RNA and Protein Interactions Modulated by Protein Arginine Methylation

JONATHAN D. GARY AND STEVEN CLARKE

The Molecular Biology Institute and The Department of Chemistry and Biochemistry
University of California, Los Angeles
Los Angeles, California 90095

This review summarizes the current status of protein arginine N-methylation reactions. These covalent modifications of proteins are now recognized in a num-

Progress in Nucleic Acid Research and Molecular Biology, Vol. 61

Copyright © 1998 by Academic Press.
All rights of reproduction in any form reserved.
0079-6603/98 $25.00

ber of eukaryotic proteins and their functional significance is beginning to be understood. Genes that encode those methyltransferases specific for catalyzing the formation of asymmetric dimethylarginine have been identified. The enzyme modifies a number of generally nuclear or nucleolar proteins that interact with nucleic acids, particularly RNA. Postulated roles for these reactions include signal transduction, nuclear transport, or a direct modulation of nucleic acid interactions. A second methyltransferase activity that symmetrically dimethylates an arginine residue in myelin basic protein, a major component of the axon sheath, has also been characterized. However, a gene encoding this activity has not been identified to date and the cellular function for this methylation reaction has not been clearly established. From the analysis of the sequences surrounding known arginine methylation sites, we have determined consensus methyl-accepting sequences that may be useful in identifying novel substrates for these enzymes and may shed further light on their physiological role. © 1998 Academic Press

I. Introduction

It has become increasingly clear that protein function is dependent on the covalent posttranslational modification of the 20 amino acid residues normally incorporated by ribosomes during protein synthesis. Some of these modifications are reversible, such as protein phosphorylation reactions, whereas others are apparently irreversible and can effectively create new types of amino acids to broaden the chemical diversity of polypeptides. In this latter group of modifications are a number of S-adenosylmethionine (AdoMet)-dependent methylation reactions that occur on the side-chain nitrogen atoms of lysine, arginine, and histidine residues (*1*). Although we understand some of the chemistry and enzymology of these reactions, less progress has been made in deciphering their functional importance.

In recent years, however, a number of advances have allowed us to get a clearer picture of the physiological role of one type of these reactions, the methylation of the guanidino group nitrogen atoms of protein arginine residues. Protein arginine methylation is found in most cells of higher eukaryotic species (*2*) as well as in lower eukaryotes, such as the trypanosomes (*3, 4*), but to date has not been found in prokaryotic organisms. As described below, it now appears that there are at least two distinct classes of protein arginine *N*-methyltranferases. The Type I enzyme catalyzes the formation of N^G-monomethylarginine and asymmetric N^G,N^G-dimethylarginine residues; whereas the Type II enzyme catalyzes the formation of N^G-monomethylarginine and symmetric N^G,N'^G-dimethylarginine residues (Fig. 1). Several *in vivo* substrates have been identified for the Type I (asymmetrically methylating) enzyme, including hnRNP A1, fibrillarin, and nucleolin. Interesting-

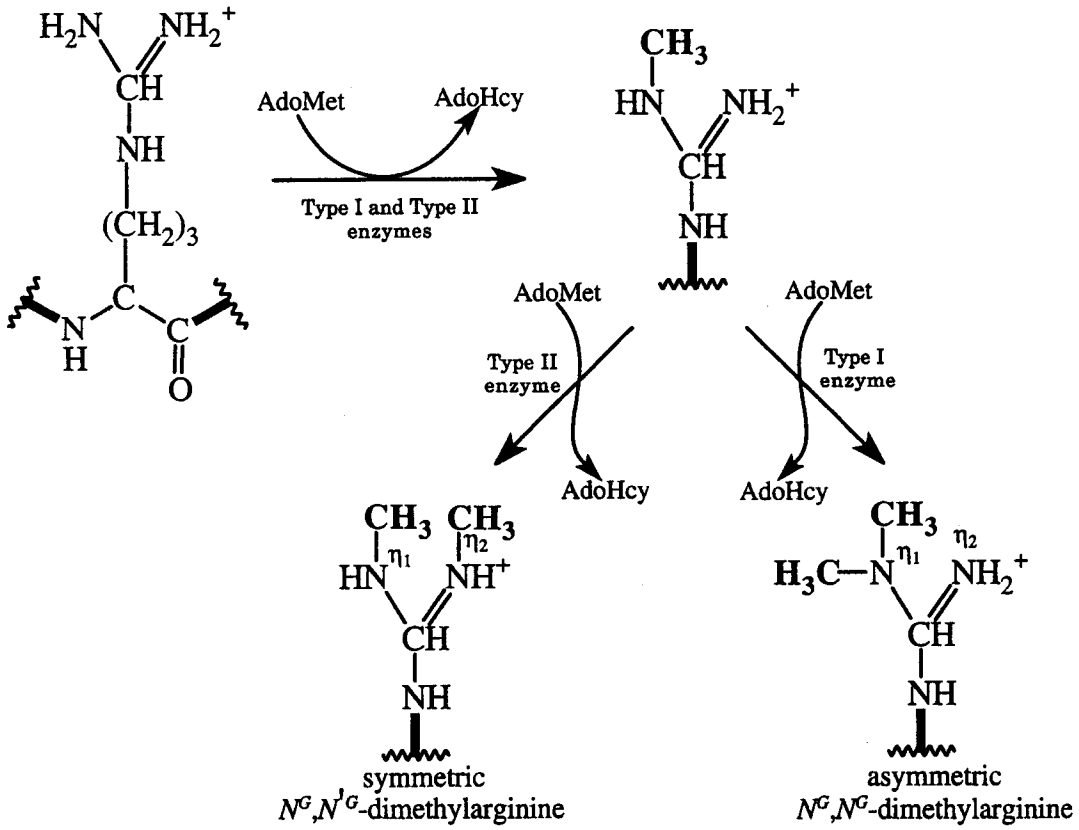


FIG. 1. The structures and stepwise synthesis of methylated arginine derivatives found in proteins. AdoMet serves as the methyl donor for two consecutive methylation events on the guanidino nitrogens of arginine residues. At least two protein arginine *N*-methyltransferases catalyze these reactions with both substrate and product specificity. One of the arginine methyltransferase activities (Type I) mono- and asymmetrically dimethylates hnRNP A1 as well as several other proteins that interact with RNA, including nucleolin and fibrillarin. The other known enzyme activity (Type II) mono- and symmetrically dimethylates myelin basic protein.

ly, all of these methyl-accepting proteins interact with RNA species, suggesting that the methylation of arginine residues might directly or indirectly modulate this interaction. In contrast, only a single *in vivo* substrate for the Type II (symmetrically methylating) methyltransferase is known: a major protein component of the myelin sheath, myelin basic protein.

A gene coding for a Type I methyltransferase has been identified in rat and yeast DNA (5–7) and apparent homologs have been found in human (8), mouse, sea urchin, *Caenorhabditis elegans, Pisolithus tinctorus,* rice, and *Arabidopsis thaliana* (9) DNA. The specific importance of asymmetric or symmetric protein arginine methylation in terms of a cellular function remains to be elucidated and current progress in this direction is addressed in this review. It is important to keep in mind that the metabolic cost of methylation is high (a net utilization of 12 ATPs is associated with the formation of

a methylated product) (*10*); therefore, the ultimate benefit of this reaction must outweigh its energetic cost for it to be retained through evolution.

II. Background

A. Guanidino-Methylated Arginine Residues in Proteins

Methylated derivatives of arginine residues were first identified as radiolabeled acid hydrolysis products of *in vitro* incubations containing calf thymus nuclei and S-adenosyl-[^{14}C-*methyl*]-L-methionine ([^{14}C]AdoMet) (*11, 12*). These products, initially termed "unknown I" and "unknown II," eluted from a cation-exchange amino acid analysis column just before and just after the elution position of free arginine (*11, 12*). In subsequent work, unknowns I and II were clearly shown to result from an enzyme activity termed "protein methylase I" that transfers methyl groups from AdoMet to the endogenous "sulfated histone" fraction of the nuclei (*12*). These products were tentatively identified as guanidino-methylated derivatives of arginine based on the production of [^{14}C]methylurea after treatment with 2 *N* NaOH and the inability of arginase to release urea from them (*12*). Unknown I was determined to be methylated only at the guanidino group, whereas unknown II was suggested to be methylated at both the guanidino and α-amino groups (however, see below) (*12*).

Similar conclusions were reached from an examination of the products of *in vitro* methylation reactions containing [^{14}C]AdoMet, calf thymus histones, and soluble proteins from rat uterine cell cytosol (*13*). Cation-exchange amino acid analysis of the acid hydrolysis products from the histone methylation reaction also revealed two major radioactive species, one eluting prior to and the second coeluting with the free arginine standard (*13*). Identically eluting peaks were also detected when poly(L-arginine) was used as the methyl acceptor (*13*).

In parallel studies, evidence for methylated arginine residues in proteins was obtained from the analysis of urine for novel amino acids (*14*). These species can result from the inability of cells to metabolize fully certain amino acid residues in proteins (*15*). Chemical degradation analysis, elemental composition, and nuclear magnetic resonance (NMR) spectra identified N^G,N^G-dimethylarginine (asymmetric, DMA) and N^G,N'^G-dimethylarginine (symmetric, DMA′) (Fig. 1) in human urine (*14*). The amounts of these amino acids in urine did not vary with dietary intake, suggesting that they resulted from the hydrolysis of cellular proteins (*14*). The existence of these amino acids as residues in cellular proteins *in vivo* was then confirmed from the direct analy-

sis of acid hydrolysates of bovine brain proteins as well as protein hydrolysates from various rat organs (*16, 17*). Synthetic standards (*14*), whose structures were determined by elemental analysis, NMR, and infrared (IR) spectroscopy (*16*) as well as various chromatographic techniques (*17*), were used for chromatographic comparison with the compounds isolated from protein hydrolysates. In rats, the levels of protein-incorporated asymmetric dimethylarginine ranged from 0.01 μmol/g of protein in blood to 1.67 μmol/g of protein in spleen; the amounts of the protein-incorporated symmetric dimethylarginine varied from 0.01 μmol/g of protein in blood to 0.31 μmol/g of protein in brain (*16*). Overall, the concentration of the asymmetric derivative was always greater than the symmetric, with the highest levels of both found in the small intestine, spleen, lung, liver, and brain (*2, 16*). In addition to asymmetric and symmetric dimethylarginine, N^G-monomethylarginine (the presumed biosynthetic intermediate) was also found in the protein hydrolysates from bovine brain (*16*).

Finally, the identities of the hydrolysis products from *in vitro* methylation reactions containing rat brain or liver proteins, with [^{14}C]AdoMet as the methyl donor, were then unambiguously identified as mono- as well as asymmetric and symmetric dimethylarginine by comparison to standards using various chromatographic techniques (*17*). The α-amino and δ-imino methylated derivatives previously suggested (*12*) were ruled out as potential *in vivo* products after alkaline hydrolysis and comparison of their chromatographic properties with standards (*16*). Taken together, these data indicate that the peak originally identified as unknown I (*11, 12*) was probably a mixture of asymmetric and symmetric dimethylarginine (*17*), whereas unknown II (*11, 12*) was most likely monomethylarginine (*17*).

The isolation and quantitation of methylarginine derivatives have been typically accomplished through the use of paper electrophoresis, chemical degradation, high-pressure liquid chromatography (HPLC) analysis, and cation-exchange chromotography (*2, 5, 6, 14, 16, 18, 19, 20, 21*). It is also possible to determine these residues directly in polypeptides by automated Edman sequencing (*22*). This latter method not only allows the positive determination of the methylated residue but also gives the position of the methylated residue within the sequence. Although it has not yet been used extensively in this area, electrospray mass spectroscopy is a potentially powerful tool to determine rapidly whether a given protein may be covalently modified, as well as the site and type of the modification. It is also possible that antibodies specific for one or more types of methylated arginine residues could be developed as general analytical probes, much as antiphosphotyrosine antibodies have been useful in identifying phosphotyrosine-containing proteins (*23*). For example, an antibody has been found that recognizes arginine methylated Npl3 isolated from yeast but not the unmodified recombi-

nant Npl3 expressed in *Escherichia coli* (*24*). Subjecting the bacterially expressed Npl3 to an *in vitro* protein arginine methylation reaction allowed antibody recognition of the protein. These results suggest that the antibody was specifically interacting with the methylated version of the protein, although it is not clear if part of the recognition epitope includes determinants aside from the methylarginine residues.

B. Protein Arginine *N*-Methyltransferases

An enzymatic activity initially termed "protein methylase I" and now generally referred to as protein arginine *N*-methyltransferase (EC 2.1.1.23) was first described in calf thymus (*11, 12*). This broad class of enzyme catalyzes the transfer of methyl groups from AdoMet to the terminal nitrogens (η1 and η2) of protein-incorporporated arginine residues to generate monomethylarginine as well as symmetric and asymmetric dimethylarginine (Fig. 1). After the identification of these three methylated arginine derivatives in proteins *in vivo* and *in vitro* (*16, 17*), two central questions concerning the enzymatic activities remained: (1) how many protein arginine *N*-methyltransferases are responsible for creating the three derivatives and (2) what is the range of possible methyl-accepting substrates acted on by each enzyme?

To answer these questions, a number of groups have pursued the purification of these enzyme activities using specific methyl-accepting proteins in *in vitro* assays. Since the initial partial purification (*12*), there have been at least 15 reports concerning the purification of these methyltransferases from various tissues (Table I).

Because the sulfated histone fraction from calf thymus was the first identified methyl-accepting substrate for a protein arginine *N*-methyltransferase (*12*), most of the initial purifications used histone fractions to track the activity at each step. Histone fractions f2a$_1$, f2b, and f3 as well as the purified histones H4, H2A, and H2B were also found to be good *in vitro* substrates for arginine *N*-methyltransferase activities isolated from rat organs and human placenta (*13, 25, 26*). For the histone preparations methylated by the rat enzyme, the exact methylated arginine derivative formed was not determined, only that it was dimethylarginine (*13, 25*). The human placental arginine methyltransferase, on the other hand, methylated a histone preparation (calf thymus type IIAS; Sigma) to give all three derivatives, asymmetric and symmetric dimethylarginine as well as monomethylarginine, in a ratio of 50:7:44, respectively (*26*). Several other proteins tested as potential substrates, including soy bean trypsin inhibitor, bovine serum albumin, lysozyme, cytochrome *c*, hemoglobin, RNase, and fibrinogen, were not found to be methyl acceptors (*13, 25, 26*).

In 1971, myelin basic protein was found to be a good *in vitro* methyl-accepting substrate for a protein arginine *N*-methyltransferase in guinea pig

brain extracts (*27*), and subsequent work confirmed this in mouse spinal cord extracts (*28*) and rat brain extracts (*29*). The products of these reactions were mono- and symmetric dimethylarginine (*30*). These *in vitro* methylation experiments were based on the results of the sequence and amino acid analyses of myelin basic protein from several species. Specifically, position 107 of bovine brain myelin basic protein is occupied by monomethylarginine (40–80%) and symmetric dimethylarginine (20%) (*18, 31, 32*). Sequence analysis also revealed the presence of mono- and symmetric dimethylarginine in endogenous myelin basic protein from rat brain (*33*) and chicken brain (*34*). Amino acid analysis of native myelin basic protein acid hydrolysates from several additional species, including bovine, frog, rabbit, monkey, and turtle, also showed the presence of mono- and symmetric dimethylarginine, indicating the general nature of this modification (*21, 35*). The presence of two methyl-accepting proteins with nonidentical but overlapping subtypes of methylated arginine residues suggested that cells may contain more than one type of protein arginine *N*-methyltransferase.

Biochemical evidence for the presence of more than one enzyme initially came from the differential ammonium sulfate fractionation of two protein arginine *N*-methyltransferase activities from rat brain that were more specific for either histones (now Type I) or myelin basic protein (now Type II) as methyl-accepting substrates (*36*). However, neither methyltransferase was completely separated from the other activity in this study. Interestingly, a second group (*37*) was not able to repeat this ammonium sulfate fractionation (*36*). They could not resolve the myelin basic protein methylating activity from the histone activity after a single chromatographic step that resulted in an 11-fold purification (*37*). From these data and the fact that myelin basic protein and histones both competitively inhibited the methyltransferase activity (*29*), they suggested that one methyltransferase recognized both substrates (*37*).

In later work, however, these two methyltransferase preparations (*36*) were further purified by gel filtration (*30*). Assays of acid-hydrolyzed myelin basic protein methylated by the gel filtration-purified Type II enzyme and [^{14}C]AdoMet indicated the presence of only radioactive monomethylarginine and symmetric dimethylarginine (*30*). In similar assays, the gel filtration-purified Type I activity only catalyzed the formation of radioactive monomethylarginine and asymmetric dimethylarginine in histone methyl-accepting substrates (*30*). This histone-methylating activity was also apparently free of any myelin basic protein methylation activity (*30*). It should be noted, however, that the Type II activity could still methylate histones, but the identity of the methylated arginine residue in these histones was not determined (*30*), allowing for the possibility of contamination by a Type I enzyme.

Additional evidence in favor of the multiple enzyme theory was provid-

TABLE I

PURIFICATIONS OF PROTEIN ARGININE *N*-METHYLTRANSFERASES

Year of purification	Ref.	Enzyme source	Fold purification	Substrate[a]	AdoMet K_m (μM)	pH optimum	Molecular mass[b]	Methylated residues[c]
Types I and II								
1968	*12*	Calf thymus	34	Histones	21	7.4	ND[d]	Guanidino-methylated arginine
1971	*13*	Rat organs	ND	Histones/ (poly)Arg	10	8.5	ND	Methylated arginine
1971	*25*	Rat thymus	35	Histones	ND	8.1	150kDa	Guanidino-methylated arginine
1975	*30*	Rat brain	250	Histones	ND	ND	ND	DMA, MMA
1978	*52*	Krebs II ascites cells	150	Histones	2.5	ND	500	DMA, MMA
1986	*39*	Bovine brain	52	Histones	ND	ND	ND	DMA, DMA′, MMA
1988	*40*	Calf brain	ND	Histone	8	ND	275 kDa	ND
1991	*26*	Human placenta	400	Histones	5	7.4–7.6	ND	DMA, DMA′, MMA
1994	*48*	Rat liver	195	hnRNP A1	6.3	7.6	450 kDa	DMA, MMA
1995	*49*	HeLa cells	550	hnRNP A1	5.8	7–7.4	450 kDa	DMA, MMA
1971	*27*	Guinea pig brain	ND	Myelin basic protein	ND	ND	ND	DMA, MMA

1974	*37*	Rat brain	6–11	Histones/myelin basic protein	ND	ND	ND	DMA, MMA
1975	*30*	Rat brain	170	Myelin basic protein	ND	ND	ND	DMA, MMA
1977	*38*	Calf brain	119	Histones/myelin basic protein	7.6	7.2	ND	DMA, DMA′, MMA
1988	*40*	Calf brain	404	Myelin basic protein	4.4	ND	500 kDa	DMA, MMA
1988	*41*	Bovine brain	ND	Myelin basic protein	7.6	ND	72 kDa[e]	ND
Type III								
1982	*50*	Wheat germ	90	Histone	5.7	9	ND	DMA, MMA
Type IV								
1985	*53*	*Euglena gracilis*	52	Cytochrome *c*	40	7	36 kDa	MMA

[a]Best substrates are listed.
[b]Determined ωy gel filtration.
[c]DMA, N^G,N^G-dimethylarginine; DMA′, N^G,N'^G-dimethylarginine; MMA, monomethylarginine.
[d]ND, not determined.
[e]Determined by SDS-PAGE.

ed by the partial purification of the Type I arginine methyltransferase from calf brain (*38*). For each active methyltransferase pool generated during the purification, the relative efficiency of histone and myelin basic protein methylation was also monitored at each step of the purification. In addition, the amount of radioactivity incorporated into each of the three methylated derivatives of arginine was determined using only histones as the methyl-accepting substrate. Throughout the purification of the Type I enzyme, the ratio of histone to myelin basic protein methylation efficiency increased from 1.6 to 8.9 (*38*), indicating the existence of at least two enzyme activities. In this same purification, the ratios of the methylated arginine products in histones did not vary greatly (DMA:DMA′:MMA = 40:5:55). From these data, the authors therefore suggested that although there are different methyltransferases recognizing different substrates, a single protein arginine *N*-methyltransferase could form all three methylated arginine derivatives. Presumably the type of posttranslationally modified arginine formed would be dependent on the local context (sequence or structure) of the arginine residue within the substrate protein, or perhaps on the presence of regulatory subunits associated with the methyltransferase.

Perhaps the best evidence for the existence of at least two different types of protein arginine methyltransferases in a single tissue came from the partial purification of a Type I methyltransferase from bovine brain capable of methylating histones and not myelin basic protein (*39*), and from the separate purifications of a myelin basic protein-specific and a histone-specific methyltransferase from calf brain (*40*) (Table I). The Type I and Type II enzymes isolated from calf brain were found to have differential stabilities to treatment with heat, *p*-chloromercuribenzoate, and guanidine-HCl (*40*); these results clearly distinguish the two enzymes from each other. Furthermore, the Type I enzyme did not methylate myelin basic protein (*40*). However, the Type II enzymes from two separate preparations (*40, 41*) had some activity toward histones, but the K_m was three orders of magnitude greater for histones than for myelin basic protein and the V_{max} was, interestingly, twofold higher for histones than for myelin basic protein as methyl acceptors (*40*). It is also possible that it is solely the affinity of substrate binding that leads to the specificity of methylation by the Type II protein arginine *N*-methyltransferase and that this enzyme is able to methylate Type I substrates at a low efficiency. Surprisingly, in this study the identities of methylated arginines incorporated into the histones by either enzyme were not determined (*40*). However, the myelin basic protein substrate was found to be specifically mono- and symmetrically dimethylated (80:20) by the Type II enzyme, and the asymmetric dimethylarginine derivative was not detected (*40*) (Table II).

The purified calf brain Type I methyltransferase has a native molecular

TABLE II
PRESENCE OF VARIOUS ISOFORMS OF METHYLATED ARGININE IN SEVERAL SUBSTRATES FROM DIFFERENT ENZYME PREPARATIONS

Year of purification	Ref.	Substrate	Asymmetric dimethylarginine	Symmetric dimethylarginine	Monomethylarginine
1975	*30*	Histones	+[a]	ND[b]	+
1977	*38*	Histones	32%	5%	63%
1978	*52*	Histones	+	ND	+
1982	*50*	Histones	ND	+	+
1986	*39*	Histones	19%	13%	69%
1991	*26*	Histones	50%	7%	44%
1994	*48*	Histones	51%	ND	49%
		hnRNP A1	41%	ND	59%
1994	*5*	hnRNP A1	+	ND	+
1994	*6*	Histones	+	ND	+
		hnRNP A1	+	ND	+
1995	*49*	hnRNP A1	+	ND	+
1971	*27*	Myelin basic protein	ND	+	+
1975	*30*	Myelin basic protein	ND	+	+
1988	*40*	Myelin basic protein	ND	20%	80%
1990	*45*	Myelin basic protein	?[c]	?[c]	+

[a]Presence of residue, not quantitated in reference.
[b]ND, none detected.
[c]Only dimethylarginine determined.

mass of 275 kDa, estimated by gel-filtration chromotography, and this enzyme activity is associated with two major polypeptides of 110 and 75 kDa by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (*40*). The purified calf brain Type II myelin basic protein-specific enzyme was determined to be about 500 kDa by gel-filtration analysis, and includes two polypeptides of 100 and 72 kDa by SDS-PAGE (*40*). However, another purification of the Type II enzyme suggested a single polypeptide with a molecular mass between 71 and 74.5 kDa (*41*). These data are consistent with both methyltransferases being multimers, although the subunit structures of this complex are far from clear at this point. The recently cloned mammalian Type I protein arginine *N*-methyltransferase gene codes for a polypeptide of approximately 40 kDa that has enzymatic activity by itself (*6*). This raises the possibility that the higher molecular weight species identified in both preparations by SDS-PAGE were contaminants, although these larger polypeptides may have regulatory roles similar to those seen in other enzymes (*42*).

In 1977, protein components of the 40S ribonuclear complex were also found to contain methylated arginine residues *in vivo* (*43, 44*). The polypeptides, heterogeneous nuclear ribonuclear proteins (hnRNPs) A1 and A2, contain almost exclusively the asymmetric dimethylarginine derivative with a minor amount of monomethylarginine, whereas none of the symmetric derivative was detected (*43, 44*). Bovine brain Type I and Type II arginine methyltransferase preparations were unable to methylate recombinant hnRNP A1; therefore, this methylation was at first considered to be of a third class (*45*). However, Type I methyltransferase preparations from rat liver and calf brain were later shown to methylate recombinant hnRNP A1 much more efficiently than histones *in vitro* (*46–48*). The K_m for hnRNP A1 is 100-fold lower and the V_{max} is at least sevenfold higher than for histones (*47*). Recombinant hnRNP A1 overexpressed in *E. coli* is a much better substrate than native hnRNP A1 protein, presumably because the latter is already nearly fully methylated (*46*). Analysis of the arginine methylation products in the recombinant hnRNP A1 after an *in vitro* reaction showed the presence of only mono- and asymmetric dimethylarginine (*48*). The fact that the methylated arginine residues found in hnRNP A1 *in vivo* and after *in vitro* reactions with a Type I methyltransferase preparation were the same as those identified from *in vitro* reactions with histones led to the conclusion that the Type I histone and hnRNP A1 methyltransferases were the same (*46, 47*). This substrate has since been used in two additional purifications of protein arginine *N*-methyltransferases (*48, 49*).

The Type I rat liver hnRNP A1 methyltransferase has an estimated molecular mass of 450 kDa by gel filtration, and SDS-PAGE analysis of the purest fraction shows an apparent single polypeptide at 110 kDa (*48*). The enzyme has maximal activity toward recombinant hnRNP A1, with relative activities of only 23 and 11% for myelin basic protein and histones, respectively (*48*). Although myelin basic protein was apparently a substrate for the purified material, amino acid analysis of the hnRNP A1 and histone substrate demonstrated the presence of only monomethylarginine and asymmetric dimethylarginine, with none of the symmetric form (*48*). These methylated products are consistent with the activity of the Type I enzyme, because myelin basic protein does not contain any asymmetric dimethylarginine (*31, 32*). It is surprising that amino acid analysis or tryptic peptide microsequencing was not done on the methylated myelin basic protein in these experiments to distinguish between the possibilities that the preparation was contaminated with a Type II methyltransferase, that the myelin basic protein was contaminated with a substrate for the Type I enzyme, or that myelin basic protein could be methylated at a site distinct from Arg-107 by the Type I enzyme (*48*).

Subsequently, a more purified form of an hnRNP A1 methyltransferase

from human ovarian carcinoma HeLa cells was prepared and its specificity for methylating recombinant hnRNP A1, histones, but not myelin basic protein was determined (*49*). This methyltransferase was apparently not purified to homogeneity because the purest fraction contained at least eight polypeptide bands by silver-stained SDS-PAGE analysis (*49*). The two most prominent bands were at 100 and 45 kDa (*49*). As with the rat liver enzyme (*48*), the HeLa cell methyltransferase also chromatographed as a 450-kDa polypeptide on gel filtration (*49*). Again, it is not clear which, if any, of these polypeptides observed in the final stages of the purifications are actually part of the native enzyme.

The above discussion has focused on the two major mammalian classes of identified protein arginine *N*-methyltransferase activities: the Type I asymmetric, hnRNP A1/histone-methylating enzyme and the Type II symmetric, myelin basic protein-specific enzyme. Other activities have been characterized from nonmammalian sources. For example, a distinct histone H4-specific methyltransferase has been characterized in plants (*50, 51*). A partially purified wheat germ protein arginine *N*-methyltransferase preparation was found to methylate endogenous histones, giving solely mono- and symmetric dimethylarginine residues (*50*). In comparison, the Type I enzyme generally catalyzes the formation of only monomethylarginine and asymmetric dimethylarginine residues when incubated with histones (*52*) (Table II). The wheat germ enzyme was also incapable of methylating myelin basic protein whose only methylated arginine residue is found as either monomethylarginine or symmetric dimethylarginine (*50*). Together with the fact that endogenous histones from wheat germ but not those from calf thymus contain symmetric dimethylarginine residues (*50*) and given the difference in the nature of the *in vitro*-methylated reaction products, these results support the conclusion that the wheat germ enzyme is distinct from its mammalian counterpart, and we designate it Type III.

A fourth type of enzyme was purified from the photosynthetic protozoan *Euglena gracilis* (*53*). This methyltransferase specifically monomethylated a single arginine residue at position 38 in cytochrome *c* from horse heart (*53*). The enzyme did not methylate histones to a great extent, but did methylate myoglobin at over half the velocity seen with cytochrome *c* (*53*). The native size of this Type IV enzyme is estimated to be 36 kDa by gel filtration (*53*) and is therefore much smaller than the native mammalian enzymes. The endogenous methyl-accepting substrates for this enzyme have not been characterized.

In summary, at least four types of protein arginine methyltransferases have been identified in various eukaryotic species, ranging from the protists, such as *Leishmania* sp. (*3*), to higher organisms, including humans (*26*). Protein arginine methylation has, however, never been observed in bacterial

species. Liu and Dreyfuss (*49*) could not detect any Type I activity from *E. coli* using recombinant hnRNP A1 as the methyl-accepting substrate, although they also could not detect the activity in *Saccharomyces cerevisiae*, which was subsequently shown to have a Type I methyltransferase (*5, 7*). Additional evidence for the lack of a Type I protein arginine *N*-methyltransferase in *E. coli* comes from the fact that recombinant hnRNP A1 is a better substrate than its endogenous form, presumably because it is fully unmethylated (*46*). Despite the number of purifications published on this activity from eukaryotic organisms, no peptide sequence results have been presented. Therefore definitive proof of two or more distinct activities, either in substrate or methylated product specificity, awaits the expression of recombinant forms of all of these enzymes with purified substrates and amino acid analysis of the reaction products.

III. Identification of Protein Arginine *N*-Methyltransferase Genes

A. Mammalian PRMT1

The first mammalian gene encoding a protein arginine *N*-methyltransferase was identified from a yeast two-hybrid screen (*54, 55*) designed to identify proteins that could interact with murine TIS21, an immediate-early/primary response gene product (*56–58*), and the closely related murine BTG1 gene product (*59, 60*). From a rat cDNA library, a single clone, 3G, was found to interact specifically with both bait proteins in the two-hybrid screen as well as in *in vitro* binding experiments (*6*). A basic local alignment search tool (BLAST) search through the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) (*61*) queried with the 3G sequence against the nonredundant protein database identified the *E. coli* ribosomal protein L11 methyltransferase as the highest scoring alignment for a protein with a known function (*6*). This bacterial methyltransferase is thought to methylate an amino-terminal alanine residue as well as an internal lysine residue (*62, 63*). The region of similarity between this protein and 3G is relatively short (29 amino acids) but encompasses two highly conserved regions common among a majority of soluble methyltransferases and AdoMet-binding proteins (*64*). These regions, I and post I, are thought to be involved in the binding of AdoMet, and recent crystallographic evidence on a catechol *O*-methyltransferase and two DNA methyltransferases demonstrates a direct interaction between these regions and AdoMet (*65*). In addition to region I and post I, 3G also was found to have two additional motifs common to methyltransferases (II and III) (*64*)(Fig. 2). From this sequence homology data, it was likely that 3G was also a methyltransferase.

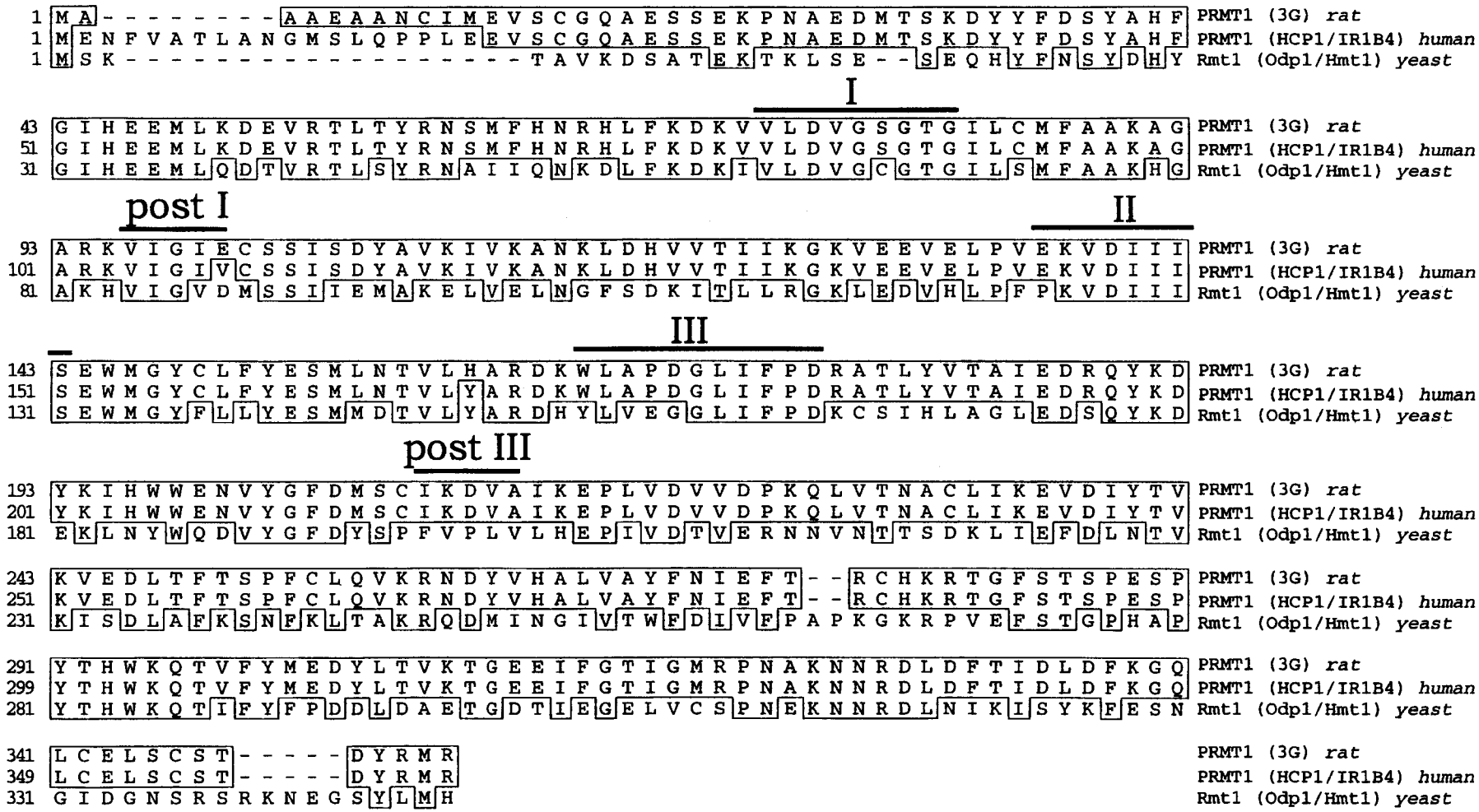


FIG. 2. Protein sequence alignment of three Type I protein arginine *N*-methyltransferases identified from the rat (GB U60882), human (GB D66904, corrected, and Ref. *8*), and yeast (GB Z35903). Those residues identical to the rat PRMT1 protein sequence are boxed. The conserved AdoMet binding regions (I, post I, II, III, and post III) common to the majority of soluble methyltransferases and AdoMet-binding proteins (*64*) are indicated by an overline.

Using an amino-terminal glutathione S-transferase fusion (GST) to 3G, various protein substrates were tested as possible substrates (*6*). Although the fusion protein was not able to methylate isoaspartyl-containing peptides (*66*) or the catalytic subunit of protein phosphatase 2A (*67, 68*), it was able to methylate peptide R1 (*69*), which contains a single methylation site for the Type I protein arginine *N*-methyltransferase (Section V,A) (*6*). Although the peptide was only monomethylated on the arginine residue, analysis of *in vitro*-methylated histones and recombinant hnRNP A1 demonstrated the presence of both mono- and asymmetric dimethylarginine (*6*). These results prove that the 3G cDNA, derived from a rat mRNA, encodes a Type I protein arginine *N*-methyltransferase, or at least its catalytic subunit. The suggestion that the hnRNP A1 and the histone arginine methyltransferases are identical is also strongly supported by these results: the identical methylated arginine products are formed from a methylation reaction with recombinantly expressed and purified GST–3G and either of the substrates. However, the affinity of the methyltransferase fusion protein for hnRNP A1 is much greater than for crude histones (*6*). In addition, myelin basic protein and cytochrome *c* were not substrates for the fusion protein. The gene encoding this protein has been termed PRMT1 (Protein aRginine *N*-Methyl Transferase. The PRMT1 protein has a calculated polypeptide molecular mass of 40.5 kDa and elutes as a native 180-kDa polypeptide species by gel-filtration chromotography (*6*). Recent results with the yeast two-hybrid system indicate that PRMT1 is able to at least dimerize with itself (J. Tang and H. Herschman, personal communication) and it is possible that the native enzyme is a homotetramer, although more complex subunit structures cannot be eliminated at this point. Northern analysis of the PRMT1 transcript shows that it is present in all tissues tested, and in RAT1 fibroblast cells it is constitutively expressed and its mRNA levels are not up-regulated after mitogen stimulation (*6*).

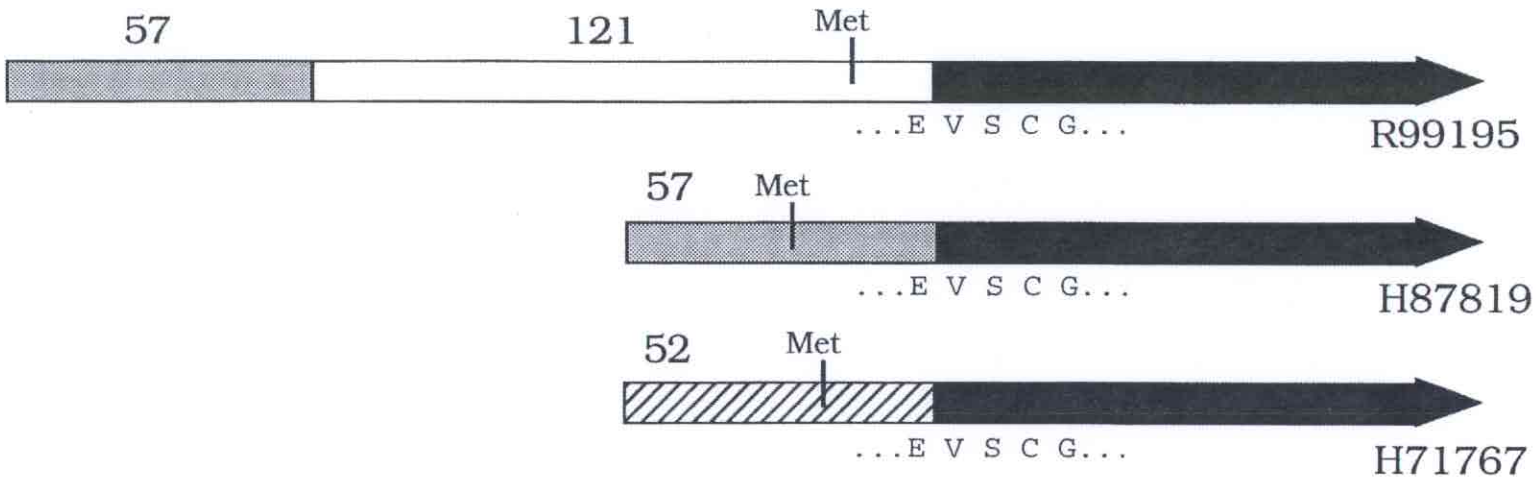
Human homologs of PRMT1 have recently been identified in two separate studies (*8, 9*). The human HCP1 (highly conserved protein) was identified as a multicopy cDNA suppressor of a yeast *ire15* mutant (*9*). This mutation causes a decrease in the expression of the inositol 1-phosphate synthase and inositol transporter genes (*70*). Suppressors of the *ire15* mutation allowed growth on inositol-free minimal agar plates (*9*). The published sequence of the human HCP1 is very similar to the rat sequence, except for a region between amino acids 147 and 175 that could be the result of a frameshift from sequencing errors (*8*). Indeed, an insertion of an A at nucleotide position 468 and an insertion of the dinucleotide GC at position 549 establishes a new open reading frame that is 96% identical in amino acid sequence to the rat PRMT1 protein (Fig. 2). The possible activity of HCP1 as a methyltranferase, however, was not assessed in this report (*9*).

A second human PRMT1 homolog was identified in a yeast two-hybrid screen designed to find proteins that interact with the intracytoplasmic domain of the interferon-α,β cytokine receptor (*8*). This human PRMT1 homolog was designated IR1B4 and is identical in sequence to the corrected HCP1 sequence (Figure 2)(*8*). The methyltransferase activity of human PRMT1 (IR1B4) was demonstrated by the fact that a histone-methylating activity could be coimmunoprecipitated using antibody to the interferon receptor IFNAR1 (*8*). In addition, extracts from a human myeloma cell line grown in the presence of an antisense construct of human PRMT1 (IR1B4) had an inhibitory effect on the level of endogenous protein arginine *N*-methyltransferase activity (*8*).

The relationship of the PRMT1 methyltransferase to the previously partially purified Type I histone and hnRNP A1 methyltransferases shown in Table I is not clear at this point. The native enzyme size of the PRMT1 enzyme appears to be smaller (180 kDa versus 275–450 kDa) than that observed in the previous preparations, and the catalytic unit has a lower molecular mass (40.5 kDa) than even the smallest of the major polypeptide species identified in any of the purifications (Section II,B).

In addition to the human PRMT1 homolog experimentally identified by both Nikawa *et al.* (*9*) and Abramovich *et al.* (*8*), searches of the database of expressed sequence tags (dbEST at NCBI) revealed three potential splicing variants of the human PRMT1, represented by human cDNA clones R99195, H87819, and H71767 (Fig. 3). We obtained and fully sequenced the insert within each of these clones (J. Gary, unpublished). With the exception of the amino terminus, the sequences of the human clones are nearly identical to the rat PRMT1 sequence. In fact, the differences seen may represent polymorphisms or errors in library construction. For example, clone H87819 (identical to the human sequences HCP1 and IR1B4) encodes a Met instead of Phe at amino acid position 77, whereas clone R99195 has a Val replacing a Leu at position 322, and clone H71767 has two Asp residues replacing two Glu residues at positions 335 and 348. The most significant differences between the human variants and the rat PRMT1 are found at the amino terminus (Fig. 3). From Fig. 3, the relatedness of R99195 to H87819 by alternative splicing is apparent. H87819 can be derived by splicing out the 121-bp segment of R99195, but the origin of the 5′ fragment of H71767 is less evident.

We have expressed all three variants as GST fusion proteins. All are able to methylate an artificial substrate, GST–GAR (J. Gary, unpublished). This substrate was constructed by fusing the coding region of GST (from pGEX-2T; Pharmacia) in-frame with the amino-terminal region of human fibrillarin (amino acids 1–148) (Section IV,B,1) (R. Kagan and J. Gary, unpublished). This region of fibrillarin is known to contain asymmetric dimethylarginine *in*




```
1  M V G V A - - - - - - - - - - - - - - - E V S C G Q A E S S E K P N A E D M T S K  R99195 protein
1  M E N F V A T L A N G M S L Q P P L E E V S C G Q A E S S E K P N A E D M T S K  H87819 (HCP1 & IR1B4) protein
1  M A A A E A A N C I M - - - - - - - - - E V S C G Q A E S S E K P N A E D M T S K  H71767 protein
```

FIG. 3. Three alternative splicing variants of the human Type I protein arginine *N*-methyltransferase identified in the database of expressed sequence tags. The upper panel is a schematic diagram of the different mRNA species; regions of identity are similarly shaded. The proposed initiator methionines (Met) were assigned because they are the most 5′-AUG codons that are in-frame with the methyltransferase sequence that follows. A translation of the sequence near the splice junction where the three forms are identical is shown under each mRNA species. The lower panel is a protein sequence alignment of the translated amino termini from the three splice variants. The H87819 spliced form has also been identified in two separate screens as IR1B4 and HCP1 (*8*, *9*).

vivo (*71*) and is therefore a substrate for the Type I protein arginine *N*-methyltransferase PRMT1. The reason for the three differently spliced forms of the human PRMT1 *in vivo* is not clear. Experiments using hypomethylated RAT1 fibroblast extracts (AdOx treated) (*6*) together with the different forms of the methyltransferase have not yet revealed any variation in methyl-acceptor substrate specificity (J. Tang and J. Gary, unpublished). Perhaps the different 5′ untranslated regions of each of the different mRNA transcripts play an important role.

In theory, the EST database should reflect the abundance of certain transcripts *in vivo:* common mRNAs should be randomly cloned proportionately more than rarer transcripts. Therefore, it is interesting to note that the H87819 splice variant has been identified twice in the literature (as HCP1 and IR1B4) in screens for specific functions but the other two clones have not, because the H87819 and R99195 amino termini are unique in dbEST and the H71767 form is present in at least five other clones.

Another study identified a rat gene encoding a protein with sequence homology to the rat PRMT1 between the methyltransferase regions I and III, but that was divergent outside of these regions (*248*). The gene is currently designated PRMT3. A GST–PRMT3 fusion enzyme can methylate GST–GAR (*248*), which classifies it as another Type I protein arginine *N*-methyltransferase.

An additional potential human arginine *N*-methyltransferase gene has been proposed based on sequence similarity alone and has been designated hHMT2 and PRMT2 (GB X99209, H. Scott, unpublished; GB U80213, N. Katsanis, unpublished). We had also identified this gene from BLAST searches of dbEST as human clones T75034 and T77642 (J. Gary, unpublished). The sequence of this potential methyltransferase (GB X99209 and U80213) is more distantly related to the rat PRMT1 than is the rat PRMT3 (27% amino acid identity between PRMT1 and PRMT2; 44% identity between PRMT1 and PRMT3), but the sequence within the conserved methyltransferase regions is very similar. The hHMT2/PRMT2 gene product has an additional interesting sequence feature: the amino-terminal portion has an SH3 domain, which is present in many signal transduction proteins and is known to bind a polyproline motif (*72*). We have expressed this protein as a GST fusion and have not detected any activity toward the synthetic Type I methyltransferase substrate GST–GAR, proteins present in the *rmt1* "hypomethylated" mutant yeast cytosol, nor myelin basic protein, the Type II substrate (A. Frankel and J. Gary, unpublished). Therefore it seems premature to designate this species as a protein arginine *N*-methyltransferase.

B. *Saccharomyces cerevisiae RMT1*

On identification of the rat PRMT1 as a Type I protein arginine *N*-methyltransferase, BLAST searches against the nonredundant protein database re-

vealed a homolog in the yeast *S. cerevisiae* with 45% amino acid identity, the hypothetical *ODP1* gene product (5). *ODP1* was originally identified as an open reading frame downstream of the gene *PDX3* (73), and then was later resequenced through the yeast genome sequencing effort (yeast open reading frame *YBR0320*) (74) (Fig. 2). It encodes a potential protein product with a calculated molecular mass of 39.8 kDa and contains the conserved methyltransferase regions I, post I, II, and III, as well as the post III region that is not found in PRMT1 or its mammalian homologs. This sequence similarity, combined with biochemical data on the rat PRMT1 gene product, strongly implicated the *ODP1* gene product as a Type I protein arginine *N*-methyltransferase as well. A previous analysis of yeast ribosomal proteins indicated the presence of methylated arginines *in vivo* (75). However, a recent attempt to detect a Type I arginine *N*-methyltransferase activity in yeast was unsuccessful (49).

The ability of a GST–Odp1 fusion enzyme to methylate the synthetic R1 peptide substrate (69) was the first experimental indication that Odp1 is a protein arginine *N*-methyltransferase (5). The fusion protein was also able to mono- and asymmetrically dimethylate histones, hnRNP A1, and to a lesser degree cytochrome *c* and myoglobin, but did not methylate myelin basic protein at all (5) (J. Gary and M. Yang, unpublished). We thus renamed *ODP1*, calling it *RMT1*, for protein aRginine *N*-Methyl Transferase, to signify its ability to mono- and asymmetrically dimethylate a broad range of substrates. *In vivo* and *in vitro* experiments using a strain in which the chromosomal copy of *RMT1* was disrupted showed that the mutant strain was viable but highly defective in the ability to form *N*-methylated arginine derivatives, whereas the parent strain produced only mono- and asymmetric dimethylarginine (5).

The *in vitro* methylation of cytosolic extracts from both the parent and mutant cells with [^{3}H]AdoMet allowed at least five proteins lacking these arginine derivatives to be visualized by SDS-PAGE analysis and fluorography (5). In addition, incubating the GST–Rmt1 fusion with cytosol from this mutant strain resulted in the specific mono- and asymmetric dimethylation of at least 13 polypeptides (5). Similar experiments with *in vivo*-labeled extract from each strain show that at least four polypeptide species (55, 42, 38, and 29 kDa) are no longer methylated in *rmt1* mutants (J. Gary, unpublished). This is a rather broad substrate specificity compared to the mammalian GST–PRMT1 fusion protein that recognizes only a single 55-kDa species in the mutant yeast extract (presumably Npl3 by SDS-PAGE migration) as well as 55- and 65-kDa polypeptides in untreated RAT1 fibroblast cytosolic fractions (5,6). This difference in substrate specificity might be explained if mammalian cells have multiple Type I methyltransferases with varying specificities instead of one enzyme that recognizes the majority of the substrates (6).

Rmt1 is responsible for over 80% of the *in vivo* and *in vitro* asymmetric arginine dimethylating activity in yeast (5). Approximately 65% of the *in vivo* monomethylating activity can be attributed to Rmt1 (5). We are currently attributing this significant *in vitro* Rmt1-independent activity to a less abundant protein arginine *N*-methyltransferase enzyme that may be released during homogenization from a cellular compartment, where it would not normally encounter such high concentrations of hypomethylated proteins (5). The protein responsible for this activity has not been identified.

RMT1 was also identified in a screen for mutants that caused synthetic lethality with a temperature-sensitive mutant allele of *NPL3* in yeast (7). The Npl3 protein has been implicated in various processes (24), including nuclear protein import (76), pre-rRNA processing (77), as well as binding and shuttling mRNA out of the nucleus (78, 79). Null mutants of *NPL3* are not viable (76); therefore, a partially impaired temperature-sensitive mutant was used to screen for mutants that might further exacerbate the phenotype of the *npl3* temperature-sensitive mutant, indicating either that the proteins have some physical interaction or are present in the same or overlapping pathways (7). The screen yielded four mutants, all in a single complementation group designated *SLN1* (synthetically lethal with *NPL3*). The *SLN1* gene was identified by complementation and found to be identical to *RMT1* (7). Importantly, a null disruption mutant of *RMT1/SLN1* is viable and does not have any of the phenotypes associated with a *npl3* mutant (7). It was shown that immunoprecipitated yeast Rmt1/Sln1 as well as recombinant Rmt1/Sln1–Myc methylated Npl3 and recombinant hnRNP A1 *in vitro* (7). With this information, the authors renamed *SLN1,* calling it *HMT1* (hnRNP methyltransferase) (7), although it seems more reasonable to now identify it as simply *RMT1* because of its broad substrate specificity. Despite the fact that the type of methylated residue formed by Hmt1 was not determined, analysis of the CNBr cleavage products of methylated Npl3 indicated that the site of methylation was in the carboxy-terminal region of Npl3 (7, 24). This region of Npl3 has a glycine- and arginine-rich (GAR) domain (80) containing the consensus methylation site for Type I protein arginine *N*-methyltransferases (Section V,A).

The subcellular localization of the Rmt1/Hmt1–Myc construct was found to be nuclear, as determined by indirect immunofluorescence, and is consistent with the presence of a weak bipartite nuclear localization signal (7). Preliminary immunofluorescence data indicate that the rat PRMT1 is present in the nucleus, as well (J. Tang and H. Herschman, personal communication). This result is interesting because PRMT1 was identified by its association with a protein, TIS21, that is completely cytosolic (58), and IR1B4/hPRMT1 was identified by its interaction with the cytoplasmic tail of a plasma membrane receptor (8). Furthermore, some of the early purification studies indicated a cytosolic localization (13, 37).

C. Reconciling the Purification and Cloning Data

The past three decades of research concerning the identities of protein arginine *N*-methyltransferases in different mammalian tissues have provided a database that is often conflicting and convoluted. From the purification data alone, there appear to be at least two major mammalian enzymes (Section II,B). However, there are discrepancies among the results of the partially purified methyltransferase preparations, and it is also unclear how the recently constructed GST fusion enzymes are related to these activities.

The Type I asymmetric arginine methyltransferase activity originally corresponded to a histone methyl-accepting activity and was also considered by some to be distinct from the Type II symmetric myelin basic protein-specific methyltransferase based on this substrate preference alone (*30, 38, 40*). However, data from various purifications of the Type I histone methyltransferase showed that arginines in histones were methylated to all three derivatives (Table II). This could be due to the substrate directing different types of methylation from the same enzyme due to the sequence or structure surrounding the methylation site (*45*). But in light of the evidence from experiments with more highly purified enzyme and substrate (*5, 6, 49*), the presence of symmetric dimethylarginine in histones may be due instead to contamination with the Type II (symmetric) myelin basic protein methyltransferase in each of these partial purifications and the subsequent action of both enzymes on histones. Histones do contain arginine residues in several different contexts as potential methylation sites and may allow varying degrees of catalysis by both types of enzymes. Myelin basic protein-specific methyltransferase preparations do methylate histones *in vitro*, but always to a lesser degree (*40, 41*). It will be very important in future work to determine whether histones are actually being methylated by the Type II enzyme and if these histones contain mostly symmetric or asymmetric dimethylarginine residues. From these results, it may be possible to distinguish between the presence of a contaminating Type I methyltransferase in Type II preparations or whether the Type II enzyme can also recognize symmetric methylation sites within histones.

Both the Type I asymmetric (RNA-binding protein-specific) and Type II symmetric (myelin basic protein-specific) arginine methyltransferases appear to exist in larger complexes than their apparent constituent polypeptide chains would suggest. By gel-filtration analysis, the masses for the enzymes range from 150 to 500 kDa (Table I). In our identification of the rat PRMT1 enzyme from RAT1 fibroblast cells, we found that the native methyltransferase eluted from a Superdex S200 gel-filtration column as a 180-kDa species (*6*). The native yeast enzyme, Rmt1, eluted at an approximate mo-

lecular mass of 74 kDa by Superdex S200 gel-filtration analysis (J. Gary, unpublished), suggesting it may be a homodimer.

At this point we cannot resolve the discrepancies seen in the native molecular masses of the mammalian enzymes. No attempts have been made so far to compare directly the gel filtration behaviors of the different enzyme preparations shown in Table I, and there is enough uncertainty in the experimental descriptions that it is difficult to say whether the enzymes are of similar size. For example, Gallwitz (25) did not mention what protein standards were used in order to make the molecular mass calculation of 275 kDa for the rat thymus enzyme. Ghosh *et al.* (*40*) did not run all of the standards simultaneously, but rather over several column runs, and stated that the runs were variable by as many as three fractions. Liu and Dreyfuss (*49*) did not show the data for their determination of the molecular mass (450 kDa), and Rawal *et al.* (*48*) did not run the standards with the sample. Finally, we used endogenous rat proteins for the column calibration and therefore ran the sample and the standards simultaneously (*6*). We, however, did not include a standard over 161 kDa and may have therefore underestimated the mass (*6*).

Another problem has been attempting to assign those polypeptides identified in the mammalian purifications with the protein arginine *N*-methyltransferase gene products that have been recently characterized. The molecular masses of the polypeptides remaining in the active pool after purification range from 45 to 110 kDa by SDS-PAGE analysis (*40, 48, 49*); however, both the mammalian and yeast proteins migrate on SDS-PAGE as approximately 40-kDa species (*5, 6*). It is clear from the methylation experiments using the GST fusion enzymes with homogeneous substrates such as recombinant hnRNP A1 or GST–GAR that PRMT1 and Rmt1 are the catalytic units of a Type I protein arginine *N*-methyltransferase.

The additional higher molecular weight polypeptides seen in many of the purifications may be contaminating proteins and/or regulatory or additional catalytic subunits. As previously mentioned, our gel-filtration and SDS-PAGE analysis data are consistent with the native rat PRMT1 being a homotetramer (homodimer for yeast), but the larger, more complex structures suggested by the results of several purifications cannot be excluded until experiments such as quantitative native immunoprecipitation of the methyltransferase can be done. It is also possible that the situation is more complex. The PRMT1 product may be a minor species within cells, and the genes encoding the actual catalytic subunits of the Type I and Type II enzymes identified in the purifications (Table I) may have not yet, in fact, been identified. This seems unlikely given that the substrate specificities of the GST–PRMT1 fusion protein and the Type I enzyme appear to be very similar, but further work is

needed in this area. Even the designations Type I and Type II are tenuous, because they are based solely on substrate and reaction product specificity. These classes are not mutually exclusive. In fact, the Type I and Type II methyltransferase complexes could feasibly differ only in their subunit organization. A similar scenario has been suggested for the phosphatase PP2A, which has a complex and interchanging subunit structure, multiple substrates and seems to affect various processes at different points in the cell cycle (*81–83*). Although much of the current progress in this area has been through the identification of new arginine methyltransferases from database searches or yeast two-hybrid analyses, it may be more practical in the future to follow the trail backward from newly or previously identified substrates.

IV. Methyl-Accepting Substrates for Protein Arginine *N*-Methyltransferases

A. General

The first *in vitro* methyl-accepting substrates for protein arginine *N*-methyltransferases were crude as well as fractionated histones from calf and rat thymus (*12, 13, 25*). However, *in vitro* methylation reactions with endogenous rat brain proteins showed that although the majority of the incorporation was into methylarginine derivatives, only 20% of the incorporation was into the crude histone fraction (*17, 84*). These results indicate that histones, although apparently they can be methylated *in vitro,* are not necessarily the major *in vivo* substrates.

The identification of endogenous arginine methylation substrates is not straightforward, because the *in vivo* methyl-accepting sites may already be partially or fully occupied in tissue extracts and in purified proteins, reducing the incorporation of radioactivity from radiolabeled AdoMet in the *in vitro* assays. For instance, the myelin basic protein from carp, which is not methylated *in vivo* (*35*), is a fivefold better substrate than myelin basic protein from cows and is 17-fold better than myelin basic protein from rabbit in terms of the [^{14}C]methyl group incorporation into the protein from [^{14}C]AdoMet (*45*). Nonetheless, specific polypeptide substrates can be detected from cell lysates by fluorography after *in vitro* labeling with radiolabeled AdoMet and a partially purified methyltransferase preparation. For example, several substrates from human placental cell cytosol and nuclei were identified using a Type I protein arginine *N*-methyltransferase partially purified from the same cells (*26*).

By incubating rat pheochromocytoma PC12 cells with adenosine dialdehyde, a specific inhibitor of S-adenosyl-L-homocysteine hydrolase, it was pos-

sible to increase the proportion of unmethylated substrates within the cells by causing an increase in the intracellular concentration of S-adenosylhomocysteine, a product inhibitor of all AdoMet-dependent methyltransferases (85). About 72% of the endogenously incorporated methyl groups from an adenosine dialdehyde-treated homogenate incubated with [^{3}H]AdoMet could be accounted for by asymmetric dimethylarginine alone (85). Similarly, specific substrates such as hnRNP A1 were identified in adenosine dialdehyde-treated RAT1 fibroblast cells incubated with [^{3}H]AdoMet and the GST–PRMT1 fusion (6). In this experiment, hnRNP A1 was immunoprecipitated from the methylation reaction containing adenosine dialdehyde-treated cytosol by either a specific monoclonal antibody or a control antibody. Analysis of protein in the immunoprecipitated pellet by SDS-PAGE and fluorography indicated that a methylated band corresponding to the migration position of hnRNP A1 was present only in the lane using the anti-hnRNP A1 antibody and not the control antibody (6). Rather than using adenosine dialdehyde with yeast, it is possible to obtain hypomethylated cytosol through genetic manipulations; yeast *rmt1* null mutants are specifically devoid of a majority of asymmetric dimethylarginine derivatized proteins (5). Therefore, cytosol from these mutant cells would be expected to contain the *in vivo* substrates in their unmethylated forms. Experiments using this hypomethylated (*rmt1*) yeast cytosol allowed the identification of approximately 13 radiolabeled polypeptide substrates using the GST–Rmt1 fusion protein and [^{3}H]AdoMet (5). The actual identities of these polypeptides have not been determined. However, a list of candidate gene products is described in Section V,B,1.

In the next section, we present the identities of known *in vivo* substrates, determined directly by sequencing or amino acid analysis, for both types of protein arginine *N*-methyltransferases. The methylation consensus sequence for the two general types of substrates parallels the distinction between their two respective methyltransferases. From this compilation of methylation sites, it is also possible to predict other potential substrates from existing protein databases that contain both characterized as well as hypothetical proteins from various genomic sequencing projects.

B. Asymmetric Dimethylarginine-Containing Substrates Identified *in Vivo*

1. Fibrillarin

Fibrillarin is a 34-kDa polypeptide component of a nucleolar small nuclear ribonuclear protein involved in the first processing step of preribosomal RNAs (86). It is associated with the U3, U8, and U13 small nuclear RNAs (87). Amino acid analysis of the rat protein indicated that approximately 13

TABLE III
LIST OF POSITIVELY DETERMINED *IN VIVO* SUBSTRATES OF ARGININE METHYLATION AND THEIR HOMOLOGS FROM OTHER SPECIES

Definition	Organism	ID[a]
Fibrillarin		
	Homo sapiens	g182591
	Mus musculus	g296547
	Xenopus laevis	g214143
	Drosophila melanogaster	g929566
	Caenorhabditis elegans	g1495012
	Schizosaccharomyces pombe	g544285
	Giardia intestinalis	g520824
	Leishmania major	g544283
	Tetrahymena thermophila[b]	g578561
Nucleolin		
	Homo sapiens	g189305
	Rattus norvegicus	g205790
	Mus musculus	g53453
	Cricetulus griseus	g387050
	Gallus gallus	g212412
	Xenopus laevis	g64936
hnRNP A1		
	Homo sapiens	g32344
	Macaca mulatta	g422724
	Rattus norvegicus	g133255
	Mus musculus	g1507693
	Xenopus laevis	g133258
	Drosophila melanogaster	g133253
Basic fibroblast growth factor	*Homo sapiens*	g183083
Myelin basic protein		
	Homo sapeins	g187408
	Rattus norvegicus	g126804
	Mus musculus	g387414
	Gallus gallus	g126798
	Bos taurus	g126796
	Cavia porcellus	g126797
	Ovis aries	g223882
	Xenopus laevis	g1816437
	Raja erinacea[c]	g1177855
	Squalus acanthias[c]	g1177857
	Heterodontus francisci[c]	g126799

[a]Unique identification number of GenBank retrieval.
[b]No GAR or RGG.
[c]No R at 107.

mol of asymmetric dimethylarginine are present per mole of fibrillarin, with no mono- or symmetric dimethylarginine (*71*). Amino-terminal sequence analysis identified six asymmetric dimethylarginine residues within the first 31 residues, and no unmodified arginine residues were detected. All of the methylated arginines were present in a glycine- and *a*rginine-*r*ich (GAR) domain (*71*) (Section V,A). Rat fibrillarin homologs are present in many different species, ranging from the *Physarum polycephalum* B-36 protein (*88*) to human fibrillarin, although their methylated arginine content has not been determined experimentally (Table III).

2. Nucleolin

Nucleolin, a 110-kDa polypeptide, is one of two major phosphoproteins from the nucleolus (*89*). It has also been implicated in several steps of ribosome biogenesis (*90*): regulation of RNA polymerase I transcription, modulation of chromatin conformation, binding to nascent rRNAs (*91, 92*), and possible shuttling between the nucleus and cytoplasm to facilitate transport of ribosomal proteins (*93*). Rat nucleolin, originally designated C23, also contains asymmetric dimethylarginine (approximately 9 mol per mole of nucleolin) and trace amounts of monomethylarginine *in vivo* (*94*). Sequence analysis of a 53-amino acid tryptic peptide derived from the carboxy-terminal region of the protein identified 10 asymmetric arginine residues, but no mono- or symmetric dimethylarginine, nor any unmodified arginine residues (*20*).

3. Heterogeneous Nuclear Ribonucleoprotein A1

Sequence analysis of a carboxy-terminal tryptic peptide fragment of calf thymus unwinding protein 1 (UP1) indicated the presence of a single asymmetric dimethylarginine residue at position 193 (194 if an initiator Met is included) out of 195 residues, and amino acid composition analysis also revealed one residue of asymmetric dimethylarginine per polypeptide (*95*). The UP1 used in this study may not have been full length, however, because its rat homolog, helix-destabilizing protein (HDP) (*96*), has an additional 124 residues at the carboxy terminus. This region in HDP contains the GAR domain, suggesting that the one residue of methylated arginine per polypeptide in UP1 may underestimate the total asymmetric dimethylarginine present in the full-length protein. Subsequently, it was determined that HDP is identical to hnRNP A1 (*97, 98*). Kumar *et al.* (*97*) determined that endogenous hnRNP A1 from calf thymus contained 3.1 asymmetric dimethylarginine residues per protein. The earlier determination of 0.31 residues of asymmetric dimethylarginine per hnRNP A1 molecule (*43*) cannot be explained assuming that full-length hnRNP A1 was analyzed.

The hnRNP A1 protein is a 34-kDa polypeptide and a major component

of the 40S ribonucleoprotein particle (*43*). A role for hnRNP A1 has been established in the differential selection of proximal or distal 5′ splice sites (*99, 100*). *In vitro,* hnRNP A1 can promote alternative exon skipping in certain constructs (*99*). As in the previous examples of *in vivo* substrates, hnRNP A1 is also nuclear localized and may shuttle to the cytoplasm as well (*101, 102*).

Recombinant hnRNP A1 has the distinction that it is the only *in vivo* substrate that has been extensively used as a methyl-accepting substrate *in vitro.* It has been used to monitor enzyme activity in several purifications of the Type I methyltransferase (Section II,B) (Table I). Analysis of the *in vitro*-methylated hnRNP A1 with various endo- and exoprotein hydrolases, including V8 protease, trypsin, and carboxypeptidase B, suggested that the site of methylation is the same arginine (position 194; SSQ**R**GRS) that was determined to be the methylated residue in UP1 (*47*). Arginine residue 196 is not methylated; carboxypeptidase B treatment of the tryptic peptide KQEMASASSSQRGR releases only free arginine, not a methylated derivative. The fact that the bond at Arg-194 is not cleaved by trypsin and that after the carboxypeptidase B treatment the radioactivity is still associated with the peptide suggests that the methylation of the only remaining arginine occurs at position 194 (*47*). The methylation status of the other four arginines in the GAR region could not be absolutely determined, but rather was suggested only by the methylation of smaller tryptic peptide fragments derived from the region of the protein carboxy-terminal to residue 196 (*47*). In a separate study, sequence analysis of a tryptic fragment corresponding to amino acids 197–215 shows an unidentified amino acid at position 206 that is encoded as an arginine residue, possibly indicating the presence of an asymmetric dimethylarginine residue at that position (*47, 97*).

Although hnRNP A1 has been the most extensively studied, the 40S ribonucleoprotein particle is composed of at least 11 additional polypeptides (*44*). In addition to hnRNP A1, five other unidentified polypeptide components of this particle were found to contain asymmetric dimethylarginine *in vivo* (*44*). Amino acid analysis of bulk ribonuclear proteins from the 40S core particle showed that 0.5 mol% of the amino acids was asymmetric dimethylarginine, and that specifically hnRNP A2 contained 1.4 asymmetric dimethylarginine residues per polypeptide (*43*).

4. Basic Fibroblast Growth Factor

The mammalian mitogenic protein, basic fibroblast growth factor (bFGF), can stimulate growth in a wide range of mesodermal and neuroectodermal cell types (*103, 104*) as well as vascular endothelial cells (*105*). Although this protein is a potent mitogen, an active mechanism for the secretion of bFGF from cells *in vivo* has remained elusive (*106*). The protein has no secretion signal sequences and has the general features of a cytosolic pro-

tein (*107*). In fact, evidence exists to support the idea that bFGF is actually only released from cells after cell death and/or cell lysis (*108–111*). Therefore, although the determined mitogenic action of bFGF seems to separate it from the other methylated RNA-binding substrates mentioned above, the true function of this polypeptide intracellularly remains to be determined.

This polypeptide has been determined to be an 18-kDa species by the isolation of human bFGF (*112*) and sequence analysis of a human cDNA clone (*113, 114*). Amino-terminally extended forms of bFGF have also been isolated, ranging from just two additional amino acid residues (*115*) to an additional 53 residues, creating a 25-kDa protein (*114*). Interestingly, these high-molecular-weight forms of bFGF are translated from alternative CUG start codons (*116*). Sequence analysis of tryptic fragments of the 25-kDa bFGF isolated from guinea pig brains suggested the presence of at least three sites of arginine methylation in the context of a GAR domain (*114*). The putative methylated phenylthiohydantoin–amino acid derivative eluted slightly earlier than a symmetric dimethylarginine derivative from myelin basic protein (*117*), suggesting that the methylated residue is an asymmetric dimethylarginine.

Although the 18-kDa form of bFGF is predominantly cytoplasmic, the larger forms of bFGF are specifically localized to the nucleus (*106, 118*), and it has been suggested that nuclear localization is dependent on the methylation of the peptide (*106*). When NIH 3T3 cells were incubated with a general inhibitor of AdoMet-dependent methyltransferases, 5′-methylthioadenosine, the larger forms of bFGF were not posttranslationally modified, as determined by differential electrophorectic mobility on SDS-PAGE (*106*). This presumably unmethylated bFGF did not accumulate in the nucleus as was found in control cells. This result was not due to a decrease in the stability of the unmodified form (*106*). In fact, treatment with methylthioadenosine actually stabilized all forms of bFGF in pulse-chase experiments and specifically increased the expression of the high-molecular-weight forms (*106*).

5. Histones?

It is important to note that there are uncertainties about the methylation of histone fractions, the original methyl-accepting substrate of the Type I enzyme in *in vitro* studies. For instance, it is still not certain exactly what types of methylated arginine residues are present in the *in vitro*-methylated preparations and whether the methyl acceptors are indeed histone polypeptides! Amino acid analysis of *in vitro* methylation reactions containing histones, [^{3}H]AdoMet, and a partially purified preparation of Type I methyltransferase indicates the presence of a small amount of symmetric dimethylarginine residues (*26, 38, 39*) (Table II). However, four additional studies did not de-

tect the symmetric form in similar methylation assays with their Type I methyltransferase preparations (*30, 48, 119*) or with a recombinant GST–PRMT1 methyltransferase fusion purified from *E. coli* (*6*) (Table II). The radiolabeled symmetric dimethylarginine residues detected in the former three studies may be due to a contaminating Type II arginine methyltransferase that methylates the exogenous histones.

Several fractionated histone preparations as well as individually purified histones have been used as substrates for the asymmetric protein arginine *N*-methyltransferase (*12, 13, 25, 26, 38, 52*). The "histone" substrates used include the "sulfated histone fraction," the arginine-rich fraction, the lysine-rich fraction, calf thymus fractions IIb–III and IV, f2a1(IV), f1, f2a, f2b, f3, and chicken fraction IV, as well as pure histone H1, H2A, H2B, and H4. The histones in these protein preparations have been assumed generally to be uncontaminated. Therefore, the experimental determination of arginine-specific methylation involved only the bulk precipitation and quantitation of the radioactivity present in the total protein of the reaction. Quantitation often involved direct determination of radiolabeled methylated arginine content by amino acid analysis of protein acid hydrolysates. However, it was also common to just count the total acid precipitable counts from the reaction and assume they corresponded to the amount of radiolabeled methylated arginine.

Given the difference in the *in vitro* activity of the methyltransferase on hnRNP A1 and histones (0.49 μg of hnRNP A1 gives an activity similar to 100 μg of histones) (*5*), it is possible that the presumed histone methylation is due to the preparation having a small amount of contamination with a true substrate and that histones are not, in fact, methyl-accepting substrates. None of the studies, except for those dealing with the Type III wheat germ histone methyltransferase (*50, 51*), analyzed tryptic fragments or performed amino acid analysis that would prove the actual arginine methylation of histones *in vitro*.

By incubating calf thymus histone fraction IIAS from Sigma, a histone preparation common to the arginine methylation literature, with the purified rat Type I methyltransferase fusion enzyme (GST–PRMT1) or yeast fusion enzyme (GST–Rmt1) and [^{3}H]AdoMet, we found only a single, major methylated polypeptide (Fig. 4). We were able to visualize the histones by separating the components of the reaction on a detergent–acid–urea gel system (*120, 121*) and Coomassie staining the gel (Fig. 4). Fluorography indicated that the radiolabeled polypeptide corresponded to the elution position of histone H2B (*120*). Another band, which is 5- to 10-fold less intense, was also observed on the fluorograph and migrates at a position similar to histone H2A or H3 (Fig. 4). Significantly, histones H1 and H4, which previously had been suggested to be substrates (*26, 38*), are present but are not methylated under the conditions used here (Fig. 4). The protein sequences of histones H2B, H2A, and H3 do not contain sequences similar to the glycine- and argi-

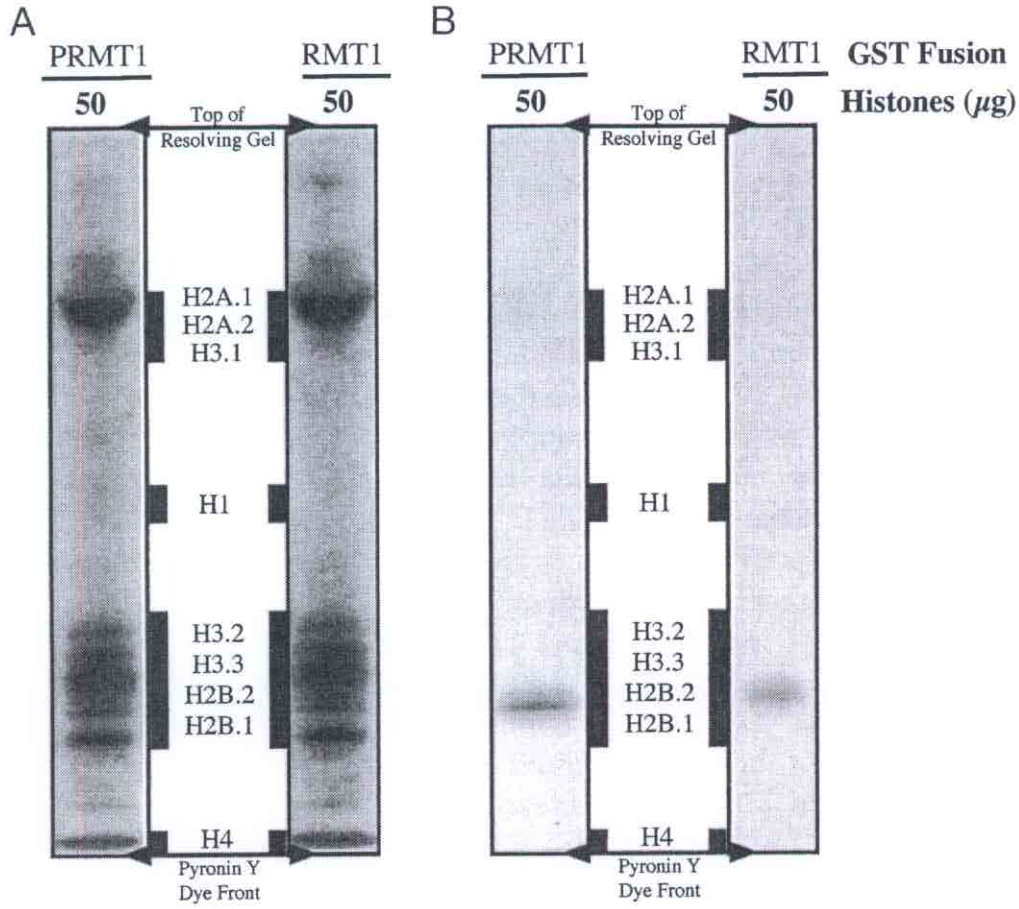


FIG. 4. Histone H2B is a candidate methyl-accepting substrate from *in vitro* arginine-methylated calf thymus histone fraction IIAS. The rat and the yeast Type I protein arginine *N*-methyltransferase fusions, GST-PRMT1 and GST-Rmt1, respectively, were incubated with 50 μg of calf thymus histone fraction IIAS (Sigma) and [^{3}H]AdoMet in a reaction identical to those previously described (*5, 6*). The reactions were stopped by the addition of sample loading buffer and boiling for 5 min. The samples were then separated on a detergent–acid–urea gel (*120, 121*) to resolve the various histone components. A photograph of the stained gel is shown (A). The gel was then prepared for fluorography as described previously (*5, 6*). The fluorograph was exposed for 3 days at −80°C (B). A species running similarly to the published migration position of H2B (*120*) was the major methyl-accepting species. A minor methylated species was also seen in the area where H2A and H3 are known to run (*120*).

nine-rich region or the consensus methylation sequence that is found in *in vivo* substrates (Section V,A). However, the histones do contain many arginine residues and they may become artifactual sites of methylation when present at high enough concentrations. In fact, even poly(arginine) has been identified as a substrate when as much as 3 mg was used per reaction (*13, 38*). As with the other reports of histone methylation *in vitro*, we cannot rule out the possibility that contaminating methyl-accepting proteins run similarly to the labeled bands and are the actual methyl-accepting substrates in the reaction. Indeed, the known *in vivo* substrates for the Type I protein arginine

N-methyltransferase are very abundant in the nucleus and may be ubiquitous contaminants of histone preparations. At the least, we can state that under our conditions, using recombinant methyltransferase fusion proteins and histone fraction IIAS, we observed the methylation of a single major and one minor polypeptide *in vitro,* and not the multiple histone species suggested by the literature (*12, 13, 25, 26, 38, 52*).

From the data described above, it is not clear which histones, if any, are methylated *in vitro* and evidence for the presence of methylated arginines *in vivo* is even more conflicting. Methylated arginines have been suggested to be present in histones *in vivo* from labeling experiments with rat tissues and mammalian cell culture using [*methyl*-^{14}C]Met (*122–124*). Unfortunately, the identities of the radiolabeled peaks, supposedly representing methylated arginine species, from the amino acid-analysis of the acid hydrolyzed proteins were not directly compared to free methylated arginine standards. It was also reported that histones isolated from rat liver nuclei contained methylated arginines, whereas those from calf thymus did not (*125*). However, in additional studies, no methylated arginines were detected in histones isolated from rat organs (*126, 127*). It is possible that, as mentioned above, the histones isolated in the experiments in which methylarginine was detected were contaminated with other nuclear or nucleolar proteins that are actual *in vivo* substrates. In fact, it was an additional purification step after the usual acid extraction of "crude histones" that demonstrated that a histone preparation could be contaminated with nonhistone proteins containing asymmetric dimethylarginine (*127*).

The situation in nonmammalian cells may be different, however. The most compelling evidence for the *in vivo* methylation of histones is a report that the methylation pattern in *Drosophila* histone H3 shifts from lysine to arginine after a heat-shock treatment at 37°C (*128*). In this case, thin-layer chromatography detected both symmetric and asymmetric dimethylarginine in acid hydrolysates of histone H3 isolated from *Drosophila* Kc III cells after heat shock (*128*). A final determination as to whether histones contain methylarginine derivatives *in vivo* awaits the application of the mass spectroscopy technique either described by Edmonds *et al.* (*129*) or used by Kouach *et al.* (*130*) on a variety of histones isolated under different conditions and from different organisms. It is certainly possible that histones are methylated in some species (flies and plants) (*51*) and not in others (mammalian).

6. Other Potential *in Vivo* GAR Domain-Containing Substrates

Two additional proteins have been suggested to contain asymmetric dimethylarginine *in vivo:* ICP27 from the herpes simplex virus and Npl3 from yeast. ICP27 (infected cell protein) is an immediate-early gene product pro-

duced by the herpes simplex virus soon after infection of a cell (*131*). The protein is essential for virus growth and it is specifically required for the viral transition to the late phase of infection (*132*), as well as the general stimulation of viral DNA replication (*133*). A direct mechanism for its mode of action has not been determined. However, it is thought to be posttranslational through the altering of the pre-mRNA processing machinery (*134, 135*). Consistent with these functions, ICP27 is found in the nucleus and also accumulates in the nucleolus (*136, 137*). The *in vivo* methylation of ICP27 was demonstrated in African green monkey kidney Vero cells labeled with [*methyl*-^{3}H]Met in the presence of protein synthesis inhibitors (*138*). The methylation of the protein was shown to be dependent on the presence of consensus asymmetric arginine dimethylation sites in the amino-terminal third of ICP27 (*138*). However, direct amino acid analysis or protein sequencing of ICP27 has not been done to confirm the identity of the methylated residues as asymmetric dimethylarginine.

Similarly, the evidence supporting the *in vivo* arginine methylation of yeast Npl3 is based on the presence of a carboxy-terminal consensus sequence for asymmetric arginine methylation (*7*). CNBr cleavage of Npl3 isolated from yeast labeled *in vivo* with [*methyl*-^{3}H]Met versus [^{35}S]Met indicated that methylation occurs in the carboxy-terminal region of the protein (*7*). Additional evidence for *in vivo* methylation came from the discovery of an antibody that was specific for Npl3 isolated from yeast, but not from *E. coli* (*24*). It was subsequently found that the antibody recognizes the bacterial recombinant Npl3 if it is incubated with yeast extracts and AdoMet. Binding was competitively inhibited by peptides corresponding to the consensus asymmetric arginine dimethylation site (R1 or R3) (*24, 69*). The authors concluded that the epitope recognized by the antibody included the asymmetrically dimethylated arginines of Npl3 (*24*). The *NPL3* gene was originally isolated because a mutant of this gene was defective in nuclear protein import (*76*). However, it is currently believed that the protein is also involved in the binding and export of mRNA from the nucleus (*139*). Temperature-sensitive mutants of *NPL3* accumulate mRNA in the nucleus (*79*), and the wild-type protein shuttles between the nucleus and the cytoplasm (*78*).

Although neither ICP27 nor Npl3 have been proved to contain asymmetric dimethylarginine *in vivo,* we have included them in this section because of the strong indirect evidence.

7. Candidate *in Vivo* Substrates

Several other possible methyl-accepting substrates have been identified *in vivo,* including the high-mobility group proteins 1 and 2 from calf thymus, heat-shock proteins HSP70A and HSP70B from chicken fibroblasts, as well as myosin from rats. The mono- and asymmetric dimethylation of these pro-

teins was determined by amino acid analysis of acid hydrolysates derived from each of the endogenously isolated proteins. These studies, however, do not address the possibility of contamination in their preparations of these putative methyl-accepting substrates. Therefore, the methylated arginines in these isolates could conceivably be due to contamination with a small amount of a highly methylated Type I protein arginine *N*-methyltransferase substrate, such as fibrillarin or hnRNP A1. It will be necessary to obtain amino acid sequence data from these possible methyl-accepting substrates in order not only to identify the type of methylated residue but also to positively determine the identity of the substrate protein.

i. High-mobility group (HMG) proteins are nonhistone proteins associated with chromosomal DNA (*140*). Although the functions of HMG1 and HMG2 remain unclear, they are known to bind DNA in the minor groove of AT tracts (*141*). It has also been shown that the proteins induce DNA bending and may promote cis-acting protein–protein interactions or be important in the compaction of DNA (*142*). Direct amino acid analysis of endogenous HMG1 and HMG2 isolated from calf thymus indicated the presence of asymmetric but not mono- or symmetric dimethylarginine (*143*).
ii. Two of the major species in the 70-kDa heat-shock protein superfamily have been shown to be methylated on arginine residues *in vivo* in chicken embryo fibroblasts (*144*) and in mouse 3T3 fibroblasts (*145*). These proteins act as molecular chaperones to ensure proper protein folding under stressed and unstressed conditions. Protein methylation in both systems is tightly linked to protein synthesis (*145*). The level of monomethylarginine in the HSC70 homolog is markedly reduced in 3T3 cells but not in chicken cells on arsenite treatment that induces the heat-shock response. The level of monomethylarginine is reduced, however, in the chicken HSP70B protein after such treatment. Significantly, peptide binding and ATPase activities are similar in purified preparations of recombinant rat HSC70 (presumably unmethylated) and bovine brain HSC70 (presumably methylated), suggesting that methylation may not affect these properties of the proteins (*146*).
iii. There is a limited literature describing the methylation of myosin on arginine residues. Only the asymmetric derivative was found in the myosin of cultured rat muscle cells labeled with [*methyl*-^{14}C]Met (*147*). Arginine-methylated myosin was also detected directly in tissues from rat fetuses and pups younger than 7 days old. Adult rats had no detectable methylated arginine residues in myosin (*147, 148*). Besides the lack of protein sequence data, no attempts have been made to determine which polypeptide chain of myosin is actually methy-

lated. This developmental regulation of methylarginine content in myosin has not been reported for the other constitutively methylated asymmetric substrates such as fibrillarin (*71*), which makes it more unlikely that the methylated arginines in the myosin preparations are derived from a contaminating protein.

C. Symmetric Dimethylarginine-Containing *in Vivo* Substrates—Myelin Basic Protein

Several studies have identified only the mono- and symmetrically dimethylated arginine, at position 107, corresponding to the bovine sequence, in myelin basic protein from human, monkey, bovine, rabbit, guinea pig, chicken, turtle, and frog species; however, there was no mono- or symmetric dimethylarginine found in carp myelin basic protein (*18, 21, 22, 31, 32, 34, 35*). Only two studies have identified asymmetric dimethylarginine in myelin basic protein (*19, 149*); these preparations may have been contaminated, however, with other proteins containing this residue (*34, 35*).

Mammalian myelin basic protein is one of the major protein components of the myelin sheath (*150*). It is located in compact myelin and interacts with opposed cytoplasmic membrane surfaces, perhaps playing an essential role in maintaining this structure (*151–153*) (see Ref. *154* for review). The protein exists in two isoforms, 18.5 and 14 kDa, generated from the same gene by differential mRNA splicing (*155*). In the rat, the smaller form of myelin basic protein is identical to the larger form, except for a 40-amino acid internal deletion, and it is also methylated on the corresponding arginine (*33*). The two isoforms can be further posttranslationally modified to create one of eight charge isomers (*155*). Other posttranslational modifications of this protein include phosphorylation, ADP-ribosylation, and conversion of arginine residues to citrulline residues (*155*).

Myelin basic protein is the only example of a protein known to contain symmetric dimethylarginine *in vivo*. To date, all of the Type II arginine methyltransferases have been isolated from the brain tissue from various species (Table I), matching the localization of the only known substrate. Methylation experiments with hypomethylated, adenosine dialdehyde-treated rat PC12 cells (Section IV,A) also suggest that there may be far fewer symmetric dimethyl-accepting arginine substrates than asymmetric ones (*85*). Amino acid analysis of these PC12 extracts after the addition of [^{3}H]AdoMet showed that only 2.9% of [^{3}H]methyl group incorporation was into symmetric dimethylarginine, compared to 71.5% for asymmetric dimthylarginine (*85*). This result should be viewed with some caution, because this type of analysis necessitates that the protein methyltransferases can recognize their substrates after completion of protein synthesis and folding. It may be that the enzymes responsible for the asymmetric versus symmetric dimethylation

of arginine residues are differentially able to methylate methyl-accepting substrates that have been released from the ribosome. In fact, Type II methyltransferase activity has been found in other tissues besides brain (*29*), suggesting that other nonmyelin basic protein substrates may exist for this methyltransferase.

V. Determination of Arginine Methyltransferase Substrates

A. Sequence Analysis of *in Vivo* Asymmetric Methylation Sites

From the four amino acid sequences in which the actual methylated residue(s) were directly determined, a total of 20 asymmetric dimethylarginine residues have been identified (Fig. 5A). Because the methylated arginine residues of fibrillarin and bFGF are found near the amino terminus and those of nucleolin and hnRNP A1 are near the carboxy terminus, it appears that the methylation site is independent of its location within the polypeptide. This is also supported by the observation that the GST–GAR fusion protein is a substrate (Section III,A) (J. Gary, unpublished).

By compiling data on the residues surrounding the methylated arginine residue, it is possible to generate a consensus sequence for asymmetric dimethylarginine formation: (G/F)GGRGG(G/F) (Fig. 6). Only the arginine and glycine at positions 0 and +1 are found in all methylation sites (Fig. 6). Whether the glycine residue at position +1 is essential for methylation is not known. We chose to limit our comparison to positions −3 to +3 among these sequences, because the percentage of residue conservation had dropped to below 50% by the third position away from the arginine (Fig. 6). For these proteins, the methylation occurs at an RGG-methylation site that is within the larger context of the GAR domain (Fig. 6). The term "RGG" is used to describe the actual site of asymmetric methylation; the Arg-Gly sequence is completely conserved for known Type I methyltransferase substrates, and the final Gly is included to distinguish this site from the myelin basic protein symmetric methylation site Arg-Gly-Leu (Section V,C)

In order to determine the general nature of the consensus generated here, we also separately compiled the residues surrounding the arginine residues in the GAR regions from the homologs of the known *in vivo* substrates from different species (Fig. 6 and Table III). Interestingly, the known *in vivo* methylation site for hnRNP A1 is not a particularly good match to the consensus (compare Figs. 5A and 6). In fact, a peptide identical to a portion of this sequence (SSQRG) is a very poor substrate for a partially purified Type

A

rat Fibrillarin	FSP**R**GGGFGG**R**GGFGD**R**GG**R**GGG**R**GG**R** (GGF)
rat Nucleolin	FGG**R**GGG**R**GGFGG**R**GGG**R**GG**R**GGFGG**R**G**R**GGFGG**R**GGF**R**GG**R**GGG
guinea pig bFGF (25 kDa-form)	VGG**R**G**R**G**R**GTA
rat UP1/hnRNP A1	SSQ**R**GR (S)

B

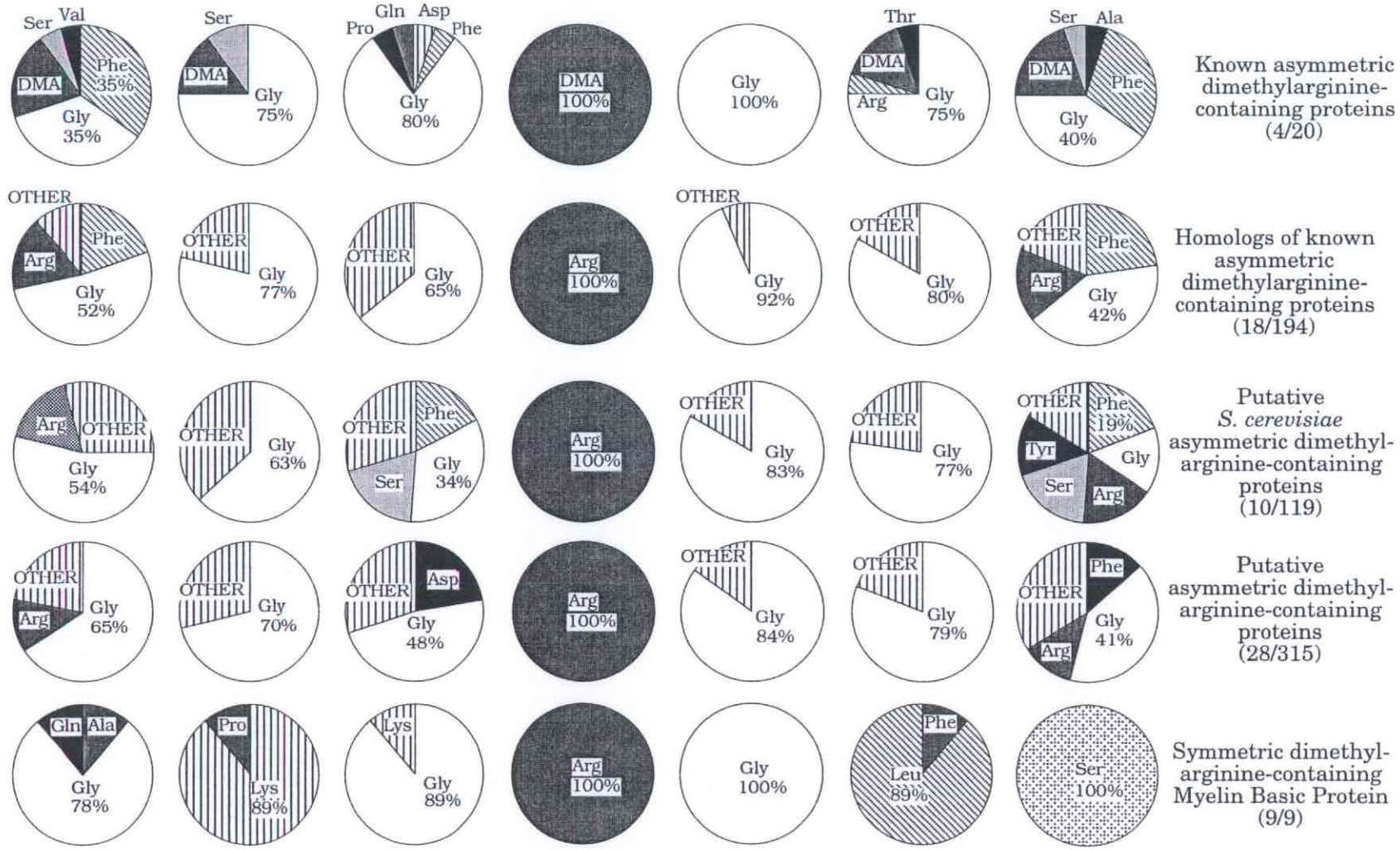
	203 ↓ ↓ 236
rat hnRNPA1	GGG**R**GGGFGGNDNFG**R**GGNFSG**R**GGFGGS**R**GGG
monkey hnRNPA1	GGG**R**GGGFGGNDNFG**R**GGNFSG**R**GGFGGS**R**GGG
mouse hnRNPA1	GGG**R**GGGFGGNDNFGGFGGS**R**GGG
Xenopus hnRNPA1	FGG**R**GGNFGGN**R**GGGGGFGN**R**GYG
Drosophila hnRNPA1	GGG**R**GGPGG**R**AGGN**R**GNMGGGNYGNQNGGGNWNNGGNNWGNN**R**GGN

FIG. 5. Sequences of proteins containing Type I protein arginine *N*-methyltransferase methylation sites. (A) The sequences surrounding arginine residues that have been directly determined to be methylated *in vivo*. These asymmetrically dimethylated arginine residues are boldfaced and underlined. Residues in parentheses were not reported in the original studies but were taken from the corresponding GenBank entry and are included here for the compilation of the consensus sequence. (B) The GAR domains of hnRNP A1 proteins from various organisms.

I protein arginine *N*-methyltransferase from rat liver and calf brain, but a peptide (GNFGGGRGGGFGG) corresponding to a portion of the GAR motif six residues away is a much better substrate (*156*). For these reasons, in the expanded compilation we included only those arginine residues that were clearly in the GAR domain (Figs. 5B and 6). A surprising feature is that the fibrillarin from *Tetrahymena thermophila* does not contain a GAR domain and is therefore presumably not arginine methylated *in vivo*. In the end, the results of this expanded comparison are not drastically different from those obtained with just the known substrates (Fig. 6).

The structure of the GAR domain has not been determined. This domain, however, could adopt higher order structures akin to other proteins that have high percentages of glycine residues, e.g., silk and collagen. The structure of the larger GAR domain may not necessarily be crucial for methylation. This short RGG consensus noted above seems to be sufficient to direct asymmetric arginine methylation, as seen in the methylation of the synthetic peptide R1 (GGFGGRGGFG) by a Type I protein arginine *N*-methyltransferase from rat PC12 cells (*69*) as well as the rat GST–PRMT1 and yeast GST–Rmt1 fusion proteins (*5*, *6*).

From this consensus sequence, the likelihood that histones, specifically H2B, H2A, and H3 (possible histone substrates identified in Fig. 4), are *in vivo* substrates of the same enzyme that methylates hnRNP A1 and fibrillar-


Ser Val Phe 35% DMA Gly 35%
Ser DMA Gly 75%
Gln Asp Pro Phe Gly 80%
DMA 100%
Gly 100%
Thr DMA Arg Gly 75%
Ser Ala DMA Phe Gly 40%
Known asymmetric dimethylarginine-containing proteins (4/20)
OTHER Phe Arg Gly 52%
OTHER Gly 77%
OTHER Gly 65%
Arg 100%
OTHER Gly 92%
OTHER Gly 80%
OTHER Phe Arg Gly 42%
Homologs of known asymmetric dimethylarginine-containing proteins (18/194)
Arg OTHER Gly 54%
OTHER Gly 63%
OTHER Phe Ser Gly 34%
Arg 100%
OTHER Gly 83%
OTHER Gly 77%
OTHER Phe 19% Tyr Gly Ser Arg
Putative S. cerevisiae asymmetric dimethyl-arginine-containing proteins (10/119)
OTHER Arg Gly 65%
OTHER Gly 70%
OTHER Asp Gly 48%
Arg 100%
OTHER Gly 84%
OTHER Gly 79%
Phe OTHER Arg Gly 41%
Putative asymmetric dimethyl-arginine-containing proteins (28/315)
Gln Ala Gly 78%
Pro Lys 89%
Lys Gly 89%
Arg 100%
Gly 100%
Phe Leu 89%
Ser 100%
Symmetric dimethyl-arginine-containing Myelin Basic Protein (9/9)
-3
-2
-1
0
+1
+2
+3

in is low. Neither of the histones contain sequences resembling the RGG consensus sequence in a GAR domain. The *in vitro* methylation of these histones may be due to an artifactual *in vitro* activity of the Type I methyltransferase on the histones or to the presence of a contaminating Type I methyl-accepting substrate in the histone preparation.

Of the other potentially arginine-methylated proteins described in Section IV,B,7, only the HMG1 and HMG2 proteins have directly corresponding protein sequences in the NCBI database that can be examined for the methylation consensus. Neither the bovine HMG1 (GB X12796) nor the pig HMG2 (GB J02895) protein contains RGG-methylation consensus sequences. For the other proteins, without experimental protein sequence data it is not clear which of the protein sequences present in the protein database correspond to the methylated HSP70A and HSP70B (*145, 157*); furthermore, the methylation site within myosin was not specifically determined to be on the light or heavy chain (*147, 148*). In these latter cases, examination of several HSP70 sequences as well as myosin heavy- and light-chain sequences revealed that these proteins also do not contain the methylation sites. It is always possible, however, that an additional methyltransferase is responsible for the activity. Interestingly, however, the myosin heavy chain 1B from the amoeba does contain consensus asymmetric arginine dimethylation sites. In either case, myosin distinguishes itself as a possible methyl acceptor for the Type I protein arginine *N*-methyltransferase because it is one of only a few potential substrates whose primary function does not involve binding to RNA.

B. RNA-Binding Proteins May Be the Major Group of Substrates for the Type I Protein Arginine *N*-Methyltransferases

1. Potential *Saccharomyces cerevisiae* Substrates

Using the methyl-accepting consensus sequence derived above for the Type I protein arginine *N*-methyltransferase, it is possible to identify other candidate proteins through protein database searches. Such a search was per-

FIG. 6. Frequency of amino acid residues that are found at positions surrounding known or proposed methylated arginine residues. For each row of pie charts, the number of sequences and all of their known or proposed methylated arginines (indicated in parentheses, respectively) were compiled in the subgroups denoted in the text on the right side of the figure. The methylated arginine residue is designated to be at position 0 and the three flanking amino acids on either side are also shown. A percentage for the amino acid that is most frequently present at a particular position is given under the amino acid abbreviation. Amino acids that are present less than 15% of the time at a particular position were included in the OTHER category.

formed by Najbauer *et al.* (*69*) with a single methylation consensus sequence (FGGRGGF) against all of the proteins in the GenPept database (release 72.0, November, 1992). This search identified the members of the nucleolin, fibrillarin, and hnRNP A1 families of known *in vivo* substrates. Four putative *S. cerevisiae* methyl-accepting substrates were also listed: Gar1, Npl3, Ssb1 (now Sbp1), and Nsr1.

Using the consensus sequence defined above, we chose to search for potential substrates in the yeast *S. cerevisiae,* because not only is the sequence of the entire genome now available, but also because a gene encoding a Type I protein arginine *N*-methyltransferase catalytic subunit (*RMT1*) was identified in this organism (*5, 7*). Our original analysis of the *rmt1* mutant yeast strain indicated that there were at least five polypeptides that were endogenously methylated *in vitro* in the parent versus the mutant strain (*5*). Additional *in vivo* labeling experiments with these two strains using [^{3}H]AdoMet show that at least four individual peptides are methylated in an *RMT1*-dependant manner (J. Gary, unpublished). Furthermore, the addition of GST–Rmt1 to *rmt1* cytosol caused the methylation of approximately 13 distinct polypeptides (*5*), suggesting that a number of different methyl-accepting substrates are present in yeast cells.

Initial searches of the *S. cerevisiae* genome were performed on the BLAST server at the Stanford Genome Database (http://genome-www.stanford.edu/Saccharomyces/) using a string of 21 amino acids composed of three repeats of the methylation site consensus determined above. Because all of the full-length *in vivo* substrates for the Type I methyltransferase contain multiple methylated arginine residues, we used this longer search sequence to eliminate putative positives that contained only a single arginine in an RGG context. To make the search exhaustive, each of the GAR domains identified in this first tier of the search was used to search the database again. The regions of similarity between the consensus search sequence and those returned by the BLAST server sometimes did not contain arginine residues in the proper context determined above, most likely due to the fact that the search string was relatively short and highly repetitive; for instance, poly(Gly) sequences containing no arginine residues were common false positives. Therefore the sequences returned by the server had to be examined manually to remove those that did not contain a GAR region or any RGG consensus sites. From these searches, 10 proteins were identified as possible substrates for the Type I protein arginine *N*-methyltransferase (Table IV).

As might be expected, the compilation of the amino acid sequences surrounding the putative arginine modification sites in the GAR domains of these potential yeast substrates shows a consensus similar to the one derived from the known substrates (Fig. 6). However, two differences were found that should be noted. At position −1, Ser and Phe residues are found with Gly as

TABLE IV
LIST OF PUTATIVE TYPE I ARGININE *N*-METHYLTRANSFERASE SUBSTRATES FROM *Saccharomyces cerevisiae*

ORF name	Common names	ID[a]	Localization[b]	Nucleic acid binding	Ref.
DR432w	*NPL3 /NOP3 / MTS1 /MTR13*	g172051	Nu/Cy	+	*77, 139*
DL014w	*NOP1*	g1430978	No	+	*161, 247*
HR089c	*GAR1*	g487935	No	+	*80*
HL034c	*SBP1 /SSBR1*	g508771	No/Nu	+	*177, 179*
NL112w	*DBP2*	g1302034	ND	ND	*186*
GR159c	*NSR1 /SHE5*	g1323271	No	+	*167, 172*
LR398c	*SKI2*	g410320	No	ND	*181*
CL011c	*RLF6 /GBP2*	g1907116	ND	+	*174*
OL123w	*HRP1 / NAB4/ NAB5*	g1420003	ND	ND	*160*
GL122c	*NAB2*	g1322681	Nu	+	*176*

[a]Unique identification number for GenBank retrieval.
[b]Nu, Nuclear; No, nucleolar; Cy, cytoplasmic; ND, not determined.

the predominant residues. At position +3, a more diverse group of amino acids was found than would have been predicted from the known substrates.

One of the substrates identified by the search of the yeast genome database is the *NPL3* gene product (also known as the genes *NOP3, NAB1, MTS1,* and *MTR13*), for nuclear protein localization (*76*). Npl3 has been implicated in a wide variety of cellular processes: nuclear protein import (*76*), rRNA processing (*77*), association with and export of poly(A+) RNA from the nucleus (*78, 79, 139*), mitochondrial protein import (*158*), and involvement in mating-type silencing (*159*). Current models propose that Npl3 shuttles between the nucleus and the cytoplasm as a carrier of mRNA (*78, 139*) and also may be involved in mRNA splicing (*160;* C. Siebel and C. Guthrie, unpublished). As described earlier, the yeast Type I protein arginine *N*-methyltransferase was identified in a synthetic lethal screen to identify genes that could exacerbate the phenotype of an *npl3* temperature-sensitive allele (*7*). In addition, Npl3 was found to be *in vivo* and *in vitro* methylated in its carboxy-terminal region, where the GAR domain is also located (*7, 24*).

Of the other nine substrates identified by the homology searches, six (Nop1, Gar1, Sbp1, Nsr1, Rlf6, and Nab2) have been experimentally shown to bind nucleic acids, and the remaining three (Dbp2, Ski2, and Hrp1) display sequence similarity to nucleic acid-binding species as well (Table IV). Three of these potential methyl-accepting substrates (Nop1, Gar1, and Nsr1) appear to be nucleolar and are involved in ribosome biogenesis. Nop1 is the

yeast homolog of fibrillarin (*161–163*), and is also found to be associated with small nucleolar RNA (*164*). This nucleolar protein is essential for yeast viability and various temperature-sensitive mutants of *NOP1* are impaired in pre-rRNA maturation and processing (*165*). Gar1 is also an essential nucleolar protein that is required for rRNA processing (*80, 166*). The gene, in yeast, was identified on a Southern blot by its hybridization to a cDNA probe derived from the GAR domain of *Xenopus* fibrillarin; as an aside, the putative methyl-accepting substrates *NOP1* and *SSB1* (now *SBP1*) were also identified in this screen (*80*). Nsr1 is a nucleolar protein (*167*) that was originally identified as the protein p67, which could specifically bind to the nuclear localization sequence of histone H2B (*168*). Sequence analysis revealed that *NSR1* is related to mammalian nucleolin (*169, 170*). Null mutants of this gene have a slow-growth phenotype and are deficient in pre-rRNA processing (*171*). The *NSR1* gene was then subsequently cloned under two separate circumstances: as a cold-shock inducible gene (*170*), as well as from a yeast genomic screen using a human hnRNP A1 cDNA as the probe (*172*). *NSR1* was also recently cloned in a screen designed to detect proteins with the ability to bind to a single-stranded telomeric repeat *in vitro;* another possible methyl-accepting substrate, Rlf6, was also found during this screen (*173*). Although the localization of Rlf6 has not been determined, its ability to bind telomeric repeats *in vitro* as well as its requirement for the proper distribution of the Rap1 telomerase-associated protein within the nucleus suggest that its localization is at least nuclear (*174*). The importance of nuclear Rap1 protein localization, and therefore Rlf6, is underscored by the fact that this protein has been shown to be responsible for telomere length regulation in the protein-counting mechanism established in yeast (*175*).

Other yeast candidates for asymmetric methylation at arginine residues are also localized to the nucleus and interact with RNA. The essential *NAB2* gene was cloned in a screen identifying proteins that interact with polyadenylated mRNA (*176*). Nab2 is a nuclear protein that can be purified from the polyadenylated mRNA fraction of yeast cells in which proteins have been cross-linked to mRNA *in vivo* by UV irradiation (*176*). Sbp1 is a nuclear/nucleolar protein that was originally identified as the single-stranded nucleic acid-binding protein Ssb1(*177–179*). Sbp1 is not essential for viability and seems to not be required for promoting mRNA splicing *in vitro,* but it may compete for mRNA binding with true splicing factors or be involved in some aspect of RNA metabolism (*179, 180*). *HRP1* was found because mutations in this essential gene suppressed the phenotype of an *npl3* temperature-sensitive allele (*160*). Hrp1 has sequence homology to the hnRNP A and hnRNP B families and is similarly localized to the nucleus (*160*).

The last two potential yeast methyl-accepting substrates identified are the putative RNA helicases Ski2 and Dbp2. One of the functions of the nucleo-

lar protein Ski2 (*181*) is to act as an antiviral protein *in vivo* by blocking the translation of viral mRNAs that are not capped or polyadenylated (*182*). Under normal conditions, it may inhibit the translation of fragmented mRNAs, which may also lack these 5′ or 3′ features (*183–185*). *DBP2* was identified in a two-hybrid screen for proteins that interact with Upf1, a protein involved in the selective degradation of nonsense codon-containing mRNAs (*186*).

All of the proposed endogenous substrates of the yeast Type I protein arginine *N*-methyltransferase appear to be nuclear or even nucleolar. All of the known substrates of the mammalian enzyme are also nuclear. From the fact that the known substrates as well as all of the putative methyl-accepting proteins in yeast interact with RNA, it is tempting to speculate that methylation of arginine residues can modulate the nucleic acid interactions of these proteins.

2. Potential Substrates from Other Organisms

We also performed database searches on the entire NCBI nonredundant protein database (February, 1997) in order to identify additional substrates in organisms other than *S. cerevisiae*. The search was done in a fashion similar to that described above for the yeast example, although the second tier of searches was not done. Non-GAR-containing sequences were again manually excluded. The 28 highest scoring sequences are compiled in Table V. The consensus sequence obtained for these proteins (Fig. 6) is very similar to that of the known substrates for the Type I arginine methyltransferase. The glycine residue at position −1 is less frequently observed, similar to the situation seen with the yeast candidate protein substrates (Fig. 6).

Several of the substrates in Table V are homologs or family members of known substrate proteins or proposed yeast substrates for the Type I protein arginine *N*-methyltransferase: hnRNPs (U and A3), helicases, nucleolar proteins (Nopp44/46 and the "nucleolin homolog" C27H5.3), and the *Schizosaccharomyces pombe* Gar2 nucleolar protein involved in ribosome biogenesis. As previously noted, the common feature of almost all of these proteins is that they have either been directly shown, or are thought on the basis of sequence similarity, to bind nucleic acids, mostly RNA. Only one protein on the list, myosin heavy chain 1B, is not known to bind nucleic acids as a primary function.

We should note that many of the same proteins were identified in a previous database search for substrates of the Type I protein arginine *N*-methyltransferase (*69*). We have, however, excluded two of the sequences originally identified in that report, based on our criteria for selecting putative substrates. Cytochrome P450 and the mouse replication protein A were not included here because the RGG-methylation consensus of these proteins was not present in the context of multiple repeating methylation sites. However,

TABLE V
LIST OF PUTATIVE TYPE I ARGININE N-METHYLTRANSFERASE SUBSTRATES FROM VARIOUS ORGANISMS

Definition	Organism	ID[a]	Nucleic acid binding[b]
RBP56	*Homo sapiens*	g1613775	By homology
EWS	*Homo sapiens*	g1666067	+
hTAFII68	*Homo sapiens*	g1628403	+
hnRNP A2	*Homo sapiens*	g500638	+
FUS/TLS RNA-binding protein	*Homo sapiens*	P35637[c]	By homology
hnRNP U	*Homo sapiens*	g32358	+
Pigpen	*Bos taurus*	Ref. 219[d]	By homology
Ribosomal protein S2	*Rattus rattus*	g57717	By homology
FLI2	*Mus musculus*	g193323	+
hnRNP A3	*Xenopus laevis*	P51968[c]	By homology
GCR 101	*Drosophila melanogaster*	S49193[e]	?
caz RNA-binding protein	*Drosophila melanogaster*	S54729[e]	+
p62 putatative ATP-dependent RNA helicase	*Drosophila melanogaster*	P1909[c]	By homology
T01C3.7	*Caenorhabditis elegans*	g1495012	By homology
ORF from yk76b9.5	*Caenorhabditis elegans*	g1255346	?
F58G11.2	*Caenorhabditis elegans*	g1742993	By homology
C27H5.3	*Caenorhabditis elegans*	g540269	By homology
GAR2	*Schizosaccharomyces pombe*	P41891[c]	+
DBP2 p68 RNA helicase	*Schizosaccharomyces pombe*	g173419	+
GAR1	*Schizosaccharomyces pombe*	Q06975[c]	+
GBP1	*Chlamydamonas rheinhardi*	g520518	+
AtGRP2b	*Arabidopsis thaliana*	g1063684	By homology
Bmsqd-1	*Bombyx mori*	g784909	By homology
Cutinase negative-acting protein	*Fusarium solani*	g1438951	+
RNA helicase, PRH75	*Spinacia oleracea*	g1488647	By homology
Nopp44/46	*Trypanosoma brucei*	g1314705	+
Glycine-rich protein	*Solanum lycopersicum*	S14985[e]	?
Glycine-rich RNA-binding protein	*Daucus carota*	Q03878[c]	By homology
EBNA-1	Epstein–Barr virus	g1334880	+
ICP27	Herpes simplex virus	P36295[c]	?
Myosin heavy chain 1B	*Acanthamoeba castellanii*	MWAXIB[e]	–

[a]Unique identification number for GenBank retrieval.
[b]As listed in the reference (ID) provided in column 3.
[c]Unique identification number for Swiss-Prot retrieval.
[d]See reference list.
[e]Unique identification number for PIR retrieval.

we want to stress the fact that the presence of a single RGG-methylation consensus in a protein without a larger GAR domain does not exclude proteins from being *in vitro* substrates and therefore possible *in vivo* methyl-accept-

ing substrates as well. In fact, three proteins that were not identified from the protein database search are methylated *in vitro*. Expressed forms of the fragile X protein, FMRP, as well as *Drosophila* PSI and SXL, can be methylated by the GST fusion of the yeast Type I protein arginine *N*-methyltransferase *in vitro* (J. Gary and S. Warren, unpublished; J. Gary and D. Rio, unpublished). Interestingly, these three proteins are also nucleic acid-binding proteins (*138, 187, 188*). In addition, PSI and SXL, like hnRNP A1, are regulators of mRNA splicing (*188–191*). Sequence analysis of the SXL and PSI proteins indicate that each contains one possible asymmetric arginine methylation site, and they are not optimal consensus sequences compared to the consensus proposed above. FMRP, on the other hand, has up to seven arginines in a small GAR domain (*131*). Perhaps only two or three of these are in an RGG context that is within the consensus parameters.

C. Analysis of the *in Vivo* Symmetric Methylation Site of Myelin Basic Protein

Biochemical evidence supports the idea that the mammalian Type II myelin basic protein-specific symmetric protein arginine *N*-methyltransferase is distinct from the Type I asymmetric methyltransferase (*6*). This conclusion is also supported by the identification of a consensus methylation site that is similar to but distinct from the asymmetric methylation site (Fig. 6). The Type II consensus site was determined by aligning the known as well as the analogous postulated methylation sites within myelin basic proteins from various organisms (Table III) in a manner similar to that noted previously. Although the arginine residue, like the Type I methyltransferase substrates, is still generally flanked on both sides by a glycine residue, additional glycine residues are apparently absent from positions −2 and +2. Instead, these positions generally contain lysine and leucine residues, respectively. Interestingly, myelin basic proteins from the marine organisms shark, skate, and dogfish do not have an arginine at the analogous position (107) to the bovine protein. These proteins have not been tested in an *in vitro* methylation reaction with a Type II methyltransferase preparation in order to determine if another arginine residue may be methylated. On the other hand, the myelin basic protein from carp is known not to be methylated *in vivo* (*35*), but it does seem to contain a methylatable arginine residue because it was found to be the best substrate for a bovine brain Type II protein arginine *N*-methyltransferase (*45*). The exact position of the *in vitro*-methylated arginine in the myelin basic protein from carp, however, was not determined.

Direct analysis of the methyl-accepting site for the Type II methyltransferase has been approached via synthetic peptide substrates. Rawal *et al.* (*156*) performed activity assays with a partially purified calf brain preparation of the Type II protein arginine *N*-methyltransferase, using various pep-

tides based on the myelin basic protein sequence surrounding Arg-107 as methyl-accepting substrates. They found that the peptide matching residues 104–109 (GKGRGL) was a good methyl acceptor, but peptides where the flanking glycine residues on either side of Arg-107 were changed independently to several other amino acids, including histidine, phenylalanine, glutamine, leucine, and aspartate, were not substrates (*156*). It is of particular interest to note that the circularized peptide (by disulfide formation) CGKGRGLC was found to be a better substrate than its linear version (*156*). This may be an important indication that tertiary structure also plays a role in substrate recognition by the Type II methyltransferase. In native full-length myelin basic protein, Arg-107 is only five residues away from a triproline sequence, thought to impose a sharp bend in the polypeptide (*155, 192*), and this feature may have been mimicked by the synthetically circularized peptide.

D. Other Protein Arginine Methyltransferase Consensus Methylation Sites

1. Plant Type III Histone H4-Specific Arginine Methyltransferase

The Type III methyltransferase was isolated from wheat germ extract and specifically mono- and symmetrically dimethylates endogenous histones as well as calf thymus histone H4 (*50*). In a subsequent study, the site of methylation was determined by amino acid analysis of peptide fragments from calf histone H4 incubated with [^{14}C]AdoMet and a preparation of the wheat germ Type III methyltransferase (*51*). A single peptide fragment corresponding to residues 24–35 was radioactive and contained only monomethylarginine (*51*). The site of methylation could be established because this peptide contains only one arginine residue at position 35. The amino acid context surrounding this methylation site (PAIRRLA) is different from all the other arginine methyltransferases and supports the recognition of this enzyme as a separate subtype.

2. Type IV Cytochrome *c*-Specific Arginine Methyltransferase

An activity able to monomethylate an arginine residue in cytochrome *c* *in vitro* has been identified in the protist *Euglena gracilis* (*53*) and designated the Type IV enzyme. The arginine methylation site in cytochrome *c* was determined to be at position 38 in the horse heart protein that was presumably used because the endogenous *E. gracilis* cytochrome *c* is fully methylated *in vivo* (*53*). The arginine is in a context distinct from the asymmetric or myelin basic protein consensus sequences. Using cytochrome *c* sequences

from horse, human, yeast, mouse, and rat cells, the consensus amino acids surrounding the analogous Arg-38 are (I/L)FG**R**(K/H) (S/T)G. It would seem from the consensus sequence data that if the modification does occur *in vivo*, the cytochrome *c* methyltransferase may also be a distinct enzyme. Interestingly, the yeast arginine methyltransferase fusion protein (GST–Rmt1), known to have a broader substrate specificity, is able to poorly methylate cytochrome *c*, whereas the mammalian GST–PRMT1 does not (*5*, *6*). Despite the activity of the GST–Rmt1 fusion protein toward cytochrome *c*, it is unable to methylate myelin basic protein (*5*), which would seem to have a methylation site closer to its own consensus sequence.

VI. Functional Roles for the Protein Arginine *N*-Methyltransferase and Arginine Methylation

A. Asymmetric Methylation

The importance of asymmetric protein arginine *N*-methylation is underscored by the number of possible substrates involved in important cellular processes (Table V) and the presence of the activity in many different organisms and tissues (*2*). By attempting to understand aspects of both the asymmetric protein arginine *N*-methyltransferase and the substrates it acts on, we may begin to ascertain the significance of this posttranslational modification in the context of a cell. In the following sections we describe possible functions for asymmetric arginine methylation.

1. Signal Transduction

Perhaps the clearest evidence to support a role for the Type I protein arginine *N*-methyltransferase in signal transduction can be seen from the ways in which the mammalian methyltransferase genes have been cloned. Rat PRMT1 was found because of its interaction with TIS21 and BTG1 in a yeast two-hybrid system (*6*). TIS21 is an early-response/immediate-early gene product, the expression of which is up-regulated independently of protein synthesis in response to a mitogenic signal such as hormones, serum, or phorbol esters (*56*, *58*). Its closely related family member BTG1 was originally identified because of its proximity to a breakpoint found in leukemias; however, it was found that BTG1 expression could also be induced by epidermal growth factor in RAT1 fibroblast cells (*6*). Interestingly, in dissociated brain cells from embryonic mice, a Type I arginine methyltransferase activity is stimulated 120 and 130% by the β-adrenergic agonist (−)isoproterenol and by epinephrine, respectively (*193*).

The first human homolog of rat PRMT1, HCP1, was identified in a screen

for genes that allowed *ire15* mutant yeast cells to overcome their inability to grow without inositol. HCP1 expression in the *ire15* yeast cells caused an increase in the transcription of *INO1* mRNA, encoding inositol 1-phosphate synthase, which is greatly reduced in the mutant cells (*9, 70*). In a third study, the human homolog of PRMT1 was also isolated as a protein that specifically interacts with the intracytoplasmic domain of the cytokine receptor INFAR1 (*8*). This domain of the receptor provides a docking site for proteins already known to be involved in signal transduction, such as tyrosine and serine/threonine kinases (*194, 195*).

Another role for Type I arginine methylation in signal transduction may be in the induction of the heat-shock response, at least in nonmammalian systems. After a heat shock or arsenite treatment, the methylation states of arginine residues within histone H3 and two of the 70-kDa heat-shock proteins are changed. Histone H3 in *Drosophila* shows an increase in methylarginine content (mono- and asymmetric dimethylarginine) (*128*) and HSP70A and HSP70B from chicken fibroblasts have decreased levels of monomethylarginine (*144, 145*). It is not clear whether the decrease in methylation of HSP70A and HSP70B is due to a true decrease in methylation or an artifact of the inability of the endogenous methyltransferase to cope adequately with the rapid overexpression of these proteins in response to a heat shock. In either case, the function of this change in the state of methylation is not understood. However, it is known that a large portion of HSP70 becomes localized to the nucleus after a heat shock (*145*).

There have been several reports that correlate signal transduction with possible changes in arginine methylation in unidentified proteins. For example, lipopolysaccharide stimulation of pre-B cells results in increased membrane protein methylation (*196*), whereas epidermal growth factor and nerve growth factor stimulate methylation within PC12 cells (*197*). However, there is no evidence that the increases in protein methylation are associated with arginine residues. Experiments designed to look specifically at the results of inhibiting both of the protein arginine *N*-methyltransferases found that various *S*-adenosylhomocysteine analogs, such as 5′-deoxy-5′-*S*-(2-methylpropyl)-5′-thioadenosine (SIBA), were good inhibitors of the Type I and Type II methyltransferases, but did not affect the protein lysine methyltransferase activity (*198*). These same analogs also inhibited foci formation in chick embryo fibroblasts transformed by the Rous sarcoma virus, circumstantially linking arginine methylation and the process of viral transformation (*198*). Sinefungin and SIBA both cause a decrease in all three methylated arginine derivatives *in vivo* when administered to murine splenic lymphocytes and inhibit the normal cellular lymphoproliferative response to lipopolysaccharide, again implicating the involvement of an arginine methyltransferase in a signaling pathway (*199*).

In fact, arginine methylation may even cooperate with protein phosphorylation events to control activity. A second posttranslational modification of the *S. cerevisiae* methyl-accepting substrate Npl3 has been proposed, whereby the serine residues adjacent to the proposed methylated arginine residues are phosphorylated (C. Siebel and C. Guthrie, unpublished). Having this phosphorylated SR domain (*24, 200*) overlap with an arginine-methylated GAR domain could give rise to multiple modes of regulation of Npl3 activity (*24*).

2. Nuclear Localization

There are several examples whereby the GAR domain and even the arginine methylation event have been directly implicated in promoting nuclear import. Almost all of the known asymmetric dimethylarginine-containing proteins are localized to the nucleus and even more specifically the nucleolus. A majority of the proposed substrates from yeast and other organisms are similarly localized (Table IV).

Many studies have been done characterizing the signal for nuclear import. Most of the nuclear proteins that are actively imported contain basic amino acid stretches that are termed nuclear localization sequences (NLSs) (*201*). Those destined for the nucleolus often contain additional sequences for this targeting (*201*). The GAR domains of many of the proposed substrates for the Type I protein arginine *N*-methyltransferase were assessed for the ability to promote nuclear or nucleolar targeting. A yeast *npl3* mutant missing the carboxy-terminal portion of the protein containing the GAR domain fails to localize to the nucleus (*78*). Import of nucleolin into the nucleus requires a NLS, but its further localization into the nucleolus requires the GAR domain and the adjacent RNA-binding domains (*202, 203*). The yeast Nsr1 protein also requires both the GAR domain and the RNA-binding domains for proper nucleolar import (*167*). Interestingly, a mutant form of the herpes simplex virus protein ICP27, lacking the RGG sites, is nuclear but not nucleolar localized (*131, 132, 249*). Conflicting results were reported when similar experiments were performed with both hnRNP A1 and Gar1. The nuclear localization signal within hnRNP A1 is contained within a 40-amino acid domain in the carboxy-terminal region, 30 amino acids downstream from the GAR domain (*204*). With Gar1, it was also found that neither of the GAR domains was sufficient for nucleolar localization, and they did not seem to be required because the central domain and an adjacent basic domain alone were sufficient for proper localization (*201*). However, having a nuclear localization signal in close proximity to or overlapping a nucleotide-binding region (Section VI,A,4) suggests a possible mechanism for the shuttling of certain proteins between the nucleus and the cytoplasm, such as with hnRNP A1 and Npl3 (*78, 101, 102, 250*). When the proteins are imported into the nucleus, they encounter a high concentration of mRNA substrate to bind.

The binding of mRNA with or near to the GAR domain could partially block the localization function of the domain, allowing the protein and its "cargo" to exit into the cytoplasm. The bound RNA species could then dissociate because of the lowered cytosolic concentration, and the empty shuttle protein would then be actively targeted back to the nucleus.

Although all of the above examples are still consistent with the GAR domain playing an accessory role in nuclear or nucleolar import, a direct role for not only the GAR domain but also the asymmetric dimethylation of arginines within the GAR domain was shown for the nuclear import of the high-molecular-weight forms of basic fibroblast growth factor. The 18-kDa form of bFGF is predominantly cytoplasmic; however, the higher molecular weight forms are preferentially found in the nucleus and contain asymmetrically dimethylated arginine residues (Section IV,B,4) (*114, 117, 118, 205*). These higher molecular weight forms are transcribed from alternative upstream CUG start codons, and the amino-terminal extension contains a GAR domain (*116*). After the addition of the methyltransferase inhibitor 5′-methylthioadenosine to 3T3 cells, the high molecular weight forms of bFGF were no longer preferentially localized to the nucleus as in the control cells (*106*). The higher molecular weight forms of bFGF from methylthioadenosine-treated 3T3 cells demonstrate increased electrophoretic mobility compared to when they are immunoprecipitated from untreated cells (*106*). This change in mobility was assumed to be due to methylation, because the shift is dependent on the presence of methionine, although no direct measurement of the amount of methylation in these higher molecular weight bFGF species from methylthioadenosine-treated cells was reported (*106*). Furthermore, methylthioadenosine did not increase the turnover rate of either form of bFGF and it actually enhanced the level of expression of the higher molecular weight form (*106*).

3. Nucleic Acid Binding

The most common feature of the proteins that contain GAR domains is their ability to bind single-stranded nucleic acids (Tables IV and V). For example, the GAR domain of the yeast Nab2 protein, when *in vitro* transcribed and translated in various protein contexts, binds poly(G), poly(U), and single-stranded DNA linked to a solid support resin (*176*). A GAR domain from the herpes simplex virus protein, ICP27, can also confer RNA-binding abilities when fused to heterologous proteins (*138*). ICP27 is critical for several viral functions, all of which affect cellular mRNA processing; perhaps the ability of the ICP27 GAR domain to bind nucleic acids may explain why RGG-deficient forms of ICP27 do not allow efficient replication of the herpes simplex virus in Vero cells (*131, 138*). In Epstein–Barr-virus infected cells, a sim-

ilar decrease in transcriptional activiation was observed when a mutant EBNA-1 protein with altered GAR domains was expressed compared to a wild-type version of the protein (*206, 207*).

These examples are, however, in proteins that have not been positively confirmed to contain methylated arginine residues. Therefore, the study of the function of the GAR domain in hnRNP proteins, where methylation is established, may be more relevant to understanding its role in RNA binding. The carboxy-terminal domain of hnRNP A1 (residues 203–236 in rat hnRNP A1; see Fig. 5B), containing the GAR domain, provides cooperative binding to single-stranded DNA cellulose and may be responsible for the protein–protein interactions important in this cooperativity (*208*). The carboxy-terminal domain is also necessary for the alternative splicing activity, the stable RNA binding, and the optimal annealing activity of hnRNP A1 (*100, 208*). Interestingly, bacterially produced hnRNP A1 constructs that have had their carboxy-terminal GAR domain removed are not functional in their ability to bind RNA or promote alternative splicing (*100*). Similarly, in hnRNP U, loss of the GAR domain abolishes single-stranded DNA binding activity *in vitro,* but an expressed form of the GAR domain of this protein alone is able to bind single-stranded DNA (*209*). It is important to note that in all of the *in vitro* studies performed with hnRNP A1, the protein was isolated from recombinant *E. coli* cells that do not have arginine methyltransferase activity. To examine the effect of methylation of the GAR domain, the nucleic acid-binding abilities of methylated versus unmethylated recombinant hnRNP A1 were compared (*210*). It was found that the methylated form of the protein eluted earlier in a salt gradient from a single-stranded DNA–cellulose column, indicating a weaker binding compared to the unmethylated form (*210*). The general association of the GAR domain with single-stranded nucleic acids is thought to be through the interaction of the positive charge on the arginine residues and the negative charge on the phosphate backbone (*211*). More specific interactions may arise through specific hydrogen bonds that are possible with the guanidino nitrogens and the nucleic acids. The methylation of these nitrogens does not affect the gross charge of the side chain, but the addition of two methyl groups onto a single nitrogen may disrupt the more specific hydrogen bond interactions (*211*). This hypothesis is consistent with the experimental data showing a decreased affinity for single-stranded DNA by the methylated form of hnRNP A1. Because only 1.45 arginines per hnRNP A1 molecule (out of a total of six potential sites) were shown to be methylated *in vitro* (*210*), the difference in binding affinity may be more substantial in more fully methylated proteins *in vivo.* A "less-specific" form of hnRNP A1 may be advantageous in light of the fact that the protein is responsible for differential 5′ splice site selection (*99, 212*).

4. Developmental Aspects

The Type I methyltransferase activity seems to be developmentally regulated. Arginine methyltransferase activity from either the whole homogenate or the cytosolic fraction is high in rat fetal brain and then declines rapidly after birth (*213*). Although the arginine methyltransferase activity is reduced by half in the first 30 days after birth (*213*), monomethylarginine residues in rat brain increase with a sigmoidal curve over the first 180 days (*2*). The initial increase begins at day 15, and the half-maximal content is reached at 45 days (*2*). During this same period, the content of asymmetric dimethylarginine residues did not vary (*2*).

Similar experiments with synchronized HeLa S-3 cells showed that Type I methyltransferase activity gradually increases after the initiation of mitosis and reaches a maximum 18 hr later in late G_2 (*214*). In a separate experiment, levels of methylated endogenous substrates, the 40S and 60S ribosomal subunits, were examined throughout the cell cycle (*215*). The levels of arginine-methylated 40S subunit were highest in late G_1, and the arginine-methylated 60S subunit was most abundant in early S phase (*215*).

5. Arginine Methylation with Unknown Consequences

There are many other substrates that are potentially asymmetrically methylated *in vivo* (see Table V and Section IV,B,7), of which several do not have a function or do not appear to be related to nucleic acid-binding proteins. Determining the function of these proteins and trying to assess the role of the arginine methylation is an obvious course of action. Among these proteins are histones, which may be methylated *in vivo* in response to a heat shock, as well as HMG1 and HMG2. The interactions of both of these proteins with DNA may be modulated by arginine methylation. A histone methyltransferase has even been suggested to be involved in the oddly shaped chromosomal structures seen in cells with chronic erythremic myelosis (*216, 217*).

Methylated arginine residues are resistant to proteolytic degradation by trypsin (*20, 27, 31*). This has led to the hypothesis that the modification may be involved in protecting certain proteins from proteolytic degradation. This resulting stabilization may be beneficial to some of the other uncharacterized substrates, including the heat-shock proteins HSP70A and HSP70B, cytochrome *c*, and myosin.

6. Conclusions

The roles discussed above for asymmetric protein arginine *N*-methylation are not mutually exclusive and they can be brought together under a rather general model described below. It is intriguing that proteins that are respon-

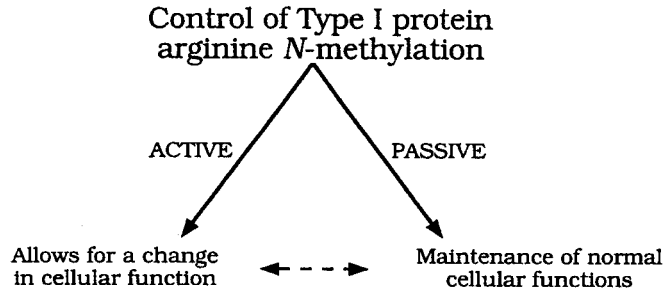


FIG. 7. A schematic of the two possible modes of *in vivo* Type I protein arginine *N*-methylation.

sible for binding single-stranded nucleic acids may be localized to the correct organelle by the same sequence that aids in the cooperative binding of the nucleic acid substrate. In addition, this system provides a way for the cell to receive extracellular signal inputs and transmit them directly to the nucleus.

In its simplest form, the model proposes that asymmetric dimethylation of arginine residues can occur in two modes, the passive and the active (Fig. 7). Passive methylation is proposed to be constitutive and occurs typically in unstimulated cells, under conditions of "normal" metabolism. For example, proteins such as fibrillarin and nucleolin have been found only in their fully methylated forms *in vivo* (*20, 71*), and these proteins as well as others may require this complete derivatization for proper localization and/or function. However, under certain conditions, "active" methylation of new, normally under- or unmethylated (or even unexpressed) substrates is required. In response to an external stimulus (cytokine, hormone, heat, etc.) the methyltransferase activity may be modulated to change substrate specificity and therefore affect cellular processes. For example, the reported levels of asymmetric dimethylarginine in hnRNP A1 *in vivo* vary (*43, 97*); therefore, the methylation state of hnRNP A1 may be changing under certain conditions to nonspecifically modify splicing patterns in mRNA transcripts. In these studies, hnRNP A1 was isolated from HeLa cells grown to cell densities that varied by as much as three orders of magnitude (*43, 97*). Perhaps the difference in the density of cell–cell contacts could trigger a change in the arginine methylation levels.

The conditions under which the asymmetric methyltransferase activity shifts from passive to active may be as general as a response to a positive growth signal, as seen in the up-regulation of methyltransferase activity in regenerating hepatic tissue in rats (*218*). This pathway could include interactions with the methyltransferase regulators TIS21 and BTG1 to effect the desired result. It is also possible that this branch of the active pathway is what

causes the induction of *INO1* mRNA in yeast *ire15* mutants. The active pathway may also be induced under conditions that induce an inhibition of growth, as in the treatment of human myeloma U266S cells with interferon-β (*8*). This active pathway is blurred with the passive arm, however, because during either type of response it is possible that the expression of passively methylated substrates, which need to be methylated for proper function, is also up-regulated. Perhaps this increase in Type I methyltransferase activity is to cope with the enhanced expression of growth-induced methyl-accepting substrates such as the pigpen protein, whose cellular levels increase during the growth and differentiation of epithelial cell during angiogenesis (*219*). Active up-regulation of the methyltransferase activity toward similar substrates would constitute the crossover between these two pathways (Fig. 7).

As with other posttranslational modifications involved in signal transduction, whether it be phosphorylation or methylation, it is important to reset the system back to the resting state by removing the modification. It is at this point that the model proposed in Fig. 7 and the role of asymmetric dimethylation of arginines in signal transduction are problematic. No demethylase has been identified for asymmetrically dimethylated arginine residues, although one has been suggested for the myelin basic protein symmetrically dimethylated arginine (*220*). Experiments with *in vivo*-labeled histones from rat tissues indicated that the turnover of the methylated arginines coincided with the degradation of the histones (*122–124*). If the mechanism for the removal of asymmetrically dimethylated proteins is specific proteolysis, it would be a very costly method of signal transduction in terms of cellular resources.

B. Symmetric Arginine Dimethylation of Myelin Basic Protein

1. Background

The relative and absolute content of mono- and symmetric dimethylarginine in myelin basic protein reported in the literature varies widely among species (*19, 21, 22, 31, 32, 34, 35*). Levels of symmetric dimethylarginine residues range from 0.033–0.6 mol per mole of myelin basic protein when isolated from human, monkey, bovine, chicken, guinea pig, rabbit, rat, frog, and turtle tissues (*19, 21, 35*). In these same myelin basic protein isolates, monomethylated arginine levels ranged from 0.17 to 0.48 mol per mole of myelin basic protein (*21, 35*). These studies also determined the *in vivo* ratios of symmetric dimethylarginine to monomethylarginine at position 107, which varied as much as from 0.14 to 1.95 (*19, 22, 32, 34, 35*). In 1977, it was hypothesized that this difference may not be due to different experimental techniques, but rather to a difference in the *in vivo* ages of the myelin

basic protein (*221*). This idea is supported by the fact that monomethylated myelin basic protein from 2-day-old chickens has a half-life of 40 days, and the symmetrically dimethylated species has an apparent half-life of less than 50 hr (*34*), whereas newly formed myelin basic protein from mouse brain, labeled with [^{3}H]leucine, has a half-life of approximately 36 days (*222*).

Further evidence indicating a difference in the extent of arginine methylation of myelin basic protein *in vivo* with age was obtained from experiments with rabbits, rats, and mice of different ages (12 days and 2 years; 8 days and 10 months; and 10 days and 6 months, respectively). Importantly, myelination in mice begins 10 days after birth and reaches a maximum at 30 days, followed by a steady decline of accumulation (*222*), whereas in rats total myelin protein begins to accumulate at 10–15 days after birth as well; however, it reaches a maximum level at 100 days that is sustained beyond 1 year after birth (*221*). In both animals, the expression of myelin basic protein also follows the general trend of myelination. From the rabbits, rats, and mice, as much as a 9-fold increase in symmetric myelin basic protein-derived dimethylarginine residues and a 14-fold increase in myelin basic protein derived-monomethylarginine were observed in the older animals versus the younger ones (*149*). Interestingly, this large difference in the absolute amounts of the two arginine derivatives is not reflected in the ratios of the two, which did not vary greatly between the age groups (*149*).

Although an explanation for the increase of arginine methylation in older animals is not obvious, it may at least suggest a functional relationship between the extent of methylation and myelination. In support of this hypothesis, the Type II methyltransferase activity from mice is maximally active toward exogenous bovine myelin basic protein at 17–40 days after birth, coinciding with the peak of myelination (*222, 223*). As expected, the levels of methylated arginine derivatives also increased during this period (*149, 221*). Furthermore, the methylation of myelin basic protein during myelination may be under hormonal control, because Type II activity from cultured mouse brain cells is stimulated by the administration of thyroid hormone (*224*).

In addition to these studies, analyses were also conducted in mutant strains of dysmyelinating mice. Jimpy mice produce normal levels of myelin basic protein, but fail to incorporate it into the myelin sheath (225). Type II protein arginine *N*-methyltransferase activity was reduced by an average of 70% in brain extracts from the 21-day-old homozygous jimpy mice compared to the normal controls; in the heterozygous mice activity levels remained similar (90%) to the control level (225). Earlier in development (12–15 days), significant differences in enzymatic activity could not be detected among the three genotypes (225). The myelin basic protein from 23-day-old heterozygous and homozygous jimpy mice contain 23 and 16%, re-

spectively, of the symmetric arginine derivative compared to the mother control, as well as 32 and 7%, respectively, for the monomethylarginine species (*149*). No explanation has been offered for the fact that heterozygous jimpy mice are phenotypically normal and have almost wild-type levels of Type II methyltransferase activity, but have significantly reduced levels of myelin basic protein-derived methylated arginine derivatives (*149, 225*). Alternatively, for the 12-day-old shiverer dysmyelinating mutant mice, the Type II protein arginine *N*-methyltransferase activity from brain extracts was greater than twofold higher than the normal controls, and fell to near normal levels by day 21 (*226*). The amounts of the specific methylated arginine derivatives were not assayed for the shiverer mice.

2. Functional Models

Attempts to understand how the methylation of the single arginine residue could effect myelination began with the modeling of the myelin basic protein structure (*192*). Myelin basic protein is quite refractory to crystallization techniques; however, early molecular modeling algorithms predicted high levels of β-sheet that could not be confirmed experimentally (*192*). The results of these modeling exercises was that a methylated versus an unmethylated arginine would make different contacts with nearby residues. Specifically, the methylated form would have favorable interactions with Pro97, stabilizing the hairpin loop created by the triproline sequence (*192*). A more recent analysis of the structure of myelin basic protein incorporated data from three-dimensional reconstructions of electron microscope analyses (*155*). This revised structure is based on the β-sheet backbone proposed earlier (*192*); however, the myelin basic protein structure model could now be constrained to fit the electron microscopy reconstruction data (*155*). In this new model, myelin basic protein has a C-shaped configuration and is still composed mainly of β-sheets. The area containing the triproline sequence and the arginine methylation site is suggested to be in a position where post-translational modifications may affect the global structure of the protein (*155*). Interestingly, however, the methylation of the arginine was not mentioned in the study (*155*) as one of the possible *in vivo* modifications to myelin basic protein.

Exogenously added myelin basic protein can induce the aggregation of unilamellar myelin vesicles *in vitro,* which can be monitored by light-scattering changes (*227*). Experiments comparing unmethylated carp myelin basic protein with the presumably *in vivo*-methylated bovine protein suggested that the arginine methylation was more efficient at promoting aggregation by 0.93 kcal/mol, whereas the unmethylated myelin basic from carp was found to be no more efficient than other cationic proteins tested, such as lysozyme or poly(histidine) (*227*). An additional increase (0.13 kcal/mol) in

efficiency for the bovine myelin basic protein was observed after it had been further methylated *in vitro* by a preparation of the Type II protein arginine *N*-methyltransferase from bovine brain (227). Interestingly, the myelin basic protein from carp was not arginine methylated *in vitro* and assayed for vesicle aggregation (227). The authors concluded that the effect of arginine methylation in the vesicle assay could not be attributed solely to the increased hydrophobicity of the methylated arginine residue, but included additional gains in free energy from more favorable ionic interactions (227).

Furthermore, incubation of mouse embryonic brain cells with the methyltransferase inhibitor sinefungin lead to an inhibition of the formation of compact myelin without decreasing the gross amount of myelin basic protein (228). From these results, it was suggested that the methylation of Arg-107 played a key role in the compaction of the myelin sheath. However, analysis of various myelin fractions at different degrees of compaction showed that the highest methyl acceptability was found in the most compact fraction of multilayered myelin, suggesting that it was the least, and not the most, methylated fraction (229). It was proposed that the methylated myelin basic protein is excluded from the compact myelin because it is less cationic than the unmethylated form, and the extent of positive charge of myelin basic protein is thought to be critical for the association of the two opposing negatively charged phopholipid membranes (229). However, in guanidine-HCl the nature of the charge on methylated myelin basic protein was not different from the unmethylated form; in fact, only after the removal of the denaturant was any difference seen (229), although this may be attributable to improper refolding *in vitro* rather than to any real difference *in vivo*.

A solution to this apparent paradox established by the vesicle aggregation experiments compared to assays on compact myelin is not at hand. It is possible that the arginine methylation of myelin basic protein is a dynamic rather than a static process. One may speculate that the methylation of Arg-107 occurs during or soon after its synthesis to prevent its rapid complexation with the membrane and allow for proper folding in a fashion similar to the proposed role of the phosphorylation of Thr-99 and Ser-166 (192). This scenario would cause a reevaluation of the data indicating that methylation induced favorable interactions with the membrane (227). The data may be viewed as artifactual if the role of the methylation is only critical soon after the protein is synthesized. Once the derivatized myelin basic protein is properly folded, it could associate with the membrane and lose its arginine methylation (or demethylation could precede membrane association). Indeed, as mentioned above, the protein may only go from symmetric dimethylarginine to monomethylarginine, as is observed in the experiment with young versus old animals (149). This aspect of the model is supported by the discovery in a crude preparation of the bovine brain myelin basic protein methyltransferase

of a demethylase activity that could not be attributed to a loss of myelin basic protein through general proteolysis (*220*).

3. Myelin Basic Protein Methylation and the Relationship to Disease

Defects in myelination have been associated with alterations in methyl group metabolism. For example, a deficiency of vitamin B_{12} (a necessary cofactor for the synthesis of methionine, a precursor to the methyl-group donor AdoMet) is associated with malformed myelin sheaths in subacute combined degeneration disease (SCD) (*27*). Another inhibitor of AdoMet synthesis, 1-aminocyclopentane carboxylic acid (cycloleucine), can induce myelin lesions similar to those seen in SCD when injected into chickens (*230*) or when given to mice (*231, 232*). Finally, nitrous oxide (an inhibitor of methionine synthetase) also causes these lesions in myelin, and the neurological as well as the histological phenotypes can be fully reversed by supplementing the diet with methionine (*233*). However, whether these effects are mediated in whole or in part via arginine methylation in myelin basic protein is not clear. Administration of cycloleucine to rats does cause a 26% decrease in myelin basic protein methylation; however, no effect was found with nitrous oxide treatment of rats, nor did a depletion of vitamin B_{12} in bats by dietary measures cause a significant decline (*234*).

The methylation of arginine residues and their eventual metabolism may also reflect on pathological conditions. Levels of asymmetric dimethylarginine in the urine of 12 patients with muscular distrophy were initially found to be 2-fold higher than in 10 normal children, and the average ratio of the asymmetric to the symmetric derivative increased from 1.4 to 2.6 (*235*). Later it was observed that both types of dimethylarginine residue levels were almost 10-fold higher in a single patient affected with muscular distrophy compared to 9 unaffected individuals, and the ratio of the two residues was 1.6 (*236*). The discrepancy between these two studies has yet to be explained; however, it may be due to differences in sample size or to the fact that children (5–16 years old) were specifically used in the earlier study (*235*), while presumably only a single adult was used in the later one, although the age of the individual is not specifically stated (*236*). This change in the absolute excretion levels of dimethylarginine may still be of use for the early diagnosis of this disease.

The levels of these dimethylated arginine derivatives in normal human urine (16–71 mmol/mg creatinine) are comparable to the levels of arginine (6–40 mmol/mg creatinine), but are significantly lower than for other methylated amino acids, such as 3-methylhistidine (120–385 mmol/mg creatinine) (*235, 236*). This difference suggests that there is an active mechanism for the metabolism of the proteolytic products of arginine-methylated proteins

In fact, several enzymes have been characterized that may be involved in the catabolism of these free methylated arginines in rats (*237–241*). Some of the by-products of these reactions include N^G-methylagmatine, ornithine, the corresponding α-keto acids, and citrulline. The importance of these catabolic enzymes may become relevant in light of the fact that methylguanidine may be related to several uremic disorders (*242*). Additionally, free monomethylarginine (*243*) is known to inhibit nitric oxide synthase *in vitro* (*244, 245*), and perhaps the dimethylarginine derivatives can inhibit this enzyme as well.

VII. Future Directions

In this review, we have provided an overview of protein arginine methylation, a field in in which there is considerable activity at this time. Current excitement comes from the identification of new protein arginine *N*-methyltransferase genes and the analysis of new potential methyl-accepting substrates.

At present, at least two mammalian genes are known that encode catalytically active arginine methyltransferase subunits, PRMT1 and PRMT3. Three splice variants of the PRMT1 gene in humans have also been described. All can be classified as Type I methyltransferases based on their activity with the synthetic substrate GST–GAR, but they still methylate only a single 55-kDa polypeptide in *rmt1* yeast extracts, compared to the 13 substrates recognized by the GST–Rmt1 yeast fusion Type I protein arginine *N*-methyltransferase (*5*) (J. Tang and J. Gary, unpublished results). There are also five "orphan" mammalian substrates (HMG1, HMG2, HSP70A, HSP70B, and myosin) that have been tentatively shown to contain methylated arginines *in vivo* but do not contain the asymmetric methylation consensus. It is unclear whether distinct gene products are required for their methylation. A gene for the Type II arginine methyltransferase has also not yet been identified. Together, these findings suggest the existence of potentially many more protein arginine *N*-methyltransferases in higher organisms with different substrate specificities. The finding of additional protein arginine *N*-methyltransferase substrates is critical for discovering new methyltransferases and determining the complexity of this family of enzymes and ultimately its role in various cellular processes. At this stage, it will be important to reinvestigate those potential substrates identified only by amino acid analysis with more purified forms of the enzymes and more sensitive means of analysis, such as mass spectroscopy, to verify the original claims. Having a larger pool of identified substrates, recombinantly expressed, gives us the ability to either purify novel methyltransferases from cellular extracts

or assess the activity of the methyltransferase fusion enzymes already constructed.

The most crucial and rewarding work will be done by those whose protein of interest is a substrate for a protein arginine *N*-methyltransferase. The ability to obtain purified methylated or unmethylated protein from recombinant bacterial cells coexpressing a protein arginine *N*-methyltransferase for use in an *in vitro* assay will open the door to understanding the function of this methylation event. The coexpression of the methyltransferase and the substrate is suggested over *in vitro* methylation based on our results attempting to stoichiometrically methylate recombinant Npl3 with GST–Rmt1. Under all of the conditions tested, such as high AdoMet concentrations and thrombin removal of the GST moiety, we were unable to achieve stoichiometric methylation (C. Seibel and J. Gary, unpublished). This result supports earlier hypotheses that the methylation event was either cotranslational or very tightly coupled to it (*246*).

As more of the proteins listed in Table V are characterized, perhaps additional *in vitro* assays will be developed and used to probe the functional advantage of arginine methylation. In the yeast system, these types of functional assays can be performed *in vivo* using an *RMT1* versus an *rmt1* strain in combination with mutations in the gene of choice. Indeed, this is how the *RMT1* gene was identified (*7*). Using this type of approach, it might be possible to ascertain what portions of the substrate proteins act in conjunction with methylation, causing an interesting phenotype that may reveal its role in nucleotide binding, signal transduction, or localization. For example, are point mutations in the conserved RNA recognition motifs more deleterious in an *rmt1* strain, or can a GAR domain deletion be rescued by including a heterologous nuclear localization signal? Making synthetic substrates by the addition of either Type I or Type II arginine methylation sites to certain proteins may also be instructive.

The past 30 years of literature concerning the symmetric and asymmetric dimethylation of arginine residues is large and often contradictory. As the specific methyltransferase genes are cloned and expressed and methods of detecting and analyzing the derivatized residues improve, it will be of great importance to revisit many of the questions raised here with the hope of unifying the record and providing a physiological rationale for this posttranslational modification reaction.

References

1. S. Clarke, *Curr. Biol.* **5,** 977 (1993).
2. Y. Kakimoto, Y. Matsuoka, M. Miyake, and H. Konishi, *J. Neurochem.* **24,** 893 (1975).

3. P. Paolantonacci, F. Lawrence, F. Lederer, and M. Robert-Gero, *Mol. Biochem. Parasitol.* **21,** 47 (1986).
4. N. Yarlett, A. Quamina, and C. J. Bacchi, *J. Gen. Microbiol.* **137,** 717 (1991).
5. J. D. Gary, W. J. Lin, M. C. Yang, H. R. Herschman, and S. Clarke, *J. Biol. Chem.* **271,** 12585 (1996).
6. W. J. Lin, J. D. Gary, M. C. Yang, S. Clarke, and H. R. Herschman, *J. Biol. Chem.* **271,** 15034 (1996).
7. M. F. Henry and P. A. Silver, *Mol. Cell. Biol.* **16,** 3668 (1996).
8. C. Abramovich, B. Yakobson, J. Chebath, and M. Revel, *EMBO J.* **16,** 260 (1997).
9. J. Nikawa, H. Nakano, and N. Ohi, *Gene* **171,** 107 (1996).
10. D. E. Atkinson, "Cellular Energy Metabolism and its Regulation." Academic Press, New York, 1977.
11. W. K. Paik and S. Kim, *Biochem. Biophys. Res. Commun.* **29,** 14 (1967).
12. W. K. Paik and S. Kim, *J. Biol. Chem.* **243,** 2108 (1968).
13. A. M. Kaye and D. Sheratzky, *Biochim. Biophys. Acta* **190,** 527 (1969).
14. Y. Kakimoto and S. Akazawa, *J. Biol. Chem.* **245,** 5751 (1970).
15. T. Teerlink and B. E. de, *J. Chromatog.* **4 91,** 418 (1989).
16. T. Nakajima, Y. Matsuoka, and Y. Kakimoto, *Biochim. Biophys. Acta* **230,** 212 (1971).
17. Y. Kakimoto, *Biochim. Biophys. Acta* **243,** 31 (1971).
18. E. H. Eylar, S. Brostoff, G. Hashim, J. Caccam, and P. Burnett, *J. Biol. Chem.* **246,** 5770 (1971).
19. S. W. Brostoff, A. Rosegay, and W. J. Vandenheuvel, *Arch. Biochem. Biophys.* **148,** 156 (1972).
20. M. A. Lischwe, R. G. Cook, Y. S. Ahn, L. C. Yeoman, and H. Busch, *Biochemistry* **24,** 6025 (1985).
21. G. E. Deibler and R. E. Martenson, *J. Biol. Chem.* **248,** 2392 (1973).
22. P. R. Young and F. Grynspan, *J. Chromatog.* **421,** 130 (1987).
23. M. Ohtsuka, S. Ihara, R. Ogawa, T. Watanabe, and Y. Watanabe, *Int. J. Cancer* **34,** 855 (1984).
24. C. W. Siebel and C. Guthrie, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 13641 (1996).
25. D. Gallwitz, *Arch. Biochem. Biophys.* **145,** 650 (1971).
26. M. K. Paik, K. H. Lee, S. S. Hson, I. M. Park, J. H. Hong, and B. D. Hwang, *Int. J. Biochem.* **23,** 939 (1991).
27. G. S. Baldwin and P. R. Carnegie, *Science* **171,** 579 (1971).
28. A. J. Crang and W. Jacobson, *Biochem. Soc. Trans.* **8,** 611 (1980).
29. N. Sundarraj and S. E. Pfeiffer, *Biochem. Biophys. Res. Commun.* **52,** 1039 (1973).
30. M. Miyake, *J. Neurochem.* **24,** 909 (1975).
31. S. Brostoff and E. H. Eylar, *Proc. Natl. Acad. Sci. U.S.A.* **68,** 765 (1971).
32. G. S. Baldwin and P. R. Carnegie, *Biochem. J.* **123,** 69 (1971).
33. P. R. Dunkley and P. R. Carnegie, *Biochem. J.* **141,** 243 (1974).
34. D. H. Small and P. R. Carnegie, *J. Neurochem.* **38,** 184 (1982).
35. G. E. Deibler and R. E. Martenson, *J. Biol. Chem.* **248,** 2387 (1973).
36. M. Miyake and Y. Kakimoto, *J. Neurochem.* **20,** 859 (1973).
37. G. M. Jones and P. R. Carnegie, *J. Neurochem.* **23,** 1231 (1974).
38. H. W. Lee, S. Kim, and W. K. Paik, *Biochemistry* **16,** 78 (1977).
39. G. H. Park, L. P. Chanderkar, W. K. Paik, and S. Kim, *Biochim. Biophys. Acta* **874,** 30 (1986).
40. S. K. Ghosh, W. K. Paik, and S. Kim, *J. Biol. Chem.* **263,** 19024 (1988).
41. P. R. Young and C. M. Waickus, *Biochem. J.* **250,** 221 (1988).
42. B. McCright, A. M. Rivers, S. Audlin, and D. M. Virshup, *J. Biol. Chem.* **271,** 22081 (1996).

43. A. L. Beyer, M. E. Christensen, B. W. Walker, and W. M. LeStourgeon, *Cell* **11,** 127 (1977).
44. J. Karn, G. Vidali, L. C. Boffa, and V. G. Allfrey, *J. Biol. Chem.* **252,** 7307 (1977).
45. S. K. Ghosh, S. K. Syed, S. Jung, W. K. Paik, and S. Kim, *Biochim. Biophys. Acta* **1039,** 142 (1990).
46. R. Rajpurohit, W. K. Paik, and S. Kim, *Biochim. Biophys. Acta* **1122,** 183 (1992).
47. R. Rajpurohit, S. O. Lee, J. O. Park, W. K. Paik, and S. Kim, *J. Biol. Chem.* **269,** 1075 (1994).
48. N. Rawal, R. Rajpurohit, W. K. Paik, and S. Kim, *Biochem. J.* **300,** 483 (1994).
49. Q. Liu and G. Dreyfuss, *Mol. Cell. Biol.* **15,** 2800 (1995).
50. A. Gupta, D. Jensen, S. Kim, and W. K. Paik, *J. Biol. Chem.* **257,** 9677 (1982).
51. S. G. Disa, A. Gupta, S. Kim, and W. K. Paik, *Biochemistry* **25,** 2443 (1986).
52. P. Casellas and P. Jeanteur, *Biochim. Biophys. Acta* **519,** 243 (1978).
53. J. Z. Farooqui, M. Tuck, and W. K. Paik, *J. Biol. Chem.* **260,** 537 (1985).
54. S. Fields and O. Song, *Nature (London)* **340,** 245 (1989).
55. S. Fields and R. Sternglanz, *Trends Genet.* **10,** 286 (1994).
56. B. S. Fletcher, R. W. Lim, B. C. Varnum, D. A. Kujubu, R. A. Koski, and H. R. Herschman, *J. Biol. Chem.* **266,** 14511 (1991).
57. A. Bradbury, R. Possenti, E. M. Shooter, and F. Tirone, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 3353 (1991).
58. B. C. Varnum, S. T. Reddy, R. A. Koski, and H. R. Herschman, *J. Cell. Physiol.* **158,** 205 (1994).
59. R. Rimokh, F. Berger, P. Cornillet, K. Wahbi, J. P. Rouault, M. Ffrench, P. A. Bryon, M. Gadoux, O. Gentilhomme, D. Germain *et al., Genes Chromosom. Cancer* **2,** 223 (1990).
60. J. P. Rouault, C. Samarut, L. Duret, C. Tessa, J. Samarut, and J. P. Magaud, *Gene* **129,** 303 (1993).
61. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J. Mol. Biol.* **215,** 403 (1990).
62. M. J. Dognin and B. Wittmann-Liebold, *Hoppe-Seylers Z. Physiol. Chem.* **361,** 1697 (1980).
63. A. Vanet, J. A. Plumbridge, and J. H. Alix, *J. Bacteriol.* **175,** 7178 (1993).
64. R. M. Kagan and S. Clarke, *Arch. Biochem. Biophys.* **310,** 417 (1994).
65. G. Schluckebier, M. O'Gara, W. Saenger, and X. Cheng, *J. Mol. Biol.* **247,** 16 (1995).
66. J. D. Lowenson and S. Clarke, *in* "Deamidation and Isoaspartate Formation in Peptides and Proteins" (D. W. Aswad, ed.), p. 47. CRC Press, Boca Raton, Florida, 1995.
67. H. Xie and S. Clarke, *J. Biol. Chem.* **268,** 13364 (1993).
68. H. Xie and S. Clarke, *J. Biol. Chem.* **269,** 1981 (1994).
69. J. Najbauer, B. A. Johnson, A. L. Young, and D. W. Aswad, *J. Biol. Chem.* **268,** 10501 (1993).
70. J. Nikawa, *Gene* **149,** 367 (1994).
71. M. A. Lischwe, R. L. Ochs, R. Reddy, R. G. Cook, L. C. Yeoman, E. M. Tan, M. Reichlin, and H. Busch, *J. Biol. Chem.* **260,** 14304 (1985).
72. N. Katsanis, M. L. Yaspo, and E. M. Fisher, *Mammal. Genome* **8,** 526 (1997).
73. A. Loubbardi, C. Marcireau, F. Karst, and M. Guilloton, *J. Bacteriol.* **177,** 1817 (1995).
74. H. Feldmann, M. Aigle, G. Aljinovic, B., andre, M. C. Baclet, C. Barthe, A. Baur, A. M. Becam, N. Biteau, E. Boles *et al., EMBO J.* **13,** 5795 (1994).
75. T. Kruiswijk, A. Kunst, R. J. Planta, and W. H. Mager, *Biochem. J.* **175,** 221 (1978).
76. M. A. Bossie, C. DeHoratius, G. Barcelo, and P. Silver, *Mol. Biol. Cell* **3,** 875 (1992).
77. I. D. Russell and D. Tollervey, *J. Cell Biol.* **119,** 737 (1992).
78. J. Flach, M. Bossie, J. Vogel, A. Corbett, T. Jinks, D. A. Willins, and P. A. Silver, *Mol. Cell. Biol.* **14,** 8399 (1994).

79. I. Russell and D. Tollervey, *Eur. J. Cell Biol.* **66,** 293 (1995).
80. J. P. Girard, H. Lehtonen, M. Caizergues-Ferrer, F. Amalric, D. Tollervey, and B. Lapeyre, *EMBO J.* **11,** 673 (1992).
81. S. Zolnierowicz, C. Csortos, J. Bondor, A. Verin, M. C. Mumby, and A. A. DePaoli-Roach, *Biochemistry* **33,** 11858 (1994).
82. C. Kamibayashi, R. Estes, R. L. Lickteig, S. I. Yang, C. Craft, and M. C. Mumby, *J. Biol. Chem.* **269,** 20139 (1994).
83. P. Turowski, A. Fernandez, B. Favre, N. J. Lamb, and B. A. Hemmings, *J. Cell Biol.* **129,** 397 (1995).
84. W. K. Paik and S. Kim, *J. Neurochem.* **16,** 1257 (1969).
85. J. Najbauer and D. W. Aswad, *J. Biol. Chem.* **265,** 12717 (1990).
86. J. P. Aris and G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 931 (1991).
87. K. Tyc and J. A. Steitz, *EMBO J.* **8,** 3113 (1989).
88. M. E. Christensen and K. P. Fuxa, *Biochem. Biophys. Res. Commun.* **155,** 1278 (1988).
89. A. W. Prestayko, M. O. Olson, and H. Busch, *FEBS Lett.* **44,** 131 (1974).
90. G. Jordan, *Nature (London)* **329,** 489 (1987).
91. B. Lapeyre, H. Bourbon, and F. Amalric, *Proc. Natl. Acad. Sci. U.S.A.* **84,** 1472 (1987).
92. M. Caizergues-Ferrer, P. Mariottini, C. Curie, B. Lapeyre, N. Gas, F. Amalric, and F. Amaldi, *Genes. Dev.* **3,** 324 (1989).
93. R. A. Borer, C. F. Lehner, H. M. Eppenberger, and E. A. Nigg, *Cell* **56,** 379 (1989).
94. M. A. Lischwe, K. D. Roberts, L. C. Yeoman, and H. Busch, *J. Biol. Chem.* **257,** 14600 (1982).
95. K. R. Williams, K. L. Stone, M. B. LoPresti, B. M. Merrill, and S. R. Planck, *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5666 (1985).
96. F. Cobianchi, D. N. SenGupta, B. Z. Zmudzka, and S. H. Wilson, *J. Biol. Chem.* **261,** 3536 (1986).
97. A. Kumar, K. R. Williams, and W. Szer, *J. Biol. Chem.* **261,** 11266 (1986).
98. F. Cobianchi, R. L. Karpel, K. R. Williams, V. Notario, and S. H. Wilson, *J. Biol. Chem.* **263,** 1063 (1988).
99. A. Mayeda, D. M. Helfman, and A. R. Krainer, *Mol. Cell. Biol.* **13,** 2993 (1993).
100. A. Mayeda, S. H. Munroe, J. F. Caceres, and A. R. Krainer, *EMBO J.* **13,** 5483 (1994).
101. S. Pinol-Roma and G. Dreyfuss, *Nature (London)* **355,** 730 (1992).
102. S. Pinol-Roma and G. Dreyfuss, *Mol. Cell. Biol.* **13,** 5762 (1993).
103. D. Gospodarowicz, G. Neufeld, and L. Schweigerer, *Cell Diff.* **19,** 1 (1986).
104. D. Gospodarowicz, A. Baird, J. Cheng, G. M. Lui, F. Esch, and P. Bohlen, *Endocrinology* **118,** 82 (1986).
105. R. Montesano, J. D. Vassalli, A. Baird, R. Guillemin, and L. Orci, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 7297 (1986).
106. G. Pintucci, N. Quarto, and D. B. Rifkin, *Mol. Biol. Cell* **7,** 1249 (1996).
107. C. Basilico and D. Moscatelli, *Adv. Cancer Res.* **59,** 115 (1992).
108. A. Jackson, S. Friedman, X. Zhan, K. A. Engleka, R. Forough, and T. Maciag, *Proc. Natl. Acad. Sci. U.S.A.* **89,** 10691 (1992).
109. S. R. Opalenik, J. T. Shin, J. N. Wehby, V. K. Mahesh, and J. A. Thompson, *J. Biol. Chem.* **270,** 17457 (1995).
110. P. L. McNeil, L. Muthukrishnan, E. Warder, and P. A. D'Amore, *J. Cell Biol.* **109,** 811 (1989).
111. P. T. Ku and P. A. D'Amore, *J. Cell. Biochem.* **58,** 328 (1995).
112. M. T. Story, F. Esch, S. Shimasaki, J. Sasse, S. C. Jacobs, and R. K. Lawson, *Biochem. Biophys. Res. Commun.* **142,** 702 (1987).

113. J. A. Abraham, J. L. Whang, A. Tumolo, A. Mergia, J. Friedman, D. Gospodarowicz, and J. C. Fiddes, *EMBO J.* **5,** 2523 (1986).
114. A. Sommer, D. Moscatelli, and D. B. Rifkin, *Biochem. Biophys. Res. Commun.* **160,** 1267 (1989).
115. A. Sommer, M. T. Brewer, R. C. Thompson, D. Moscatelli, M. Presta, and D. B. Rifkin, *Biochem. Biophys. Res. Commun.* **144,** 543 (1987).
116. H. Prats, M. Kaghad, A. C. Prats, M. Klagsbrun, J. M. Lelias, P. Liauzun, P. Chalon, J. P. Tauber, F. Amalric, J. A. Smith, *et al., Proc. Natl. Acad. Sci. U.S.A.* **86,** 1836 (1989).
117. W. H. Burgess, J. Bizik, T. Mehlman, N. Quarto, and D. B. Rifkin, *Cell Reg.* **2,** 87 (1991).
118. M. Renko, N. Quarto, T. Morimoto, and D. B. Rifkin, *J. Cell. Physiol.* **144,** 108 (1990).
119. P. Casellas and P. Jeanteur, *Biochim. Biophys. Acta* **519,** 255 (1978).
120. W. M. Bonner, M. H. West, and J. D. Stedman, *Eur. J. Biochem.* **109,** 17 (1980).
121. A. Zweidler, *Meth. Cell Biol.* **17,** 223 (1978).
122. P. Byvoet, *Biochim. Biophys. Acta* **238,** 375a-376b (1971).
123. P. Byvoet, G. R. Shepherd, J. M. Hardin, and B. J. Noland, *Arch. Biochem. Biophys.* **148,** 558 (1972).
124. P. Byvoet, *Arch. Biochem. Biophys.* **152,** 887 (1972).
125. W. K. Paik and S. Kim, *J. Biol. Chem.* **245,** 88 (1970).
126. J. A. Duerre and S. Chakrabarty, *J. Biol. Chem.* **250,** 8457 (1975).
127. M. T. Tuck and R. Cox, *Carcinogenesis* **3,** 431 (1982).
128. R. Desrosiers and R. M. Tanguay, *J. Biol. Chem.* **263,** 4686 (1988).
129. C. G. Edmonds, J. A. Loo, R. D. Smith, A. F. Fuciarelli, B. D. Thrall, J. E. Morris, and D. L. Springer, *J. Toxicol. Environ. Health* **40,** 159 (1993).
130. M. Kouach, D. Belaiche, M. Jaquinod, M. Couppez, D. Kmiecik, G. Ricart, D. A. Van, P. Sautiere, and G. Briand, *Biol. Mass Spectrom.* **23,** 283 (1994).
131. W. E. Mears, V. Lam, and S. A. Rice, *J. Virol.* **69,** 935 (1995).
132. M. K. Hibbard and R. M. Sandri-Goldin, *J. Virol.* **69,** 4656 (1995).
133. A. M. McCarthy, L. McMahan, and P. A. Schaffer, *J. Virol.* **63,** 18 (1989).
134. W. R. Hardy and R. M. Sandri-Goldin, *J. Virol.* **68,** 7790 (1994).
135. R. M. Sandri-Goldin, M. K. Hibbard, and M. A. Hardwicke, *J. Virol.* **69,** 6063 (1995).
136. D. M. Knipe, D. Senechek, S. A. Rice, and J. L. Smith, *J. Virol.* **61,** 276 (1987).
137. S. A. Rice, L. S. Su, and D. M. Knipe, *J. Virol.* **63,** 3399 (1989).
138. W. E. Mears and S. A. Rice, *J. Virol.* **70,** 7445 (1996).
139. M. S. Lee, M. Henry, and P. A. Silver, *Genes Dev.* **10,** 1233 (1996).
140. G. H. Goodwin and E. W. Johns, *Eur. J. Biochem.* **40,** 215 (1973).
141. J. F. Maher and D. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 6716 (1996).
142. T. T. Paull, M. J. Haykinson, and R. C. Johnson, *Genes Dev.* **7,** 1521 (1993).
143. L. C. Boffa, R. Sterner, G. Vidali, and V. G. Allfrey, *Biochem. Biophys. Res. Commun.* **89,** 1322 (1979).
144. C. Wang, E. Lazarides, C. M. O'Connor, and S. Clarke, *J. Biol. Chem.* **257,** 8356 (1982).
145. C. Wang, J.-M. Lin, and E. Lazarides, *Arch. Biochem. Biophys.* **297,** 169 (1992).
146. C. Wang and M.-R. Lee, *Biochem. J.* **294,** 69 (1993).
147. M. Reporterer and J. L. Corbin, *Biochem. Biophys. Res. Commun.* **43,** 644 (1971).
148. R. Helm, Z. Deyl, and O. Vancikova, *Exp. Gerontol.* **12,** 245 (1977).
149. N. Rawal, Y. J. Lee, W. K. Paik, and S. Kim, *Biochem. J.* **287,** 929 (1992).
150. L. F. Eng, F. C. Chao, B. Gerstl, D. Pratt, and M. G. Tavaststjerna, *Biochemistry* **7,** 4455 (1968).
151. E. E. Golds and P. E. Braun, *J. Biol. Chem.* **251,** 4729 (1976).
152. F. X. Omlin, H. D. Webster, C. G. Palkovits, and S. R. Cohen, *J. Cell Biol.* **95,** 242 (1982).
153. J. F. Poduslo and P. E. Braun, *J. Biol. Chem.* **250,** 1099 (1975).

154. R. E. Martenson, *in* "Biochemistry of Brain" (S. Kumar, ed.), p. 49. Pergamon, Oxford, 1980.
155. R. A. Ridsdale, D. R. Beniac, T. A. Tompkins, M. A. Moscarello, and G. Harauz, *J. Biol. Chem.* **272,** 4269 (1997).
156. N. Rawal, R. Rajpurohit, M. A. Lischwe, K. R. Williams, W. K. Paik, and S. Kim, *Biochim. Biophys. Acta* **1248,** 11 (1995).
157. C. Wang, J. M. Lin, and E. Lazarides, *Arch. Biochem. Biophys.* **297,** 169 (1992).
158. E. M. Ellis and G. A. Reid, *Gene* **132,** 175 (1993).
159. S. Loo, P. Laurenson, M. Foss, A. Dillin, and J. Rine, *Genetics* **141,** 889 (1995).
160. M. Henry, C. Z. Borland, M. Bossie, and P. A. Silver, *Genetics* **142,** 103 (1996).
161. R. Henriquez, G. Blobel, and J. P. Aris, *J. Biol. Chem.* **265,** 2209 (1990).
162. D. Tollervey, H. Lehtonen, M. Carmo-Fonseca, and E. C. Hurt, *EMBO J.* **10,** 573 (1991).
163. R. P. Jansen, E. C. Hurt, H. Kern, H. Lehtonen, M. Carmo-Fonseca, B. Lapeyre, and D. Tollervey, *J. Cell Biol.* **113,** 715 (1991).
164. T. Schimmang, D. Tollervey, H. Kern, R. Frank, and E. C. Hurt, *EMBO J.* **8,** 4015 (1989).
165. D. Tollervey, H. Lehtonen, R. Jansen, H. Kern, and E. C. Hurt, *Cell* **72,** 443 (1993).
166. B. Lubben, P. Fabrizio, B. Kastner, and R. Luhrmann, *J. Biol. Chem.* **270,** 11549 (1995).
167. C. Yan and T. Melese, *J. Cell Biol.* **123,** 1081 (1993).
168. W. C. Lee, Z. X. Xue, and T. Melese, *J. Cell Biol.* **113,** 1 (1991).
169. K. Kondo and M. Inouye, *J. Biol. Chem.* **267,** 16252 (1992).
170. K. Kondo, L. R. Kowalski, and M. Inouye, *J. Biol. Chem.* **267,** 16259 (1992).
171. W. C. Lee, D. Zabetakis, and T. Melese, *Mol. Cell. Biol.* **12,** 3865 (1992).
172. C. Gamberi, G. Contreas, M. G. Romanelli, and C. Morandi, *Gene* **148,** 59 (1994).
173. J. J. Lin and V. A. Zakian, *Nucleic Acids Res.* **22,** 4906 (1994).
174. L. M. Konkel, S. Enomoto, E. M. Chamberlain, P. McCune-Zierath, S. J. Iyadurai, and J. Berman, *Proc. Natl. Acad. Sci. U.S.A.* **92,** 5558 (1995).
175. S. Marcand, E. Gilson, and D. Shore, *Science* **275,** 986 (1997).
176. J. T. Anderson, S. M. Wilson, K. V. Datar, and M. S. Swanson, *Mol. Cell. Biol.* **13,** 2730 (1993).
177. A. Y. Jong, R. Aebersold, and J. L. Campbell, *J. Biol. Chem.* **260,** 16367 (1985).
178. A. Y. Jong and J. L. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 877 (1986).
179. A. Y. Jong, M. W. Clark, M. Gilbert, A. Oehm, and J. L. Campbell, *Mol. Cell. Biol.* **7,** 2947 (1987).
180. M. E. Cusick, *Biochim. Biophys. Acta* **1171,** 176 (1992).
181. S. G. Lee, I. Lee, S. H. Park, C. Kang, and K. Song, *Genomics* **25,** 660 (1995).
182. D. C. Masison, A. Blanc, J. C. Ribas, K. Carroll, N. Sonenberg, and R. B. Wickner, *Mol. Cell. Biol.* **15,** 2763 (1995).
183. S. P. Ridley, S. S. Sommer, and R. B. Wickner, *Mol. Cell. Biol.* **4,** 761 (1984).
184. A. W. Johnson and R. D. Kolodner, *Mol. Cell. Biol.* **15,** 2719 (1995).
185. S. M. Noble and C. Guthrie, *Genetics* **143,** 67 (1996).
186. F. He and A. Jacobson, *Genes Dev.* **9,** 437 (1995).
187. E. Sakashita and H. Sakamoto, *J. Biochem.* **120,** 1028 (1996).
188. C. W. Siebel, A. Admon, and D. C. Rio, *Genes Dev.* **9,** 269 (1995).
189. M. D. Adams, R. S. Tarng, and D. C. Rio, *Genes Dev.* **11,** 129 (1997).
190. J. Wang and L. R. Bell, *Genes Dev.* **8,** 2072 (1994).
191. J. Valcarcel, R. Singh, P. D. Zamore, and M. R. Green, *Nature (London)* **362,** 171 (1993).
192. G. L. Stoner, *J. Neurochem.* **43,** 433 (1984).
193. S. G. Amur, G. Shanker, and R. A. Pieringer, *J. Neurosci. Res.* **16,** 377 (1986).
194. C. Abramovich, L. M. Shulman, E. Ratovitski, S. Harroch, M. Tovey, P. Eid, and M. Revel, *EMBO J.* **13,** 5871 (1994).

195. H. Yan, K. Krishnan, A. C. Greenlund, S. Gupta, J. T. Lim, R. D. Schreiber, C. W. Schindler, and J. J. Krolewski, *EMBO J.* **15,** 1064 (1996).
196. R. E. Law, J. B. Stimmel, M. A. Damore, C. Carter, S. Clarke, and R. Wall, *Mol. Cell. Biol.* **12,** 103 (1992).
197. D. A. Kujubu, J. B. Stimmel, R. E. Law, H. R. Herschman, and S. Clarke, *J. Neurosci. Res.* **36,** 58 (1993).
198. J. Enouf, F. Lawrence, C. Tempete, M. Robert-Gero, and E. Lederer, *Cancer Res.* **39,** 4497 (1979).
199. B. R. Kim and K. H. Yang, *Toxicol. Lett.* **59,** 109 (1991).
200. E. Birney, S. Kumar,, and A. R. Krainer, *Nucleic Acids Res.* **21,** 5803 (1993).
201. J. P. Girard, C. Bagni, M. Caizergues-Ferrer, F. Amalric, and B. Lapeyre, *J. Biol. Chem.* **269,** 18499 (1994).
202. M. S. Schmidt-Zachmann and E. A. Nigg, *J. Cell Sci.* **105,** 799 (1993).
203. B. Messmer and C. Dreyer, *Eur. J. Cell Biol.* **61,** 369 (1993).
204. H. Siomi and G. Dreyfuss, *J. Cell Biol.* **129,** 551 (1995).
205. R. Z. Florkiewicz, A. Baird, and A. M. Gonzalez, *Growth Factors* **4,** 265 (1991).
206. D. K. Snudden, J. Hearing, P. R. Smith, F. A. Grasser, and B. E. Griffin, *EMBO J.* **13,** 4840 (1994).
207. J. L. Yates and S. M. Camiolo, *Cancer Cells* **6,** 197 (1988).
208. S. G. Nadler, B. M. Merrill, W. J. Roberts, K. M. Keating, M. J. Lisbin, S. F. Barnett, S. H. Wilson, and K. R. Williams, *Biochemistry* **30,** 2968 (1991).
209. M. Kiledjian and G. Dreyfuss, *EMBO J.* **11,** 2655 (1992).
210. R. Rajpurohit, W. K. Paik, and S. Kim, *Biochem. J.* **304,** 903 (1994).
211. B. J. Calnan, B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel, *Science* **252,** 1167 (1991).
212. A. Mayeda and A. R. Krainer, *Cell* **68,** 365 (1992).
213. W. K. Paik, S. Kim, and H. W. Lee, *Biochem. Biophys. Res. Commun.* **46,** 933 (1972).
214. H. W. Lee, W. K. Paik, and T. W. Borun, *J. Biol. Chem.* **248,** 4194 (1973).
215. F. N. Chang, I. J. Navickas, C. Au, and C. Budzilowicz, *Biochim. Biophys. Acta* **518,** 89 (1978).
216. L. Kass and C. J. Zarafonetis, *Proc. Soc. Exp. Biol. Med.* **145,** 944 (1974).
217. L. Kass and D. Munster, *Biochim. Biophys. Acta* **524,** 497 (1978).
218. H. W. Lee and W. K. Paik, *Biochim. Biophys. Acta* **277,** 107 (1972).
219. M. C. Alliegro and M. A. Alliegro, *Dev. Biol.* **174,** 288 (1996).
220. P. R. Young and C. M. Waickus, *Biochem. Biophys. Res. Commun.* **142,** 200 (1987).
221. M. O. Aspillaga and J. R. McDermott, *J. Neurochem.* **28,** 1147 (1977).
222. L. P. Chanderkar, W. K. Paik, and S. Kim, *Biochem. J.* **240,** 471 (1986).
223. A. J. Crang and W. Jacobson, *J. Neurochem.* **39,** 244 (1982).
224. S. G. Amur, G. Shanker, and R. A. Pieringer, *J. Neurochem.* **43,** 494 (1984).
225. S. Kim, M. Tuck, M. Kim, A. T. Campagnoni, and W. K. Paik, *Biochem. Biophys. Res. Commun.* **123,** 468 (1984).
226. S. Kim, M. Tuck, L. L. Ho, A. T. Campagnoni, E. Barbarese, R. L. Knobler, F. D. Lublin, L. P. Chanderkar, and W. K. Paik, *J. Neurosci. Res.* **16,** 357 (1986).
227. P. R. Young, D. A. Vacante, and C. M. Waickus, *Biochem. Biophys. Res. Commun.* **145,** 1112 (1987).
228. S. G. Amur, G. Shanker, J. M. Cochran, H. S. Ved, and R. A. Pieringer, *J. Neurosci. Res.* **16,** 367 (1986).
229. S. K. Ghosh, N. Rawal, S. K. Syed, W. K. Paik, and S. D. Kim, *Biochem. J.* **275,** 381 (1991).
230. D. H. Small, P. R. Carnegie, and R. M. Anderson, *Neurosci. Lett.* **21,** 287 (1981).
231. W. Jacobson, G. Gandy, and R. L. Sidman, *J. Pathol.* **109,** 13 (1973).

232. G. Gandy, W. Jacobson, and R. Sidman, *J. Physiol.* **233,** 1P- (1973).
233. J. M. Scott, J. J. Dinn, P. Wilson, and D. G. Weir, *Lancet* **2,** 334 (1981).
234. R. Deacon, P. Purkiss, R. Green, M. Lumb, J. Perry, and I. Chanarin, *J. Neurol. Sci.* **72,** 113 (1986).
235. M. F. Lou, *Science* **203,** 668 (1979).
236. N. Rawal, Y. J. Lee, J. N. Whitaker, J. O. Park, W. K. Paik, and S. Kim, *J. Neurol. Sci.* **129,** 186 (1995).
237. W. K. Paik, M. Abou-Gharbia, D. Swern, P. Lotlikar, and S. Kim, *Can. J. Biochem. Cell Biol.* **61,** 850 (1983).
238. T. Ogawa, M. Kimoto, H. Watanabe, and K. Sasaoka, *Arch. Biochem. Biophys.* **252,** 526 (1987).
239. T. Ogawa, M. Kimoto, and K. Sasaoka, *Biochem. Biophys. Res. Commun.* **148,** 671 (1987).
240. T. Ogawa, M. Kimoto, and K. Sasaoka, *J. Biol. Chem.* **264,** 10205 (1989).
241. T. Ogawa, M. Kimoto, and K. Sasaoka, *J. Biol. Chem.* **265,** 20938 (1990).
242. P. P. DeDeyn, P. Robitaille, M. Vanasse, I. A. Qureshi, and B. Marescau, *Nephron* **69,** 411 (1995).
243. S. Ueno, A. Sano, K. Kotani, K. Kondoh, and Y. Kakimoto, *J. Neurochem.* **59,** 2012 (1992).
244. N. M. Olken and M. A. Marletta, *Biochemistry* **32,** 9677 (1993).
245. N. M. Olken, Y. Osawa, and M. A. Marletta, *Biochemistry* **33,** 14784 (1994).
246. M. Miyake and Y. Kakimoto, *Metab. Clin. Exp.* **25,** 885 (1976).
247. T. Dandekar and D. Tollervey, *Nucleic Acids Res.* **21,** 5386 (1993).
248. J. Tang, J. D. Gary, S. Clarke, and H. R. Herschman, *J. Biol. Chem.* **273,** in press (1998).
249. R. M. Sandri-Goldin, *Genes Develop.* **12,** 868 (1998).
250. E. C. Shen, M. F. Henry, V. H. Weiss, S. R. Valenti, P. A. Silver, and M. S. Lee, *Genes Develop.* **12,** 679 (1998).