Nucleic Acids Research, 1996, Vol. 24, No. 20 **4063–4070**

hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding

Norzehan Abdul-Manan and Kenneth R. Williams[1,*]

Department of Molecular Biophysics and Biochemistry and [1]Howard Hughes Medical Institute, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8024, USA

Received May 13, 1996; Revised and Accepted July 18, 1996

ABSTRACT

To understand the range of possible and probable A1 functions in pre-mRNA biogenesis, it is important that we quantify the relative ability (or inability) of A1 to bind high affinity RNA target sequences and/or structures. Using a fluorescence competition assay we have determined apparent binding affinities for a wide range of 20mer oligos containing putative and possible A1 targets including the high affinity 'winner' sequence identified by selection/amplification [Burd,C.G and Dreyfuss,G. (1994) *EMBO J.* 13, 1197–1204], AUUUA sequences found in 3′-UTRs of labile mRNAs, 5′- and 3′-splice sites and telomeric sequences. With the exception of a 20mer 'winner' sequence, all other 20mers examined bind A1 with a narrow, ~10-fold range of affinities extending from 3.2×10^6 to 4.2×10^7 M^{-1}. Studies with homo-oligomers suggest this range reflects nucleotide base rather than sequence specificity and hence, it was possible to predict reasonably accurate affinities for all other 20mers examined except for the 'winner', whose unusually high affinity of 4.0×10^8 M^{-1} results from a unique higher order structure and sequence. Since there is no known physiological role for the 'winner' 20mer sequence, these data suggest A1 generally binds indiscriminately to all available pre-mRNA sequences. Both the large abundance of A1 *in vivo* and its binding properties are thus consistent with it playing a structural role in pre-mRNA biogenesis.

INTRODUCTION

Heterogeneous nuclear ribonucleoproteins (hnRNP) consist of a group of at least 20 abundant proteins that are primarily found in the eukaryotic cell nucleus associated with pre-mRNA transcripts (1–5). Six of the more abundant hnRNPs (A1, A2, B1, B2, C1 and C2) have the unique ability to package pre-mRNA into a repeating array of 40S ribonucleoprotein particles (1). Of these six 'core' proteins, A1 hnRNP is the best characterized. The amino acid sequence of A1 suggested it contained two domains, the 1–195 region, and the glycine-rich, 196–319 C-terminal domain. Within the N-terminal domain of A1 is a region of internal sequence homology such that when residues 3–93 are aligned with 94–194, 32% of the residues are identical (6). The high degree of conservation of basic and aromatic residues in these ~90 residue regions (62% and 80% respectively), suggested these internal repeats represent independent nucleic acid binding domains (6). Since homologous RNA binding domains (RBDs) were found subsequently in the yeast poly(A) binding protein (7) and then in more than 100 other eukaryotic RNA binding proteins (8), it is clear that the two A1 RBDs provide a prototype for a widely distributed RNA binding motif. This ~90 residue domain has been referred to as the RNP motif RNA binding domain (2,9) or the RNA recognition motif (RRM) (10) and proteins contain one (type C hnRNP) to as many as four [nucleolin and the poly(A) binding protein] of these domains (11–14). Although the extent of sequence identity among RNP motifs is low, they all appear to share a common β1-α1-β2-β3-α2-β4 structure that results in a four-stranded β-sheet 'platform' backed by two helices (15–18).

Each of the isolated A1 RBDs bind nucleic acids (19–22). Although the binding energies of the two A1 RBDs are not additive (14,22), together they contribute ~50% of the free energy of A1 binding to an extended single-strand lattice (22). Under physiological salt concentrations this corresponds to an affinity of only about 5×10^4 M^{-1} for poly r(∈A) (21), which is too low to be detected by most non-equilibrium binding assays. The remaining 50% of the overall A1 binding energy derives from cooperative A1:A1 and direct A1:nucleic acid interactions contributed by the glycine-rich C-terminal domain (20–22).

In addition to its presumed role in pre-mRNA packaging and transport, A1 hnRNP has other activities that might be of biological importance. Both *in vitro* and *in vivo* studies demonstrate A1 has the potential to influence 5′-splice site selection in pre-mRNAs that contain multiple 5′-splice sites (23–25). Several reports (26–28) also demonstrate that A1 promotes renaturation of complementary single-stranded nucleic acids. This strand-annealing activity of A1 is localized in its C-terminal domain

* To whom correspondence should be addressed at present address: W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, PO Box 9812, 295 Congress Avenue, New Haven, CT 06536–0812, USA

(27,28) and is modulated by phosphorylation of serine 198 within this domain (29).

To critically evaluate the full range of possible and probable A1 functions in pre-mRNA biogenesis, it is important that we quantify the ability of A1 to selectively bind high affinity RNA targets and that we elucidate determinants of this specificity. Previous (non-equilibrium) binding studies found that hnRNP A1 exhibits preferential binding to a wide range of targets including splice sites (30–33), the reiterated AUUUA sequences found in the 3′-untranslated region of many labile mRNAs (34) and human telomeric DNA and analogous RNA sequences (32). Recently, Burd and Dreyfuss (33) utilized selection/amplification from pools of random sequence RNA to identify a consensus high affinity A1 binding site, UAGGGA/U, that has some resemblance to consensus sequences for vertebrate 5′- and 3′-splice sites. The highest affinity, 'winner', sequence identified in this study (33) contained a duplication of this binding site separated by two nucleotides. While the high affinity of A1 for the 20mer 'winner' sequence was confirmed recently via the use of a competition fluorescence assay (35), this assay also demonstrated that under equilibrium conditions A1 cannot specifically recognize a β-globin 3′-splice site oligo that a UV cross-linking study suggested represented a high affinity A1 target (31). In the present study we have used this same fluorescence assay to quantify the equilibrium binding affinity of A1 for several other putative and potential high affinity targets. In addition, by quantifying the affinity of A1 for a series of homo-oligomers we have shown it is possible to use nucleotide base compositions to predict with reasonable accuracy the non-sequence specific affinity of A1 for a wide variety of other oligonucleotides. As demonstrated by the studies that follow, comparison of the predicted and observed affinity for A1 provides a valuable criterion to quickly differentiate specific from non-specific binding.

MATERIALS AND METHODS

Purification of A1 hnRNP

A1 was expressed and purified as described (36).

Nucleic acids

Oligonucleotides were synthesized in the HHMI Biopolymer Laboratory/W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University and purified by reverse phase HPLC as described (37). Following reverse phase HPLC, aliquots of all oligos were subjected to anion exchange HPLC and those that contained greater than 10% failure sequences were further purified via preparative anion exchange HPLC. In both cases, final desalting was accomplished via either gel filtration or dialysis. Oligos were quantified via absorbance at 260 nm using the following nucleotide base extinction coefficients ($M^{-1}cm^{-1}$) from Pharmacia LKB Biotech. Inc: rU (9350), rG (10 400), rA (9800), rC (6200), dT (8520), dG (7400), dA (8600), dC (7400), ∈A (3700). Unless otherwise mentioned, oligo concentrations are in terms of phosphate concentration.

Fluorescence spectroscopy

'Forward' competition fluorescence titrations were used to quantify binding affinities. This approach, which has been described (35), was carried out by titrating a fixed concentration of oligo [d(∈A)$_{20}$] with A1 in the presence of an oligonucleotide competitor. Fluorescence titrations were carried out in 2 ml, temperature-regulated and continuously stirred cuvettes on an SLM 8000C spectrofluorometer interfaced to an HP Vectra computer. Three 10 second acquisitions were averaged for each data point. Titrations were performed in duplicate or triplicate with excitation and emission wavelengths of 315 and 400 nm, respectively. Unless otherwise noted titrations were carried out in 10 mM Tris–HCl, pH 7.4, 0.1 M NaCl, 0.1 mM EDTA and 1 mM dithiothreitol. Corrections were made for background fluorescence and dilution effects due to addition of protein. The overall average K_{app} and corresponding standard deviation determined in this study was $1.7\pm0.64\times10^7$ M^{-1}. Thus, we consider differences in affinity of less than 2-fold insignificant.

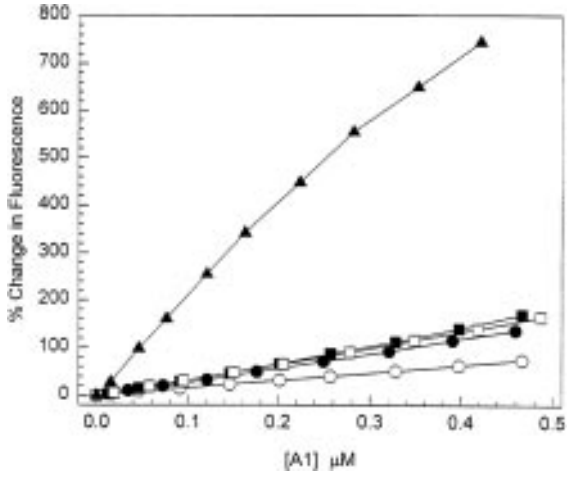

Figure 1. Competition fluorescence assay of A1 hnRNP binding to 20mer RNA oligos containing a human influenza 5′-splice site(oligo 3, open circles, 25 μM), a human V_κ gene intron (oligo 10, filled circles, 30 μM), a β-globin 3′-splice site (oligo 4, open squares, 20 μM) and a random oligo synthesized with equal amounts of each nucleotide base at each position (oligo 20, filled squares, 25 μM). As described in Materials and Methods, a fixed concentration (1.0μM) of oligo d(∈A)$_{20}$, which provides the fluorescence signal that is monitored, and the oligo being examined are titrated with increasing amounts of A1. The titration carried out in the absence of competing oligo is indicated by the curve containing the filled triangles. The resulting affinities are listed in Table 2.

An occluded site size (*n*) of 20 was assumed for A1 in 100 mM NaCl [as determined in figure 4 of Nadler *et al* (21)]. To calculate the K_{app} for the competing oligonucleotide it was assumed that the free A1 concentration is the same at the same extent of fluorescence enhancement of oligo [d(∈A)$_{20}$] both in the absence and presence of the oligo competitor. This assumption allows for the determination of the apparent affinity of the competing oligonucleotide (K_{comp}) using the expression:

$$K_{comp} = [comp]_{bound} \times \{oligo\ [d(\in A)_{20}]_{free}\} \times K_{oligo[d(A)20]} / \{[comp]_{free} \times oligo\ [d(\in A_{20})]_{bound}\}$$

In this expression, the $[comp]_{bound}$ is calculated from the difference between $[protein]_{total}$ and the sum of $[protein]_{free}$ and oligo $[d(\in A)_{20}]_{bound}$ (calculated in terms of oligo concentration bound at that point in the titration). The $[protein]_{free}$ at that point of fluorescence enhancement is obtained from an identical titration carried out in the absence of competitor and the apparent affinity of A1 for oligo d(∈A)$_{20}$ was determined by analysis of double reciprocal plots. The oligo [d(∈A)$_{20}$] concentration was

1 μM (phosphate) and the competing oligo concentrations ranged from 10 to 70 μM (phosphate) as indicated in the figure legends. As expected, the apparent affinity for the competing oligo was independent of its concentration.

RESULTS

Binding of hnRNP A1 to naturally occurring RNA sequences

As shown in Table 1, 12 different RNA sequences (oligos 1–12) have been selected to better document the range of affinities of A1 for naturally occurring RNAs. In terms of high affinity A1 targets, these include oligos containing 5′-splice site (oligo 1) and 3′-splice site (oligos 4, 5 and 9) sequences reported to bind preferentially to A1 (30,31,33) as well as the high affinity, AUUUA-rich sequence from the 3′-untranslated region (UTR) of a short lived mRNA (oligo 12 in ref. 34). Control oligos include five other splice site sequences whose base composition ranges from A-rich (oligos 2 and 3) to C- (oligo 8) and U-rich (oligo 7) and whose splice site position ranges from being near the 5′- (oligo 2) to near the 3′-end (oligos 5 and 9) of the oligo. In addition, other controls include two intron sequences (oligos 10 and 11), one of which (oligo 11) had been reported to bind A1 with low affinity (33). Figure 1 illustrates representative data obtained from fluorescence competition assays and Table 2 gives the corresponding apparent binding affinities. Qualitatively, the titrations shown in Figure 1 suggest there is relatively little difference in the ability of the 5′- and 3′-splice site sequences on the one hand and an intron derived sequence and a random oligo [synthesized with equal amounts of each nucleotide base at each position (oligo 20)] on the other hand to compete with the oligo $d(\in A)_{20}$ probe for binding to A1. Overall, the apparent affinities determined for naturally occurring oligo RNA sequences ranged from a low of ~1.1×10^6 M^{-1} for a 19mer (oligo 9), corresponding to a β-globin 3′-splice site, to a high of ~4.2×10^7 M^{-1} for a 20mer (oligo 1), corresponding to a β-globin 5′-splice site. For those four oligos (see oligos 1, 4, 11 and 18 in Table 2) whose affinities had previously been estimated by a non-equilibrium filter binding assay (33), there was good agreement generally between the two sets of data (compare last two columns in Table 2). Hence, in all four instances shown in Table 2 there was less than 5-fold difference between affinities estimated by these two approaches and there did not seem to be any consistent error in that some previously reported affinities (ie., oligo 4 and 18) were higher and one (ie., oligo 1) was lower than found in the present work. The average affinity determined in this study for these four oligos was 1.1×10^8 M^{-1} as compared with the value of 2.6×10^8 M^{-1} reported previously (33). Thus, these two very different approaches for evaluating apparent binding affinities are in reasonably good agreement.

Table 1. Oligonucleotides used for fluorescence binding studies with A1 hnRNP

No.	Description	Sequence[a]	Ref.
1	β-globin 5′-splice site	CCCUGGGCAG/GUUGGUAUCA	33
2	Ad2 E1a 5′-splice site	UACA/GUAAGUGAAAAUUAUG	38
3	Human influenza (segment 7) 5′-splice site	AAGCAG/GUAGAUAUUGAAAG	38
4	β-globin 3′-splice site	CCACCCUUAG/GCUGCUGGUG	33
5	Adenovirus 3′-splice site	GUCCCUUUUUUUUCCACAG/C	30
6	Human V^{κ} gene 3′-splice site	UAUUUCCAAUCUCAG/GUGCC	38
7	Ad2 E1a 3′-splice site	UGAUUUUUUUAAAAG/GUCCU	38
8	Chicken lysozyme 3′-splice site	UCUCCCUCCGCCCAG/GGUCG	38
9	β-globin 3′-splice site	UCUAUUUUCCCACCCUUAG/	35
10	Human V^{κ} gene intron	AAUUUACUCAGCCCAGUGUG	39
11	β-globin intron	GAUCACUUGUGUCAACACAG	33
12	AUUUA sequence from 3′-untranslated region of a short lived GM-CSF mRNA	CAUUUAUUUAUUUAUUUAAG	34
13	Human telomeric DNA	TTAGGGTTAGGGTTAGGGTTAGGG	32
14	RNA analogue of oligo 13	UUAGGGUUAGGGUUAGGGUUAGGG	32
15	Oligo 14 with U to C mutation at first position	CUAGGGCUAGGGCUAGGGCUAGGG	32
16	Oligo 14 with U to C mutation at second position	UCAGGGUCAGGGUCAGGGUCAGGG	32
17	Oligo 14 with A to G mutation at third position	UUGGGGUUGGGGUUGGGGUUGGGG	32
18	'Winner' high affinity sequence	UAUGAUAGGGACUUAGGGUG	33
19	Randomized winner sequence no. 1	UGCUGAUGUUGAUGAGAGAG	35
20	Random oligo	$(rX)_{20}$	this work
21	Oligo G	$(rG)_{20}$	this work
22	Oligo A	$(rA)_{20}$	this work
23	Oligo C	$(rC)_{20}$	this work
24	Oligo U	$(rU)_{20}$	this work

[a]The slash mark indicates the location of 3′- and 5′-splice sites.

Table 2. Decreasing order of affinity of different classes of oligonucleotides for A1 hnRNP

Oligo no.	Description	Splice position	K_{app} (0.1 M NaCl)	
			Burd and Dreyfuss (ref. 33)	This work[a]
18	'Winner' high affinity sequence		1×10^9	(4.0×10^8)[b]
1	β-Globin 5′-splice site	10–11	1.4×10^7	4.2×10^7
3	Human influenza (segment 7) 5′-splice site	6–7		3.2×10^7
2	Ad2 E1a 5′-splice site	4–5		1.7×10^7
	5′-splice site average K_{app}			2.7×10^7
13	Human telomeric DNA			2.6×10^7
12	AUUUA sequence from a short lived GM-CSF mRNA			2.5×10^7
14	RNA analogue of oligo 13			2.3×10^7
16	Oligo 14 with U to C mutation at position 2 in repeat			2.2×10^7
17	Oligo 14 with A to G mutation at position 3 in repeat			2.2×10^7
15	Oligo 14 with U to C mutation at position 1 in repeat			1.9×10^7
	RNA telomere average K_{app}			2.1×10^7
21	$(rG)_{20}$			4.8×10^7
22	$(rA)_{20}$			1.8×10^7
24	$(rU)_{20}$			3.8×10^6
23	$(rC)_{20}$			2.0×10^6
	Homo-oligo average K_{app}			8.8×10^6
10	Human V^{κ} gene intron			1.0×10^7
11	β-Globin intron		3.3×10^6	4.2×10^6
	Intron average K_{app}			6.4×10^6
7	Ad2 Ela 3′-splice site	15–16		1.1×10^7
8	Chicken lysozyme 3′-splice site	15–16		1.1×10^7
6	Human V^{κ} gene 3′-splice site	15–16		9.0×10^6
5	Adenovirus 3′-splice site	19–20		6.1×10^6
4	β-Globin 3′-splice site	10–11	1.4×10^7	3.2×10^6
9	β-Globin 3′-splice site	19		1.1×10^6
	3′-splice site average K_{app}			5.2×10^6
19	Randomized winner sequence 1			5.0×10^6
20	Random oligo			4.9×10^6
	Random sequence average K_{app}			5.0×10^6

[a]Average affinities for groups of oligonucleoties were determined from the average free energies of binding.
[b]Because the affinity of the 'winner' sequence is too high in 0.1 M NaCl to accurately measure via fluorescence, it was calculated from the salt sensitivity data in Table 4 of Abdul-Manan *et al.* (35).

Although A1 hnRNP appeared to have slightly higher (ie., about 4-fold) affinity for the 5′-splice site and AUUUA-rich sequences examined than for intron and 3′-splice site sequences, less than two standard deviations separate the average affinities for the 5′- and 3′-splice site and intron sequences listed in Table 2. In addition, since the affinities of two of the 3′-splice site (oligos 7 and 8) and one of the intron (oligo 10) sequences examined are not significantly less than that for the 5′-splice site sequence contained in oligo 2, the most reasonable conclusion from the data in Table 2 is there is generally not any significant difference between the affinity of A1 hnRNP for 5′- or 3′-splice sites on the one hand versus that for intron sequences on the other. As previously (33,35), the 'winner' sequence identified by selection/amplification appeared to be in a class by itself in that its affinity for A1 was 10-fold higher than that for any other oligo examined (Table 2).

Binding of A1 hnRNP to human telomeric DNA and to its RNA analogue: effect of point mutations in the RNA analogue

As shown in Figure 2, A1 binds tightly and with approximately equal affinity to both the $(TTAGGG)_4$ telomeric DNA sequence and its $(UUAGGG)_4$ RNA analogue. In contrast to the finding of Ishikawa *et al.* (32), substitution of any of the first three bases in the repeating UUAGGG sequence did not significantly decrease the apparent binding affinity (Fig. 2). Since Ishikawa *et al.* (32) used a non-equilibrium binding assay and carried out their studies with a partially purified fraction containing proteolytic fragments of the A1 and A2/B1 hnRNP proteins (as well as at least one other 55 kDa protein), it is probable that one of these factors accounts for the different results obtained in this earlier study.

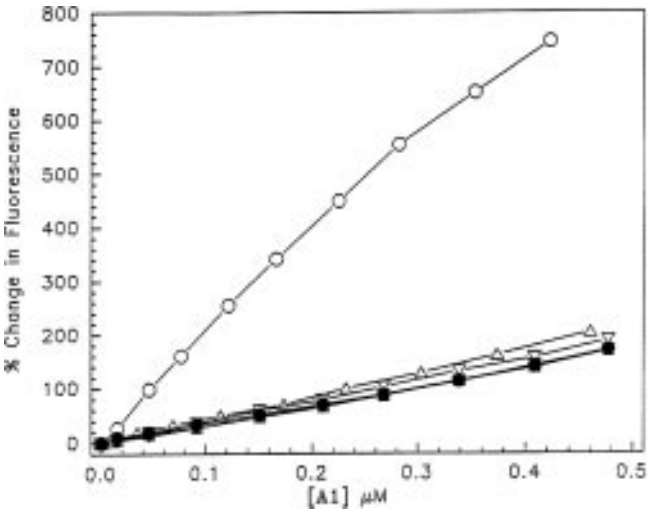

Figure 2. Competition fluorescence assay of A1 hnRNP binding to 20mer oligos containing a DNA telomere sequence (oligo 13, open triangles) and four RNA analogues of this sequence, with all being present at a concentration of 10 μM. Three of these oligos (oligos 15–17) carried single mutations in each of the four internal sequence repeats. While the curve for the RNA analogue with a U→C mutation in position 1 (oligo 15) can be seen above (open inverted triangles), the remaining three curves for oligos 14, 16 and 17) are nearly coincident. The curve with the open circles shows a titration for the oligo $d(\in A)_{20}$ probe (1.0 μM) in the absence of competitor.

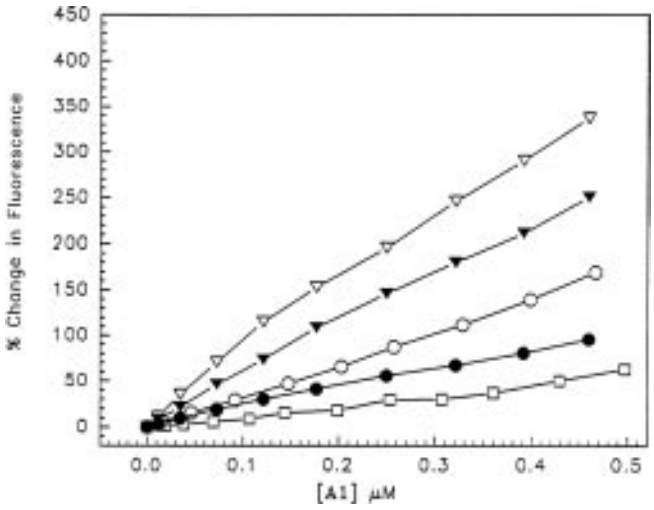

Figure 3. Competition fluorescence assay of A1 hnRNP binding to homo-oligo 20mers and to a random 20mer synthesized to contain equal fractions of each of the four nucleotide bases at each position. Details concerning the assay may be found in the legend for Figure 1 and in Materials and Methods. The following oligos were used in this study: $r(C)_{20}$, (open inverted triangles); $r(U)_{20}$, (closed inverted triangles); random oligo, (open circles); $r(A)_{20}$, (filled circles); and $r(G)_{20}$, open squares. With the exception of $r(G)_{20}$, which was present at 20 μM, the concentration of the other competing oligos was 25 μM. Note that this graph has been expanded somewhat (in comparison with Figures 1 and 2) to better illustrate the relatively small differences in affinity seen for these oligonucleotides.

Binding of A1 hnRNP to homo-oligonucleotides: effect of base composition on affinity

To determine whether the ~10-fold difference in affinities observed in Table 2 for naturally occurring 20mers (ie., the observed affinities varied from 3.2×10^6 M^{-1} for a β-globin 3′-splice site (oligo 4) to 4.2×10^7 M^{-1} for a β-globin 5′-splice site (oligo 1), might reflect differences in base specificity rather than in sequence specificity *per se*, binding affinities were determined for all four homo-ribo-oligonucleotides and for an oligo RNA synthesized to contain equal ratios of each of the four bases at each position. As shown in Figure 3, the following order of affinities was observed for these 20mers: G>A>U>C and, as expected, the random 20mer had intermediate affinity. This order of relative affinities is similar to the U>G>A>C order observed by Swanson and Dreyfuss (30) except for the relative affinity of U. In the previous study (30) binding was monitored under non-equilibrium conditions using Sepharose-immobilized polynucleotides and the order of homo-polynucleotide binding affinities was based on the relative salt sensitivity of the respective A1:polynucleotide complexes. Since the present study was carried out under equilibrium conditions on 20mers and it measured the absolute A1 affinities at a given salt concentration (ie., 0.1 M), there are several possible explanations (in addition to those resulting from immobilization of the polynucleotides onto a solid support) for the difference in the relative affinities of U-containing oligo/polynucleotides. For instance, the inability of U-containing polynucleotides to assume significant higher order structure in solution might especially favor cooperative A1 binding to this particular polynucleotide whereas this effect would not be seen in the present study which was limited to oligos that were sufficiently short to preclude cooperative A1 binding (21).

Since oligo $(rA)_{20}$ and $(rG)_{20}$ have affinities for A1 that are nearly 10-fold higher than that of oligo $(rU)_{20}$ or $(rC)_{20}$, A1 should generally bind more tightly to A/G-rich oligos. Indeed, this appears to be the case as demonstrated by the direct correlation observed in Table 3 between A/G content and binding affinity. Hence, those six oligos that had the lowest average affinity (4.1×10^6 M^{-1}) also had the lowest A/G content (41%). Similarly, the eight oligos with the highest average affinity (2.7×10^7 M^{-1} in Table 3) also had the highest A/G content (61%). Qualitatively consistent with the high affinity of the winner 20mer sequence is the high A/G (65%) content of this oligonucleotide (Table 3).

Use of homo-oligonucleotide binding data to estimate A1 binding affinities for other oligonucleotides

If the higher order structures that may be assumed by some homo-oligonucleotides [eg., oligo G is known to form tetrad-like structures (40)] do not substantially alter affinity for A1, then it should be possible to use the affinities of the individual homo-oligonucleotides to calculate the 'non-specific' affinity of any other oligonucleotide. To do this we have calculated the free energy of binding of each of the four homo-oligonucleotides and then used this data to calculate the free energy of A1 binding to each of the four nucleotide bases. This leads to the following equation which can be used to estimate the non-specific free energy of A1 binding (kcal) to any 20mer oligonucleotide at 25°C:

$$-\Delta G = 0.53\,(\#G) + 0.50\,(\#A) + 0.43\,(\#C) + 0.45\,(\#U)$$

The predicted A1 affinities shown in the last column of Table 3 were calculated from the predicted free energy changes as determined from the above relationship. In general, there is good agreement between the predicted and observed affinities. With the exception of the 'winner' oligo, all other predicted and observed affinities are within 4-fold of each other and hence appear to result primarily from non-specific binding of A1. Assuming differences in affinity that are less than 2-fold are not significant (see Materials and Methods), 75% of the predicted affinities in Table 3 are within experimental error of the observed values.

Table 3. Influence of oligonucleotide base composition on affinity for A1

Oligo no.	No. res	GGG	Base composition					Observed	Predicted
		Sequence?	A	G	C	U	AG%	K_{app}	K_{app}[a]
18	20	yes	5	8	1	6	65%	4.0×10^8	1.5×10^7
1	20	yes	3	7	5	5		4.2×10^7	1.0×10^7
3	20	no	9	6	1	4		3.2×10^7	1.7×10^7
12	20	no	6	1	1	12		2.5×10^7	6.9×10^6
14	24	yes	4	12	0	8		2.3×10^7	1.6×10^7
16	24	yes	4	12	4	4		2.2×10^7	1.7×10^7
17	24	yes	0	16	0	8		2.2×10^7	2.3×10^7
Average			4.3	9.0	1.8	6.8	61%	2.6×10^7	1.4×10^7
15	24	yes	4	10	3	3		1.9×10^7	1.7×10^7
2	20	no	9	4	1	6		1.7×10^7	1.3×10^7
7	20	no	5	3	2	10		1.1×10^7	7.9×10^6
8	20	yes	1	5	10	4		1.1×10^7	5.6×10^6
10	20	no	5	4	5	6		1.0×10^7	8.1×10^6
6	20	no	4	3	6	7		9.0×10^6	6.4×10^6
Average			4.7	4.8	4.5	6.0	48%	1.2×10^7	8.9×10^6
5	20	no	2	2	7	9		6.1×10^6	4.6×10^6
19	20	no	5	8	1	6		5.0×10^6	1.5×10^7
20	20	no	5	5	5	5		4.9×10^6	9.2×10^6
11	20	no	6	4	5	5		4.2×10^6	8.8×10^6
4	20	no	2	6	7	5		3.2×10^6	7.6×10^6
9	19	no	3	1	7	8		1.1×10^6	2.1×10^6
Average			3.8	4.3	5.3	6.3	41%	3.6×10^6	6.7×10^6

[a]Predicted affinities were calculated from the estimated free energy of A1 binding to each of the component nucleotide bases as described in the Results. For those telomeric analogues that were 24mers, the two bases at the 5′- and 3′-end of each oligo were not considered when deriving the predicted affinities. Average affinities for each group of oligos were calculated from the respective average free energies of binding.

DISCUSSION

Since all of the functions so far ascribed to A1 hnRNP can be explained in terms of its affinity for nucleic acids, understanding the possible and probable functions of A1 requires that its ability to differentiate specific from non-specific RNA targets be quantified. For instance, the range of possible functions for A1 in splicing is greater if A1 can specifically bind splice sites than if it binds 'non-specifically' to any 20mer pre-mRNA sequence that is available. As shown in Table 2, we found that (in 0.1 M NaCl) A1 has only an ~10-fold range in affinities for the 'naturally occurring' 20mer RNA sequences examined in this study. Since the average affinity of A1 for six, 3′-splice sites was the same (~5×10^6 M^{-1}) as that for a random mixture of 20mer oligos that was synthesized with equal fractions of each of the four bases at each position, it is unlikely A1 can function in any activity that requires that it generally recognize 3′-splice sites. If we assume A1 has an affinity of ~5×10^6 M^{-1} for an 'average' 20mer of average base composition, the data in Table 2 suggest A1 might have 5-fold higher affinity for an average 5′-splice site. The question then becomes whether such a difference, which only represents about two standard deviations (see Materials and Methods), might be physiologically important. To help put this in some perspective, if we assume an average pre-mRNA is 8 kb (4), that it contains 10 introns and that A1 has an occluded binding site of 20 nucleotides (21) then the ratio of potential non-5′-splice/5′-splice sites is about 40. Hence, for A1 to have equal probability of binding to a 5′-splice site as opposed to some other pre-mRNA site it would need to have a 40-fold preference for 5′-splice sites, which is a degree of preference not supported by the data in Table 2. Although our analysis is naive in that it ignores the overlapping A1 binding sites available in non-5′-splice site regions of pre-mRNA and it does not take into account either the higher order structure that may be present in pre-mRNA nor the possibility of competition with other pre-mRNA binding proteins, it nonetheless strongly suggests that by itself, A1 cannot specifically target 5′-splice sites. Similar logic leads to the conclusion that A1 cannot specifically recognize the AUUUA sequences found in the 3′-untranslated regions of many short lived mRNAs, telomeric DNA and analogous RNA sequences and, as previously suggested by studies on a β-globin 3′-splice site (35), A1 also cannot specifically recognize 3′-splice sites. The inability of A1 to recognize these putative high affinity targets is also consistent with their sequences sharing only 2–3 nucleotides (out of a total of 20–24) in common (Table 4). Although both the *in vitro* selected winner sequence and the RNA analogue of the human telomeric DNA sequence listed in Table 4 share two copies of the UAGGGA/U consensus high affinity A1 binding site (33) in common, overall, the sequences listed in this table only share the AG(G) sequence. A sequence of only 2–3 nucleotides would seem to be too short and too widely distributed to permit specific recognition by any protein. One interesting finding with respect to telomeric sequences is that A1 has similar affinities for both telomeric DNA and analogous RNA sequences (Table 2) which suggests A1 might be able to play a role in telomeric DNA metabolism—providing that role requires only that A1 bind, as opposed to specifically target DNA telomeres.

We hypothesize that many previous reports dealing with the identification of high affinity binding sites for A1 resulted from the non-equilibrium techniques that were employed. Hence,

Table 4. Putative high affinity targets for hnRNP A1 share little direct sequence homology

Description	Sequence	Ref.
AUUUA Repeat from 3'-UTR OF mRNA	C A U U U A U U U A U U U A U U U A A G	34
In vitro selected "winner" sequence	U A U G A U A G G G A C U U A G G G U G	33
β-Globin 5'-splice site	C C C U G G G C A G G U U G G U A U C A	33
β-Globin 3'-splice site	C C A C C C U U A G G C U G C U G G U G	33
DNA Analogue of β-Globin 3'-splice site	T C T A T T T T C C C A C C C T T A G G T	31
Human Telomeric DNA Sequence	T T A G G G T T A G G G T T A G G G T T A G G G	32
RNA Analogue of Human Telomeric Sequence	U U A G G G U U A G G G U U A G G G U U A G G G	32

LeStourgeon *et al.* (1) offered an alternative interpretation of a ribonuclease T1 digestion/immunopurification study that found high affinity binding of A1 to the 3′-end of introns (30). Based on their analysis (1), the earlier data of Swanson and Dreyfuss (30) actually argued in favor of A1 not being able to recognize the 3′-end of introns, which is thus in agreement with the data in Table 2 and 3. Similar concerns may be raised with the use of UV cross-linking to detect specific binding of A1 to AUUUA-rich regions in the 3′-UTRs of short lived mRNAs (34). Firstly, A1 cross-links extremely well—with the multiple A1 sites of cross-linking (19) perhaps accounting for the fact that a fragment corresponding to the N-terminal two-thirds of A1 cross-links with 85% efficiency — as compared with less than 30% for three other single-strand binding proteins (41). Secondly, there is often a several order of magnitude range in how well different nucleic acid bases cross-link to proteins. Hence, adenine does not cross-link to A1 hnRNP (19) and both adenine and guanine cross-link poorly (if at all) to two other single strand binding proteins (41,42). In contrast, uridine cross-links extremely well to several nucleic acid binding proteins (42–44). While the cross-linking study of Hamilton *et al.* (34) and Table 2 in this work certainly indicates that A1 binds AUUUA-rich regions, we suggest it is quite possible that other protein(s) may bind more tightly and more specifically than A1 does but have eluded detection by Hamilton *et al.* (34) due to their lower cross-linking efficiency. As a precedent, we note that even though many of the more than 50 proteins that make up the *Escherichia coli* ribosome contact rRNA, irradiation of this complex with UV light cross-links only two of these proteins in reasonable yield to rRNA (43,44).

Despite the relatively small difference in affinity A1 has for the naturally occurring oligo sequences given in Table 1, this difference appears to correlate with A1 base specificity as determined with homo-oligonucleotides. Thus, those oligos with the highest purine content generally have the highest affinity for A1. This finding allows for the reasonably accurate prediction of non-specific affinities for 20mer oligos and again, points out the fact that based on the oligos examined, the 20mer 'winner' sequence identified by selection/amplification is unique in terms of its unusually high affinity for A1. Hence, its observed affinity is an order of magnitude above that for any other oligo examined in Table 3 and is 25-fold above the affinity predicted on the basis of the homo-oligonucleotide studies. Some of this increased affinity results from the ability of the winner sequence to form a higher order aggregate in solution that is probably a G-quartet (35). In fact, G-tetrad formation could well account for the ability of A1 to select RNA sequences that contain either one or two UAGGGA/U consensus high affinity A1 binding sites. Based on our previous work (35) we believe that the 20mer winner sequence forms an antiparallel tetra-G like structure that involves the interaction of two, 20mer winner sequences that are each in the form of hairpins (see figure 8 in ref. 35). In the case of those selected RNAs that contain only a single UAGGGA/U sequence, we postulate an analogous G-quartet structure forms from the parallel interaction of four RNA molecules. Regardless of whether the G-tetrad is formed from the interaction of two hairpins or from the interaction of four individual RNAs, in both cases there would be an identical stacking of three G-quartets. Hence, A1 might be expected to select RNAs that contain either one or two UAGGGA/U sequences as indeed was the case (33).

Although the ability of the winner sequence to form a higher order aggregate is an important determinant of its high affinity for A1, binding studies on the (anion exchange isolated) monomeric winner sequence (35) and the affinity reported in Table 2 for oligo $(rG)_{20}$ both suggest there is also an important sequence-specific component of A1 binding to the winner sequence. That is, since the affinity of A1 for $(rG)_{20}$, which forms G-tetrads in solution (40), is only 3-fold above that for $(rA)_{20}$, which does not form a corresponding structure in solution, G-tetrad formation by itself does not seem to be able to account for the unusually high affinity of the 'winner' sequence. Although the molecular basis for recognition of the 20mer winner sequence by A1 is intriguing, its physiological significance remains uncertain as database searches failed to show any apparent pattern to the location of these sequences within the context of pre-mRNAs (33).

In summary, with the exception of the 20mer winner oligo containing two copies of the consensus UAGGGA/U high affinity binding site, A1 appears to have a relatively narrow range of affinities for naturally occurring RNA oligo sequences examined and this range can be accounted for almost entirely in terms of nucleic acid base rather than sequence specificity. Based on the data in Table 3, nucleotide base compositions may be used to predict reasonably accurate A1 affinities for virtually all other oligonucleotide sequences that do not contain the UAGGGA/U high affinity binding site. This finding should greatly facilitate studies directed at understanding the range of possible and probable functions for this interesting hnRNP protein. With regard to the latter, the high abundance of A1 *in vivo* [ie., Kiledjian *et al.* (4), estimate there are 7×10^7 molecules of both A1 and core histones per HeLa cell] and its ability to bind promiscuously to oligo RNA sequences are both consistent with

A1 playing a structural role in pre-mRNA biogenesis and serving as a pre-mRNA 'chaperone' as postulated by Herschlag (45).

The techniques and approaches used in this and in a previous study (35), including the use of fluorescence to determine equilibrium binding affinities and the use of homo-oligonucleotides and sequence-randomization to discriminate nucleotide base from sequence and/or structure specificity, are applicable to virtually all protein–nucleic acid interactions and serve to illustrate limitations that are inherent in some non-equilibrium binding assays. In this regard, however, we note there is generally good agreement between the apparent affinities reported in Table 2 for four oligos whose affinities had also previously been determined by a non-equilibrium filter binding assay (33).

ACKNOWLEDGEMENTS

We especially thank the staff of the Oligonucleotide Synthesis Section of the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for synthesizing the custom oligonucleotides used in this work. This study was supported by NIH grant GM31539, NSF grant MCB9514179 and by a Special Stipend from the Yale University School of Medicine to KRW.

REFERENCES

1 Le Stourgeon, W., Barnett, S.F. and Northington, S.J.(1990) in Strauss, P. and Wilson, S. (eds) *The Eukaryotic Nucleus, Molecular Biochemistry and Macromolecular Assemblies.* Telford Press, London Vol 2. p 477.
2 Dreyfuss, G., Matunis, M.J., Pinol Roma, S. and Burd, C.G. (1993) *Annu. Rev. Biochem.* **269,** 289–321.
3 Keene, J.D. and Query, C.C. (1991) in Cohn, W. and Moldave, K. (eds) *Progress in Nucleic Acid Research and Molecular Biology.* Academic Press, Vol 41, pp. 179–202.
4 Kiledjian, M., Burd, C.G., Görlach, M., Portman, D.S. and Dreyfuss, G. (1994) in Mattaj, I. and Nagai, K. (eds) *Protein–RNA Interactions-Frontiers in Molecular Biology.* Oxford University Press, Oxford, pp. 127–149.
5 Piñol-Roma, S., Choi, Y.D., Matunis, M.J. and Dreyfuss, G. (1988) *Genes Dev.* **2,** 215–227.
6 Merrill, B.M., LoPresti, M.B., Stone, K.L. and Williams, K.R. (1986) *J. Biol. Chem.* **261,** 878–883.
7 Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6,** 2932–2943.
8 Birney, E., Kumar, S. and Krainer, A.R. (1993) *Nucleic Acids Res.* **21,** 5803–5816.
9 Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) *Trends Biochem. Sci.* **13,** 6–91.
10 Query, C.C., Bentley, R.C. and Keene, J.D. (1989) *Cell* **57,** 89–101.
11 Bandzilius, R.J., Swanson, M.S. and Dreyfuss, G. (1989) *Genes Dev.* **3,** 431–437.
12 Merrill, B.M. and Williams, K.R. (1990) In Strauss, P. and Wilson, S. (eds) *The Eukaryotic Nucleus, Molecular Biochemistry and Macromolecular Assemblies.* Telford Press, London Vol. 2, pp. 579–604.
13 Kenan, D.J., Query, C.C. and Keene, J.D. (1991) *Trends Biochem. Sci.* **16,** 214–220.
14 Shamoo, A.Y., Abdul-Manan, N. and Williams, K.R. (1995) *Nucleic Acids Res.* **23,** 725–728.
15 Nagai, K. Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) *Nature* **348,** 515–520.
16 Hoffman, D.W., Query, C.C., Golden, B.L., White, S.W. and Keene, J.D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2495–2499.
17 Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. and Mueller, L. (1992) *Biochemistry* **31,** 6254–6265.
18 Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. and Nagai, K. (1995) *Nature* **372,** 42–438.
19 Merrill, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H. and Williams, K.R. (1988) *J. Biol. Chem.* **263,** 3307–3313.
20 Kumar, A., Casas-Finet, J.R., Luneau, C.J., Karpel, R.L., Merrill, B.M., Wiliams, K.R. and Wilson, S.H. (1990) *J. Biol. Chem.* **265,** 17094–17100.
21 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) *Biochemistry* **30,** 2968–2976.
22 Shamoo, Y., Abdul-Manan, N. Patten, A.M., Crawford, J.K, Pellegrini, M.C. and Williams, K.R. (1994) *Biochemistry* **33,** 8272–8281.
23 Mayeda, A. and Krainer, A.R. (1992) *Cell* **68,** 365–375.
24 Cáceras, J., Stamm, S., Helfman, D.M. and Krainer, A.R. (1994) *Science* **265,** 1706–1709.
25 Yang, X., Bani, M-R., Lu, S-J., Rowan, S., Ben-David, Y. and Chabot, B. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6924–6928.
26 Pontius, B.W. and Berg, J.M. (1992) *J. Biol. Chem.* **267,** 13815–13818.
27 Kumar, A. and Wilson, S.H. (1990) *Biochemistry* **29,** 10717–10722.
28 Munroe, S.H. and Dong, X.F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 895–899.
29 Cobianchi, F., Calvo, C., Stoppini, M., Buvoli, M. and Riva, S. (1993) *Nucleic Acids Res.* **21,** 949–955.
30 Swanson, M.S. and Dreyfuss, G. (1988) *EMBO J.* **11,** 3519–3529.
31 Buvoli, M., Cobianchi, F., Biamonti, G. and Riva, S. (1990) *Nucleic Acids. Res.* **18,** 6595–6600.
32 Ishikawa, F., Matunis, M.J., Dreyfuee, G. and Cech, T.R. (1993) *Mol. Cell. Biol.* **13,** 4301–4310.
33 Burd, C.G. and Dreyfuss, G. (1994) *EMBO J.* **13,** 1197–1204.
34 Hamilton, B.J., Nagy, E., Malter, J.S., Arrick, B.A. and Rigby, W.F.C. (1993) *J. Biol. Chem.* **268,** 8881–8887.
35 Abdul-Manan, N., O'Malley, S.M. and Williams, K.R. (1996) *Biochemistry* **35**, 3545–3554.
36 Cobianchi, R., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1988) *J. Biol. Chem.* **263,** 1063–1071.
37 Webster, K.R., Shamoo, A.Y., Konigsberg, W. and Spicer, E. (1991) *Biotechnology* **11,** 658–661.
38 Mount, S.M. (1982) *Nucleic Acids Res.* **10,** 459–473.
39 Bentley, D. and Rabbits, T., (1980) *Nature* **288**, 730–733.
40 Pearson, A.M., Rich, A. and Krieger, M. (1993) *J. Biol. Chem.* **268**, 354–355.
41 Williams, K.R. and Konigsberg, W.H. (1991) In Sauer, R. (ed.) *Methods in Enzymology.* Academic Press, Inc., New York Vol 208, pp 516–539.
42 Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R. and von Hippel, P.H. (1986) *J. Biol. Chem.* **261,** 3512–3518.
43 Zweib, C. and Brimacombe, R. (1979) *Nucleic Acids Res.* **6,** 1775–1790.
44 Maly, P., Rinke, J., Ulmer, E., Zwieb, C. and Brimacombe, R. (1980) *Biochemistry* **19,** 4179–4188.
45 Herschlag, D. (1995) *J. Biol. Chem.* **36**, 20871–20874.