figure 4.4, all of our oligonucleotides with modifications at Gua10 and -11 abrogated G-quadruplex formation, thus allowing a reliable comparison of their relative affinities.

A crystal structure of the human telomeric repeat suggests that Ade9 is not directly involved in the formation of the planar G-quadruplex interactions, but rather loops out into solvent (Parkinson *et al.*, 2002). Our CD studies agree with the structure in that modifications to Ade9 did not prevent formation of a stable G-quadruplex (Figure 4.4(b)), whereas modifications of Gua10 and -11 did. Substitution of Ade9 did reduce 295 nm peak intensity to approximately half of that seen for d(TTAGGG)₂, suggesting that the modifications at Ade9 somewhat reduce the stability of the G-quadruplex.

An hTR with inosine placed in the Gua11 position was synthesized to measure UP1 binding to an hTR where no hydrogen bonds are affected but the ability of the oligonucleotide to form G-tetrads was. The crystal data show no significant differences from the UP1:hTR wild-type structure, but the strength of binding is slightly higher ($K_d = 70$ nM for wild-type hTR and 60 nM for G(11)Ino). This small difference allows reasonable comparisons of UP1 binding data collected from oligonucleotides that form quadruplexes and those that do not.

Figure 4.4 Circular dichroism spectra of oligonucleotides. The human telomeric repeats $d(\text{TTAGGG})_n$ have a strong tendency to form G-quadruplex structures. The G-quadruplexes formed by these repeats have a characteristic spectrum with strong positive peaks at 295 and 250 nm and a shallow negative region at 266 nm. As shown in (a) the G-quadruplex can be denatured at high temperature to a single-stranded species with a single positive peak at 260 nm and loss of the peak at 295 nm. 25°C ($\square \square$), 35°C ($\square \bullet \square$), 45°C ($\square \square$). Panels (b-d) show spectra obtained for each modified oligonucleotide by position. Panel (b) Adenine-9 substitutions: A(9)7deaza ($\square \square$), A(9)Neb ($\square \square$); (c) Guanine-10 substitutions: G(10)7deazaG ($\square \square$), G(10)Ino ($\square \bullet \square$), G(10)2AP ($\square \bullet \bullet \square$), G(10)Neb ($\square \square$), G(10)A ($-\cdot-\cdot-$); and (d) Guanine-11 substitutions: G(11)Ino ($\square \square$), G(11)2AP ($\square \square$).

4.6 Structure-based Incorporation of a Fluorescent Guanine Analog as a Tool for Studying UP1:hTR Binding

It was important to have a sensitive assay to measure the changes in binding that result from the removal of one or two hydrogen bonds since each hydrogen bond represents slightly more than 2% of hydrogen bonding, stacking, and electrostatic interactions in the UP1:hTR complex. We needed to be able to see changes as small as two-fold reductions in binding in the K_d range of 50 to 500 nM, and, since numerous fluorescence binding studies were performed on hnRNP A1 and UP1 in the past, we investigated the utility of fluorescence for our own studies.

Preliminary binding experiments with UP1 and TR2, the 12-nucleotide human telomeric repeat (hTR), utilized intrinsic tryptophan fluorescence to monitor binding. UP1 has a single tryptophan at position 37 in the $\alpha 1$ helix of RRM1 that is not directly involved in binding, although the environment of this residue changes upon binding as indicated by a change in fluorescent properties. This property of Trp37 was exploited in early binding studies with "reverse titrations" of hnRNP A1 where oligonucleotides were titrated into fixed concentrations of protein and a gain of signal was recorded with excitation and emission wavelengths of 280 and 325 nm respectively (Abdul-Manan and Williams, 1996). We attempted similar experiments with UP1 and also observed a gain in signal when UP1 was titrated with TR2; however, the weak intrinsic fluorescence resulted in noisy data that were difficult to interpret. This result was in part due to the high inner filter absorbance that occurs when titrating nucleic acid into a protein sample that is excited at 280 nm. We were also concerned that inner filter corrections obtained from titrations of N-acetyltryptophanamide (NATA) with TR2 would be inaccurate since

TR2 forms G-tetrads in solution and changes structure when binding to UP1. Thus, the total excitation signal absorbance of TR2 would be expected to change depending on whether the oligonucleotide is free in solution or bound to UP1, and this cannot be determined by a NATA titration. Errors encountered in the light scattering experiments that were used to determine the molecular weight of the UP1:TR2 complex in solution are likely the result of the same phenomenon.

It was possible to significantly reduce the inner filter effect by exciting UP1 at 295 nm, but this technique required a combination of higher UP1 concentrations and more sensitive settings of the spectrophotometer photo multiplier tubes (PMTs). This situation was again problematic since higher UP1 concentrations required greater amounts of TR2 in order to achieve maximum fluorescence. With sufficiently high concentrations of UP1, inner filter absorbance again made it difficult to estimate total binding. Increasing the sensitivity of the PMTs allowed titrations at lower UP1 concentrations and consequently lower concentrations of TR2 at the total binding point; however, increasing sensitivity also increased signal noise, which led to significant errors when fitting the data. Some of the noise was the result of scattering, and, while polarizing filters improved the quality of the data, they did so at the expense of further reducing the emission signal.

We were able to find conditions that optimized the intrinsic tryptophan fluorescence of UP1 that also limited the inner filter effect and reduced the noise associated with high sensitivity. However, we were concerned that since we had optimized for UP1:TR2 binding, oligonucleotides designed to decrease binding might not behave well in similar conditions. We were also concerned that the assay would not be

sensitive enough to see small reductions in binding, such as the loss of a single hydrogen bond. Because of these limitations, we decided to explore the possibility of using a fluorescent dye that would allow us to use an excitation signal downfield of the absorption range of both the protein and nucleic acid and have little effect, if any, on UP1 binding. There are numerous oligonucleotide modifications available, such as fluorescein, that affect only the 5'-end, but we were concerned that these dyes would not crystallize with UP1. The structure of UP1 bound to a fluorescently labeled oligonucleotide was necessary in order to validate its use in our binding studies. Dr. S. Singleton and Dr. A. Roca suggested we investigate the use of pteridine compounds that have been used as Ade and Gua analogs in duplex DNA, and we used the crystal structure of UP1 to determine an appropriate position for the incorporation of a fluorescent Gua analog into the hTR.

4.7 Use of 6-Methyl-8-(2-deoxy- β -ribofuranosyl)isoxanthopteridine (6MI) in hTRs as a Reporter Molecule for UP1:hTR Interactions

We were able to show that the proper positioning of 6MI as a reporter can be highly effective for structure-function studies when there is a high-resolution structure available. Based upon previous studies suggesting that 6MI is an excellent fluorescent guanosine analog in duplex DNA, we synthesized two singly substituted oligonucleotides with two hTR repeats: TR2-6F and TR2-11F. Unlike dsDNA-binding proteins, single-stranded nucleic acid binding proteins can use nearly all the sugar-phosphate backbone and base positions for specificity, and some care is necessary when using modified nucleotides (Allers and Shamoo, 2001). 6MI closely resembles guanine (Figure 4.5) and

Figure 4.5 Comparison of Gua and the fluorescent Gua analog used in this study, 6MI. 6MI has been shown to form Watson-Crick base pairs in duplex DNA without affecting the T_m of the oligonucleotides (Hawkins *et al.*, 1997), and the arrows indicate hydrogen bond donor (pointing away) and acceptor (pointing toward) groups. We used 6MI to replace Gua6 in the TR2 sequence for our fluorescence studies and also showed that 6MI can bind in the Gua11 binding pocket of UP1.

has a position analogous to N7 but the presence of a methyl group at the pteridine C6 would make this position less accessible. Other steric and electrostatic differences induced by the methyl and carbonyl groups at position six and seven respectively could undermine sequence-specific protein:ssDNA interactions such that its utility may be limited. This situation prompted us to use the UP1:hTR structure as a guide for positioning of 6MI as a reporter for binding.

Using the program ENTANGLE (Allers and Shamoo, 2001) to inspect the UP1-hTR co-structure, we found that the guanine base at position six makes no direct contacts with UP1 and is solvent accessible upon complex formation. The substitution of Gua11 with 6MI was also chosen to test the utility of 6MI as a probe for binding in those cases

where the fluorescent base remains stacked upon binding. As before, 6MI was substituted at a position where interactions with UP1 are limited, although in this case the Gua11 makes a hydrogen bond to the main chain amide of Lys183 and stacks between Gua10 and Gua12. In neither case were substantive changes were seen in the crystal structures of the 6MI substituted oligonucleotides bound to UP1 compared to wild-type, suggesting that we were successful in an innocuous placement of the modified nucleosides (Figure 4.6).

The oligonucleotides TR2-6F and TR2-11F were then evaluated as reporters for UP1-hTR complex formation by exciting at 340 nm and measuring the gain-of-signal at 430 nm. The two positions represent different approaches to employing 6MI. The first substitution at position six was chosen on the assumption that the purine-rich hTR remains partially stacked as a single-stranded oligonucleotide. DNA melting studies have shown that incorporation of 6MI at either the six or eleven position abrogate quadruplex formation under the solution conditions used for our binding studies. This result is not surprising considering the strict geometry of quadruplexes and the fact that the analogous N7 of 6MI would not be positioned properly to support tetrad formation.

Figure 4.6 (a) RIBBONS diagram (Carson, 1991) of UP1 bound to the sequence d(TTAGGGTTAGGG). The position of the 6MI substitutions in oligonucleotides TR2-6F and TR2-11F are indicated (circled). (b) 2F_o-F_c electron density from a composite omit map of 6MI substituted for Gua11 (TR2-11F) contoured at 1.2 sigma. (c) Electron density of 6MI from a 2F_o-F_c composite omit map generated using phases obtained from molecular replacement of UP1 bound to d(TTAGGGTTAGGG) contoured at 1.2 sigma. The model shows the position of 6MI in TR2-6F (blue). The electron density corresponding to the phosphodiester and deoxyribose atoms at position six are of good quality, whereas the highly mobile base appears only weakly. In both the original and substituted DNA, the base at position 11 stacks with Gua10 and makes interactions through the base O6 (O4 in 6MI) to the mainchain of UP1. In contrast to TR2-6F, the electron density of 6MI in TR2-11F is moderately well ordered.

4.8 6MI Gain-of-Signal Depends Strongly upon Change of Environment During Complex Formation

The gain of fluorescent signal seen in Figure 4.7 is consistent with the unstacking of bases upon complex formation with UP1, and its strength permitted us to decrease the concentration of TR2-6F in our binding studies to 10 nM to obtain binding isotherms closer to equilibrium tight binding conditions. Figure 4.7 shows the fluorescence emission signal exchange of TR2-6F and TR2-11F at a concentration of 10 nM during UP1 binding. TR2-6F shows a 3-4 fold increase in fluorescence emission at 430 nm upon complex formation, whereas TR2-11F shows only a modest 1.2 fold increase in intensity. This result is in marked contrast to either intrinsic tryptophan fluorescence of UP1 or ribo-ethanlylated RNA that requires micromolar quantities of reporter, thus limiting the ability to measure binding of UP1 to high affinity targets such as the hTR. The differences in magnitude of the changes in intensity for the TR2-6F and TR2-11F are consistent with their structures when bound to UP1: 6MI in TR2-6F is completely solvent accessible and that in TR2-11F remains stacked with Gua10.

Estimates of equilibrium dissociation rates were made for both oligonucleotides using a forward titration of 10 nM TR2-6F and TR2-11F with increasing amounts of UP1. Binding isotherms were fitted to the data (Materials and Methods) and displayed stoichiometric binding to TR2-6F (~50 nM) at salt concentrations approximating those found *in vivo* (100-200 mM). In order to determine accurate dissociation constants, it was necessary to evaluate binding as a function of salt concentration and pH.

Figure 4.7 Fluorescence titration of UP1 with TR2-6F (●) and TR2-11F (■). Fluorescent signal of 6MI is strongly influenced by its environment. In TR2-6F, 6MI becomes fully unstacked and solvent exposed upon UP1 binding. In TR2-11F, 6MI stacks with neighboring Gua10 when bound to UP1. Excitation and emission wavelengths were 340 nm and 430 nm, respectively, and both titrations were performed in 20 mM MES, pH 6.0, and 100 mM NaCl at 25 °C. Total gain of signal for TR2-6F was 3.3 times the baseline and for TR2-11F it was 1.2 times baseline.

4.9 Fluorescence Binding Studies of UP1 to TR2-6F as a Function of Salt

Concentration and pH

Figure 4.8 shows the binding of UP1 to TR2-6F as a function of NaCl concentration (Table 4.1). The binding of UP1 is moderately sensitive to salt concentration, consistent with its role in both specific and non-sequence specific binding of single-stranded nucleic acids. The slope of the $\log K_a$ versus the log of salt concentration (Figure 4.9) was used to estimate the electrostatic contribution for TR2-6F binding to UP1 (Mascotti and Lohman, 1990; Record, 1976). Data were fitted by a linear least squares fit to produce a slope = -3.8 and y-intercept = 16. A slope of -3.8 is in good agreement with UP1 binding to ssRNA although the overall binding to TR2-6F is

$y = X/(M1+X)$		
	Value	Error
m1	1.5966	0.016673
Chisq	0.0007391	NA
R	0.99971	NA

$y = X/(M1+X)$		
	Value	Error
m1	5.343	0.082878
Chisq	0.0008726	NA
R	0.99907	NA

$y = X/(M1+X)$		
	Value	Error
m1	13.097	0.30037
Chisq	0.00098045	NA
R	0.9964	NA

Figure 4.8 Fluorescence titrations of .250 μM TR2-6F with UP1 at 600 mM, 800 mM, and 900 mM NaCl in 20 mM MES, pH 6.0. The concentration of fluorescent oligonucleotide was roughly an order of magnitude below the K_d or less, and complete saturation was determined by performing titrations under tight binding conditions (data not shown). The concentration of UP1 is the total amount titrated in μM and M1 is the estimated K_d in the equations above. Error bars represent 5% at each data point.

Figure 4.9 Log K_a of titrations performed at 400 mM, 600 mM, 800 mM, and 900 mM NaCl plotted against the log [NaCl] mM. Error bars represent 5% error and the line through these data points has a slope = -3.8 with a y-intercept of 16. The number of ionic interactions was estimated to ~ 5 and was determined by dividing the slope by the thermodynamic constant for the fraction of counterions bound per phosphate for ssDNA ($\psi = 0.71$) (Record, 1976) as described by Mascotti and Lohman (1990).

~1000-fold stronger than values measured for poly r(ϵ A) and poly (U) (Nadler *et al.*, 1991; Shamoo *et al.*, 1994). The number of ionic interactions was estimated to be ~ 5, as determined by dividing the slope by the thermodynamic constant for the fraction of counterions bound per phosphate for ssDNA ($\psi = 0.71$) (Record *et al.*, 1976) as described by Mascotti and Lohman (1990). This number is in good agreement with the crystal structure of the UP1:TR2 complex. ENTANGLE identifies 4 such interactions, further supporting the similarities of the complex in solution with that of the crystal. Figures 4.10-4.13 and Table 4.2 show that pH has little effect on UP1:TR2-6F complex formation over a range of pH 6.0 – 7.5.

Table 4.1 Dissociation Constants measured for UP1 binding to TR2-6F at increasing concentrations of NaCl.

[NaCl] mM	K_d μ M
600	1.6
800	5.3
900	13

Table 4.2 Dissociation Constants measured for UP1 binding to TR2-6F as a function of pH.

pH	K_d nM
6.0	150
6.5	130
7.0	150
7.5	180

Figure 4.10 Fluorescence titration of 10 nM TR2-6F with UP1 at pH 6.0 in 300 mM NaCl, 20 mM MES. The concentration of fluorescent oligonucleotide was less than an order of magnitude of the K_d , and the concentration of UP1 represents the total amount titrated in μM . The value of M1 is the K_d and error bars represent 5% error at each data point.

Figure 4.11 Fluorescence titration of 10 nM TR2-6F with UP1 at pH 6.5 in 300 mM NaCl, 20 mM MES. The concentration of fluorescent oligonucleotide was less than an order of magnitude of the K_d , and the concentration of UP1 represents the total amount titrated in μ M. The value of M1 is the K_d and error bars represent 5% error at each data point.

Figure 4.12 Fluorescence titration of 10 nM TR2-6F with UP1 at pH 7.0 in 300 mM NaCl, 20 mM HEPES. The concentration of fluorescent oligonucleotide was less than an order of magnitude of the K_d , and the concentration of UP1 represents the total amount titrated in μM . The value of $M1$ is the K_d and error bars represent 5% error at each data point.

Figure 4.13 Fluorescence titration of 10 nM TR2-6F with UP1 at pH 7.5 in 300 mM NaCl, 20 mM HEPES. The concentration of fluorescent oligonucleotide was less than an order of magnitude of the K_d , and the concentration of UP1 represents the total amount titrated in μM . The value of M1 is the K_d and error bars represent 5% error at each data point.

4.10 Competition Binding Studies

In light of our findings that TR2-6F and TR2-11F did not support G-tetrad formation, we measured the affinity of UP1 for the sequence d(TTAGGG)₂ in a competition binding assay. The stable G-tetrad structure observed for the hTR and MN DNA by thermal denaturation studies would presumably have to be denatured by UP1 to form the extended structure seen in the crystallographic studies. The binding of UP1 to TR2 at 100 mM NaCl was found to be ~5 nM at pH 6.0 in the competition assay, although direct measurements using TR2-6F yield a $K_d \sim 15$ nM. The fluorescent analog prevents G-tetrad formation, but hTR tetrad structures have reduced stability in shorter oligonucleotides and may be essentially absent at low concentrations. The weaker binding of the non-tetrad forming TR2-6F may reflect the energy required to unstack 6MI from neighboring bases and make it solvent accessible. Competition binding experiments using electrophoretic mobility-shift assay (EMSA) carried out by Fukuda *et al.* (2002) with UP1 and mouse MN DNA as well as hTR DNA showed similar binding affinities to longer repeats. UP1 bound a d(GTTAGG)₅ and d(GTCAGG)₅ sequence with 3 nM and 4 nM affinity respectively at pH 7.4 in 150 mM NaCl.

4.11 Competition Binding Experiments of UP1:TR2-6F and hTRs that Contain Modified Bases

The affinity of UP1 for modified DNAs was determined by competition assays in which UP1 bound to TR2-6F (both at 150 nM) was challenged by non-fluorescent oligonucleotides of interest in 300 mM NaCl at pH 7.4. Higher concentrations of 6MI

were used so that polarizing filters could be utilized to reduce signal errors caused by light scattering. Higher salt concentrations also helped reduce error and lowered binding affinities to the wild-type hTR sequence by roughly an order of magnitude. However, because equal concentrations of UP1 and TR2-6F were used, it was not possible to determine the free concentration of TR2-6F. Therefore, the binding constants derived by this method are only useful when comparing the competition assays against each other. Formation of the UP1:TR2F complex resulted in a ~200% increase in emission at 430 nm when excited at 340 nm and the strength of the emission was advantageous for recording the moderately weaker binding isotherms we expected from our modified oligonucleotides.

Using the native d(TTAGGG)₂ sequence as a competitor, we measured an affinity of 70 nM in 300 mM NaCl, 20 mM HEPES pH 7.4 (Figure 4.10), which is in good agreement with earlier measurements of UP1 affinity for telomeric DNA sequences (Fukuda *et al.*, 2002; Myers *et al.*, 2003). The results of the competition binding assays of UP1 for each modified sequence are shown in Figures 4.14-4.16 and Table 4.3. The range of measured affinities varied from an increase in affinity for G(10)Ino (60 nM) to a nearly 10-fold decrease for G(10)A. Few general trends can be ascertained from these binding data without an appropriate structure of the complex upon which to base them. Given that the affinity of UP1 for the sequence d(TTAGGG)₂ is approximately 1000-fold stronger than that for a comparable non-specific single-stranded sequence (Nadler *et al.*, 1991) and involves many contacts to the DNA bases, it is not surprising that our intentionally modest changes have not produced dramatic changes in affinity. The range

of affinities observed is consistent with the effects expected from the loss of 1-2 non-ideal hydrogen bonding interactions

Table 4.3 Dissociation constants for UP1 binding to oligonucleotides from competition assays.

Oligonucleotide	K _d (nM)
Wild-type	70
A(9)7deazaA	280
A(9)Neb	90
G(10)7deazaG	250
G(10)Ino	80
G(10)2AP	160
G(10)Neb	270
G(10)Ade	~500-1000 ^a
G(11)Ino	60
G(11)2AP	180

^a Very weak binding of this oligonucleotide precluded an accurate estimate of affinity at 0.3 M NaCl.

y = 1/(1+10 ^{^(log(X)-log(M1))} ...		
	Value	Error
m1	158.47	2.8071
m2	1.7236	0.062263
Chisq	0.0095442	NA
R	0.99767	NA

y = 1/(1+10 ^{^(log(X)-log(M1))} ...		
	Value	Error
m1	601.65	13.584
m2	0.9004	0.018329
Chisq	0.0076729	NA
R	0.99862	NA

y = 1/(1+10 ^{^(log(X)-log(M1))} ...		
	Value	Error
m1	198.24	6.5502
m2	1.2088	0.045135
Chisq	0.0073655	NA
R	0.99794	NA

Figure 4.14 Competition binding assay of modified oligonucleotides versus the fluorescently labeled oligonucleotide d(TTAGGG(6MI)TTAGGG). Competitor oligonucleotides were titrated into a cuvette containing 150 nM UP1 and 150 nM d(TTAGGG(6MI)TTAGGG) in 300 mM NaCl and 20 mM HEPES pH 7.4. Data were fitted to the Hill equation where “M1” is the IC₅₀ point and “M2” is the “Hill slope” or the slope of the curve at the IC₅₀ point. Each data point represents the average of five measurements and the standard deviation of each point is shown with error bars.

Figure 4.15 Competition binding assay of modified oligonucleotides versus the fluorescently labeled oligonucleotide d(TTAGGG(6MI)TTAGGG). Competitor oligonucleotides were titrated into a cuvette containing 150 nM UPI and 150 nM d(TTAGGG(6MI)TTAGGG) in 300 mM NaCl and 20 mM HEPES pH 7.4. Data were fitted to the Hill equation where “M1” is the IC50 point and “M2” is the “Hill slope” or the slope of the curve at the IC50 point. Each data point represents the average of five measurements and the standard deviation of each point is shown with error bars.

y = 1/(1+10 ^{((log(X)-log(M1))...}		
	Value	Error
m1	158.47	2.8071
m2	1.7236	0.062263
Chisq	0.0095442	NA
R	0.99767	NA

y = 1/(1+10 ^{((log(X)-log(M1))...}		
	Value	Error
m1	126.36	1.6409
m2	1.2363	0.021483
Chisq	0.0035177	NA
R	0.99918	NA

y = 1/(1+10 ^{((log(X)-log(M1))...}		
	Value	Error
m1	382.72	19.55
m2	1.0342	0.052007
Chisq	0.013006	NA
R	0.99622	NA

Figure 4.16 Competition binding assay of modified oligonucleotides versus the fluorescently labeled oligonucleotide d(TTAGGG(6MI)TTAGGG). Competitor oligonucleotides were titrated into a cuvette containing 150 nM UP1 and 150 nM d(TTAGGG(6MI)TTAGGG) in 300 mM NaCl and 20 mM HEPES pH 7.4. Data were fitted to the Hill equation where “M1” is the IC50 point and “M2” is the “Hill slope” or the slope of the curve at the IC50 point. Each data point represents the average of five measurements and the standard deviation of each point is shown with error bars.

4.12 Conclusions from Competition Binding Experiments

The strength of UP1 binding to hTR sequences is consistent with a physiologically relevant role for hnRNP A1/UP1 in binding available purine-rich sequences (nYAGGn) found in consensus splice sites, hTR, and MN sequences. UP1 binds stoichiometrically to d(TTAGGG)₂, with an estimated K_d of ~5 nM in 100 mM NaCl at pH 6.0. UP1 binding is only moderately sensitive to salt and pH (Tables 4.1 and 4.2). Electrostatic contributions make up ~15% of the overall affinity of UP1 for hTR DNA, similar to that measured for non-sequence specific UP1 binding to polynucleotides. Record *et al.* (1976) have shown that the number of ionic interactions is correlated to the slope of data fitted to a log K_a versus log [NaCl] plot. In the case of UP1, the number of interactions predicted is ~ 5, and the crystal structure shows that four residues, Arg55, Arg92, Arg140 and Arg146, are within 5 Å of phosphodiester oxygen atoms. Since one proposed role for hnRNP A1 *in vivo* is the generalized transport of mature pol-II mRNAs, these data suggest that the transport of mRNA may proceed through non-specific and largely electrostatic interactions, whereas the high affinity association of UP1 and hnRNP1 A1 to purine-rich sequences containing the consensus sequence (nYAGGn) will be dominated by hydrogen bonding, stacking, and hydrophobic interactions as seen in the crystal structure.

CHAPTER 5. ANALYSIS OF Ade9 BINDING SURFACE USING 7-DEAZA-ADENINE AND NEBULARINE

5.1 Ade9 Binding Pocket

In the native UP1:d(TTAGGG)₂ structure, Ade9 is stacked over Phe108 and makes hydrogen bonds from the N7 to the Arg178 guanidinium group (2.7 Å), from N6 to the main chain carbonyl of Lys179 (3.0 Å), and from N1 to the Leu181 amide (2.7 Å) (Figure 5.2a). Arg178 also makes a hydrogen bond to the O2 of Thy8. We originally thought that the Arg178 to Ade N7 interaction gave the binding pocket specificity for purines and that a pyrimidine in the binding site would not be able to interact with Arg, thereby disrupting the Arg178 to Thy8 interaction. We also thought that, while the O6 position of Gua would clash electrostatically with the main chain carbonyl at Lys179, the distance between the two hydrogen bond acceptors may be great enough (3.00 Å) to accommodate the clash. The interaction between Ade N1 and the Leu181 main chain amide may provide specificity for Ade and act to exclude Gua, since a steric and electrostatic hydrogen bond donor-donor clash would be made between the base and the rigid main chain. Additional clashes may occur between a Gua N2 position and the main chain and to the neighboring Gua10. We utilized a combination of modified bases in the Ade9 binding site to examine base specificity and we used our high resolution crystal structures as a guide for suggesting whether or not the binding pocket can accommodate bases other than Ade.

5.2 Substitution of Ade9 with Nebularine

Substitution of Ade9 with nebularine (Figure 5.1) eliminates the N6 amino group but retains the N7 and N1 hydrogen bond acceptors characteristic of Ade (Figure 5.2b). This substitution has only modest effects on structure and UP1 affinity. The loss of the N6 to Lys179 main chain carbonyl interaction reduced the K_d from 70 nM (wild-type) to 109 nM. Any interpretation of this change in affinity must also take into account the observation that loss of the N6 amino group allows the nebularine to make a closer contact with Arg178 (2.5 Å) and presumably helps compensate for loss of the main chain interaction. Additionally, the slight loss in binding affinity only reflects the loss of a hydrogen bond that is favorable in the UP1:hTR complex and does not address the loss in affinity that would occur if a carbonyl group, such as a Gua O6, were to occupy the same position.

5.3 Substitution of Ade9 with 7-deaza-adenine

Compared to the Neb substitution, significantly different results were obtained upon substitution of Ade9 with 7-deaza-adenine. 7-Deaza-adenine changes the N7 to a carbon and eliminates the ability of the base to serve as a hydrogen bond acceptor at this position (Figure 5.1). This substitution changes the binding affinity from 70 nM to 280 nM and moves the Arg178 side chain away from the base into an alternate conformation that allows the guanidinium group to interact with Glu93 more than 9.5 Å away (Figure 5.2c). The void left by the movement of Arg178 is filled, in part, by a water while the position of the 7-deaza-adenine is nearly identical to that of the Ade9 wild-type structure.

Affinity for Ade appears to be mostly associated with the Arg178 to adenine N7 bond as opposed to the N6 to Lys179 carbonyl hydrogen bond; however, it is clear that the combination of main chain to N6 hydrogen bond and stacking of Ade9 on Phe108 are sufficient to restrain the base into its wild-type position without the interaction of Arg178. These results point out the importance of the stacking and main chain interactions for ligand positioning.

The dramatic repositioning of Arg178 suggests that a steric clash between the 7-deaza carbon and Arg178 guanidinium group, together with the loss of the favorable interaction to the adenine N7, induce Arg178 to find a new and more favorable interaction with Glu93.

Figure 5.1. Guide to the bases used in the crystallography studies. Arrows indicate positions that are good hydrogen bond donors or acceptors.

Figure 5.2 Structures of UP1-oligonucleotide complexes for substitution of Adenine-9. All maps are 2Fo-Fc composite omit electron density maps contoured at 1.25 σ . Chevrons indicate the hydrogen bonding network. (a) Wild-type structure with Ade9 shown making hydrogen bonding contacts to the Arg-178 guanidinium group through its N7 (2.7 Å) and from N6 to the main chain carbonyl of Lys-179 (3.0 Å). Ade9 is stacked directly over the conserved Phe-108 of the RNP2 consensus sequence. (b) The UP1-A(9)Neb structure showed no substantive changes in either protein or DNA structures. The absence of the N6 amino group allowed the base to move slightly closer to Arg-178 (2.5 Å). (c) UP1-A(9)7deazaA structure shows a large conformational rearrangement of the Arg-178 side chain. The Arg-178 guanidinium group shifted 9.1 Å away from its position in the wild-type structure, where it makes contacts to the O2 of Thy8 and N7 of Ade9 to make a new set of contacts to Glu-93. All electron density figures were made using PYMOL (DeLano Scientific, CA).

5.4 Substitutions that Failed to Produce Crystals

5.4a Substitution of Ade9 with Inosine and 4-methylindole

Inosine and 4-methylindole (Figure 5.3) were inserted into the hTR sequence at position A9 to study the effects of placing a carbonyl and methyl group in the position occupied by the Ade N6 which donates a hydrogen bond to the main chain Lys179.

Inosine, like Gua, can donate a hydrogen bond from the N1 position which would disrupt the N1 to main chain hydrogen bond made by Ade. Ino was used to simulate the effects of Gua in the Ade binding pocket without the possible steric exclusion of the Gua N2. It was expected that the electrostatic repulsion of the Ino O6 and Lys179 main chain carbonyl group would decrease binding without inducing structural changes in the main chain or side chain groups that make up the binding surface. The close hydrogen bond donor-donor pairing between N1 and the Leu181 main chain (2.7 Å in the native UP1:TR2 complex) was expected cause a significant reduction in binding and possible exclusion of Ino altogether. 4-Methylindole was chosen to examine the steric effects of placing a methyl group in close proximity to the RRM2 main chain. It was expected that the main chain would retain rigidity and the modified base would be forced into a new position, possibly breaking the Arg178 hydrogen bond but perhaps maintaining the Phe108 stacking interaction given the hydrophobic nature of the purine with a methyl group.

Both modifications produced crystals that were morphologically similar to those produced by wild-type UP1:hTR complexes, but none were large enough for data collection. A more rigorous sampling of crystallization conditions will likely yield

positive results. Binding data were not collected on these oligonucleotides, and future work should include fluorescence studies in the absence of high resolution structures.

5.4b Substitution of Ade9 with Guanine and Thymine

Crystallization experiments were tried with UP1 bound to hTRs that contained Gua and Thy in the Ade9 position, but none of the conditions yielded crystals. Thy was also placed in the Ade3 position which contacts RRM1 in a manner similar to that of Ade9 to RRM2. Protein:nucleic acid ratios ranged from 0.5:1 to 1:2, and it is possible that binding is sufficiently weak to require higher concentrations of the oligonucleotides. It is questionable whether the N2 of Gua would be tolerated in the binding pocket since, when superimposed in place of Ade, the distance from N1 to the Leu181 main chain amide would place these two hydrogen bond donors only 2.80 Å apart and the Gua N2 would clash with the neighboring Gua10 N1, at only 2.21 Å apart (Figure 5.3b). The Gua N2 would also clash with the Leu181 main chain carbonyl at only 2.13 Å apart in the superposition of Gua in the binding site. It appears unlikely that the base or protein main chain would be able to move to accommodate the potential hydrogen between the two.

Thy may be more sterically tolerable in the Ade9 binding site; however, it is not easy to predict how a pyrimidine would be accommodated in a binding pocket that is specific for a purine. When Thy is superimposed in the Ade9 binding pocket in the *anti* conformation, the 5-methyl group is positioned only 1.13 Å from the Arg178 guanidinium group (Figure 5.5). Thus, either Thy would be forced to assume the *syn* conformation, or the Arg178 side chain would be forced to break the Thy8 O2 interaction and assume a different conformation. A hydrogen bond from Arg178 to the Thy O4 may

be quite favorable with Thy in the *syn* conformation (donor-acceptor distance is 2.8 Å when Thy is superimposed in the binding site) and main chain interactions would likely be too far away from the base to either contribute to or disrupt binding. An intrastrand hydrogen bond from the Thy O2 to the Gua10 N2 is seen in the hypothetical superposition of Thy in *anti* conformation. Ura lacks the Thy 5-methyl group, and it is likely that *anti* Ura could occupy this binding pocket. The Phe108 stacking interaction would probably not provide the same stability to a pyrimidine as it does for a purine, but there are numerous examples of stacking interactions from Phe to pyrimidines in the structural data base. Cyt would probably reduce binding, since the Cyt N4 group would clash with the Arg178 and possibly disrupt the Arg178 to Thy8 interaction in a manner similar to what was seen in the 7-deaza-A example.

5.5 Conclusions

Although binding data were not collected on complexes that did not crystallize, the crystal structures and binding data collected from the A(9)7DA and A(9)Neb hTRs are sufficient to draw some conclusions about the specificity for Ade in the Ade9 binding pocket. Arg178 is strongly associated with the purine N7 of Ade9 and pyrimidine O2 of Thy8 in the wild-type UP1:hTR complex. Gua might be expected to maintain these interactions in the Ade9 binding site, but it is less clear how a pyrimidine would behave in the same site. The Thy 5-methyl group would clash with the Arg178 guanidinium group but in the absence of this group, as in Ura, the O4 would be positioned to accept a hydrogen bond from Arg178 (Figure 5.5). An additional hydrogen bond from the pyrimidine O2 to the neighboring Gua10 N2 would also help stabilize Ura in the Ade9

binding pocket. Cyt would not be favored in the Ade9 binding site since the Cyt N4 would clash with the Arg178 guanidinium group. Fluorescence binding studies should be conducted in the future on those complexes that did not crystallize, and crystallization trials should be conducted with a A(9)U hTR oligonucleotide.

Phe108 is located in a highly conserved region of β -strand 1 in RRM2 called the RNP2 consensus region. A sequence alignment of RNP2s in the Conserved Domain Database shows that this position is occupied by Phe in 42% of the domains, and Tyr which is found 30% of the time [Ding, 1999; Jacobs, 2002]. Aromatic residues in RNP2 may play a role in stabilizing bases in most RRM:nucleic acid interactions, as seen in the crystal structures of several other RRM:nucleic acid complexes in which base stacking interactions are made to purines and pyrimidines. Although base stacking is more common among purines in the available structures, there is no reason to assume that a pyrimidine would not be stabilized at this position in UP1. Thus, Phe108 may provide stability and perhaps some selectivity for purines at this position, but it probably does not discriminate against pyrimidines.

The Lys179 main chain carbonyl contributes little to binding affinity, as indicated by a small reduction in K_d and small changes in the structure when Ade9 was replaced with Neb. The main chain does, however, form a tight barrier around the Watson-Crick edge of the base, and it is likely that the main chain creates a steric hindrance that excludes Gua on the basis of the N1 and N2 positions (Figure 5.4b). These positions are hydrogen bond donors in Gua. Examples of this type of exclusion with tight binding pockets are seen in the crystal structure of poly-(A) binding protein (PDB 1CVJ) (Deo *et al.*, 1999). Gua might avoid steric clashes by assuming the *syn* conformation, and Gua4

and Gua10 bind UP1 in this manner. If a glycosidic flip is favorable, it would likely be seen in the Ino substitution. Hopefully, we will be able to grow diffraction quality crystals of the UP1:A(9)Ino complex in the near future. Based on the data collected to date, it appears that the highly conserved Phe108 provides base stability, Arg178 provides purine specificity and possibly Ura O4 specificity, and the main chain residues provide Ade specificity through steric exclusion of Gua N1 and N2 donor groups.

Figure 5.3. Examples of bases used to replace Ade9 in crystallization trials that failed to produce crystals. Arrows represent good hydrogen bond donors and acceptors.

Figure 5.4. A, Ade in the Ade9 binding pocket of UP1. Ade makes hydrogen bonds from N7 to the Arg178 side chain, N6 to the Lys179 main chain, and N1 to the Leu181 main chain. B, Gua superimposed in the Ade9 binding pocket of UP1. Electrostatic clashes would exist between O6 and the Lys179 main chain carbonyl and between N1 and the Leu181 main chain amide. A steric clash would exist between N2 and the Leu181 main chain carbonyl. An additional clash would occur between the N2 position and the neighboring Gua10 N1 (not shown) where.

Figure 5.5. Superposition of Thy in the Ade9 binding pocket. A favorable hydrogen bond might be made from Arg178 to the Thy O4, however, a steric clash with the Thy methyl group is apparent. With the absence of the methyl group, Ura might be able to occupy this binding site. The distance between the O4 and the main chain carbonyl would be too great to affect binding and a hydrogen bond between the O2 and the neighboring Gua10 N2 would also help stabilize a pyrimidine in this position.

CHAPTER 6. ANALYSIS OF Gua10 BINDING SURFACE USING 7-DEAZA-GUANINE, INOSINE, 2-AMINOPURINE, AND NEBULARINE

6.1 Gua10 Binding Pocket

Gua10 is the only base that occupies its binding site in the *syn* conformation in RRM2. Gua4 also assumes the *syn* conformation in RRM1. These bases and binding sites are equivalent in the sense that they represent the first Gua in the hTR, and their binding pockets are made from the same RRM positions. The nucleotide recognition surface of UP1 for Gua10 is comprised of a stacking interaction from Phe150 and several base specific hydrogen bond contacts from the Lys106 side chain and Leu181 main chain. The ϵ -amino group of Lys106 makes bidentate hydrogen bonds to N7 (3.0 Å) and O6 (3.5 Å) of Gua10. Gua10 also makes a hydrogen bond from N1 to the main chain carbonyl of Leu181 (2.6 Å). The carbonyl of Leu181 might also make a more distant (3.3 Å) interaction to the N2 position, but the geometry is not optimal. Although Lys106 is proximal to both the O6 and N7 of Gua10, a lower map contour shows that stronger electron density exists for the O6 to Lys106 hydrogen bond (not shown). Gua10 was replaced with four different modified bases in the UP1:hTR complex in order to better understand the elements of base specificity (Figure 6.1). We expected that the Lys106 bidentate hydrogen bonds would be important in excluding Ade from the binding site, since a clash between the Ade N6 and the Lys amino group might prevent the Lys106 to N7 interaction. We therefore used 7-deaza-guanine and 2-aminopurine substitutions to better understand these interactions. 2-Aminopurine was also used, in addition to a nebularine substitution, to examine the effects of placing a hydrogen bond acceptor in

close proximity to the Leu181 main chain carbonyl, which hydrogen bonds with the Gua N1 in the UP1:hTR complex.

Figure 6.1. Examples of modified bases used to replace Gua10 in crystal structures of the UP1:hTR complex. Arrows indicate good hydrogen bond donor and acceptor groups.

6.2 Substitution of Gua10 with 7-deaza-guanine

The structure and binding studies of UP1 bound to G(10)7deazaG showed a substantial decrease in affinity from 70 nM in the native UP1:TR2 complex to 250 nM, with no significant change in position of the bases or UP1 (Figures 6.1 and 6.2b, Table 6.1). The ϵ -amino group of Lys106 is better poised to make contact with the O6 (2.7 Å from 3.5 Å in the wild-type structure), but the loss of the bidentate interaction to the

purine N7 led to a measurable decrease in affinity. Despite the lowered affinity, the combination of the single Lys106 interaction to O6, Leu181 main chain to N1, and stacking interaction with Phe150 are able to maintain the position of 7-deaza-guanine.

Table 6.1. Dissociation constants for UP1 binding to oligonucleotides from competition assays.

Oligonucleotide	K _d (nM)
wild-type	70
G(10)7deazaG	250
G(10)Ino	80
G(10)2AP	160
G(10)Neb	270
G(10)Ade	~500-1000 ^a

^a Weak binding of this oligonucleotide precluded an accurate estimate of affinity at 0.3 M NaCl.

6.3 Substitution of Gua10 with 2-aminopurine and Nebularine

The structures of UP1 complexed with G(10)2AP and G(10)Neb both exhibit dramatic changes in base position (Figure 6.1 and 6.2c,d). In the case of G(10)2AP, loss of the Gua O6 and change of N1 to a hydrogen bond acceptor causes 2AP to flip into the *anti* conformation relative to the glycosidic bond. Despite rotating nearly 180° about the glycosidic linkage, binding is reduced only 2.4-fold (Table 6.1), largely because the ε-amino group of Lys106 can now make a good hydrogen bond to the N1 of 2-aminopurine

(2.7 Å). All potential contacts from N1 and N2 to the carbonyl of Leu181 are lost as a new set from the N2 to Glu135 are made (2.7 Å). A water-mediated hydrogen bond from N7 to the carbonyl of Leu181 is also formed.

Substitution of Gua10 with nebularine addresses the potential contribution of the N1, N2, and O6 for positioning Gua10. The UP1-G(10)Neb structure suggests that the nebularine base becomes substantially more mobile than G(10)2AP and may flip between *syn* and *anti* conformers. Figure 6.2d shows that the electron density for nebularine is substantially disordered and indicates increased base mobility. The G(10)Neb has a 3.6-fold decrease in binding, presumably because there is no O6 for Lys106, or N1 for Leu181, to contact as previously shown in the *syn* conformation. Like G(10)2AP, the nebularine base does have a viable N1 hydrogen bond acceptor and shows a similar positioning of the ϵ -amino group of Lys106; however, without the N2 group, Neb cannot hydrogen bond with Glu135. Steric clashes make it unlikely that Lys106 could contact an adenine N1 *in vivo*, since atomic clashes between the Ade N6 and N1 position in the *anti*, or the N7 position in the *syn* conformation, would seriously limit adenine binding to the RRM at this position. Although we did not produce crystals of the UP1-G(10)A complex, we were able to show that binding is reduced five to ten-fold, the largest of any of the modified bases. Therefore Lys106 is an excellent candidate to discriminate against Ade in either the *syn* or *anti* conformer.

6.4 Substitution of Gua10 with Inosine

Substitution of Gua10 with inosine removes the potential Leu181 carbonyl to N2 interaction. Our studies showed a modest increase in affinity over the native UP1:hTR complex, 60 nM for G(10)Ino compared to 70 nM for TR2, but no effect on overall base position. These data suggest that loss of the putative N2 contact (3.3 Å) has a minimal affect (Figure 6.2e and Table 6.1). It is likely that the N1 position, which is closer (2.6 Å) and in a better geometric position to contact the carbonyl of Leu181, is more important to recognition than N2. The increase in affinity is likely the result of the inability of G(10)Ino to form stable G-quadruplexes. A shift in the G(10)Ino to a more single-stranded population would undoubtedly facilitate UP1 binding to the ligand and account for the higher affinity.

Figure 6.2 Structures of UP1-oligonucleotide complexes for substitution of Guanine-10. All maps are 2Fo-Fc composite omit electron density maps contoured at 1.25 σ . (a) Wild-type structure with UP1 shown making a bidentate hydrogen bond from the amino group of Lys-106 to N7 (3.0 Å) and O6 (3.5 Å) and from the carbonyl group of Leu-181 to N1 (2.6 Å) of Gua-10. (b) UP1-G(10)7deazaG shows no overall change in structure with Lys-106 in good position to contact O6 (2.7 Å). (c) The UP1-G(10)2AP structure shows that the base rotates nearly 180° from the *syn* to *anti* conformation. The amino group of Lys-106 makes a new contact to the N1 position and all contacts to the main chain carbonyl of Leu-181 are lost. (d) UP1-G(10)Neb shows greatly weakened electron density for nebularine and indicates increased base mobility consistent with the *syn* and *anti* conformers. (e) The UP1-G(10)Ino structure is essentially unchanged compared to wild-type complex.

6.5 Gua10 Binding Pocket Conclusions

The Gua10 binding pocket of UP1 is selective for Gua in the *syn* conformation through one stacking interaction and hydrogen bonds from two amino acids, one of which is a bidentate interaction. Although binding sites specific for Ade have the potential to exclude Gua sterically at the N2 position, Gua binding pockets must utilize well-positioned hydrogen bonds to exclude Ade. Hydrogen bond donor-donor pairs may be used to exclude Ade in the Gua position, where a rigid donor in close proximity to the Gua O6 would prevent an Ade N6 from occupying the same position. In the case of Gua10, a Lys side chain donates a bidentate hydrogen bond to the O6 and N7 positions of Gua. A main chain interaction might be expected to provide greater rigidity and more specificity than that of a long flexible side chain like Lys; however, a second hydrogen bond to the N1 of Gua from a main chain carbonyl may act to provide the necessary stability. Our crystal structures show that removal of the Lys106 to N7 hydrogen bond had little effect on the structure of the binding pocket and the positioning of the base, although a greater than three-fold reduction in affinity was observed. Removal of the O6 carbonyl group, however, induced a base flip from *syn* to *anti* and binding was reduced more than two-fold compared to the wild-type UP1:TR2 structure. The base flip was likely the result of the difference in electronegativity at the N1 position since Gua can donate a hydrogen bond from this position and the 2-aminopurine substitution cannot. Although a greater reduction in affinity might be expected, the base flip allows new hydrogen bonds to be made: Lys106 side chain to N1, Leu181 main chain to N7 (water mediated), and Glu135 side chain to N2.

The glycosidic flip from *syn* to *anti* raises questions as to whether an Ade could occupy the Gua10 binding site given the flexibility of the Lys106 side chain and the water mediated hydrogen bond to N7 of 2-aminopurine in the *anti* conformation. Unlike 2-aminopurine, Ade would not be able to hydrogen bond to Glu135, since it lacks an N2. The result of this loss was seen in the nebularine substitution. The electron density around nebularine indicated high mobility, although it did suggest that the base occupied the *anti* conformation a great deal of the time. The position and electron density surrounding Lys106 supports such a claim. There was, however, no clear evidence of a water-mediated hydrogen bond to the N7 position which is not surprising considering the base is mobile. The nebularine substitution resulted in the weakest binding of all the modified bases used in the Gua10 position. Although the Lys106 side chain does have some flexibility, and it was seen to donate hydrogen bonds to either the N7 of Gua or the N1 of 2AP, we hypothesized that the Ade N6 would clash with Lys106 and prevent a hydrogen bond to N1. As expected, binding was reduced five to ten-fold with Ade in the Gua10 binding pocket. Crystallization trials have not yet been attempted on this complex and it will be interesting to see if crystals will form, given the comparatively weak binding as compared to the UP1:TR2 complex. To date, all attempts to crystallize UP1 bound to oligonucleotides that contain modified bases in the Gua10 position have produced diffraction quality crystals.

For the purpose of identifying residues that could be mutated in order to engineer a binding site with high affinity for bases other than Gua, it is important to consider both the *syn* and *anti* conformations of the base. The close contact made to Gua from the Leu181 main chain is not detrimental for Ade if the base behaves like 2AP and assumes

the *anti* conformation. If the Ade N6 position is stabilized, a water mediated hydrogen bond from Leu181 to N7 would contribute to overall binding. N6 might be stabilized by replacing Lys106 with Gln (Figure 6.3). It is interesting to note that in poly(A)-binding protein, the equivalent RNP2 positions are Ser12 and Asn100 in RRM1 and RRM2, respectively, and both accept hydrogen bonds from the N6 positions of Ade.

Alternatively, a hydrogen bond might be gained by replacing Glu135 with Lys (Figure 6.4). In the UP1:G(10)2AP crystal structure, where 2-aminopurine assumes the *anti* conformation, Glu135 accepts a hydrogen bond from N2. In the absence of N2 in Ade, the N1 position would be accessible to a Lys side chain. A double mutation of Lys106 and Glu135 to Gln and Lys respectively would likely cause steric and electrostatic clashes, and it may be more reasonable to mutate Lys106 to Ala in a Glu135 to Lys mutant. The equivalent residue to Glu135 in PABP is Lys129, which donates a hydrogen bond to the Ade O2' position as opposed to the N1; however, a hydrogen bond is made to the N1 of the same Ade from a Ser two residues N-terminal to this position, Ser127.

We superimposed Thy in the Gua10 binding pocket to estimate how a pyrimidine would be tolerated (Figure 6.5). Lys106 appeared to clash sterically with the Thy 5-methyl group when the base was placed in the *syn* conformation, and an electrostatic clash between the Lys106 amino group and Thy N3 hydrogen bond donor would occur with Thy in the *anti* conformation. A Lys rotamer search in PYMOL did not indicate that the side chain would be able to interact with the Thy O4 position. However, Ura might occupy the pocket in the *syn* conformation, since a steric exclusion by the Thy 5-methyl group would not occur. Cyt would probably not be able to bind in the Gua10 binding site since the Cyt N4 group would clash with Lys106. However, a Lys106 to Gln

mutation, as suggested for increasing affinity for Ade, might allow Cyt to occupy the pocket.

6.6 Comparison of the Gua10 and Gua4 Binding Pockets

Although we were unable to grow crystals that contained modified bases bound to RRM1, the crystallographic and binding data gathered for RRM2 provide clues to RRM1 specificity. The binding pocket of Gua4 in RRM1 is similar to that of Gua10 in RRM2 in utilization of “equivalent” residues to accomplish binding. Not only are these residues found in the same folds and loops within their respective RRMs, they are separated by exactly 91 amino acids and they interact with Gua in the same manner: Lys15 and Lys106 donate hydrogen bonds from their side chains to N7, Phe59 and Phe150 are involved in base stacking, and Val90 and Leu181 accept hydrogen bonds from Gua N1 to their main chain carbonyl groups (Figure 6.6). Two additional hydrogen bonds are made in the Gua4 binding pocket that are not found in the Gua10 pocket; Gln12 side chain to O6 and Ser95 side chain to N2. The Gln12 interaction is interesting in that separate bonds are made to the O6 and N7 positions of Gua4 in RRM1, whereas a bidentate interaction from Lys106 to both O6 and N7 of Gua10 is sufficient for strong binding.

We had thought that Ade could potentially bind to RRM1 in the Gua4 position based on the following premises: (a) the electrostatic nature of the N7 position would not change, (b) there would be no steric hindrance from the lack of an N2 hydrogen bond donor to side chain interaction or from the lack of an N1 to main chain carbonyl hydrogen bond, and, (c) most importantly, Gln12 could potentially rotate its side chain 180 to form a stable interaction with N6. In order to test whether the Gln12 side chain

could flip, we placed 2,6-diaminopurine in the Gua4 binding pocket, since the only interaction we thought to be affected was the N1 to Val90 side chain hydrogen bond. Like Ade, the N1 of 2,6-diaminopurine has the potential to act as a hydrogen bond acceptor, and we anticipated that the electrostatic repulsion at this position would not be sufficient to abrogate binding. There was also the potential for a steric clash between the electropositive N6 and Lys15 side chain; however, hydrogen bonding to the Ade N7 from Lys is not uncommon in the structural protein database and we did not expect a major clash. The complex failed to crystallize and the inability to form crystals may provide clues as to the disruptive nature of inserting an Ade or 2,6-diaminopurine in the Gua4 position. Fluorescence binding studies will provide a better understanding of the effects of placing a base other than Gua in the Gua4 binding pocket.

The structural data gleaned from modified bases in the Gua 10 binding pocket prompted us to superimpose Ade in the Gua4 position in the *anti* conformation (not shown). Surprisingly, while 2AP occupies the Gua10 binding site in the *anti* conformation without intrastrand clashes, the C2 of *anti* Ade would clash with the N6 of the neighboring Gua5 at a distance of only 1.6 Å. Thus, although the two RRM of UP1 have similar binding surfaces and bind specifically to identical sequences, the geometry between the bases is not identical in the local stereochemistry of the Gua4 and Gua10 pockets.

Thy was also superimposed in the Gua4 binding pocket, and, like the Thy superposition in the Gua10 site, it appears that steric and electrostatic clashes would exclude Thy (Figure 6.7). A rotamer search indicates that the hydrogen bond donor-acceptor distance between Lys15 and the Thy O4 is too short for a proper hydrogen bond,

and conformations with more acceptable distances tend to clash with the 5-methyl group when Thy is shown in the *syn* conformation. With Thy in the *anti* conformation, an electrostatic clash would potentially disrupt the Lys15 to O4 interaction as well. Ura may be able to bind in the *syn* conformation, since the absence of a 5-methyl group would allow Lys15 to assume a favorable orientation for hydrogen bonding with the Ura O4. Cyt would probably not be accepted in the binding site without a Lys15 mutation to Gln, and it appears that the only potential interaction that could be made from this binding surface to a pyrimidine would have to be made to either the O4 of Ura or the N4 of Cyt. Molecular dynamics would be a valuable tool in predicting the behavior of alternative bases in Gua binding sites.

Figure 6.3. A, Gua in the UP1 Gua10 binding pocket. Gua occupies the binding pocket in the *syn* conformation and makes two hydrogen bonds with UP1; one from Gua N1 to the Leu181 main chain carbonyl and a second from Gua N7 to the Lys106 side chain. B, Ade in the UP1 Gua10 binding pocket in the *anti* conformation with a proposed mutation of Lys106 to Gln. It is possible that Gln would make a bidentate hydrogen bond with N1 and N6 of Ade. An additional water mediated hydrogen bond might be made from the N7 position to the Lue181 carbonyl as was seen in the crystal structure with 2-aminopurine in the Gua10 binding pocket (Chapter 7.3).

Figure 6.4. Ade superimposed in the Gua10 binding pocket in the *anti* conformation with a proposed Glu135 to Lys mutation in addition to the proposed Lys106 to Gln mutation. Lys135 would have the potential to hydrogen bond with the Ade N1 although the distance might be too great considering the geometry of the interaction. A potential hydrogen bond donor-donor clash between Lys135 and Gln106 (shown in red) may render the double mutation untenable.

Figure 6.5. Superposition of Thy in the Gua10 binding pocket. A, Thy in the *syn* conformation. The Lys106 and Thy O4 hydrogen bond donor-acceptor pairing would be too close to form a stable interaction, and a rotamer search in PYMOL indicated that the Lys would not be able to position its side chain to make a favorable interaction. B, Thy in the *anti* conformation. The Lys amino group and Thy O4 distance is somewhat greater, and a rotamer search indicated that Lys would be more likely to interact favorably with the base; however, an electrostatic clash with the Thy N3 would likely occur in all cases.

Figure 6.6. A, Gua in the Gua4 binding pocket of UP1. Gua occupies the binding site in the *syn* conformation and makes hydrogen bonds from N7 to the Lys15 side chain, N6 to the Gln12 side chain, N1 to the Val90 main chain carbonyl, and from N2 to the Ser95 side chain. B, Ade superimposed in the Gua4 binding pocket of UP1 in the *syn* conformation. An electrostatic clash would exist between N1 and the Val90 main chain carbonyl, but the N7 to Lys15 hydrogen bond may not be affected. The Gln12 side chain may flip 180° to accept a hydrogen bond from N6. The distance between Ser95 and N1, 4.12 Å, is probably too great to affect binding.

Figure 6.7. A, Thy superimposed in the Gua4 binding pocket in the *anti* conformation. Lys15 has the potential to act as a hydrogen bond donor to the Thy O4 position; however, a simple superposition shows that the distance would be too short (2.29 Å) to provide a good interaction, and a potential clash would exist between Lys and the Thy N3. B, Thy superimposed in the Gua4 binding pocket in the *syn* conformation. A steric clash would exist between the Lys15 side chain and the Thy methyl group, and, as in the proposed structure above, the distance between Lys15 and O4 would be too short to make a hydrogen bond. These superpositions assume no movement of the Lys15 side chain, and a rotamer search in PYMOL found no positions that would be more favorable.

CHAPTER 7. ANALYSIS OF Gua11 BINDING SURFACE USING INOSINE AND 2-AMINOPURINE

7.1 Gua11 Binding Pocket

Unlike the Ade9 and Gua10 binding pockets, which are defined by numerous main chain and side chain hydrogen bonds and stacking interactions, the Gua11 binding pocket is comparatively simple. The only direct contact between Gua11 and UP1 is from the main chain amide of Lys183 to the O6 atom of the base (2.7 Å) (Figure 7.2a). Additionally, rather than stacking upon the conserved aromatic residues of the RNP motif, Gua11 makes intrastrand stacking arrangements by stacking between Gua10 and Gua12 (Figure 7.3a). The disruption of the Lys183 hydrogen bond was investigated by substituting 2AP in the Gua11 binding site, and Ino was substituted in the binding pocket in an effort to disrupt G-tetrad formation without affecting binding (Chapter 4 and Figure 7.1). This site was also used to test the ability of 6MI to occupy a Gua binding pocket and to study the effects of intrastrand stacking on 6MI fluorescence.

7.2 Substitution of Gua11 with 2-aminopurine

The UP1-G(11)2AP structure shows a clear void where the O6 was present in the wild-type structure, and electron density suggests that the base is mobile about the glycosidic bond (Figure 7.2c). The stacking of Gua11 between Gua10 and Gua12 partially stabilizes its position in the absence of any direct contacts from UP1; however, the electron density for 2-amino-purine is much weaker, and loss of this single main chain hydrogen bond allows significant base mobility. This enhanced flexibility is readily

seen in the average temperature factors for the 2AP ($B_{\text{aver}} = 61.3 \text{ \AA}^2$) compared to the wild-type Gua ($B_{\text{aver}} = 28.4 \text{ \AA}^2$). Electron density for Gua12 in the G(11)2AP structure is also significantly weaker and suggests that Gua11 mobility partially disrupts stacking with the neighboring base. The binding of G(11)2AP was reduced from 70 nM in the native UP1:hTR complex to 180 nM, presumably from the loss of the O6 to main chain amide interaction. This reduction in affinity makes clear that nonionized main chain contacts can be important determinants in binding.

Figure 7.1. Guide to modified bases used to replace Gua11 in the UP1:hTR crystal studies. 6MI is a fluorescent pteridine compound that has hydrogen bonding properties similar to that of Gua and was used in position Gua6 in the fluorescent binding studies. The Gua11 binding pocket is characterized by a single main chain hydrogen bond from the Lys183 main chain amide to the Gua O6, and 2AP was used to eliminate this interaction. Ino was used to study the effects of disrupting G-tetrad prior to UP1 binding without affecting the main chain hydrogen bond.

7.3 Substitution of Gua11 with Inosine and 6MI

Ino was substituted in the Gua11 position to abrogate the formation of G-tetrads without affecting base-specific interactions at the protein:nucleic acid interface. This substitution was used to compare the difference in binding between hTRs that are essentially identical with the exception that the wild-type hTR forms tertiary structures, and CD data show that the G(11)Ino oligonucleotide does not form G-tetrads under the same conditions as TR2 (Chapter 4). As expected, no discernable changes in the crystal structure were observed compared to wild-type (Figure 7.2), and UP1 had slightly higher affinity for G(11)Ino over wild-type; $K_d = 60$ nM for G(11)Ino and 70 nM for wild-type hTR.

6MI was substituted for Gua11 in the TR2-11F oligonucleotide to study the behavior of a fluorescent pteridine analog of Gua in a protein:nucleic acid interaction. 6MI has been shown to be effective in replacing Gua in duplex DNA, where it forms Watson-Crick basepairs to Cyt (Hawkins, 2001). This base was expected to behave similarly to Gua in the Gua11 binding pocket. The analogous O6 position of 6MI hydrogen bonds with the Lys183 main chain, and the overall r.m.s.d. from the superposition of the wild-type and TR2-11F C α s was 0.29 Å. This observation suggests that there were no global changes to the complex structure upon substitution of 6MI at the Gua11 position. The structure of TR2-11F bound to UP1 shows that Gua12 is disordered, whereas in the native structure, Gua12 stacks with Gua11 with no direct contacts between Gua12 and UP1. Thus, 6MI behaves similarly to Gua11 in terms of providing Watson-Crick hydrogen bonds, but it does not provide the same level of stability in stacking arrangements with Gua12.

Figure 7.2. Structures of UP1-oligonucleotide complexes for substitution of Guanine-11. Maps are 2Fo-Fc composite omit electron density maps contoured at 1.25σ (a-b) and 1.0 (c). (a) Wild-type structure shows Gua11 stacks between Gua10 and -12 and has only one direct contact to UP1 through the Lys-183 main chain amide to O6 ($2.7 \text{ \AA}$). (b) UP1-G(11):Ino remains stacked and maintains high affinity. (c) UP1-G(11)2AP also remains stacked between adjacent guanines, despite loss of the one contact from O6 to the main chain amide; however, the mobility of G(11)2AP and Gua12 have increased as demonstrated by lower electron density and commensurately increased temperature factors. Gua11 is positioned by the favorable stacking interactions with Gua10 and Gua12, but loss of the main chain interaction with G(11)2AP results in significantly increased base mobility and an overall decrease in affinity.

7.4 Conclusions

Given that only one hydrogen bond is made to Gua11, it was surprising to find that eliminating this interaction resulted in a 2.6-fold reduction in binding, especially since the 2-amino-purine substitution did not induce steric clashes. Much of the reduction in binding affinity is likely to be caused by destabilizing the Gua12 stacking interaction. Molecular dynamics simulations may provide quantitative estimates of this destabilization. In any case, it is likely that Ade would not be tolerated in the Gua11 binding pocket in terms of both steric and electrostatic clashes of the N6 and Leu181 hydrogen bond donor-donor pairing (Figure 7.3). The fact that it is a main chain hydrogen bond that dictates specificity for Gua presents difficulties for engineering affinity for a different base in this position.

Pyrimidines may also be sterically excluded from the Gua11 binding site, and Figure 7.4 shows the superposition of Thy in the binding pocket. Although the Thy O4 and the Lys183 main chain amide make good hydrogen bond acceptor and donor pairs, the distance between these groups with Thy in either the *anti* or *syn* conformation (2.54 Å) may be too close to allow a hydrogen bond. Main chain to base hydrogen bonds with distances as short as 2.62 Å do exist in the UP1:hTR crystal structure, and it would be interesting to see if the base and main chain can move to accommodate Thy in the Gua11 pocket. Molecular dynamics simulations may provide a better understanding of how Thy might interact with UP1 in this position, and diffraction quality crystals of a UP1:G(11)T complex would prove invaluable in understanding the flexibility of this site. It is unlikely that Cyt would be able to occupy the Gua11 pocket since the Cyt N4 would clash

sterically and electrostatically with the Lys183 amide in either the *anti* or *syn* conformation.

The Gua11 binding pocket is substantially different from its homolog in RRM1 which binds Gua5. Gua11 makes only one hydrogen bond to UP1 through an O6 to Lys183 main chain and the base presumably achieves some stability by stacking with Gua10 and Gua12. Gua5 on the other hand hydrogen bonds with the “equivalent” Lys183 residue in RRM2, which is Arg91 in RRM1, but the bond is made from N7 to the Arg91 side chain. An additional set of hydrogen bonds is made from the Asp42 to N1 and N2, and a stacking interaction is also seen to Met46. The crystal structure may provide clues as to why Gua is bound in such a different manner in binding pockets that are quite similar.

The bases that follow 3' to Gua5 and Gua11 are both Gua, and neither make contacts to the RRM binding surfaces. As mentioned previously, Gua12 stacks with Gua11 and strong density is usually seen around the base in UP1:hTR complexes. Gua6, however, does not stack with neighboring bases but is instead solvent exposed, which is why this position was chosen for insertion of a fluorescent base for binding studies. Electron density is rarely seen around Gua6 in UP1:hTR complexes and, when it is present, it is only visible at low contours and is highly disordered. Gua6 is pushed out into solvent because the hTR oligonucleotide must twist in order for bases Thy8 through Gua11 to contact RRM2. Thus, Gua5 may require additional hydrogen bonds and a stacking interaction to achieve stability similar to Gua11.

Figure 7.3. A, Gua in the Gua11 binding site of UP1. Gua interacts with UP1 in the *anti* conformation and makes a single hydrogen bond from O6 to the Lys183 main chain amide. B, an electrostatic clash between N6 and the Lys183 main chain amide is shown when Ade is superimposed in the Gua11 binding pocket of UP1 in the *anti* conformation. C, an electrostatic clash would exist with Ade in the *syn* conformation between the same groups (N6 and the Lys183 main chain amide) that would induce a clash in the more common *anti* conformation. Although Gua11 is held in this binding site by only one main chain hydrogen bond, this interaction appears to be sufficient for exclusion of Ade.

Figure 7.4. Superposition of Thy in the Gua11 binding pocket shown here in the *anti* conformation. The O4 to Lys183 main chain amide distance is roughly the same when Thy is superimposed in the *syn* conformation, and the distance may be too close to form an optimal hydrogen bond. An electrostatic and steric clash between the amide, and Cyt N4 would exclude Cyt from the Gua11 binding site.

CHAPTER 8. SUMMARY AND CONCLUSIONS

8.1 Comparison of UP1:TR2 Complex with other RRM:nucleic acid Complexes

In addition to the UP1:TR2 structure (2UP1), five crystal structures of RRM containing proteins complexed with nucleic acid have been solved and their coordinates deposited in the protein data bank (Figure 8.1). The *Drosophila* protein sex-lethal (Sxl) contains two RRMs, and the structure has been solved to 2.6 Å bound to the U-rich sequence r(GUUGUUUUUUUU) (PDB 1B7F) (Handa *et al.*, 1999). Poly-A binding protein (PABP) contains two RRMs, and the structure of this protein bound to a poly-A oligonucleotide, r(A)₁₁, was solved to 2.6 Å resolution (PDB 1CVJ) (Deo *et al.*, 1999). The structure of the two RRM nucleolin protein Rbd12 (PDB 1FJE) was solved by NMR bound to an Snre RNA loop, r(GGCCGAAAUCCCGAAGUAGGCC) (Allain *et al.*, 1996). Hud is also a two RRM protein, and the structure bound to two different AU-rich elements has been solved: Hud bound to tumor necrosis factor αRNA r(UAUUUUAUUUA) to 2.3 Å (PDB 1G2E) and Hud bound to C-Fos RNA r(UUUUAUUUU) to 1.8 Å (PDB 1FXL) (Zhang *et al.*, 2001). U1A is from U1 small ribonucleoprotein that, unlike UP1, Sxl, PABP, Rbd12, and Hud, contains only one RRM, and the structure of U1A bound to an RNA hairpin loop, r(AAUCCAUUGCACUCCGGAUUU), was solved to 1.92 Å resolution (PDB 1URN) (Oubridge *et al.*, 1994). Each of these proteins utilize highly conserved RRMs to bind nucleic acid sequences that are quite different in terms of purine and pyrimidine content and, in the case of Rbd12 and U1A, secondary structure. Additionally, while UP1 binds two strands of DNA in an anti-parallel fashion where each oligonucleotide is bound by

two RRMs of two copies of UP1. The target sequences of Sxl, PABP, Hud, and Rbd12 are bound by the RRMs of a single copy of protein. ENTANGLE was used to compare and contrast the hydrogen bond and stacking interactions of these RRM-containing proteins to gain a better understanding of the elements that give rise to sequence specificity in these proteins.

A sequence alignment shows that nucleic acid binding is accomplished through roughly equivalent RRM residues (Figure 8.2 and Table 8.1). No hydrogen bonds or stacking interactions are made from the $\alpha 1$ or $\alpha 2$ helices. The regions N-terminal and C-terminal of $\alpha 2$ are also absent of hydrogen bonding interactions in all five structures, and the majority of interactions are made from the $\beta 4$ sheet and the residues C-terminal of $\beta 4$. A total of nine residues are involved in hydrogen bonding interactions from $\beta 4$, and 29 residues are involved in nucleic acid recognition from the non-structured region C-terminal of $\beta 4$. Sixteen residues in $\beta 1$ are involved in nucleic acid binding, and 8 residues between $\beta 1$ and $\alpha 1$ are used for binding. UP1, PABP, and Sxl utilize a single non-conserved residue N-terminal to $\beta 2$ for base recognition, and only PABP, Sxl, and U1A use residues in $\beta 2$ for binding with a total of 11 amino acids. Out of the six structures, all but Hud make interactions through the region between $\beta 2$ and $\beta 3$ with a total of 11 residues. All six complexes make interactions from $\beta 3$ with a total of 15 residues.

Of the 102 amino acids involved in nucleic acid binding in the six structures, 56 are conserved and 46 are not conserved. It is not surprising that slightly more than half of the interactions are made through conserved residues given the high degree of similarity between the structures, but is interesting to note that in the absence of stacking

Figure 8.1. Comparison of available RRM:nucleic acid complexes. A, Sex lethal is a two RRM protein that binds a U-rich sequence (PDB 1B7F). B, Poly-(A) binding protein contains two RRMs that bind A-rich RNAs (PDB 1CVJ). C, Rbd12 is a two RRM protein that binds an RNA loop (PDB 1FJE). D, Hud binds AU-rich elements through two RRMs, and the structures of Hud bound to two different AU-rich polynucleotides have been solved (PDB 1G2E and 1FXL). E, U1A is a single RRM protein that has been shown to bind an RNA hairpin loop (PDB 1URN). F, UP1 is made from two RRMs and is a crystallographic dimer when complexed with two strands of human telomeric repeat ss(DNA) (PDB 2UP1). These structures show that RRM-containing proteins interact with nucleic acid in a variety of ways, although a sequence alignment shows that sequence specific binding is accomplished through roughly the same regions of the RRM motif.

Table 8.1. Summary of the number of base-stacking and hydrogen bonding interactions identified by ENTANGLE from the available RRM:nucleic acid complexes. The majority of interactions are made from the non-structured regions C-terminal to $\beta 4$. Interactions made through conserved amino acids outnumber those from non-conserved residues, and 18 conserved interactions are from the base-stacking elements in the $\beta 1$ and $\beta 3$ regions.

	$\beta 1$		$\alpha 1$		$\beta 2$		$\beta 3$		$\alpha 2$		$\beta 4$		Totals
UP1	4	1	0	1	0	3	2	0	0	0	5	6	22
PABP	3	1	0	1	5	1	5	0	0	0	0	7	23
Sxl	2	2	0	1	3	3	2	0	0	0	1	4	18
Hud	2	2	0	0	0	0	3	0	0	0	2	3	12
Rbd12	1	0	0	0	0	2	2	0	0	0	0	3	8
U1A	4	2	0	0	3	2	1	0	0	0	1	6	19
Totals	16	8	0	3	11	11	15	0	0	0	9	29	102
Conserved	9	2	0	3	5	5	14	0	0	0	2	16	56
Non-conserved	7	6	0	0	6	6	1	0	0	0	7	13	46

8.2 RRM Hydrophobic Core

Of the 11 RRMs compared with ENTANGLE (two RRMs in UP1, Sxl, PABP, Hud, and Rbd12 and one in U1A), all have a hydrophobic cleft made from conserved $\beta 1$ and $\beta 3$ residues and base stacking interactions are made from the majority of these positions (Figure 8.2). The conserved residues in $\beta 3$ are either Phe or Tyr, and all are involved in base stacking in the RRMs compared. Seven of the 11 conserved $\beta 1$ amino acids stack with nucleic acid bases and all residues involved in stacking are either Phe or Tyr. With the exception of RRM1 in Rbd12, stacking interactions are seen in all but one case where Phe or Tyr occupies the hydrophobic cleft. In three cases, Ile occupies the conserved $\beta 1$ position and Leu is present in one example. Out of the 18 conserved stacking residues, 11 are made with Phe and seven are from Tyr. Interestingly, the

hydrophobic cleft in the first RRM of the complexes that have two RRMs (UP1, Sxl, PABP, Hud, and Rbd12) have a conserved hydrophobic residue at the N-terminus of $\beta 3$, but this is not seen in $\beta 3$ of the second RRM. Additional binding affinity is also found in some cases where the hydroxyl group of Tyr donates hydrogen bonds to the ribophosphate backbone. In PABP, Tyr14 and Tyr56 from $\beta 1$ and $\beta 3$ of RRM1 respectively both form hydrogen bonds to phosphate backbone oxygens (O1P of Ade6 and O2P of Ade8), and the $\beta 1$ Tyr214 in RRM2 of Sxl donates a hydrogen bond to the O2' position of U3. Sxl, Hud, and Rbd12 do not form base stacking interactions from their RRM1 $\beta 1$ sheets.

Two conserved residues involved in base stacking from $\beta 1$ of UP1 are Phe17 and Phe108 of RRM1 and RRM2 respectively (Figure 8.3). Both form base stacking interactions with Ade in the UP1:TR2 structure and both help make up a hydrophobic cleft with two other conserved residues in $\beta 3$ of RRM1 and RRM2; Phe59 and Phe150.

Figure 8.3. The hydrophobic clefts of UP1. A, the hydrophobic cleft of RRM1 is made from the conserved Phe17 ($\beta 1$) and Phe59 ($\beta 3$) which stack with bases in the UP1:TR2 complex. Phe57 ($\beta 3$) is also highly conserved in RRM1 of the complexes compared here, but is not involved in base stacking. B, The hydrophobic cleft of RRM2 is made from Phe108 ($\beta 1$) and Phe150 with bases in the UP1:TR2 complex. Phe148 ($\beta 3$) is also in the hydrophobic cleft but is not highly conserved in the second RRM of the complexes compared here.

Figure 8.4. The hydrophobic clefts of Sxl. A, In RRM1, the hydrophobic cleft is made from Ile128 (β 1), and the β 3 residues Tyr168 and Phe170. B, Tyr168 and Phe170 both make stacking interactions with Ura.

Figure 8.5. The hydrophobic clefts of Hud. A, The hydrophobic cleft of RRM1 consists of Ile42 from β 1 and Tyr82 and Phe84 of β 3. Phe84 is the only core involved in base stacking. B, The hydrophobic cleft in RRM2 is made from β 1 residue Tyr128 and the β 3 residues Val168 and Phe170. Tyr128 and Phe170 are both involved in base stacking.

The $\beta 3$ residues are also involved in base stacking interactions with Gua in the UP1:TR2 structure. Phe57 and Phe148 are $\beta 3$ residues that make up the hydrophobic cleft in RRM1 and RRM2 respectively, but neither are involved in base stacking. Interestingly, Phe57 is fairly conserved when compared with the first RRM of other proteins in the structural data bank (Figure 8.2), but Phe148 in RRM2 is less conserved. One additional stacking interaction is made outside the hydrophobic cleft in UP1 from the $\beta 4$ Arg92 to Thy7.

The hydrophobic cleft in RRM1 of Sxl is made from Ile128 in $\beta 1$ and Phe168 and Phe170 in $\beta 3$ (Figure 8.4). Ile128 does not interact with bases in the Sxl:RNA complex, while Tyr168 and Phe170 stack with Ura. An additional stacking interaction is made from Tyr131 in $\beta 1$ which stacks with Ura. The hydrophobic cleft of RRM2 consists of Tyr214 in $\beta 1$ and Phe256 in $\beta 3$ and both are involved in base stacking. Tyr214 stacks with a Gua, the only purine in the complex, and Phe256 stacks with Ura. Tyr214 makes an additional interaction in the complex by donating a hydrogen bond to O2' of Ura. Like UP1, Sxl uses a $\beta 4$ Arg to stack with a pyrimidine, Arg195 to Ura, and both align within one residue in the primary sequence.

The primary sequence of Hud resembles that of Sxl, and the $\beta 1$ residues in RRM1 are identical (Figure 8.2). The equivalent hydrophobic residues in $\beta 1$ and $\beta 3$ are Ile42 and Phe84, respectively, and Phe84 stacks with Ura (Figure 8.5). Tyr82 also contributes to the hydrophobic cleft from the $\beta 3$ sheet but does not interact with the polynucleotide in the Hud:RNA complex. As with Tyr131 in Sxl, Tyr45 in Hud is involved in base stacking, and it stacks with Ura. However, unlike Sxl, this residue does not interact with

the nucleic acid phosphate backbone. The RRM2 β 1 residues of Hud are nearly identical to those of Sxl and differ by only one amino acid. The hydrophobic cleft in this motif is formed by Tyr128 in β 1 and Val168 and Phe170 in β 3. Tyr128 stacks with Ade and Phe170 stacks with Ura.

The hydrophobic cleft of RRM1 in PABP uses Tyr14 in β 1 and Tyr54 and Tyr56 in β 3, and both Tyr54 and Tyr56 stack with Ade (Figure 8.6). The hydroxyl groups of these stacking residues donate hydrogen bonds to the phosphate oxygens of the nucleic acid backbone. The hydrophobic cleft in RRM2 is made from the conserved Phe102 and Phe142 in β 1 and β 3, respectively, in addition to Tyr140 in β 3. As with RRM1, these conserved residues are involved in base stacking with Ade. Like UP1, PABP makes an additional base stacking interaction outside the hydrophobic cleft with a β 4 Arg, Arg94, which stacks with Ade.

U1A is a single RRM protein that utilizes Tyr13 and Phe56 in the hydrophobic cleft of β 1 and β 3 respectively. Tyr13 stacks with Cyt and Phe56 stacks with Ade. Unlike the β 1 and β 3 Tyr in PABP and the β 1 Tyr in Sxl, Tyr13 does not donate hydrogen bonds to the nucleic acid phosphate backbone.

The β 1 and β 3 hydrophobic regions of the Rbd12 RRMs are similar to UP1 in terms of primary sequence (Figure 8.2). Phe17 in β 1 of RRM1 is conserved, and, like Phe17 in UP1, this residue is involved in base stacking in the Rbd12:RNA structure (Figure 8.8). Unlike the other conserved hydrophobic β 1 residues in this comparison, the equivalent β 1 residue in RRM2 of Rbd12 is the only example that uses Leu. While Ile is used in Sxl and Hud, base stacking interactions in these proteins are made from other β 1

tyrosines. In Rbd12, no stacking interactions are made from $\beta 1$ amino acids, although stacking interactions are made from conserved $\beta 3$ residues. The N-terminal $\beta 3$ Phe56 in RRM1 is conserved with UP1 and does not stack with bases in the Rbd12:RNA complex. Tyr58 from the conserved $\beta 3$ region of RRM1 stacks with Gua, and Tyr140 in RRM2 stacks with Cyt. An additional stacking interaction is made from Arg127 in $\beta 2$, and, although UP1, Sxl, and PABP also utilize Arg in stacking interactions, these examples use Arg from the RRM $\beta 4$ sheet.

The RRM hydrophobic cleft has a role in the construction of the RRM, while at the same time providing a surface for base stacking to occur in most cases. In all the structures studied, the conserved $\beta 3$ hydrophobic residue is either a Phe or Tyr and in all cases the $\beta 3$ residue is involved in base stacking. The hydrophobic cleft $\beta 1$ residue is

Figure 8.8. The hydrophobic clefts of Rbd12. A, The hydrophobic cleft of RRM1 consists of $\beta 1$ residue Phe17 and $\beta 3$ residues Phe56 and Tyr58. Phe17 and Tyr58 are both involved in base stacking. B, The hydrophobic cleft of RRM2 consists of Leu103 from $\beta 1$ and Tyr140 from $\beta 3$. Tyr140 stacks with Cyt in the Rbd12:RNA complex.

less conserved, but, when the residue is a Phe or Tyr, base stacking is observed in the majority of cases. The binding targets of the six RRM proteins are composed of a roughly equal number of purines and pyrimidines (45 purines and 43 purines), but twice as many stacking interactions are made to purines. Twelve purine stacking interactions were observed compared to only six made with pyrimidines. All four RNA bases were represented in the complexes, and there appears to be no correlation between specific bases and the amino acid side chains they stack with. We could not identify a correlation between the primary sequences of RRM hydrophobic clefts and whether or not those residues are involved in stacking or to which bases they likely stack.

8.3 UP1 Consensus Binding Sequence

The structure of UP1 complexed with TR2 determined by Ding *et al.* (1999) and our own structure studies indicate that a likely consensus binding sequence is (nYAGGn). In the crystal structure, Thy1 of the repeat is highly disordered and does not bind with RRM1 and is not included in the crystallographic model. ENTANGLE shows that Thy7 (the first Thy in the second repeat) also makes no interactions with RRM2, although it is stabilized by a stacking interaction with Arg92 which also stacks with Gua4. The crystal structure shows that Thy2 and Thy8 could potentially be replaced with Cyt as in the mouse MN repeats. In both cases the O2 position, nonvariant in Ura, Thy, or Cyt, accepts a hydrogen bond from a side chain amide: Lys87 in Thy2 and Arg178 in Thy8. Thus, it appears that the Thy2 and Thy8 binding sites in UP1 might be nonspecific among pyrimidines but discriminate against purines.

Figure 8.9. Summary of base-stacking and base specific hydrogen bonds between UP1 and TR2. Dotted lines indicate hydrogen bonds, and solid arrows indicate base-stacking. Residues enclosed in single-line boxes make interactions through side chains, and those enclosed in double-line boxes contact DNA with main chain amides or carbonyls.

Multiple interactions occur between the RRM of UP1 and the next three bases, AGG. In addition to its presence in hTR and mouse MNs, this repeat is found in the SELEX “winner” sequence (Burd and Dreyfuss, 1994), and the 5’ and 3’ distal splice site consensus sequence to which the full length protein, hnRNP A1, has been suggested to bind in its role in pre-mRNA packaging, splicing, and transport (Abdul-Manan and Williams, 1996). Nasim *et al.* (2002) have shown that positioning high affinity hnRNP A1 binding sites to promote loop formation within pre-mRNAs can attenuate splice site selection, and the binding of nucleic acid to UP1 in an anti-parallel fashion is consistent with this role in alternative splicing. Both RRMs provide multiple stacking interactions and base specific hydrogen bonds from protein side chains and main chains.

The final base in the hTR sequence corresponds to Gua6 and Gua12 in the crystal structure and neither interact with UP1. The solvent exposed nature of the Gua6 position was exploited for placement of 6MI in the fluorescence studies, and, while Gua12 stacks with Gua11, Gua12 makes no direct contacts to UP1.

8.3a Characterizing the Pyrimidine Binding Pocket

Based on the crystal structure of the UP1:TR2 complex and the high affinity of UP1 for hTRs and mouse MN repeats, we have suggested that UP1 does not discriminate between cytosine and thymine in the Thy2 and Thy8 binding pockets of UP1.

Oligonucleotides have been made to test our theory in both RRM1 and RRM2; however, only changes to the Thy2 binding pocket of RRM1 are tolerated crystallographically.

This result is somewhat surprising given the similarity of the binding pockets and somewhat ironic since we were unable to crystallize UP1 bound to oligonucleotides that

contained modified purine bases in the RRM1 binding pocket. The situation with the pyrimidine binding pocket is reversed.

T(2)5MC is an oligonucleotide based on the hTR sequence that utilizes 5-methyl-cytosine in place of Thy2, and TR2-2C is a similar oligonucleotide that uses cytosine in place of Thy2. These changes were made to test our theory that the consensus binding sequence of UP1 is (nYAGGn)_n and that a change from Thy to Cyt would cause the Glu85 side chain to move from the pyrimidine N3 position (hydrogen bond donor in Thy, hydrogen bond acceptor in Cyt) to the amino group N4 (O4 in Thy). We also speculated that the Lys87 side chain would move from O2 to a position that would allow a bidentate hydrogen bond interaction to the O2 and N3 positions of 5-methyl-cytosine and Cyt, thus supporting the idea that the first position in the consensus binding sequence need only be a pyrimidine. Although the T(2)5MC structure is not yet fully refined, initial observations show good electron density in the 5-methyl-cytosine position and side chain movement is visible. The UP1:TR2-2C data is fully refined and side chain movements to accommodate Cyt are clearly visible.

In the UP1:TR2-2C structure, the Glu85 side chain moves from 2.65 Å to 2.85 Å at the N3 position where a hydrogen bond is no longer allowed and moves from 3.90 Å to 2.78 Å at the N4 position where a hydrogen bond is allowed. Lys87 moves from 4.71 Å to 2.96 Å at the N3 position where a hydrogen bond is now available, and the distance between it and the O2 position also decreases as a result, from 3.03 Å to 2.71 Å. Similar movements are expected in the UP1:T(2)5MC complex, and these structures indicate that, while the pyrimidine methyl group does not induce steric clashes when Thy or 5-MC are in the binding pocket, the methyl group is also not crucial in binding. These

observations shed light on the findings of Fukuda *et al.* (2002) that UP1 can bind the telomeric repeat sequence as well as the mouse MN sequence with high specificity. Unfortunately, UP1 bound to a sequence where Thy8 is replaced with Cyt has yet to crystallize, but the same movements of Glu176 and Arg178 side chains are expected as those seen with Glu85 and Lys87. Future binding experiments will also help to determine pyrimidine specificity in both positions in the absence of crystal data.

8.3b UP1 Adenine Binding Pocket

In both RRM, Ade makes main chain hydrogen bonds from N6 to equivalent RRM positions. Arg88 and Lys179 accept hydrogen bonds from N6 of Ade3 and Ade9, respectively, from main chain carbonyls. As mentioned previously, stacking interactions are made from the hydrophobic cleft β 1 residues Phe17 and Phe108. Unlike Ade3, Ade9 interacts with UP1 through the N7 position to Arg178, which also donates a hydrogen bond to O2 of Thy8.

It is unlikely that the Ade binding sites can be engineered to accept alternative bases since both are defined by main chain interactions. Steric exclusions from the N2 position would almost certainly prevent Gua from occupying the pocket. Val90 (RRM1) and Leu181 (RRM2) donate main chain hydrogen bonds to N1 of Ade3 and Ade9 respectively. The close proximity of the Ade N1 with the protein main chain would likely exclude the N1 hydrogen of Gua from the Ade9 pocket (Figure 5.4). Pyrimidines may not be sterically excluded and Gua in the *syn* conformation could conceivably occupy the Ade position. However, without altering the size and electrostatic nature of the pocket, Ade will most certainly have higher selectivity. Given the contribution of

Arg178 in binding affinity as seen in the competition experiments, it may be possible to drive affinity in the Ade9 binding pocket toward acceptance of a pyrimidine. The equivalent residue in RRM1 is Lys87, which, like Arg178 in RRM2, hydrogen bonds with the O2 of the 5' Thy (Thy2 in RRM1 and Thy8 in RRM2).

By mutating Arg178 to Glu, the Ade9 pocket may gain affinity for Cyt. The Cyt N4 could potentially donate a hydrogen bond to the electrostatic Glu side chain, and steric clashes with the UP1 main chain could be avoided. Affinity for the Thy8 O2 from Arg178 would be lost, however, and it is difficult to predict the contribution of this bond to total binding. This proposed change would be an excellent candidate for a molecular dynamics simulation. A similar mutation of Lys87 in RRM1 may drive affinity for Cyt in the Ade3 position. It would also be interesting to experiment with increasing the affinity for Ade in the A3 pocket by mutating Lys87 to Arg to see if the guanidinium group makes a hydrogen bond with the Ade3 N7 position as it does with Ade9.

8.3c UP1 Guanine Binding Pockets

The Gua4 and Gua10 binding pockets were described in detail in Chapter 7, and we have shown that the Lys106 to Gua O6 hydrogen bond is important in Gua recognition in the Gua10 binding site. Binding experiments with Ade in the Gua10 position showed a five to ten-fold reduction in binding, and crystal structures with modified bases in place of Gua10 showed that, when an O6 was absent, the nucleotide occupied the pocket in the *anti* conformation. Lys106 is able to donate a hydrogen bond to the purine N1 when a base assumes the *anti* conformation, but when Ade is superimposed in the binding pocket, the N6 clashes with the Lys106 side chain. If

Lys106 were replaced with Gln (Figure 6.3), it is likely that two strong contacts would be made to the Ade N6 and N1. In the crystal structure with 2-aminopurine in the Gua10 binding pocket, a water-mediated hydrogen bond was made from the Leu181 carbonyl to the purine N7, and it is likely that a similar interaction would give binding affinity to Ade.

The Gua4 binding site is similar to that of Gua10, but two additional hydrogen bonds are made to the base through protein side chains (Figure 6.6). A similar main chain interaction to the Gua N1 from Val90 in the Gua4 binding pocket suggests that an electrostatic clash would exclude Ade in the *syn* conformation. The electrostatic clash would, however, be between two hydrogen bond acceptors, and there would not be a steric clash as would be expected if two hydrogen bond donors were placed at a similar distance. While Lys106 makes a bidentate hydrogen bond to the N7 and O6 of Gua10, the equivalent RRM1 residue, Lys15, makes a single hydrogen bond to N7 of Gua4, and Gln12 donates a hydrogen bond to the O6 of Gua4. With Ade in the same position, the N7 interaction would be unaffected, and the Gln12 side chain may flip 180° and accept a hydrogen bond from the Ade N6 to the Gln carbonyl. In the absence of crystal structures with modified bases, it is difficult to predict how Ade would be accepted in the Gua4 binding site, but a superposition of Ade in the *anti* conformation suggests that a steric clash between Ade and neighboring Gua5 would force Ade to assume the *syn* conformation.

The Gua4 and Gua10 binding sites appear to be equally exclusive of pyrimidines (Figures 6.5 and 6.7). Depending on the flexibility of the Lys15 and Lys106 side chains in the Gua4 and Gua10 binding pockets, it is plausible that Ura may be able to accept

hydrogen bonds to the O4 position from these Lys amino groups, but the geometry and distances would not be optimal. Steric and electrostatic exclusions of Thy and Cyt would likely prevent these bases from binding altogether.

The final Gua in the (nYAGGn)_n consensus binding sequence occupies the Gua5 and Gua11 binding pockets of RRM1 and RRM2, respectively. Gua5 is bound by an Arg92 to N7 hydrogen bond through the Arg guanidinium group and by two hydrogen bonds from the Arg42 side chain to the Gua N1 and N2 position. An additional stacking interaction from Met46 helps stabilize the base. Gua11 is bound by only one hydrogen bond from the Lys183 main chain to O6, and no protein to base stacking interactions are seen, although Gua11 stacks with Gua10 and Gua12.

The Gua5 binding site presents a couple of opportunities for engineering affinity for Ade. An electrostatic clash between N1 and Asp42 is seen when Ade is superimposed in the Gua5 binding pocket but the Arg92 to N7 interaction would probably not be affected (Figure 8.10). Mutating Asp42 to Asn would provide a hydrogen bond donor to the Ade N1 position. An additional Lys15 to Gln mutation would likely provide a hydrogen bond acceptor for the Ade N6 while still providing the N7 hydrogen bond to Gua4 that Lys15 makes.

The Gua5 binding site may also be modified to accept pyrimidines by mutating Asp42 to Gln. The Met46 stacking interaction is probably most stable when a purine occupies the binding pocket but an Asp42 to Gln mutation would probably exclude purines. Like the proposed Gln12 flip that could accept either a Gua O6 or Ade N6 in the Gua4 binding pocket, Gln42 would be able to hydrogen bond with either a Cyt N4 or a Thy or Ura O4 with these bases in the *anti* conformation.

The Gua11 binding site offers little in terms of engineering affinity for other bases. When Ade is superimposed in the Gua11 binding pocket, it is clear that a hydrogen bond donor to donor clash from the Lys183 main chain amide to the Ade N6 would disrupt binding (Figure 7.3). If Ade flipped from the *anti* to the *syn* conformation the clash would be greater since the N6 would be about 0.35 Å closer to the amide (Figure 7.4). Thy may be able to occupy this site by accepting a hydrogen bond from the Lys183 main chain amide to the Thy O4, but Gua probably will still have higher selectivity since it stacks with neighboring purines and the O6 to amide distance is more favorable.

Figure 8.10. A, Gua in the Gua5 binding pocket of UP1. B, Proposed effects of mutating Lys15 to Gln and Asp42 to Asn in the Gua5 binding pocket of UP1 in order to give the binding site affinity for Ade. The Lys15 side chain makes a hydrogen to Gua4 and replacing Lys with Gln would likely retain this affinity, while at the same time providing a hydrogen bond for the Ade N6. A hydrogen bond might also be made at the Ade N1 position by replacing Asp42 with Asn.

REFERENCES

- Abdul-Manan, N., and Williams, K. R. (1996). hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res* 24, 4063-4070.
- Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., and Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* 380, 646-650.
- Allers, J., and Shamoo, Y. (2001). Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J Mol Biol* 311, 75-86.
- Andrade, M. A., and Bork, P. (1995). HEAT repeats in Huntington's disease protein. *Nat Gen* 11, 115-116.
- Bai, Y., Lee, D., Yu, T., and Chasin, L. A. (1999). Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res* 27, 1126-1134.
- Balagurumoorthy, P., and S.K., B. (1994). Structure and stability of human telomeric sequence. *J Biol Chem* 269, 21858-21869.
- Barabino, S. M., Blencowe, B. J., Ryder, U., Sproat, B. S., and Lamond, A. I. (1990). Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* 63, 293-302.
- Beyer, A. L., Christensen, M. E., Walker, B. W., and LeSturgeon, W. M. (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* 11, 127-138.
- Birney E., K. S., Krainer A.R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Research* 21, 5803-5816.
- Blanchette, M., and Chabot, B. (1997). A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *Rna* 3, 405-419.

Blanchette, M., and Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *Embo J* 18, 1939-1952.

Boffa, L. C., Karn, J., Vidali, G., and Allfrey, V. G. (1977). Distribution of NG, NG, - dimethylarginine in nuclear protein fractions. *Biochem Biophys Res Comm* 74, 969-976.

Bonifaci, N., Moroianu, J., Radu, A., and Blobel, G. (1997). Karyopherin beta-2 mediates nuclear import of a mRNA binding protein. *Proc Natl Acad Sci USA* 94, 5055-5060.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst D* 54, 905-921.

Bryan, T. M., and Cech, T. R. (1999). Telomerase and the maintenance of chromosome ends. *Curr Opin Cell Biol* 11, 318-324.

Burd, C. G., and Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo J* 13, 1197-1204.

Buvoli, M., Cobiainchi, F., and Riva, S. (1992). Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly. *Nucleic Acids Res* 20, 5017-5025.

Caceres, J. F., Stamm, S., Helfman, D. M., and Krainer, A. R. (1994). Regulation of Alternative Splicing in Vivo by Overexpression of Antagonistic Splicing Factors. *Science* 265, 1706-1709.

Calvio, C., Neubauer, G., Mann, M., and Lamond, A. I. (1995). Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *RNA* 7, 724-733.

Carson, M. (1991). Ribbons 2.0. *J Appl Cryst* 24, 958-961.

Cartegni, L., Maconi, M., Morandi, E., Cobiainchi, F., Riva, S., and Biamonti, G. (1996). hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol* 259, 337-348.

- Castrignano, T., Chillemi, G., Varani, G., and Desideri, A. (2002). Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration. *Biophys J* 83, 3542-3552.
- Cheng, A. C., Chen, W. W., Fuhrmann, C. N., and Frankel, A. D. (2003). Recognition of Nucleic Acid Bases and Base-pairs by Hydrogen Bonding to Amino Acid Side-chains. *J Mol Biol* 327, 781-796.
- Cheng, A. C., Frankel, A.D. (2004). Ab Initio Interaction Energies of Hydrogen-Bonded Amino Acid Side Chain-Nucleic Acid Base Interactions. *J Amer Chem Soc* 126, 434-435.
- Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. (1986). Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science* 231, 1534-1539.
- Chook, Y. M., and Blobel, G. (1999). Structure of the nuclear transport complex karyopherin-beta2-RanGppNHp. *Nature* 399, 230-237.
- Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M., and Riva, S. (1993). Phosphorylation of human hnRNP protein A1 abrogates in vitro strand annealing activity. *Nucleic Acids Res* 21, 949-955.
- Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., and Wilson, S. H. (1986). Structure of rodent helix-destabilizing protein revealed by cDNA cloning. *J Biol Chem* 261, 3536-3543.
- Collins, K. (2000). Mammalian telomeres and telomerase. *Curr O Cell Biol* 12, 378-383.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., and Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J*, 11, 1921-1929.
- Counter, C. M., Hirte, H. W., Bacchetti, S., and Harley, C. B. (1994). Telomerase activity in human ovarian carcinoma. *Proc Natl Acad Sci U S A* 91, 2900-2904.

Dallaire, F., Dupuis, S., Fiset, S., and Chabot, B. (2000). Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J Biol Chem* 275, 14509-14516.

Daneholt, B. (1997). A look at messenger RNP moving through the nuclear pore. *Cell* 88, 585-588.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of Polyadenylate RNA by the Poly(A)-Binding Protein. *Cell* 98, 835-845.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13, 1102-1115.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3, 195-205.

Dreyfuss, G., MJ, M., S, P.-R., and CG, B. (1993). hnRNP proteins and the biogenesis of mRNA. *Ann Rev Bioc* 62, 289-321.

Evans, S. K., Bertuch, A. A., and Lundblad, V. (1999). Telomeres and telomerase: at the end, it all comes together. *Trends Cell Bio* 9, 329-331.

Eversole, A., and Maizels, N. (2000). In vitro properties of the conserved mammalian protein hnRNP D suggest a role in telomere maintenance. *Mol Cell Biol* 20, 5425-5432.

Faustino, N. A., and Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes and Devel* 17, 419-437.

Fiset, S., and Chabot, B. (2001). hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res* 29, 2268-2275.

Ford, L. P., Wright, W. E., and Shay, J. W. (2002). A model for heterogenous nuclear ribonucleoproteins in telomere and telomerase regulation. *Oncogene* 21, 580-583.

Fridell, R. A., Truant, R., Thorne, L., Benson, R. E., and Cullen, B. R. (1997). Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta. *J Cell Sci* 110 (*Pt 11*), 1325-1331.

- Fukuda, H., Katahira, M., Tsuchiya, N., Enokizono, Y., Sugimura, T., Nagao, M., and Nakagama, H. (2002). Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc Natl Acad Sci U S A* 99, 12685-12690.
- Gallouzi, I. E., and Steitz, J. A. (2001). Delineation of mRNA export pathways by the use of cell-permeable peptides. *Science* 294, 1895-1901.
- Gattoni, R., Mahe, D., Mahl, P., Fischer, N., Mattei, M. G., Stevenin, J., and Fuchs, J. P. (1996). The human hnRNP-M proteins: structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res* 24, 2535-2542.
- Gavathiotis, E., and Searle, M. S. (2003). Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)₄ containing the human telomeric repeat: evidence for A-tetrad formation from NMR and molecular dynamics simulations. *Org Biomol Chem* 1, 1650-1656.
- Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* 253, 157-163.
- Hall, K. B. (2002). RNA-protein interactions. *Curr O Struct Biol* 12, 283-288.
- Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* 4, 430-444.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* 398, 579-585.
- Hawkins, M. E. (2001). Fluorescent pteridine nucleoside analogs: a window on DNA interactions. *Cell Biochem Biophys* 34, 257-281.
- Hawkins, M. E., Pfeleiderer, W., Balis, F. M., Porter, D., and Knutson, J. R. (1997). Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Anal Biochem* 244, 86-95.

- Hoffman, M. M., Khrapov, M. A., Cox, J. C., Yao, J., Tong, L., and Ellington, A. D. (2004). AANT: the amino acid-nucleotide interaction database. *Nucleic Acids Res* 32, D174-181.
- Huang, M., Rech, J. E., Northington, S. J., Flicker, P. F., Mayeda, A., Krainer, A. R., and LeStourgeon, W. M. (1994). The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol Cell Biol* 14, 518-533.
- Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B. (2002). Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* 277, 29745-29752.
- Ishikawa, F., Matunis, M. J., Dreyfuss, G., and Cech, T. R. (1993). Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol Cell Biol* 13, 4301-4310.
- Izaurrealde, E., Jarmolowski, A., Beisel, C., Mattaj, I. W., Dreyfuss, G., and Fischer, U. (1997). A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export. *J Cell Biol* 137, 27-35.
- Katahira, M., Fukuda, H., Kawasumi, H., Sugimura, T., Nakagama, H., and Nagao, M. (1999). Intramolecular quadruplex formation of the G-rich strand of the mouse hypervariable minisatellite Pc-1. *Biochem Biophys Res Commun* 264, 327-333.
- Kataoka, N. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* 6, 673-682.
- Kim, S., Merrill, B. M., Rajpurohit, R., Kumar, A., Stone, K. L., Papov, V. V., Schneiders, J. M., Szer, W., Wilson, S. H., Paik, W. K., and Williams, K. R. (1997). Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly/Gly/Phe is a preferred recognition motif. *Biochemistry* 36, 5185-5192.
- Kim, V. N., and Dreyfus, G. (2001). Nuclear mRNA binding proteins couple pre-mRNA splicing and post-splicing events. *Mol Cells* 12, 1-10.
- Kleywegt, G. J. (1995). Dictionaries for Heteros. CCP4/ESF-EACBM Newsletter on Protein Crystallography 31, 45-50.

- Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Curr O Cell Bio* 11, 363-371.
- LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M. R., Wellinger, R. J., and Chabot, B. (1998). Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet* 19, 199-202.
- LeHir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO* 19, 6860-6869.
- Lingner, J., and Cech, T. R. (1998). Telomerase and chromosome end maintenance. *Curr O Genet Dev* 8, 226-232.
- Lohman, T. M., and Bujalowski, W. (1991). Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: spectroscopic approaches to monitor binding. *Methods Enzymol* 208, 258-290.
- Lorsch, J. R. (2002). RNA Chaperones Exist and DEAD Box Proteins Get a Life. *Cell* 109, 797-800.
- Lou, H., Helfman, D.M., Gagel, R.F., Berget, S.M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol Cell Biol* 19, 78-85.
- Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stevenin, J., and Branlant, C. (2002). A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. *J Mol Biol* 323, 629-652.
- Mascotti, D. P., and Lohman, T. M. (1990). Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. *Proc Natl Acad Sci U S A* 87, 3142-3146.
- Mayeda, A., and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68, 365-375.

McAfee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyengar, S., and Lestourgeon, W. M. (1997a). The packaging of pre-mRNA. Cold Spring Harbor, NY: Oxford University Press, 68-102.

McAfee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyengar, S., and Lestourgeon, W. M. (1997b). The packaging of pre-mRNA. In *Eukaryotic mRNA Processing* (Cold Spring Harbor, NY, Oxford University Press), pp. 68-102.

Melchior, F., Guan, T., Yokoyama, N., Nishimoto, T., and Gerace, L. (1995). GTP hydrolysis by Ran occurs at the nuclear pore complex in an early step of protein import. *J Cell Bio* 131, 571-591.

Merrill, B. M., Lopresti, M. B., Stone, K. L., and Williams, K. R. (1987). Amino acid sequence of UP1, an hnRNP-derived single-stranded nucleic acid binding protein from calf thymus. *Int J Pept Protein Res* 29, 21-39.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., and Williams, K. R. (1988). Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J Biol Chem* 263, 3307-3313.

Michael, W. M., Choi, M., and Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* 83, 415-422.

Minvielle-Sebastia, L., Katrin Beyer, Annette M. Krecic, Ron E. Hector, Maurice S. Swanson, Walter Keller (1998). Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO* 17, 7454-7468.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., *et al.* (2003). The InterPro database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 31, 315-318.

Munroe, S. H., and Dong, X. F. (1992). Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA.RNA annealing. *Proc Natl Acad Sci U S A* 89, 895-899.

Myers, J. C., Moore, S. A., and Shamoo, Y. (2003). Structure-based incorporation of 6-methyl-8-(2-deoxy-beta-ribofuranosyl)isoxanthopteridine into the human telomeric repeat DNA as a probe for UP1 binding and destabilization of G-tetrad structures. *J Biol Chem* 278, 42300-42306.

- Myers, J. C., and Shamoo, Y. (2004). Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM). *J Mol Biol* 342, 743-756.
- Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., and Williams, K. R. (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry* 30, 2968-2976.
- Nakielnny, S., and Dreyfuss, G. (1996). The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J Cell Bio* 134, 1365-1373.
- Nakielnny, S., and Dreyfuss, G. (1997). Nuclear export of proteins and RNAs. *Curr Opin Cell Biol* 9, 420-429.
- Nakielnny, S., and Dreyfuss, G. (1999). Transport of Proteins and RNAs in and out of the Nucleus. *Cell* 99, 677-690.
- Nasim, F. U., Hutchison, S., Cordeau, M., and Chabot, B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *Rna* 8, 1078-1089.
- Neidle, S., and Parkinson, G. N. (2003). The structure of telomeric DNA. *Curr Opin Struct Biol* 13, 275-283.
- Nichols, R. C., Wang, X. W., Jie Tang, B., Hamilton, J., High, F. A., Herschman, H. R., and Rigby, W. F. C. (2000). The RGG Domain in hnRNP A2 Affects Subcellular Localization. *Exper Cell Res* 256, 522-532.
- Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M., and Hentze, M. W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* 89, 597-606.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372, 432-438.

Pandolfo, M., Valentini, O., Biamonti, G., Morandi, C., and Riva, S. (1985). Single stranded DNA binding proteins derive from hnRNP proteins by proteolysis in mammalian cells. *Nucleic Acids Res* 13, 6577-6590.

Parkinson, G. N., Lee, M. P., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417, 876-880.

Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Cryst - Sect D* 55, 1718-1725.

Phan, A. T., and Mergny, J. L. (2002). Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res* 30, 4618-4625.

Pinol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* 355, 730-732.

Pollard, A. J., Sparey, C., Robson, S. C., Krainer, A. R., and Europe-Finner, G. N. (2000). Spatio-temporal expression of the trans-acting splicing factors SF2/ASF and heterogeneous ribonuclear proteins A1/A1B in the myometrium of the pregnant human uterus: a molecular mechanism for regulating regional protein isoform expression in vivo. *J Clin Endocrin Metab* 85, 1928-1936.

Pollard, V. W., Michael, W. M., Nakielnny, S., Siomi, M. C., Wang, F., and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. *Cell* 86, 985-994.

Record, M. T., Jr. (1976). Ion Effects on Ligand-Nucleic Acid Interactions. *J Mol Biol* 107, 145-158.

Record, M. T., Jr., Woodbury, C. P., and Lohman, T. M. (1976). Na⁺ effects on transition of DNA and polynucleotides of variable linear charge density. *Biopolymers* 15, 893-915.

Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* 108, 523-531.

Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K. R., Multhaup, G., Beyreuther, K., Werr, H., and et al. (1986). Mammalian single-stranded DNA binding protein UP I is derived from the hnRNP core protein A1. *Embo J* 5, 2267-2273.

Ruby, S. W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242, 1028-1035.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* 59, 349-358.

Shafer, R. H., and Smirnov, I. (2001). Biological Aspects of DNA/RNA Quadruplexes. *Biopolymers* 56, 209-227.

Shahied-Milam, S., Soltaninassab, R., Iyer, G. V., and LeSturgeon, W. M. (1998). The Heterogeneous Nuclear Ribonucleoprotein C Protein Tetramer Binds U1, U2, and U6 snRNAs through Its High Affinity RNA Binding Domain (the bZIP-like Motif). *J Biol Chem* 273, 21359-21367.

Shamoo, Y., Abdul-Manan, N., Patten, A. M., Crawford, J. K., Pellegrini, M. C., and Williams, K. R. (1994). Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry* 33, 8272-8281.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R., and Steitz, T. A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol* 4, 215-222.

Shih, S. C., and Claffey, K. P. (1999). Regulation of human vascular endothelial growth factor mRNA stability in hypoxia by heterogeneous nuclear ribonucleoprotein L. *J Biol Chem* 274, 1359-1365.

Showalter, S. A., and Hall, K. B. (2004). Altering the RNA-binding Mode of the U1A RBD1 Protein. *J Mol Bio* 335, 465-480.

Simard, M. J., and Chabot, B. (2000). Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Mol Cell Biol* 20, 7353-7362.

Siomi, M. C., Eder, P. S., Kataoka, N., Wan, L., Liu, Q., and Dreyfus, G. (1997). Transportin-mediated Nuclear Import of Heterogeneous Nuclear RNP Proteins. *J Cell Biol* 138, 1181-1192.

Stewart, M., Kent, H. M., and McCoy, A. J. (1998). The structure of the Q69L mutant of GDP-Ran shows a major conformational change in the switchII loop that accounts for its failure to bind nuclear transport factor2 (NTF2). *J Mol Biol* 284, 1517-1527.

Sun, S., and Shamoo, Y. (2003). Biochemical characterization of interactions between DNA polymerase and single-stranded DNA-binding protein in bacteriophage RB69. *J Biol Chem* 278, 3876-3881.

Treger, M., Westhof, E. (2001). Statistical analysis of atomic contacts at RNA-protein interfaces. *J Mol Rec* 14, 199-214.

Tseng, S. S., Weaver, P. L., Liu, Y., Hitomi, M., Tartakoff, A. M., and Chang, T. H. (1998). Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *EMBO* 17, 2651-2662.

Tuerk, C. (1997). Using the SELEX Combinatorial Chemistry Process to Find High Affinity Nucleic Acid Ligands to Target Molecules. In *PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering* (Totowa, NJ: Humana Press Inc.). 219-230.

Visa, N., Alzhanova-Ericsson, A. T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T., and Daneholt, B. (1996). A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell* 84, 253-264.

Weilbaecher, R. G., and Lundblad, V. (1999). Assembly and regulation of telomerase. *Curr O Chem Biol* 3, 573-577.

Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996). Evidence for a new step in telomere maintenance. *Cell* 85, 423-433.

Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* 5, 559-570.

Zhang, X., Lyle, C., Wang, Y., and Zeng, L. (2001). Role of hnRNP A1 in coronavirus RNA synthesis. *Adv Exp Med Biol* 494, 437-446.

Zhu, J., Mayeda, A., and Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8, 1351-1361.

Zimm, B. H. (1948). J Chem Phys *16*, 1099-1116.