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†

Sangduk Kim,‡ Barbara M. Merrill,§,|| Ramesh Rajpurohit,‡ Amalendra Kumar,⊥,# Kathryn L. Stone,§,▽ Vladimir V. Papov,§ Jennifer M. Schneiders,§ Wlodzimierz Szer,⊥ Samuel H. Wilson,○ Woon Ki Paik,‡ and Kenneth R. Williams*,§,▽

Fels Institute for Cancer Research and Molecular Biology, Temple University School of Medicine, Philadelphia, Pennsylvania 19140, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510, Department of Biochemistry, New York University Medical Center, New York, New York 10016, Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, Texas 77555, and Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06536

Received October 10, 1996; Revised Manuscript Received January 31, 1997⊗

ABSTRACT: Three sites of N^G,N^G-arginine methylation have been located at residues 205, 217, and 224 in the glycine-rich, COOH-terminal one-third of the HeLa A1 heterogeneous ribonucleoprotein. Together with the previously determined dimethylated arginine at position 193 [Williams et al., (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5666−5670], it is evident that all four sites fall within a span of sequence between residues 190 and 233 that contains multiple Arg-Gly-(Gly) sequences interspersed with phenylalanine residues. These RGG boxes have been postulated to represent an RNA binding motif [Kiledjian and Dreyfuss (1992) *EMBO J. 11*, 2655−2664]. Dimethylation of HeLa A1 appears to be quantitative at each of the four positions. Arginines 205 and 224 have been methylated *in vitro* by a nuclear protein arginine methyltransferase using recombinant (unmethylated) A1 as substrate. This suggests A1 may be an *in vivo* substrate for this enzyme. Examination of sequences surrounding the sites of methylation in A1 along with a compilation from the literature of sites that have been identified in other nuclear RNA binding proteins suggests a methylase-preferred recognition sequence of Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe, with the COOH-terminal flanking glycine being obligatory. Taken together with data in the literature, identification of the sites of A1 arginine methylation strongly suggests a role for this modification in modulating the interaction of A1 with nucleic acids.

Several naturally occurring proteins contain unusual guanidino (N^G) methylated arginine derivatives, N^G-monomethylarginine, N^G,N^G-(asymmetric)-dimethylarginine, and N^G,$N^{G'}$-(symmetric)-dimethylarginine, whose formation is catalyzed by protein-arginine methyltransferase (*S*-adenosylmethionine:protein-arginine *N*-methyltransferase; EC 2.1.1.23; protein methylase I) (Kim et al., 1990; Paik & Kim, 1985). These post-translational modifications have been found in nucleoplasmic, nucleolar, and cytoplasmic proteins, e.g., hnRNP[1] proteins, nucleolin, fibrillarin, HMG chromosomal protein, histones, myelin basic protein, heat shock proteins, tooth matrix protein, ribosomal protein, and myosin (Kim et al., 1990; Paik & Kim, 1985, 1993). They are particularly prevalent in hnRNP proteins, which contain about 65% of the total N^G,N^G-dimethylarginine (Dma) in the cell nucleus (Boffa et al., 1977). In addition to the type A and B hnRNP proteins (Beyer et al., 1977; Wilk et al., 1985), which are major components of 40S core particles (Beyer et al., 1977), at least 10 other hnRNP proteins also contain Dma (Liu & Dreyfuss, 1995). So far, only one hnRNP protein methylation site has been reported and that is at arginine 193 in the NH_2-terminal proteolysis fragment of the A1 hnRNP protein (Williams et al., 1985). Since only 12% of the arginine residues in hnRNP proteins are methylated (Boffa et al., 1977), the requirements for substrate recognition by protein-arginine methyltransferase must be rather stringent. Utilizing several synthetic peptides whose sequences are homologous to the arginines that are methylated in A1 hnRNP and other proteins, we have begun to systematically investigate the structural requirements for substrate peptides. Our initial studies indicate that the shortest peptide that can be recognized is a hexamer, that a COOH-terminal flanking glycine is obligatory, and that an NH_2-terminal flanking glycine is strongly favored (Ghosh et al., 1990; Rawal et al., 1995). However, further work has shown that the

† This research was supported by NIH Grants GM31539 (K.R.W.), ES06839 (S.H.W.), AM09602 (W.K.P.), and GM23705 (W.S.) and NSF Grant MCB 9514179 (K.R.W.).

* Corresponding author: W. M. Keck Foundation Biotechnology Resource Lab, Boyer Center for Molecular Medicine, 295 Congress Ave., Room 302, P.O. Box 9812, New Haven, CT 06536-0812. Phone: (203) 737-2206. Fax: (203) 737-2638. E-mail: WilliamsKR@MASPO1.MAS.Yale.EDU.

‡ Temple University School of Medicine.

§ Department of Molecular Biophysics and Biochemistry, Yale University.

|| Present Address: Glaxo Wellcome Inc., Research Triangle Park, NC, 27709.

⊥ New York University Medical Center.

Deceased.

▽ Howard Hughes Medical Institute, Yale University School of Medicine.

○ University of Texas Medical Branch.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

[1] Abbreviations: Ado, adenosyl; hnRNP, heterogeneous ribonucleoprotein; Dma, dimethylarginine (unless indicated otherwise, this abbreviation refers to the asymmetric, N^G,N^G derivative); HMG, high-mobility group; MBP, myelin basic protein; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; Pth, phenylthiohydantoin; PSD, post-source-decay.

S0006-2960(96)02550-0 CCC: $14.00

determinants of specificity of protein-arginine methyltransferase are more stringent than an arginine flanked by glycine residues.

We have purified two distinct subtypes of protein-arginine methyltransferase from calf brain, one specific for myelin basic protein (MBP), the other for histones (Ghosh et al., 1988) and A1 hnRNP (Rajpurohit et al., 1994). Although the latter enzyme was initially detected using histones as substrate (Ghosh et al., 1988), the subsequent finding that it has a 100-fold lower K_m for A1 led to it being named nuclear protein/histone methylase I (Rajpurohit et al., 1994). Since both MBP and A1 are methylated at Gly-Arg-Gly sequences, the failure of the nuclear protein/histone-arginine methyltransferase to methylate MBP (Ghosh et al., 1988) indicates that the recognition signal is more complex than a tripeptide within a hexamer. As part of a continuing effort to elucidate the requirements for substrate recognition by nuclear protein/histone methylase I, we have in this report identified additional sites of *in vivo* and *in vitro* methylation in A1.

Protein A1 is one of the six "core" proteins originally shown to make up the 40S hnRNP particle (Beyer et al., 1977). The mammalian A1 protein contains 319 residues and begins with a blocked serine (Buvoli et al., 1988; Cobianchi et al., 1986; Merrill et al., 1987; Williams et al., 1985). Analysis of the primary structure of A1 (Merrill et al., 1986) indicated that it contains two major domains, the 1−195 region and the glycine-rich, 196−319 COOH-terminal domain. The 1−195 region contains a region of internal sequence homology such that when residues 3−93 are aligned with 94−184, 32% of the amino acid residues are identical (Merrill et al., 1986). These two internal repeats were postulated to represent two independent nucleic acid binding domains (Merrill et al., 1986), and similar domains subsequently have been found in more than 100 other eukaryotic RNA binding proteins (Adam et al., 1986; Bandziulis et al., 1989; Birney et al., 1993; Kenan et al., 1991; Merrill & Williams, 1990). This ~90 residue domain is referred to as the RNP motif RNA-binding domain (RBD) (Dreyfuss et al., 1988) or the RNA recognition motif (RRM) (Query et al., 1989), and proteins have been found that have from one (type C hnRNP) to four [nucleolin and the poly-(A) binding protein] of these domains (Bandziulis et al., 1989; Kenan et al., 1991; Merrill & Williams, 1990; Shamoo et al., 1995).

The COOH-terminal one-third of A1 (residues 185−319) contains 45% glycine residues, and it is not homologous with the NH_2-terminal two-thirds. Within the COOH-terminal domain is an interesting region spanning residues 190−233 that contains several Arg-Gly-(Gly) (RGG box) sequences that have been proposed to constitute a conserved RNA-binding motif (Dreyfuss et al., 1993; Kiledjian & Dreyfuss, 1992). This region of A1 contains six arginines, five of which occur in Arg-Gly-(Gly) repeat boxes. Numerous other RNA-binding proteins such as splicing factors, hnRNPs, and RNA helicases contain RGG repeats (interspersed with aromatic amino acids) that have a similar spacing within the RNA-binding motif as that found in A1 (Birney at al., 1993; Dreyfuss et al., 1993). In hnRNP U, the 22-residue RGG motif may represent the entire RNA-binding domain (Kiledjian & Dreyfuss, 1992).

In this report we identify three sites of *in vivo* methylation of human A1 (Arginines 205, 217, and 224) in addition to the previously determined site at arginine 193 (Williams et al., 1985). We also show that these sites include two major sites of *in vitro* methylation (Arginines 205 and 224) of recombinant A1 with nuclear protein/histone methylase I. This observation lends support to the notion that A1 is one of the substrates of this enzyme *in vivo*. Furthermore, the finding that all sites of A1 methylation occur within its RGG domain lends strong support to previous data (Rajpurohit et al., 1994) suggesting that methylation modulates the nucleic acid binding properties of A1.

EXPERIMENTAL PROCEDURES

Materials. *S*-Ado-L-[*methyl*-^{3}H]methionine (specific activity, 78.5 Ci/mmol) was obtained from Dupont NEN.

Purification of Protein A1-Specific Arginine Methyltransferase. The purification procedure was essentially as described (Ghosh et al., 1988; Rajpurohit et al., 1994). Briefly, calf brain was homogenized in 4 volumes of 5 mM sodium phosphate (pH 7.2) containing 5 mM EDTA and 0.32 M sucrose and the homogenate was centrifuged at 78500*g* for 60 min. The supernatant was then subjected to ammonium sulfate precipitation, and the 40−70% fraction was chromatographed on DE-52 followed by Sephadex G-200. The purified enzyme transferred approximately 90 pmol of methyl groups per min per mg of enzyme protein at 37 °C under the assay conditions described (Ghosh et al., 1988; Rajpurohit et al., 1994).

Purification of HeLa hnRNP Protein A1. The HeLa A1 protein was purified under nondenaturing conditions as described (Kumar et al., 1986).

Purification of Recombinant hnRNP Protein A1. The protein was isolated from *Escherichia coli* harboring plasmid pEXII and purified as described (Cobianchi et al., 1988). The resulting protein was concentrated by dialysis against polyethylene glycol 20 000.

Preparation of Recombinant [methyl-^{3}H]Protein A1. Recombinant protein A1 (1 mg) was incubated in 0.1 M phosphate buffer (pH 7.6) in the presence of 120 μM Ado [*methyl*-^{3}H]Met and purified calf brain nuclear protein/histone methylase I (123 units) in a total volume of 10 mL for 2 h at 37 °C. The reaction mixture was loaded onto a ssDNA-cellulose column (1 × 5 cm). After extensive washing with phosphate buffer containing 0.4 M NaCl to remove all the unreacted Ado [*methyl*-^{3}H]Met, the [*methyl*-^{3}H]protein A1 was eluted with the same buffer containing 1.0 M NaCl. The methylated A1 thus obtained contained 1.45 mol of [*methyl*-^{3}H] groups per mol of protein and gave a single band of 34 kDa on SDS− PAGE [see Figure 8, lane 1 of Rajpurohit et al. (1994)]. This level of methylation is significantly above that observed for other proteins that have been methylated *in vitro* previously [i.e., the maximal level of *in vitro* methylation of lysine 72 in cytochrome *c* was only 0.2% (Durban et al., 1978)].

Trypsin Digestion of hnRNP A1. Following TCA precipitation, the A1 proteins (650 pmol of HeLa and 450 pmol of *in vitro* methylated recombinant A1 containing 24 880 cpm, which in both cases were mixed with an equal amount of the recombinant A1 controls) were dissolved in 50 μL of 8 M urea, 0.4 M NH_4HCO_3, reduced with dithiothreitol, carboxymethylated with iodoacetic acid, and then digested with a 1:25 (W:W) ratio of trypsin [Boehringer Mannheim (modified)] as described (Stone & Williams, 1993).

Reverse Phase HPLC. Enzymatic digests were injected onto a Vydac C18 column (2.1 or 4.6 × 250 mm as indicated) that was equilibrated with 0.06% trifluoroacetic

acid and 2% buffer B (0.052% trifluoroacetic acid, 80% acetonitrile) and that was eluted as follows (Stone & Williams, 1993): 0−60 min (2−37.5% buffer B), 60−90 min (37.5−75% B), and 90−105 min (75−98% B). The 2.1 mm i.d. column was eluted at 150 μL/min on a HP1090 HPLC instrument while the 4.6 mm i.d. column was eluted at a flow rate of 0.7 mL/min on a Millipore/Waters HPLC instrument.

Mass Spectrometry. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was carried out as described previously (Williams et al., 1996), using a Fisons Instruments VG Tofspec SE mass spectrometer. HPLC fractionated peptides were prepared for MALDI analysis by dilution with a 10 g/L matrix solution of α-cyano-4-hydroxycinnamic acid (Sigma, St. Louis, MO) in 50% aqueous acetonitrile/0.05% trifluoroacetic acid to a final concentration between 100 and 1000 fmol/μL. About 1 μL of this peptide/matrix solution was applied to the target and allowed to evaporate. A Laser Science (Cambridge, MA) nitrogen laser operated at a repetition rate of 2 Hz was used to produce protonated peptide $(M + H)^+$ ions which were then accelerated to 25 kV kinetic energy. Typically, scans from 60 to 100 laser shots were summed to increase signal to noise. Complete post-source-decay (PSD) (Kaufman et al., 1993) mass spectra were computer generated using 8−10 individual reflectron mass spectra acquired by stepping the reflectron/acceleration voltage ratios in order to detect less energetic, lower molecular weight PSD fragment ions.

Other Analytical Methods. Protein-arginine methyltransferase (protein methylase I) activity was determined as described (Ghosh et al., 1988; Rajpurohit et al., 1994) using 0.1 mg of recombinant protein A1 in a total incubation mixture of 0.25 mL. Protein concentrations were estimated by acid hydrolysis followed by ion exchange amino acid analysis on a Beckman 6300 Analyzer or by the Coomassie Blue method of Bradford (1976) as modified by Pierce Chemical Co. Symmetric (N^G,$N^{G'}$) and asymmetric (N^G,N^G) dimethylarginine (Dma), respectively, were from Calbiochem (La Jolla, CA) and were reacted with phenyl isothiocyanate to form the phenylthiohydantoin derivatives used as standards for Edman degradation sequencing of Dma-containing peptides. Pth-(N^G,N^G)-Dma eluted between Pth-tryptophan and Pth-phenylalanine using sodium acetate buffers (Hunkapiller, 1988) and 100% acetonitrile as buffer B. Pth-(N^G,$N^{G'}$)-Dma eluted with Pth-phenylalanine using the same buffer conditions. When 12.5% 2-propanol is added to buffer B, Pth-(N^G,N^G)-(asymmetric)-Dma elutes just before diphenylthiourea (DPTU) and Pth-(N^G,$N^{G'}$)-(symmetric)-Dma, just after DPTU. Pth-methylarginine eluted prior to Pth-proline using the 2-propanol/acetonitrile buffer B.

RESULTS

Identification of Arginine Methylation Sites in Human A1 hnRNP. Comparative HPLC tryptic peptide mapping of native HeLa A1 (Figure 1A) versus (nonmethylated) recombinant A1 made it possible to locate three sites of methylation within the COOH-terminal domain of A1. Since methylated arginine is not cleaved by trypsin (Baldwin & Carnegie, 1971; Merrill et al., 1987), overlapping HeLa A1 tryptic peptides containing sites of methylation will not have counterparts in the recombinant A1 HPLC tryptic peptide profile. Similarly, the two tryptic peptides that make up each of these overlapping HeLa cell A1 tryptic peptides will be

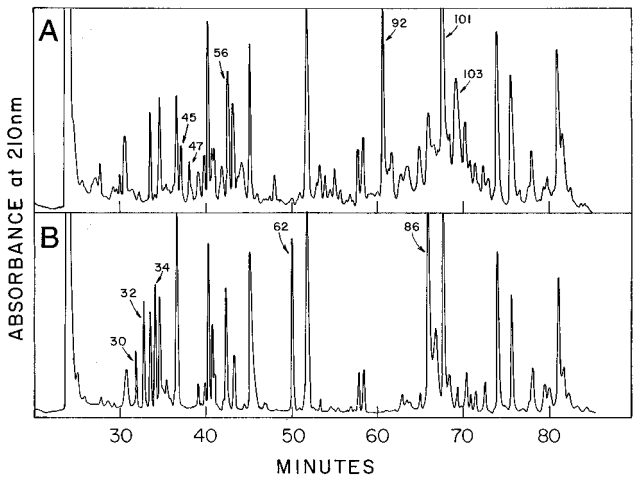

FIGURE 1: Comparative HPLC tryptic peptide maps of A1 hnRNP protein from HeLa cells (panel A) and recombinant A1 expressed in *E. coli* (panel B). See text and Table 1 for details.

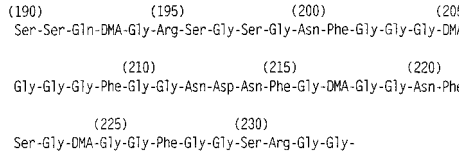

FIGURE 2: Amino acid sequence of the RGG box region of human A1 hnRNP protein that contains four N^G,N^G-dimethylarginine residues. The sequence is predicted from the cDNA sequence of Buvoli et al. (1988). See the text for identification of Dma residues at positions 205, 217, and 224, and see references (Merrill et al., 1987; Williams et al., 1985) for the identification of Dma 193.

missing from the HeLa map. Hence, Edman degradation sequencing of the most intense native A1 tryptic peptide peak that is absent in recombinant A1 (peak 92 in Figure 1A) provided a sequence that matched residues 196−230 (see Figure 2 and Table 1) in the predicted A1 sequence except that all three arginines were missing. Instead, at each of the cycles where arginine was predicted (i.e., residues 205, 217, and 224), there was a broad peak (which is a normal characteristic of Pth-arginine and its derivatives) that eluted just prior to diphenylthiourea (DPTU). The HPLC retention time of this peak was the same as that for the Pth-derivatve of asymmetric Dma and was about 1 min earlier than that for the Pth derivative of symmetric Dma, which eluted just after DPTU. On the basis of these results and previous Pth-Dma elution data (Merrill et al., 1987), we concluded that this unusual peak is the Pth derivative of asymmetric Dma and that arginines 205, 217, and 224 are dimethylated in HeLa A1. The finding that all three of these derivatives [as well as the previously identified Dma at position 193 (Merrill et al., 1987)] are asymmetric (as opposed to one or more being symmetric) Dma agrees well with previous amino acid analyses carried out on HeLa A1, which indicate that it contains 3.5 asymmetric Dma residues and no detectable symmetric Dma (Wilk et al., 1986). Furthermore, MALDI-TOF mass spectrometry of this HPLC fraction indicated a single ion at $(M + H)^+ = m/z$ 3410.1 (along with a doubly charged ion at m/z 1706.0). This value is in close agreement with the calculated value (3410.57) for the $(M + H)^+$ ion of peptide 196−231 with Dma rather than arginine at positions 205, 217, and 224. Unfortunately, it was not possible to obtain a PSD spectrum of sufficiently good quality [most likely due to the rather high glycine content (50%) of this peptide] to independently verify Dma at these positions. Consequently, smaller peptides for PSD were generated by

Table 1: Summary of Sequencing Data from HeLa and *E. coli* Expressed A1 hnRNP Tryptic Peptides[a]

source of A1	fraction	residues sequenced	yield (pmol)	comments
HeLa	45	300−317	25.0	arginine 317 not methylated
	47		∼2	low level mix of two partial sequences that could not be matched to A1
	56a	3−13	40.2	matches expected sequence
	56b	179−194	11.8	residue 193 identified as asymmetric dimethylarginine
	92	196−230	45.0	residues 205, 217, and 224 identified as asymmetric dimethylarginine
	101	31−51	128	control peptide that coelutes with a corresponding peptide in the *E. coli* digest
	103b	106−128	14.5	matches expected sequence
	103a	196−230	43.3	residues 205, 217, and 224 identified as asymmetric dimethylarginine
E. coli	30	183−192	11.2	matches expected sequence/corresponding peptide missing in HeLa digest
	32	218−224	335	matches expected sequence/corresponding peptide missing in HeLa digest
	34	196−205	330	matches expected sequence/corresponding peptide missing in HeLa digest
	62	206−217	567	matches expected sequence/corresponding peptide missing in HeLa digest
	86	232−249	79.2	matches expected sequence/corresponding peptide in HeLa digest appears to be present in lower yield

[a] Selected peptide peaks from the tryptic digests shown in Figure 1 were subjected to amino acid sequencing as described in the Experimental Procedures. Since some peptides listed were not completely sequenced, the last residue identified may not correspond to the COOH-terminus of the peptide.

subdigestion of peptide 196−231 with *endo*-Asp-N which, as determined by MALDI-TOF-MS, resulted in cleavage between asparagine 212 and aspartic acid 213, yielding peptides 196−212 and 213−231 with $(M + H)^+$ ions at *m/z* 1469.5 and 1957.8 (calculated values = 1469.65 and 1957.94, respectively). Both these data as well as further PSD analysis of these peptides provided additional confirmation of the presence of Dma in peptide 196−231. The presence of these three dimethylarginines explains why three peptides, labeled fractions 32, 34, and 62 in Figure 1B, that would be predicted to result from cleavage after arginines 205, 217, and 224, respectively (see Figure 2 and Table 1), are missing from the HeLa A1 digest (Figure 1A). The absence of all three of these peaks from the HeLa A1 digest is consistent with, in each case, *in vivo* dimethylation having proceeded to completion. Similarly, the absence in Figure 1A of a HeLa A1 peak corresponding to recombinant A1 peak 30, which spans residues 183−193 (see Table 1), provides further support for our previous finding that arginine 193 is fully dimethylated. This conclusion was further strengthened by sequencing of fraction 56 (Figure 1A), which identified a Dma at position 193 (Table 1). Since the major peptide in fraction 56 corresponds to residues 3−13, in this instance, a corresponding peak is seen in the recombinant chromatogram in Figure 1B. Sequencing of the HeLa A1 peptide in fraction 103 (Figure 1A and Table 1) indicated it is an approximately 3:1 mixture of two peptides starting at residues 196 and 106, respectively. The major component was sequenced from serine 196 to serine 230 (see Figure 2) by Edman degradation and confirmed the three dimethylarginine residues at positions 205, 217 and 224. The peptide in fraction 103 is most likely due to incomplete cleavage at Arginine 231, resulting in the much longer tryptic peptide spanning residues 196−276 and explaining its later elution than HeLa A1 peak 92, which extends to arginine 231. Following this reasoning, the HeLa peptide that spans residues 232−249 should be present in lower amounts in the HeLa as compared to in the recombinant digest, and indeed, as shown in Figure 1, this is the case as the HeLa counterpart for recombinant peak 86, which spans residues 232−249 (Table 1), is present in much reduced relative yield (Figure 1).

Identification of Sites of in Vitro Methylation of Recombinant A1 Expressed in E. coli. We previously have shown that nonmethylated recombinant A1 can be methylated *in vitro* with nuclear protein/histone methylase I (Rajpurohit et al., 1994). A low level of incorporation of [^{3}H]methyl groups was located at arginine 193, and the majority of the methylation was shown to occur at one or more positions within the COOH-terminal domain spanning residues 196−319. The latter conclusion derives from Figure 8 in Rajpurohit et al. (1994), where it is clear that the autoradiograph intensity (i.e., extent of incorporation of [^{3}H]methyl) of the COOH-terminal peptide spanning residues 185−319 (lane 4) is much greater than that for the NH_2-terminal fragment spanning residues 1−195 (lane 8). To determine the location of the major sites of *in vitro* methylation by this enzyme, we have employed the same comparative peptide mapping approach used above with HeLa A1. As shown in Figure 3A and Table 2, only five fractions (25, 45, 47, 57, and 58) contained ^{3}H-labeled peptides. All these fractions corresponded to peaks that were not present in the control digest (Figure 3B). The amino acid sequence of the peptide in fraction 45, which was the largest radioactively labeled absorbance peak, matched residues 218−231 except that arginine 224 was missing. Instead, at this cycle there was a broad peak in the HPLC profile that eluted just prior to Pth-Pro. Since this peak eluted prior to the elution position for dimethylarginine, but retained the characteristic broadness of Pth-Arg and its derivatives, we postulated it represented

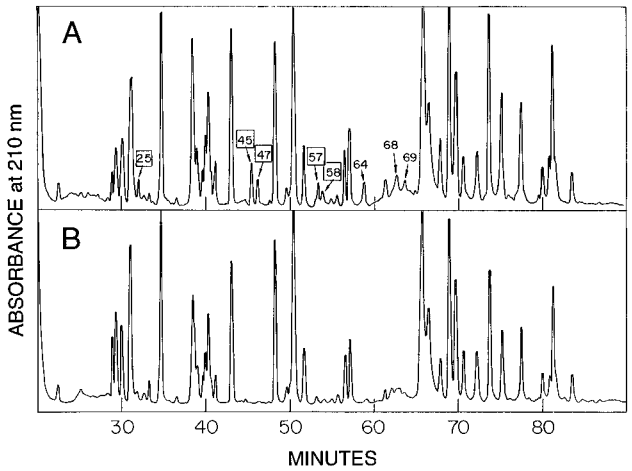

FIGURE 3: Comparative HPLC tryptic peptide maps of recombinant A1 before (panel B) and after (panel A) *in vitro* methylation with nuclear protein/histone arginine methyltransferase. Arrows indicate fractions subjected to NH_2-terminal sequencing and/or MALDI mass spectrometry. Numbers that are boxed in panel A indicate those fractions that contained ^{3}H-labeled peptides. See text and Table 2 for details.

Table 2: Characterization of Selected Tryptic Peptides from *E. coli* Expressed A1 that has been Methylated In Vitro

fraction[a]	cpm	residues sequenced	yield (pmol)	observed *m/z*	predicted *m/z*	mass error (%)	comments
25	490	179−193	16.8	1598.0	1597.7	0.02	predicted *m/z* assumes methionine 185 is oxidized
45	2290	218−231	34.1	1327.0	1327.4	0.03	instead of the expected arginine, residue 224 contained a broad peak that eluted just prior to Pth-Pro; predicted *m/z* assumes arginine 224 is monomethylated
47	2440	218−231	8.0	1341.2	1341.4	0.02	residue 224 identified as asymmetric dimethylarginine; predicted *m/z* assumes arginine 224 is dimethylated
57	810	196−217	10.3	2045.8	2046.1	0.01	instead of the expected arginine, residue 205 contained a broad peak that eluted just prior to Pth-Pro; predicted *m/z* assumes arginine 205 is monomethylated
58	510	196−215	0.8	2059.9	2060.1	0.01	residue 205 identified as asymmetric dimethylarginine; predicted *m/z* assumes this peptide extends to arginine 217; this fraction also contained three other sequences that matched A1
64	0	130−139	36.8	1235.5	1235.4	0.01	predicted *m/z* assumes methionine 136 is oxidized
68	0	232−244	10.3	ND			peptide stopped sequencing with a positive call at tyrosine 243 and a very weak call at asparagine 244
69	0	232−244	15.7	4428.2 2711.2	4424.4	0.09	peptide stopped sequencing with a positive call at tyrosine 243 and a very weak call at asparagine 244; predicted *m/z* assumes peptide spans residues 232−276

[a] Peptide peaks from the tryptic digest of *in vitro* methylated recombinant A1 shown in Figure 3A were subjected to amino acid sequencing and MALDI mass spectrometry as described in the Experimental Procedures. Since some peptides listed were not completely sequenced, the last residue identified may not correspond to the COOH-terminus of the peptide.

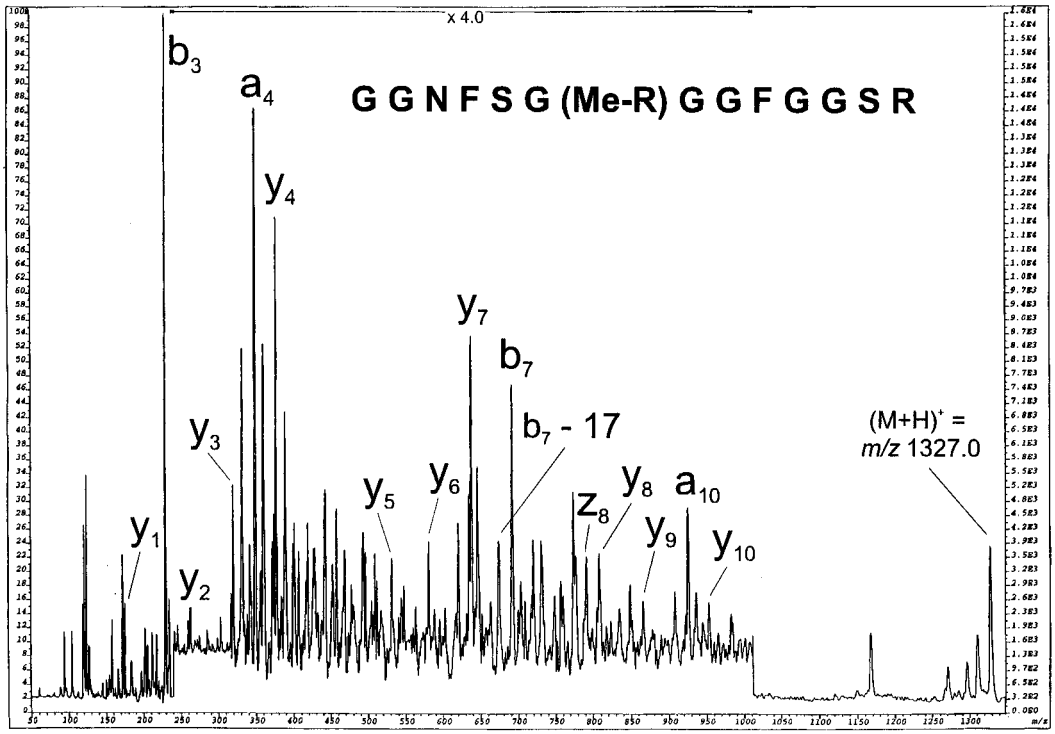

FIGURE 4: Post-source-decay (PSD) mass spectrum of an A1 tryptic peptide containing *in vitro* methylated arginine. Approximately 3% of fraction 45 (HPLC chromatogram in Figure 3A) was utilized for this MALDI-MS analysis. The assigned sequence for this peptide, Gly-Gly-Asn-Phe-Ser-Gly-(methyl-Arg)-Gly-Gly-Phe-Gly-Gly-Ser-Arg was confirmed by the y ion series fragments. The b7, b7−17, y8, and z8 ions together identified methylarginine at position 224 (see the text).

the Pth derivative of monomethylarginine. The $(M + H)^+$ ion generated from MALDI-TOF-MS of fraction 45 at *m/z* 1327.0 agreed closely with the predicted value of 1327.40 for peptide 218−231 (Table 2) with monomethylated arginine at position 224. Further confirmation of this site of methylation was determined by MALDI-TOF-MS using PSD (Figure 4). The fragment ions provide further support for the sequence assignment of this peptide [a discussion of the nomenclature used in identifying fragment ions can be found in Biemann (1990)]. Finally, the b7−17, b7, z8, and y8 ions (*m/z* 673.7, 691.4, 791.2, and 808.3, respectively) together indicate that monomethyl arginine is at position 224. Sequencing of the peptide in fraction 47, one of the two most radioactive fractions obtained (Table 2), provided the same sequence and, once again, the expected Pth-Arg derivative at residue 224 was missing and a peak appeared at the position expected for asymmetric Dma. This identification was again confirmed by the excellent agreement between the observed *m/z* (within 0.02% of the calculated value) for the peptide in fraction 47 with that predicted from the known sequence assuming residue 224 is N^G,N^G-dimethylarginine. On the basis of similar studies, fractions 57 and 58 were identified as containing the methyl- and N^G,N^G-dimethylarginine derivative of the peptide spanning residues 196−217, thus identifying arginine 205 as a second site of *in vitro* methylation.

The peptide in fraction 25, which was the only other radioactive fraction obtained from this digest, provided a sequence that matched residues 179−193. Since the observed *m/z* for the $(M + H)^+$ ion of this peptide was 1598.0 or 16.3 units higher than the predicted value of 1581.7, our conclusion was that methionine 185 is oxidized but was reduced back to methionine by the dithiothreitol included in some of the reagents used for automated Edman sequencing. In addition to the major species at *m/z* 1598.0, the MALDI mass spectrum of fraction 25 also contained a much less intense doublet of peaks at *m/z* 1407.8 and 1424.7. The mass difference between these peaks (16.9 units) and the fact that a MALDI mass spectrum of the same fraction obtained after several months storage contained only the the ion at *m/z* 1425.4 (in addition, of course, to the expected ion for the 179−193 peptide with oxidized Met) strongly suggests the 1407.8/1424.7 ions are related by methionine oxidation. Since the calculated value for the $(M + H)^+$ ion of the monomethylated/oxidized methionine peptide 183−195 is 1425.50, we presume a small amount of this peptide is present in this fraction, thus accounting for the radioactivity in this fraction and confirming the previous finding that a low level of methylation of arginine 193 does indeed occur *in vitro* with nuclear protein/histone methylase I (Rajpurohit et al., 1994). The fact this peptide was not detected during sequencing (Table 2) may result either from its low level or from cyclization of its NH_2-terminal glutamine residue during Edman degradation.

Minor peaks 64, 68, and 69 (Figure 3A) either did not appear to be present or were present at reduced levels in the control as opposed to in the *in vitro* methylated A1 and did not contain radioactivity. On the basis of amino acid sequencing and MALDI-MS data, the peptides in these fractions are located outside the RGG region (Table 2) and their presence in the digest is due to unrelated post-translational modifications including methionine oxidation and asparagine−glycine cyclization. Since these latter modifications were present in the recombinant protein that was subject to *in vitro* methylation but were absent from the batch of recombinant protein used as the control (Figure 3B) and that was purified in a similar fashion, these two post-translational modifications appear to be dependent upon either the age of the recombinant A1 or upon slight alterations in the purification protocol.

DISCUSSION

Using a comparative HPLC tryptic peptide mapping approach, three sites of arginine methylation have been located in the glycine-rich COOH-terminal domain of human hnRNP A1 protein. When taken together with previous data (Merrill et al., 1987; Williams et al., 1985), it is apparent that arginine residues 193, 205, 217, and 224 are fully methylated in A1 hnRNP isolated from HeLa cells via a procedure designed to purify native hnRNP proteins (Kumar et al., 1986). The assignment of these positions is consistent with a previous *in vitro* mutagenesis study (Liu & Dreyfuss, 1995), indicating that all sites of A1 arginine methylation are within the 41 amino acid stretch of sequence that spans residues 193−233 and that contains an RGG binding motif [see Figures 2 and 5, Dreyfuss et al. (1993) and Kiledjian and Dreyfuss (1992)]. Note that there is at least one phenylalanine between each of the Dmas.

In terms of the specific protein methylase that may modify A1 *in vivo,* the previously determined much higher affinity

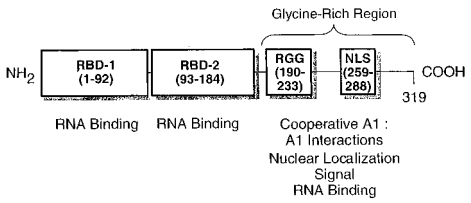

FIGURE 5: Schematic depiction of structural and functional domains in the A1 hnRNP protein. The location of the ~90 residue conserved RNA binding domains (RBDs) in the NH_2-terminal region is based on the alignment in Merrill et al. (1986), and the RGG domain (Kiledjian & Dreyfuss, 1992) has been extended to include residues 190−233 (see Figure 2). The location of the nuclear localization signal (NLS) is based on Siomi and Dreyfuss (1995) and Weighardt et al. (1995).

of nuclear protein/histone-arginine methyltransferase for recombinant hnRNP A1 as compared to histone suggested that A1 might be one of the preferred natural substrates of this transferase (Rajpurohit et al., 1994). This supposition is now strengthened by the finding that both of the major sites of *in vitro* methylation of recombinant A1 by this enzyme (arginines 205 and 224) occur at arginine residues that are methylated *in vivo*.

In an effort to derive a possible recognition sequence for nuclear protein/histone arginine methyltransferase, we have compiled in Table 3 a listing of known N^G,N^G-methylation sites in nuclear proteins. This compilation suggests a symmetrical Gly/Phe-Gly-Gly-Arg-Gly-Gly-Gly/Phe preferred sequence with an absolute requirement for a COOH-terminal flanking glycine. The latter may explain why arginine 193 in A1 is methylated while the nearby arginine 195, which is followed by serine (Table 3), is not. Although glycine is certainly preferred in the $n - 1$ position, there are instances where phenylalanine, aspartic acid, proline, and glutamine are found in this position (Table 3). Serine, however, does not appear to be accommodated in the $n - 1$ position since two Ser-Arg-Gly sequences beginning at serines 53 and 230 (the latter in the RGG binding motif) are not methylated. The other (apparently) unmethylated arginines of A1, i.e., 20 out of a total of 24, lack the COOH-terminal flanking glycine except at positions 139 (Asp-Arg-Gly) and 145 (Lys-Arg-Gly). Lack of methylation of these two sites may result either from the effect of the $n - 1$ amino acid or from the highly ordered structure of the RBDs in the NH_2-terminal domain of A1 precluding access to methylase. Of the sites listed, arginine 193 in A1 differs most from the preferred sequence given in Table 3 in that it only has the COOH-terminal flanking glycine. Indeed, on the basis of results contained within this work as well as our earlier studies (Merrill et al., 1987; Rawal et al., 1995), arginine 193 does not appear to represent a preferred site of *in vitro* methylation. This raises the possibility that a different enzyme may methylate arginine 193. The preferred sequence shown in Table 3 is consistent with our previous synthetic peptide analogue studies which demonstrated an obligatory requirement for a COOH-terminal flanking glycine and a preference for an NH_2-terminal flanking glycine (Rawal et al., 1995). It also provides a plausible explanation for the finding that HeLa protein-arginine *N*-methyltransferase appears to be specific for proteins containing the RGG box RNA binding domain (Liu & Dreyfuss, 1995). It also appears from Table 3 and from Kiledjian and Dreyfuss

Table 3: Sites of N^G,N^G-Methylarginine Residues in RNA Binding Proteins

		relative position							
protein	residues	$n-3$	$n-2$	$n-1$	n	$n+1$	$n+2$	$n+3$	ref
human A1 hnRNP	190−196	Ser	Ser	Gln	Dma	Gly	Arg	Ser	this work
	202−208	Gly	Gly	Gly	Dma	Gly	Gly	Gly	
	214−220	Asn	Phe	Gly	Dma	Gly	Gly	Asn	
	221−227	Phe	Ser	Gly	Dma	Gly	Gly	Phe	
CHO nucleolin	652−658	Phe	Gly	Gly	Dma	Gly	Gly	Gly	(Lapeyre et al., 1986)
	656−662	Gly	Gly	Gly	Dma	Gly	Gly	Phe	
	662−668	Phe	Gly	Gly	Dma	Gly	Gly	Gly	
	666−672	Gly	Gly	Gly	Dma	Gly	Gly	Gly	
	670−676	Gly	Gly	Gly	Dma	Gly	Gly	Phe	
	676−682	Phe	Gly	Gly	Dma	Gly	Dma	Gly	
	678−684	Gly	Dma	Gly	Dma	Gly	Gly	Phe	
	684−690	Phe	Gly	Gly	Dma	Gly	Gly	Phe	
	688−694	Gly	Gly	Phe	Dma	Gly	Gly	Dma	
	691−697	Dma	Gly	Gly	Dma	Gly	Gly	Gly	
rat nucleolin	656−662	Phe	Gly	Gly	Dma	Gly	Gly	Gly	(Lischwe et al., 1985)
	660−666	Gly	Gly	Gly	Dma	Gly	Gly	Phe	
	666−672	Phe	Gly	Gly	Dma	Gly	Gly	Gly	
	670−676	Gly	Gly	Gly	Dma	Gly	Gly	Dma	
	673−679	Dma	Gly	Gly	Dma	Gly	Gly	Phe	
	679−685	Phe	Gly	Gly	Dma	Gly	Dma	Gly	
	681−687	Gly	Dma	Gly	Dma	Gly	Gly	Phe	
	687−693	Phe	Gly	Gly	Dma	Gly	Gly	Phe	
	691−697	Gly	Gly	Phe	Dma	Gly	Gly	Dma	
	694−700	Dma	Gly	Gly	Dma	Gly	Gly	Gly	
rat fibrillarin	5−11	Phe	Ser	Pro	Dma	Gly	Gly	Gly	(Lischwe et al., 1985)
	12−18	Phe	Gly	Gly	Dma	Gly	Gly	Phe	
	18−24	Phe	Gly	Asp	Dma	Gly	Gly	Dma	
	21−27	Dma	Gly	Gly	Dma	Gly	Gly	Gly	
	25−31	Gly	Gly	Gly	Dma	Gly	Gly	Dma	
	28−34	Dma	Gly	Gly	Dma				
physar. fibrillarin	2−8	Phe	Glu	Gly	Dma	Gly	Gly	Phe	(Christensen & Fuxa, 1988)
	8−14	Phe	Gly	Gly	Dma	Gly	Gly	Gly	
	13−18	Gly	Gly	Asp	Dma	Gly	Gly	Dma	
	16−22	Dma	Gly	Gly	Dma	Gly	X	Gly	
Artemia P38	Frag 6	Pro	Pro	Met	Dma	Gly	Gly	Dma	(Pype et al., 1994)
	Frag 6	Dma	Gly	Gly	Dma	Gly	Gly	Leu	
	Frag 12	Gly	Pro	Ala	Dma	Gly	Gly	Lys	
	Frag 13	Gly	Pro	Thr	Dma	Gly	Gly	Lys	
frequency		G(14) F(14) Dma(7) X(3)	G(28) S(3) P(3) Dma(2) X(2)	G(29) F(2) D(2) X(5)	Dma(38)	G(38)	G(33) Dma(2) R(1)	G(14) F(11) Dma(7) K(2) X(3)	
preferred		G/F	G	G	Dma	G	G	G/F	
% preferred residues		74%	74%	76%	100%	100%	87%	74%	

(1992) that RGG boxes are clustered within a span of 15−40 amino acids.

The comparative HPLC tryptic peptide maps in Figure 1 indicate that purified human A1 is fully methylated at each of the four arginine residues that we have identified. Although these data suggest human A1 is also fully methylated *in vivo*, it is possible that unmethylated forms exist *in vivo* and they were separated during purification. Our finding that all sites of A1 methylation map within a domain (RGG) that has been shown to represent a conserved nucleic acid binding motif (Kiledjian & Dreyfuss, 1992) suggests strongly that methylation modulates nucleic acid binding. Indeed, Rajpurohit et al. (1994) have shown previously that even a very low level of *in vitro* methylation (i.e., mole ratio of incorporated methyl groups to A1 of less than 2 as opposed to 8 for native A1) decreases the apparent affinity of A1 for ssDNA cellulose and for MS2RNA. It is also intriguing that a known site of A1 phosphorylation, serine 198 (Cobianchi et al., 1993), is very close to two methylation sites (arginines 194 and 205). Since phosphorylation of serine 198 abrogates the ability of A1 to facilitate the characteristic *in vitro* strand annealing activity of A1, this provides further support for the notion that this region plays an important role in modulating the interaction of A1 with nucleic acids. In this regard, it would obviously be of interest to determine the impact of these two post-translational modifications on each other and whether their effects on A1 function are additive. An intriguing suggestion (Najbauer et al., 1993) is that methylation might help to "lock" proteins like A1 into a "nonspecific" binding mode by preventing the formation of arginine-dependent hydrogen bonds of the type that have been proposed to account for specific interactions, e.g., the HIV-1 Tat protein interaction with TAR RNA (Calnan et al., 1991). Finally, since the C-terminal A1 domain that extends from methionine 185 to phenylalanine 319 contributes all or most of the cooperativity of A1 binding (Kumar et al., 1990), it is possible that arginine methylation might effect binding by modifying the cooperative protein:protein interactions that occur between adjacent A1 molecules bound to a nucleic acid lattice.

Just as identification of sites of A1 methylation suggests a role for this modification in nucleic acid binding, it also

suggests methylation is probably not involved in nuclear localization of A1, which is one of the possible functions suggested by Gary et al. (1996). That is, the A1 nuclear localization domain (residues 259−288) found by Weighardt et al. (1995) and Siomi et al. (1995) is distant from the stretch of A1 sequence (193−224) that contains all four dimethylarginines. In addition, the arginine residue (282) within the nuclear localization domain lacks the apparently obligatory $n + 1$ glycine and, as expected, does not appear to be methylated in native A1.

The functional importance of arginine methylation is suggested not only by the large number of proteins that have been found to contain this modification (see the introduction) but also by the recent finding that the mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with and modulate the activity of protein-arginine methyltransferase (Lin et al., 1996). Modulation of arginine methyltransferase activity by two transiently expressed proteins that play important roles in mediating ligand-induced biological responses suggests that this intriguing post-translational modification may provide an additional and novel mode of signal transduction (Lin et al., 1996).

ACKNOWLEDGMENT

We especially thank the staff of the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for assistance with the protein chemistry studies.

REFERENCES

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., & Dreyfuss, G. (1986) *Mol. Cell. Biol. 6*, 2932−2943.
Baldwin, G. S., & Carnegie, P. R. (1971) *Science 171*, 579−581.
Bandziulis, R. J., Swanson, M. S., & Dreyfuss, G. (1989) *Genes Dev. 3*, 431−437.
Beyer, A. L., Christensen, M. D., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell 11*, 127−138.
Biemann, K. (1990) in *Methods in Enzymology* (McCloskey, J., Ed.) Vol. 193, Appendix 5, pp 886−887, Academic Press, San Diego, CA.
Birney, E., Kumar, S., & Krainer, A. R. (1993) *Nucleic Acids Res. 21*, 5803−5816.
Boffa, L. C., Karn, J., Vidali, G., & Allfrey, V. G. (1977) *Biochem. Biophys. Res. Commun. 74*, 969−976.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
Buvoli, M., Biamonte, G., Tsoulfas, P., Bassi, M., Ghetti, A., Riva, S., & Morandi, C. (1988) *Nucleic Acids Res. 16*, 3751−3770.
Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. D. (1991) *Science 252*, 1167−1171.
Christensen, M. E., & Fuxa, K. P. (1988) *Biochem. Biophys. Res. Commun. 155*, 1278−1283.
Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., & Wilson, S. H. (1986) *J. Biol. Chem. 261*, 3536−3543.
Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., & Wilson, S. H. (1988) *J. Biol. Chem. 263*, 1063−1071.
Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M., & Riva, S. (1993) *Nucleic Acids Res. 21*, 949−955.
Dreyfuss, G., Swanson, M. S., & Piñol-Roma, S. (1988) *Trends Biochem. Sci. 13*, 86−91.
Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., & Burd, C. G. (1993) *Annu. Rev. Biochem. 62*, 289−321.
Durban, E., Nochumson, S., Kim, S., Paik, W., & Chan, S. (1978) *J. Biol. Chem. 253*, 1427−1435.
Gary, J. D., Lin, W. J., Yang, M. C., Herschman, H. R. & Clarke, S. (1996) *J. Biol. Chem. 271*, 12585−12594.
Ghosh, S. K., Paik, W. K., & Kim, S. (1988) *J. Biol. Chem. 263*, 19024−19033.
Ghosh, S. K., Syed, S. K., Jung, S., Paik, W. K., & Kim, S. (1990) *Biochim. Biophys. Acta 1039*, 142−148.
Hunkapiller, M. (1988) in *Protein/Peptide Sequence Analysis: Current Methodologies* (Bhown, A. S., Ed.) pp 87−117, CRC Press, Inc., Boca Raton, FL.
Kaufman, R., Spengler, B., & Lützenkirchen F. (1993) *Rapid Commun. Mass Spectrom. 7*, 902−910.
Kenan, D. J., Query, C. C., & Keene, J. D. (1991) *Trends Biochem. Sci. 16*, 214−220.
Kiledjian, M., & Dreyfuss, G. (1992) *EMBO J. 11*, 2655−2664.
Kim, S., Ghosh, S. K., & Chanderkar, L P. (1990) in *Protein Methylation* (Paik, W. K., & Kim, S., Eds.) CRC Press, Boca Raton, FL.
Kumar, A., Williams, K. R., & Szer, W. (1986) *J. Biol. Chem. 261*, 11266−11273.
Kumar, A., Casas-Finet, J., Luneau, C. J., Karpel, R. L., Merrill, B. M., Williams, K. R., & Wilson, S. H. (1990) *J. Biol. Chem. 265*, 17094−17100.
Lapeyre, B., Amalric, F., Ghaffari, S. H., Venkataram Rao, S. V., Dumbar, T. S., & Olson, M. O. J. (1986) *J. Biol. Chem. 261*, 9167−9173.
Lin, W.-J., Gary, J. D., Yang, M. C., Clarke, S., & Herschman, H. R. (1996) *J. Biol. Chem. 271*, 15034−15044.
Lischwe, M. A., Cook, R. G., Ahn, Y. S., Yeoman, L. C., & Busch, H. (1985a) *Biochemistry 24*, 6025−6028.
Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M., & Busch, H. (1985b) *J. Biol. Chem. 260*, 14304−14310.
Liu, Q., & Dreyfuss, G. (1995) *Mol. Cell. Biol.* 15, 2800−2808.
Merrill, B. M., LoPresti, M. B., Stone, K. L., & Williams, K. R. (1986) *J. Biol. Chem. 261*, 878−883.
Merrill, B., LoPresti, M. B., Stone, K. L., & Williams, K. R. (1987) *Int. J. Pept. Protein Res. 29*, 21−39.
Merrill, B. M., & Williams, K. R. (1990) in *The Eukaryotic Nucleus: Molecular Biochemistry & Macromolecular Assemblies* (Strauss, P. R., & Wilson, S. H., Eds.) Vol. 2, pp 579−604, Telford Press, Caldwell.
Najbauer, J., Johnson, B. A., Young, A. L., & Aswad, D. W. (1993) *J. Biol. Chem. 268*, 10501−10509.
Paik, W. K., & Kim, S. (1985) in *Enzymology of Post-Translational Modification of Proteins* (Freeman, R. B., & Hawkins, H. C., Eds.) Vol. 2, pp 187−228, Academic Press, London.
Paik, W. K., & Kim, S. (1993) *Amino Acids 4*, 267−286.
Pype, S., Slegers, H.,Moens, L., Merlevede, W., & Goris, J. (1994) *J. Biol. Chem. 269*, 31457−31465.
Query, C. C., Bentley, R. C., & Keene, J. D. (1989) *Cell 57*, 89−101.
Rajpurohit, R., Lee, S. O., Park, J. O., Paik, W. K., & Kim, S. (1994a) *J. Biol. Chem. 269*, 1075−1082.
Rajpurohit, R., Paik, W. K., & Kim, S. (1994b) *Biochem. J. 304*, 903−909.
Rawal, A. N., Rajpurihit, R., Lischwe, M. A.,Williams, K. R., Paik, W. K., & Kim, S. (1995) *Biochim. Biophys. Acta 1248*, 11−18.
Shamoo, Y., Abdul-Manan, N., & Williams, K. R. (1995) *Nucleic Acids Res. 23*, 725−728.
Siomi, H., & Dreyfuss, G. (1995) *J. Cell Biol. 129*, 551−560.
Stone, K. L., & Williams, K. R. (1993) in *A Practical Guide to Protein & Peptide Purification for Microsequencing* (Matsudaira, P. T., Ed.) 2nd ed., pp 43−69, Academic Press, New York.
Weighardt, F., Biamonti, G., & Riva, S. (1995) *J. Cell Sci. 108*, 545−555.
Wilk, H., Werr, H., Friedrich, D., Kiltz, H., & Schafer, K. P. (1985) *Eur. J. Biochem. 146*, 71−81.
Williams, K. R., Stone, K. L., LoPresti, M. B., Merrill, B. M., & Planck, S. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5666−5670.
Williams, K. R., Samandar, S. M., Stone, K. L., Saylor, M., & Rush, J. (1996) in *The Protein Protocols Handbook* (Walker, J. M., Ed.) (in press).

BI9625509