View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Volume 220, number 1, 8–14 FEB 04952 August 1987

Review Letter

N-terminal methylation of proteins: structure, function and specificity

Ann Stock*, Steven Clarke°, Catherine Clarke+ and Jeff Stock

Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

Received 4 June 1987

A common site for the posttranslational modification of proteins is at the N-terminal α-amino group. Here we consider the enzymatic addition of one or more methyl groups that has been found to occur in several proteins. Although the methylated proteins have different overall functions, they all appear to be involved in large macromolecular structures such as ribosomes, myofibrils, nucleosomes, pilins, or flagella. Structural features at the N-termini of these methylated proteins suggest that sequences in this region may serve as recognition sites for only a few different types of methylating enzymes. Thus, we propose that three enzymes could account for the *N*-methylated species so far identified in bacteria, the hypothetical MAK, QP, and pilin methyltransferases, and a single additional enzyme, the hypothetical PK methyltransferase, could account for all of the α-amino methylations observed in eukaryotic cells. Finally, we discuss criteria that could be used in conjuction with primary sequence data to predict proteins that might be subject to methylation at their amino termini.

Protein methylation; N-terminus; Methylmethionine; Methylalanine; Methylproline; Methylphenylalanine

1. INTRODUCTION

In recent years it has been found that one of the many ways in which proteins are covalently modified after their biosynthesis is by alteration of the free α-amino group at their N-termini [1]. There are many examples in which this group is acylated. It has been estimated that at least half of the cytosolic proteins in eukaryotic cells are acetylated at their amino termini [2,3]. Another type of reaction is alkylation of the α-amino group. This reaction, although much less common, appears to be widely distributed in nature. This review represents the first overview of our current knowledge of N-terminal methylation. In bringing together data from a number of laboratories concerning a wide variety of proteins from many different organisms, common features have begun to emerge. Sequence comparisons suggest that the first few amino-terminal residues play an important role in the specificity of the methylating enzymes.

Correspondence address: J. Stock, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

* Present address: Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

° Present address: Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90024, USA

+ Present address: Division of Cardiology, UCLA School of Medicine, Los Angeles, CA 90024, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)

00145793/87/$3.50 © 1987 Federation of European Biochemical Societies

Table 1

Proteins that are methylated at their N-terminus

N-terminal residue	Protein	Species	Reference
Me-Met	IF-3	*E. coli*	[7]
Me-Met	L16	*E. coli*	[4,8]
Me-Met	CheZ	*E. coli, S. typhimurium*	[9]
Me-Met, Me-Ala	L33	*E. coli*	[5,6,8,10]
Me-Ala	S11	*E. coli*	[8]
Me-Phe	pilin	*Pseudomonas, Bacteroides, Neisseria, Moraxella*	[11–15]
Me_2-Ala, Me_3-Ala	L11	*E. coli*	[16–19]
Me_3-Ala	myosin light chains LC-1 and LC-2	rabbit[a]	[21,22]
Me_3-Ala	histone H2B	*Tetrahymena* (protozoan)	[20]
Me_2-Pro	histone H2B	*Asterias* (starfish)	[25]
Me_2-Pro	cytochrome *c*-557	*Crithidia* (protozoan)	[23,24]

[a] A 3.23 ppm NMR signal attributable to the alanyl trimethyl groups has been detected in chicken and coelacanth LC-1 and LC-2 chains, in human alkali light chains, and in bovine alkali and LC-2 chains

2. PROTEINS METHYLATED AT THEIR AMINO TERMINI

The first reports of protein methylation at the amino terminus appeared in 1976 when the *Escherichia coli* ribosomal proteins L16 [4] and L33 [5,6] were shown to be *N*-methylated (*N*-monomethyl-Met and *N*-monomethyl-Ala, respectively). At the present time the number of different proteins known to be *N*-methylated has grown to 11 (table 1). All of these proteins appear to function by interacting with other proteins, often as part of extensive macromolecular complexes. These complexes include the *E. coli* ribosome (proteins L11, L16, L33, S11, and the associated IF-3), bacterial pili (pilin), chromatin (histone H2B), myofibrils (myosin light chains LC-1 and LC-2), respiratory chains (cytochrome *c*-557), and the bacterial chemotaxis-flagellar apparatus (CheZ protein). There is evidence in the case of the myosin light chains that the methylated amino terminus may be at the site of actin binding [26], and in addition, the amino terminus of CheZ appears to be part of a domain involved in the polymerization of this protein [27].

3. SITES OF α-AMINO METHYLATION

N-methylation at phenylalanine residues has been observed in a variety of bacterial pilins, the subunits of the pili that extend from the cell walls of many bacterial species. The N-terminal sequences of all of the methylated pilins are identical (table 2), and it seems likely that essentially the same enzyme is responsible in each cell type. Outside their N-termini the pilins tend to be much less conserved as if this region performed a special role in pilus function or assembly [13]. The methylation reaction does not, however, appear to be essential, since the pilin from *E. coli*, which contains a distinct N-terminal sequence, is not methylated [13].

It is possible to divide the proteins containing monomethylmethionine into 2 classes. One class is comprised of the ribosomal protein L16 and the chemotaxis CheZ protein. The two have similar amino-terminal sequences with Gln-Pro at positions 3 and 4 (table 2). This raises the possibility that a common methyltransferase, designated the QP transferase, may methylate both proteins. A search of 4791 proteins in the National Biomedical Research Foundation Protein Sequence Bank [28] revealed one other protein with Gln-Pro in this position, *E. coli* maltodextrin phosphorylase, with Met-Ser-Gln-Pro. Amino-terminal sequence analysis has revealed that, as expected for *E. coli* proteins [2], this initiator methionine preceding a serine residue is cleaved leaving an unmodified

Table 2

Sequences at the N-termini of proteins with methylated α-amino groups

Protein	N-terminal sequence	Reference
Pilin	Phe-Thr-Leu-Ile-Glu-Leu-Met-	[11–15]
L16	Met-Leu-Gln-Pro-Lys-Glu-Arg-	[8]
CheZ	Met-Met-Gln-Pro-Ser-Ile-Lys-	[27]
IF-3	Met-Lys-Gly-Gly-Lys-Arg-Val-	[7]
L33	(Met)-Ala-Lys-Gly-Ile-Arg-Glu-Lys-	[10]
S11	Ala-Lys-Ala-Pro-Ile-Arg-Ala-	[8]
L11	Ala-Lys-Lys-Val-Gln-Ala-Tyr-	[16]
Myosin light chain LC-1	Ala-Pro-Lys-Lys-Asn-Val-Lys-	[21]
Histone H2B (*Tetrahymena*)	Ala-Pro-Lys-	[20]
Histone H2B (*Asterias*)	Pro-Pro-Lys-Pro-Ser-Gly-Lys-	[25]
Cytochrome *c*-557	Pro-Pro-Lys-Ala-Arg-Glu-Pro-	[23]

serine at the N-terminus [29]. Thus it appears that an N-terminal sequence of uncleaved methionine followed by Gln and Pro in the third and fourth positions may be sufficient for recognition by the methyltransferase.

The role of amino-terminal methylation in L16 and CheZ function is not understood. The L16 protein is a central component required for the peptidyl transfer reaction [30]. CheZ is an essential element of the chemotaxis apparatus in *E. coli* and *Salmonella typhimurium*. Genetic studies suggest a direct interaction between CheZ and elements at the flagellar motor [31,32]. The native protein has been isolated as a large homopolymer (>500 kDa) composed of 24-kDa subunits [27]. Approximately a third of the CheZ monomers within a cell are methylated immediately after translation, and the remainder are not susceptible to modification [9]. Apparently, in CheZ polymers the amino termini are inaccessible to the modification enzyme. Results obtained with protease treatments tend to confirm the conclusion that the amino terminus of polymerized CheZ is relatively inaccessible [27].

The second class of *N*-monomethylmethionines includes the translational initiation factor IF-3 and the ribosomal protein L33. For both cases different forms of the proteins containing various N-terminal sequences have been found. These are probably derived from differential cleavage of a single translation product. Whereas only the uncleaved form of IF-3 is methylated [7,10], the situation for L33 is more complex. In *E. coli* K strains almost 25% of the protein is methylated at the initial methionine and the remainder is methylated at the Ala residue exposed by the expected cleavage of the methionine preceding this alanine. In *E. coli* B strains, however, only the methylated alanine form appears to be present [8]. The amino-terminal sequences of IF-3 and L33 are similar in that the principal methylated residue, either Met or Ala, is followed by Lys and Gly (table 2). This consensus may be recognized by another methyltransferase, the hypothetical MAK enzyme.

The MAK methyltransferase may also function to methylate another ribosomal protein, S11. S11 is similar to L33 at its N-terminus in that, following the methylated alanine, the proteins exhibit the sequence Ile-Arg-X-(Lys,Arg)-X-(Lys,Arg) at residues 5–9 and 4–8, respectively.

Amino-terminal alanine and proline residues can be fully methylated (trimethylalanine and dimethylproline) in several proteins including the *E. coli* large ribosomal subunit protein L11, and eukaryotic proteins including some histone H2Bs, some myosin light chains, and at least one cytochrome *c*-557. The function of this reaction is not known, but one line of evidence suggests a role in the protection of exposed N-terminal sequences to attack by aminopeptidases within the cell [23]. For example, the amino-terminal region of histone H2B is thought to extend away from the histone octamer core [33]. Of the three types of homologous myosin light chains, only the two that

contain a 41 residue N-terminal extension are methylated [21]. Physical studies have indicated that this portion of the light chains extends toward a locus of reversible association with the actin filament in muscle [26]. Finally, cytochrome *c*-557 has a 10 amino acid N-terminal extension that is not found in cytochromes *c* that are not methylated [23].

Each of the eukaryotic methylated proteins has a methylated alanine or proline at the amino terminus that is followed by the amino acids Pro-Lys, suggesting that a common methyltransferase is involved, the hypothetical PK enzyme. The PK enzyme would presumably recognize the primary sequence Ala-Pro-Pro-Lys in the amino-terminal position. Precedence exists for such recognition; a monoclonal antibody raised against myosin head fragments cross reacts with histone H2B due to a shared amino-terminal epitope that has been mapped to the first 4 amino acids of the methylated myosin light chains [34].

The bacterial protein, L11, has an amino-terminal sequence that shares features of both the putative *E. coli* MAK consensus and the eukaryotic PK consensus. Thus L11 has the *N*-methyl-Ala-Lys characteristic of MAK, but as in the putative PK substrates of eukaryotic cells, the alanine in L11 is trimethylated and the third residue is a lysine. Of perhaps even greater significance is the fact that this common lysine residue in the third position is trimethylated in L11 just as in the myosin light chains and cytochrome *c*-557. Thus, for these proteins a total of five to six methyl groups are added to the first three residues, two or three at the α-imino or α-amino of the N-terminal Pro or Ala and three at the ϵ-amino group of the lysine.

These results suggest that in eukaryotic cells proteins with N-terminal Pro or Ala residues followed by Pro-Lys are good candidates for *N*-methylation. However, there are two proteins in the Dayhoff sequence collection [28] that contain an N-terminal Pro-Pro-Lys sequence and are apparently not methylated. These proteins are both histone H2Bs, one from *Drosophila* and the other from the mollusc *Patella granatina* (table 3). Thus, either the PK methyltransferase is not universally distributed or other recognition factors are important in determining sites of modification besides the first few residues. In addition, a search of the 4791 proteins in this sequence collection revealed a bacterial cytochrome c_3 with an N-terminal sequence, (Met)-Ala-Pro-Lys, that is not methylated [38]. The only other example of such an N-terminal sequence is the VP-1 coat protein from mouse polyoma virus. It has not been determined whether this protein is methylated or not.

4. ENZYMOLOGY OF N-TERMINAL METHYLATION

From similarities between sequences of proteins containing similar N-terminal modifications it is apparent that common methyltransferase enzymes may be responsible for their modification. Along these lines we have speculated that the observed methylations of *E. coli* proteins, except perhaps for L11, can be accounted for by two hypothetical enzymes, designated the QP and MAK methyltransferases. All of the N-terminal methylations that have been described in eukaryotic cells could be accounted for by a single hypothetical enzyme which we designated the PK methyltransferase (see above). Unfortunately there are very few data

Table 3

Sequences at the amino termini of histone H2B proteins from a variety of different species

Organism	Amino terminus of H2B	Reference
Tetrahymena	$(Me)_3$Ala-Pro-Lys-	[20]
Asterias	$(Me)_2$Pro-Pro-Lys-Pro-Ser-Gly-Lys-Gly-	[25]
C. elegans	Ala-Pro-Pro-Lys-Pro-Ser-Ala-Lys-Gly-	[35]
Patella	Pro-Pro-Lys-Val-Ser-Ser-Lys-Gly-	[36]
Drosophila	Pro-Pro-Lys-Thr-Ser-Gly-Lys-Ala-	[37]
Calf thymus	Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Ala-Pro-	[35]

available concerning these enzymatic activities at the present time.

Mutations in the *prmA* gene of *E. coli* have been found to result in defective methylation of ribosomal protein L11, but no readily observed phenotype [39]. Because both methylation on the lysine side chain as well as at the N-terminal alanine was affected, it was suggested that a single enzyme was responsible for both modifications [40]. On the other hand, preliminary studies have suggested that distinct enzymes are involved in the methylation of the amino-terminal proline and the internal lysine in cytochrome *c*-557 in the protozoan *Crithidia oncopelti* [41].

5. STOICHIOMETRY OF N-TERMINAL METHYLATION REACTIONS

Many of the amino-terminal methylation reactions go to completion and fully modify the protein. Evidence has been presented, for example, that the pilins from *Pseudomonas* [11] and *Moraxella* [14] and the ribosomal protein L16 [8] contain one methyl group per polypeptide chain. A similar situation occurs with the protozoan cytochrome *c*-557 where each N-terminal proline residue contains two methyl groups. On the other hand, only about half of the pilin molecules from *Neisseria* are methylated at their amino termini [13], and it has been reported that the *E. coli* ribosomal protein L11 can be partially modified to give a dimethylalanine derivative rather than the fully methylated trimethylalanine product [16]. One possibility is that partially methylated species may represent intermediates along the modification pathway. Alternatively, there may be competing posttranslational pathways that are mutually exclusive and irreversible. This is probably the case for L33 in *E. coli* K strains where methylation can occur at either an N-terminal Met or Ala [10]. Presumably the initiator methionine can be either cleaved or methylated. If the N-terminal Met is cleaved before it is methylated, then the resulting N-terminal alanine is methylated; if the N-terminal methionine is methylated first then it probably cannot be cleaved. Thus the partition between methyl-Met and methyl-Ala depends on the rate of cleavage versus the rate of methyl transfer. This raises the possibility that the nature of the L33 N-terminus could be controlled by levels of *S*-adenosylmethionine, rates of protein synthesis, etc. Similar considerations apply for the methylation of the CheZ protein (see above and [9]).

6. CHEMICAL CONSEQUENCES OF N-TERMINAL METHYLATION

Two types of N-terminal methylations can be distinguished on the basis of their different effects on the chemistry of the N-terminus. The first group includes all the monomethylations (*N*-monomethylmethionine, *N*-monomethylalanine, *N*-monomethylphenylalanine). In these cases the chemical effect is probably only a small change in the pK_a of the α-amino group and a slight reduction in its reactivity due to the steric effects of the methyl group. For the model compounds shown in table 4, the conversion of a primary amine (unmodified α-amino group) to a secondary amine by a monomethylation reaction results in an increase in the pK_a of up to about 0.5 units. Although there are few data available on the effect of

Table 4

Effect of N-methylation on the pK_a value for the α-amino groups in amino acids and peptides[a]

pK_a of the amino group					
Primary	pK_a	Secondary	pK_a	Tertiary	pK_a
Methylamine	10.64	dimethylamine	10.64	trimethylamine	9.76
Alanine	9.69	monomethylalanine	10.19		
		proline	10.60	monomethylproline	10.01
Gly-Gly	8.13	monomethyl-Gly-Gly	8.66	dimethyl-Gly-Gly	8.09
Gly-Leu	8.41	monomethyl-Gly-Leu	8.67	dimethyl-Gly-Leu	7.78

[a] Data from Jencks and Regenstein [42]

monomethylation on reactivity of α-amino groups in peptides, from comparisons of primary versus secondary methylamines only very small changes would be expected. For instance, there is only about a 20% decrease in the second order rate constant of *p*-nitrophenyl phosphate aminolysis by methylamine versus dimethylamine [43].

In contrast, the second class of methyl modification, the quaternization of the amino-terminal nitrogen to produce dimethylproline and trimethylalanine, produces a profound change in the chemistry of the amino terminus. In this case methylation produces a permanent positive charge on the amino group and abolishes the nucleophilicity of the α-amino nitrogen. The associated trimethylation of an adjacent lysine side chain leads to the production of two fixed positive charges at the N-terminus, independent of pH and environment. At the same time the potential reactivity of the amino groups is removed. These properties would allow the amino terminus to function as an effective counterion for nucleoside di- and triphosphates, for example. Alternatively, this modification would permit the amino cation to be maintained in a hydrophobic environment under conditions where unmodified amino groups would be deprotonated even at neutral pH. An example of this may be found in the quaternary trimethyllysine residue at the active site of myosin [44].

REFERENCES

[1] Wold, F. (1981) Annu. Rev. Biochem. 50, 783–814.

[2] Tsunasawa, S., Stewart, J.W. and Sherman, F. (1985) J. Biol. Chem. 260, 5382–5391.

[3] Tsunasawa, S. and Sakiyama, F. (1984) Methods Enzymol. 106, 165–170.

[4] Brosius, J. and Chen, R. (1976) FEBS Lett. 68, 105–109.

[5] Chang, C.N., Schwartz, M. and Chang, F.N. (1976) Biochem. Biophys. Res. Commun. 73, 233–239.

[6] Wittmann-Liebold, B. and Pannenbecker, R. (1976) FEBS Lett. 68, 115–118.

[7] Brauer, D. and Wittmann-Liebold, B. (1977) FEBS Lett. 79, 269–275.

[8] Chen, R., Brosius, J. and Wittmann-Liebold, B. (1977) J. Mol. Biol. 111, 173–181.

[9] Stock, A., Schaeffer, E., Koshland, D.E. jr and Stock, J. (1987) J. Biol. Chem. 262, in press.

[10] Chang, F.N. and Budzilowicz, C. (1977) J. Bacteriol. 131, 105–110.

[11] Frost, L.S., Carpenter, M. and Paranchych, W. (1978) Nature 271, 87–89.

[12] McKern, N.M., O'Donnell, I.J., Inglis, A.S., Steward, D.J. and Clark, B.L. (1983) FEBS Lett. 164, 149–153.

[13] Hermodson, M.A., Chen, K.C.S. and Buchanan, T.M. (1978) Biochemistry 17, 442–445.

[14] Froholm, L.O. and Sletten, K. (1977) FEBS Lett. 73, 29–32.

[15] Maars, C.F., Schoolnik, G., Koomey, J.M., Hardy, J., Rothbard, J. and Falkow, S. (1985) J. Bacteriol. 163, 132–139.

[16] Alix, J.-H., Hayes, D., Lontie, J.-F., Colson, C., Glatigny, A. and Lederer, F. (1979) Biochimie 61, 671–679.

[17] Lederer, F., Alix, J.-H. and Hayes, D. (1977) Biochem. Biophys. Res. Commun. 77, 470–480.

[18] Dognin, M.J. and Wittmann-Liebold, B. (1977) FEBS Lett. 84, 342–346.

[19] Dognin, M.J. and Wittmann-Liebold, B. (1980) Eur. J. Biochem. 112, 131–151.

[20] Nomoto, M., Kyogoku, Y. and Iwai, K. (1982) J. Biochem. 92, 1675–1678.

[21] Henry, G.D., Dalgarno, D.C., Marcus, G., Scott, M., Levine, B.A. and Trayer, I.P. (1982) FEBS Lett. 144, 11–15.

[22] Henry, G.D., Trayer, I.P., Brewer, S. and Levine, B.A. (1985) Eur. J. Biochem. 148, 75–82.

[23] Pettigrew, G.W. and Smith, G.M. (1977) Nature 265, 661–662.

[24] Smith, G.M. and Pettigrew, G.W. (1980) Eur. J. Biochem. 110, 123–130.

[25] Martinage, A., Briand, G., Van Dorsselaer, A., Turner, C.H. and Sautier, P. (1985) Eur. J. Biochem. 147, 351–359.

[26] Henry, G.D., Winstanley, M.A., Dalgarno, D.C., Scott, G.M.M., Levine, B.A. and Trayer, I.P. (1985) Biochim. Biophys. Acta 830, 233–243.

[27] Stock, A. and Stock, J. (1987) J. Bacteriol., in press.

[28] Dayhoff, M.O. (1978) Atlas of Protein Sequence and Structure (electronic data base released March 1987) Natl. Biomed. Res. Found., Silver Spring, MD.

[29] Schiltz, E., Palm, D. and Klein, H.W. (1980) FEBS Lett. 109, 59–62.

[30] Moore, V.G., Atchinson, R.E., Thomas, G., Moran, M. and Noller, H.F. (1975) Proc. Natl. Acad. Sci. USA 72, 844–848.

[31] Parkinson, J.S. and Parker, S.R. (1979) Proc. Natl. Acad. Sci. USA 76, 2390–2394.

[32] Yamaguchi, S., Aizawa, S.-I., Kihara, M., Isomura, M., Jones, C.J. and Macnab, R.M. (1986) J. Bacteriol. 168, 1172–1179.

[33] Rosenberg, N.L., Smith, R.M. and Rill, R.L. (1986) J. Biol. Chem. 261, 12375–12383.

[34] Tokunaga, M., Suzuki, M., Saeki, K. and Wakabayashi, T. (1987) J. Mol. Biol. 194, 245–256.

[35] Vanfleteren, J.R., Van Bun, S.M., Delcambe, L.L. and Van Beeumen, J.J. (1986) Biochem. J. 235, 769–773.

[36] Van Helden, P.D., Strickland, W.N., Brandt, W.F. and Von Holt, C. (1979) Eur. J. Biochem. 93, 71–78.

[37] Elgin, S.C.R., Schilling, J. and Hood, L.E. (1970) Biochemistry 18, 5679–5685.

[38] Shinkai, W., Hase, T., Yagi, T. and Matsubata, H. (1980) J. Biochem. 87, 1747–1756.

[39] Colson, C. and Smith, H.O. (1977) Mol. Gen. Genet. 154, 167–173.

[40] Colson, C., Lhoest, J. and Urlings, C. (1979) Mol. Gen. Genet. 169, 245–250.

[41] Valentine, J. and Pettigrew, G.W. (1982) Biochem. J. 201, 329–338.

[42] Jencks, W.P. and Regenstein, J. (1976) in: Handbook of Biochemistry and Molecular Biology, 3rd edn, Physical and Chemical Data, vol.I (Fasman, G.D. ed.) pp.305–351.

[43] Kirby, A.J. and Jencks, W.P. (1965) J. Am. Chem. Soc. 87, 3209–3216.

[44] Okamoto, Y. and Yount, R.G. (1985) Proc. Natl. Acad. Sci. USA 82, 1575–1579.