

N-terminal acetylation of ectopic recombinant proteins in *Escherichia coli*

Elodie Charbaut^a, Virginie Redeker^{b,1}, Jean Rossier^b, André Sobel^{a,*}

^aU440 INSERM/UPMC, Institut du Fer à Moulin, 17 rue du Fer à Moulin, 75005 Paris, France

^bCNRS UMR 7637, ESPCI, 10 rue Vauquelin, 75005 Paris, France

Received 22 August 2002; accepted 31 August 2002

First published online 18 September 2002

Edited by Julio Celis

Abstract N-terminal acetylation is a protein modification common in eukaryotes, but rare in prokaryotes. Here, we characterized five mammalian stathmin-like subdomains expressed in *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and nanoESI Q-TOF tandem mass spectrometry. We revealed that RB3_{SLD} and RB3'_{SLD} are N^α-acetylated, whereas SCG10_{SLD} and SCLIP_{SLD}, although identical up to residue 6, are not, as well as stathmin. To assess the influence of the N-terminal sequences on N^α-acetylation, we exchanged residues 7 and 8 between acetylated RB3_{SLD} and unacetylated SCG10_{SLD}, and showed that it reversed the acetylation pattern. Our results demonstrate that ectopic recombinant proteins can be extensively N^α-acetylated in *E. coli*, and that the rules governing N^α-acetylation are complex and involve the N-terminal region, as in eukaryotes.

© 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: N-terminal acetylation; Mass spectrometry; Ectopic protein expression; Stathmin family; *Escherichia coli*

1. Introduction

Acetylation of the amino-terminal group of proteins is very common in eukaryotes since 70% of soluble proteins is estimated to bear this modification. It occurs either co-translationally or post-translationally, after the possible removal of the N-terminal methionine. Although the widespread occurrence of N^α-acetylation suggests an important biological role, its exact function is not known. Several hypotheses have been proposed, including an influence on protein stability (for reviews see [1,2]).

The rules governing N^α-acetylation are not clear, unlike those governing the cleavage of the N-terminal methionine, which are well defined and conserved from prokaryotes to higher eukaryotes. The removal of the N-terminal methionine almost entirely depends on the nature of the second residue, a small residue allowing the cleavage and a bulky one preventing it [3,4]. N^α-acetylation in eukaryotes also primarily de-

pends on the nature of the second encoded residue: among those which are found to be N^α-acetylated, Ala and Ser largely predominate (after N-terminal methionine cleavage), as well as the uncleaved N-terminal Met when followed by either Asp or Glu [5]. This specificity correlates with the identification of two classes of N^α-acetyl transferase activities, one acetylating substrates beginning with Ala/Ser/Thr/Gly, and the other acetylating substrates beginning with Met (for a review see [1]). However, systematic mutagenesis in several model proteins reveals some variability in the influence of the second encoded residue on N^α-acetylation, depending on the protein studied [4,6,7]. Other determinants likely contribute to define the specificity of N^α-acetylation, also because not all proteins bearing the preferred N-terminal residues are acetylated. Experimental evidence exists that acetylation can be influenced by at least the first six residues [7]. Moreover, a computer analysis including about a hundred protein sequences suggests that the information contained in at least the first 40 residues may be necessary to distinguish between acetylated and unacetylated proteins [8].

In prokaryotes, N^α-acetylation is a much rarer event. We found only five endogenous proteins reported to be N^α-acetylated in *Escherichia coli* [9–12]. Moreover, eukaryotic proteins which are normally acetylated endogenously are frequently unacetylated when expressed in *E. coli*. In contrast, we report here the mass spectrometric characterization of several recombinant proteins corresponding to subdomains of mammalian stathmin family proteins, which we had expressed in *E. coli* and purified in a previous study [13]. Unexpectedly, two of them are N^α-acetylated, whereas the others are not, although they have very similar N-terminal sequences. Further analysis of the N^α-acetylation of two chimeric proteins mutated on residues 7 and 8 give clues concerning the rules of N^α-acetylation in *E. coli*.

2. Materials and methods

2.1. Recombinant protein expression in *E. coli*

The stathmin-like domains (SLDs) of stathmin family proteins were produced in *E. coli* BL21(DE3) and purified as described in [13,14]. The nucleotide triplets corresponding to residues 7 and 8 in SCG10_{SLD} and RB3_{SLD} were mutated with the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) and checked by sequencing (Genome Express, Montreuil, France). The resulting mutant proteins SCG10_{SLD}-IE and RB3_{SLD}-KQ (Table 2) were expressed and purified like the other proteins [13] except that the anion exchange chromatography step was omitted.

2.2. Proteolysis

Extensive trypsin proteolysis of RB3_{SLD}, RB3'_{SLD}, and SCG10_{SLD}-IE was performed overnight at 37°C in either 50 mM Tris-HCl, pH

*Corresponding author. Fax: (33)-1 45 87 61 32.

E-mail address: sobel@ifm.inserm.fr (A. Sobel).

¹ Present address: CNRS UPR 9063, Laboratoire d'Enzymologie et de Biochimie Structurales, 91198 Gif-sur-Yvette Cedex, France.

Abbreviations: MS, mass spectrometry; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; ESI, electrospray ionization; Q-TOF, quadrupole time-of-flight; TFA, trifluoroacetic acid; SLD, stathmin-like domain; *m/z*, mass-to-charge ratio

8.5, or phosphate-buffered saline (PBS), pH 7.4, 1 mM EGTA, 1 mM dithiothreitol (DTT), containing the recombinant protein (2.5 μ M), 1 mM CaCl_2 and trypsin 1/100 (w/w enzyme-to-substrate ratio). When these conditions did not allow detection of the N-terminal peptides, limited proteolysis conditions were optimized: SCG10_{SLD} (2.5 μ M) was incubated overnight at 37°C with trypsin 1/5000 in PBS, pH 7.4, 1 mM CaCl_2 , and RB3_{SLD}-KQ (2.5 μ M) was incubated for 2 h at room temperature with GluC 1/5000 in PBS, pH 7.4, 1 mM EGTA, 1 mM DTT, after which the reactions were stopped by trifluoroacetic acid (TFA) at a final concentration of 1%.

2.3. N-terminal sequencing

Samples containing a mixture of 400 pmol stathmin and 400 pmol of either RB3_{SLD} or RB3'_{SLD} in 20 μ l were desalted on a ProSorb cartridge (Applied Biosystems) and sequenced by automated Edman degradation using a Procise pulse-liquid protein sequencer (model 494, Perkin-Elmer Applied Biosystems Division, Foster City, CA, USA).

2.4. Mass spectrometry characterization of recombinant proteins

The molecular masses of intact proteins were determined as described in [13] by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a MALDI-TOF mass spectrometer Voyager-STR biospectrometry workstation (PerSeptive Biosystems, Framingham, MA, USA). They were confirmed by nano-electrospray ionization (nanoESI) quadrupole time-of-flight (Q-TOF) MS using a Q-TOF 2 mass spectrometer (Micromass, Manchester, UK), after desalting the samples on a C18 Zip-Tip (Millipore, Bedford, MA, USA) eluted in methanol, water and acetic acid (50/50/0.5).

Modifications of the proteins were characterized by mass spectrometry of the peptides resulting from their proteolysis. The peptide solution was diluted to about 5 pmol/ μ l in 0.5% TFA, and analyzed after co-crystallization with a matrix of 2,5-dihydroxybenzoic acid (saturated solution in 0.1% aqueous TFA) by MALDI-TOF MS in the positive and reflectron mode. The N-terminal peptides were identified by comparing experimental masses with the theoretical masses of the non-modified peptides released from tryptic or Glu-C digestion of the proteins. The position of the acetylated residue was determined by tandem mass spectrometry analysis of the peptide bearing the 42 Da mass increment, using the Q-TOF 2 mass spectrometer.

3. Results

3.1. MALDI-TOF MS reveals a putative acetylation of mammalian proteins expressed in *E. coli*

In a previous study, we expressed in *E. coli* and purified five mammalian SLDs, which display 70–80% amino acid identity along their sequence (about 140 residues) [13]. The molecular masses of the expressed proteins were determined by MALDI-TOF MS. For stathmin, SCG10_{SLD} and SCLIP_{SLD}, they corresponded to the values calculated for the proteins after N-terminal methionine cleavage (Table 1). For RB3_{SLD} and RB3'_{SLD}, the experimental masses were 43 and 43.9 Da larger than expected, which suggested that these two splice variants, identical up to residue 124, may bear an acetylation, responsible for a 42 Da mass increment. The relative difference between the experimental mass determination and the calculated

average mass plus 42 Da (0.006% and 0.013%) is consistent with the accuracy of MALDI-TOF MS in linear mode, and a 42 Da mass increment was confirmed by nanoESI Q-TOF MS (data not shown).

3.2. RB3_{SLD} and RB3'_{SLD} are acetylated on their N-terminal group

To demonstrate that RB3_{SLD} and RB3'_{SLD} are actually acetylated, as well as to determine the modified positions (which could be either the α -amino or a lysine ϵ -amino group), we submitted the proteins to tryptic proteolysis and analyzed the resulting peptide mixtures by MALDI-TOF MS. The peptide mass mapping covered more than 90% of each protein sequence. In both cases, only traces of the non-modified N-terminal peptide 2–11 could be detected (mass-to-charge ratio, m/z 1161.50), whereas an ion at m/z 1203.50 was consistently detected (note that we use a residue numbering including N-terminal methionine, although it is not present in the mature proteins). By nanoESI Q-TOF MS/MS, this ion was shown to correspond to peptide 2–11 acetylated on the N-terminal alanine (Fig. 1A,B for RB3_{SLD}, not shown for RB3'_{SLD}) and not to peptide 134–142 from RB3_{SLD} (theoretical mass 1203.63), since no fragment ion corresponding to this sequence was identified in the MS/MS spectrum. Consistently, a less intense peak with an additional 16 Da increment (m/z 1219.50) was often detected and identified by tandem mass spectrometry as the acetylated peptide 2–11 with a sulfoxide methionine.

The N $^\alpha$ -acetylation of RB3_{SLD} and RB3'_{SLD} was further confirmed by N-terminal sequencing on equimolar mixtures of either RB3_{SLD} or RB3'_{SLD} with stathmin. As the N-terminus of stathmin is not modified and has a different sequence, comparing the yield of RB3_{SLD} (respectively RB3'_{SLD}) residue obtained at each cycle with the corresponding yield of stathmin residue provides a quantitative estimation for the amount of N-terminally blocked RB3_{SLD} (respectively RB3'_{SLD}). In that way, 86% RB3_{SLD} and 91% RB3'_{SLD} were estimated to be blocked (data not shown). Altogether, our results demonstrate that about 90% RB3_{SLD} or RB3'_{SLD} expressed in *E. coli* is N $^\alpha$ -acetylated.

3.3. Mutation of residues 7 and 8 reverses the acetylation pattern of SCG10_{SLD} and RB3_{SLD}

N $^\alpha$ -acetylation of proteins is thought to be driven by the nature of their N-terminal residues, at least in eukaryotes. It was thus striking that RB3_{SLD} and RB3'_{SLD} were N $^\alpha$ -acetylated, whereas SCG10_{SLD} and SCLIP_{SLD} were not, although these four proteins have the same N-terminal sequence up to residue 6 (see Table 1). We thus tested whether exchanging N-terminal residues between SCG10_{SLD} and RB3_{SLD} may change their respective N $^\alpha$ -acetylation. We expressed two mu-

Table 1

Calculated mass, MALDI-TOF MS experimental mass, difference between experimental and calculated mass (in Da), and N-terminal sequence of SLDs expressed in *E. coli*

Name	Calculated mass ^a	Experimental mass	Δm	N-terminal sequence
Stathmin	17 171.3	17 172.0	0.7	MASSDIQVKELE KRASGQAFEL...
SCG10 _{SLD}	16 346.7	16 347.5	0.8	MA--DMEVKQIN KRASGQAFEL...
SCLIP _{SLD}	16 723.8	16 723.2	−0.6	MA--DMEVKQLD KRASGQSFEV...
RB3 _{SLD}	16 679.9	16 722.9	43.0	MA--DMEVIELN KCTSGQSFEV...
RB3' _{SLD}	15 023.1	15 067.0	43.9	MA--DMEVIELN KCTSGQSFEV...

^aWithout N-terminal methionine.

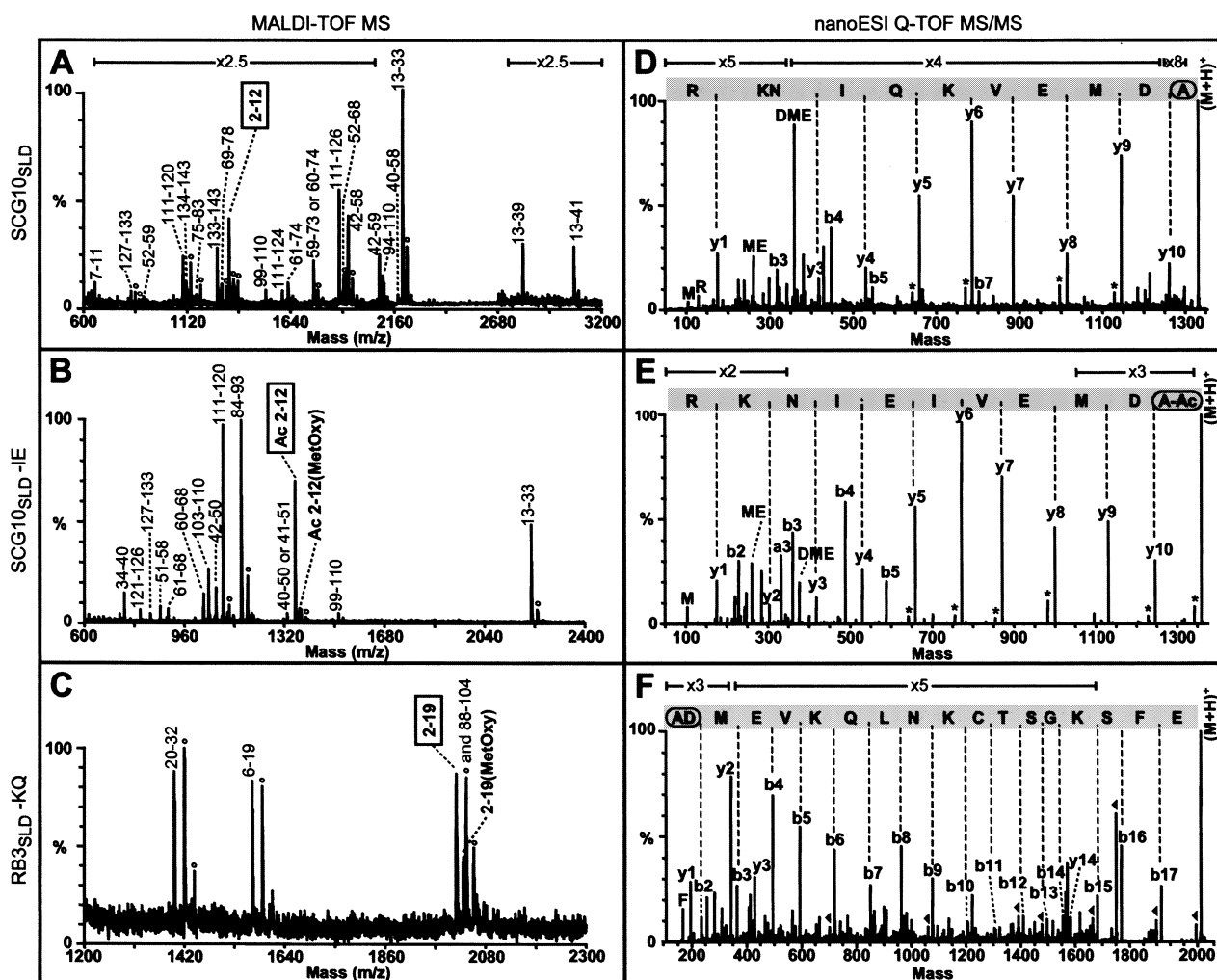
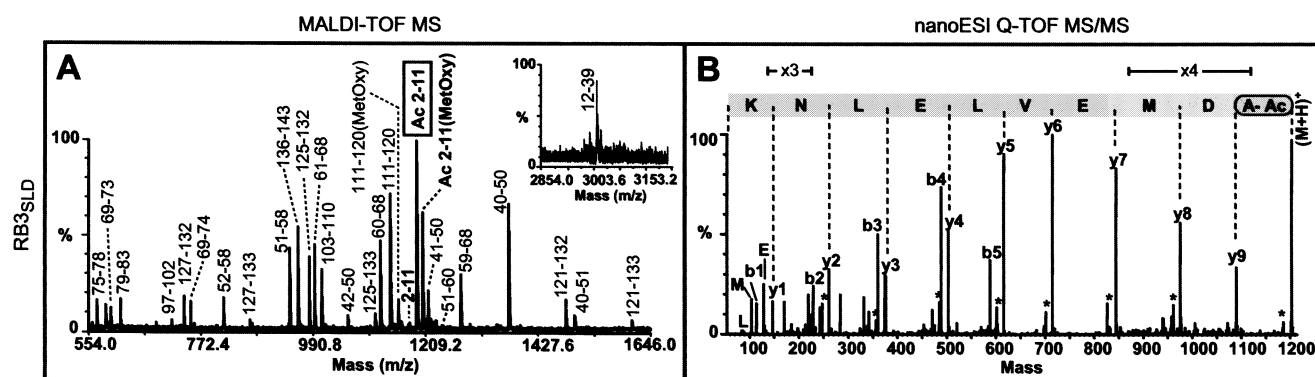


Fig. 2. MALDI-TOF mass spectra of peptides resulting from proteolysis of SCG10_{S_{LD}} by trypsin (A), SCG10_{S_{LD}}-IE by trypsin (B), and RB3_{S_{LD}}-KQ by GluC (C). Circles indicate sodium adducts. The N-terminal peptides are in bold. Tandem mass spectra resulting from fragmentation of the corresponding N-terminal peptides (boxed), showing that SCG10_{S_{LD}}-IE is N^α-acetylated (E), whereas SCG10_{S_{LD}} (D) and RB3_{S_{LD}}-KQ (F) are not. Peaks corresponding to major fragmentation products (y and b series) are indicated, as well as internal fragments (by their sequence), z fragments (asterisks), and loss of water (arrows). The deduced sequence is indicated (N-terminus circled).

Table 2

N-terminal sequence, calculated mass, experimental mass as measured by MALDI-TOF MS, and difference between experimental and calculated mass (in Da) for the mutant proteins SCG10_SLD-IE and RB3_SLD-KQ

Name	N-terminal sequence	Calculated mass ^a	Experimental mass	Δm
SCG10 _S LD-IE	MADMEVIEIN KRASGQAFEL...	16 332.7	16 377.0	44.3
RB3 _S LD-KQ	MADMEVKQLN KCTSGQSFEL...	16 693.9	16 694.6	0.7

^aWithout N-terminal methionine.

tant proteins: SCG10_SLD-IE, in which residues 7 and 8 (KQ) have been mutated into IE (residues 7 and 8 of RB3_SLD), and RB3_SLD-KQ, in which residues 7 and 8 (IE) have been mutated into the corresponding residues of SCG10_SLD (KQ).

The experimental molecular masses of the mutant proteins, measured by mass spectrometry, suggested that SCG10_SLD-IE was acetylated, whereas RB3_SLD-KQ was not (Table 2). Indeed, a tryptic digest of SCG10_SLD-IE contained an ion at m/z 1359.70, as measured by MALDI-TOF MS, which was shown by nanoESI Q-TOF MS/MS to correspond to SCG10_SLD-IE peptide 2–12 acetylated on its N-terminal alanine (Fig. 2C,F). We similarly confirmed that the non-mutated protein SCG10_SLD was not acetylated: an ion at m/z 1331.69 was detected by MALDI-TOF MS in a corresponding tryptic digest, and was shown by nanoESI Q-TOF MS/MS to correspond to the unacetylated peptide 2–12 (Fig. 2A,D). For the mutant protein RB3_SLD-KQ, the N-terminal peptide could not be detected in the tryptic peptide mass map obtained by MALDI-TOF MS. However, the analysis of a limited GluC digest revealed a peak corresponding to the unacetylated peptide 2–19 (m/z 2015.17), as was also verified by nanoESI Q-TOF MS/MS (Fig. 2B,E).

Altogether, our results show that exchanging residues 7 and 8 between SCG10_SLD and RB3_SLD is sufficient to inverse their N-terminal acetylation pattern, which suggests that the nature of residues within the N-terminal region strongly determines N^α-acetylation also in prokaryotes.

4. Discussion

We report the N^α-acetylation pattern of five mammalian stathmin-like subdomains expressed in *E. coli*. We show that despite the strong resemblance between these proteins, only two of them are N^α-acetylated, and that directed mutagenesis of residues 7 and 8 in one acetylated and one unacetylated protein reverses their acetylation pattern.

Whereas the cleavage of the N-terminal methionine was expected (since second encoded residues are alanines), the acetylation of the α -amino group was not. Indeed, N^α-acetylation is very rare in *E. coli*, being described to our knowledge for five proteins: the three ribosomal subunits L7/L12, S5, S18 [9,10], the elongation factor EF-Tu [11], and very likely the chaperone SecB [12]. Moreover, ectopic proteins recombinantly expressed in *E. coli* are frequently reported not to be acetylated (in contrast to their endogenous N^α-acetylation in eukaryotes) or only very partially [15,16].

We analyzed to what extent the N^α-acetylation pattern of the SLDs expressed in *E. coli* could match previous knowledge of N^α-acetylation rules in eukaryotes. First, the N^α-acetylated residue in RB3_SLD, RB3'_SLD and SCG10_SLD-IE is an alanine, which is with serine one of the residues most frequently acetylated in eukaryotes. Similarly, *E. coli* proteins S5 and S18 are acetylated on an alanine, whereas L7/L12,

EF-Tu and possibly SecB are acetylated on a serine. Moreover, downstream residues in the N-terminal region have been shown to influence N^α-acetylation in eukaryotes [7,8]. We show here that residues 7 and 8 of RB3_SLD and RB3'_SLD have a primary influence on their N^α-acetylation in *E. coli*, since exchange of these two residues between unacetylated SCG10_SLD and acetylated RB3_SLD reverses their acetylation pattern. As the nature of residues 7 and 8 is sufficient to switch the acetylation state, the observed modification seems strongly dependent on sequence specificity. Yet, as the SLDs share 80% of their amino acids, other residues may contribute to N^α-acetylation, particularly among the first six identical ones.

We also examined the presence in SLDs of residues specifically enriched or excluded from the N-terminus of acetylated proteins [17]. N^α-acetylated proteins display an overrepresentation of acidic residues at positions 2 and 4, which is the case for SLDs (excepted stathmin). N^α-acetylated proteins also display an overrepresentation of lysine and isoleucine in their N-terminal region. An isoleucine is found at position 7 in the acetylated SLDs (RB3_SLD, RB3'_SLD, SCG10_SLD-IE), whereas a lysine is found at the same position in the non-acetylated ones (SCG10_SLD, SCLIP_SLD, RB3_SLD-KQ). Thus, although the residues present in the N-terminal region of the SLDs are roughly compatible with N^α-acetylation, the difference due to residues 7 and 8 is not explained by the statistical study. However, N^α-acetylation is likely to be influenced not only by residues with a positive effect, but also by residues with a negative inhibitory effect [18]. In our case, either IE residues have a positive effect or KQ residues have an inhibitory influence. Further analysis of N^α-acetylation determinants of SLDs in *E. coli* may provide new insights into the rules governing this modification.

The difficulty to determine N^α-acetylation rules may also come from the existence of various acetylation enzymes with different specificity requirements. Indeed, in *E. coli*, three N^α-acetyltransferases have already been identified, namely *rimI*, *rimJ*, *rimL*, each being responsible for the modification of a different protein, S18, S5 and L12/L7, respectively [9,10]. SLDs may be modified by one of these enzymes, or by another one, still unidentified. But the lack of sequence homology of the SLDs in their N-terminal region with any protein from *E. coli* does not give any clue to this question.

Altogether, our results demonstrate that N^α-acetylation can occur in *E. coli*, even on ectopic overexpressed proteins. Although some general rules governing N^α-acetylation seem to be conserved between prokaryotes and eukaryotes, it is still impossible to predict that a protein is N^α-acetylated. Nevertheless, N^α-acetylation has been demonstrated in some cases to be important for protein activity. For example, unacetylated muscle α -tropomyosin expressed in *E. coli* polymerizes poorly, and binds weakly to F-actin [19]. Similarly, unacetylated rat glycine M-methyltransferase expressed in *E. coli* does

not exhibit the cooperative behavior of the native enzyme [20]. The way by which N^α-acetylation influences specific interactions with protein partners may be the neutralization of the N-terminus positive charge. Such an effect may also influence the correct folding of polypeptides during synthesis in other cases. A practical consequence of our results is that each recombinant protein expressed in ectopic cells should be carefully characterized by mass spectrometry at the protein level, as is currently done at the nucleic level, since modifications can occur extensively even in *E. coli*, thereby modifying the properties and homogeneity of protein preparations used in long time consuming functional studies.

Acknowledgements: We thank J.P. Le Caer and V. Labas for N-terminal sequencing. This work was supported by funds from INSERM, ARC and AFM. E.C. received a fellowship from FRM.

References

- [1] Bradshaw, R.A., Brickey, W.W. and Walker, K.W. (1998) Trends Biochem. Sci. 23, 263–267.
- [2] Arfin, S.M. and Bradshaw, R.A. (1988) Biochemistry 27, 7979–7984.
- [3] Sherman, F., Stewart, J.W. and Tsunasawa, S. (1986) BioEssays 3, 27–31.
- [4] Huang, S., Elliott, R.C., Liu, P.-S., Koduri, R.K., Weickmann, J.L., Lee, J.-H., Blair, L.C., Ghosh-Dastidar, P., Bradshaw, R.A., Bryan, K.M., Einarson, B., Kendall, R.L., Kolacz, K.H. and Saito, K. (1987) Biochemistry 26, 8242–8246.
- [5] Flinta, C., Persson, B., Jörnval, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
- [6] Boissel, J.-P., Kasper, T.J. and Bunn, H.F. (1988) J. Biol. Chem. 263, 8443–8449.
- [7] Lee, F.-J.S., Lin, L.-W. and Smith, J.A. (1990) J. Biol. Chem. 265, 11576–11580.
- [8] Augen, J. and Wold, F. (1986) Trends Biochem. Sci. 11, 494–497.
- [9] Yoshikawa, A., Isono, S., Sheback, A. and Isono, K. (1987) Mol. Gen. Genet. 209, 481–488.
- [10] Tanaka, S., Matsushita, Y., Yoshikawa, A. and Isono, K. (1989) Mol. Gen. Genet. 217, 289–293.
- [11] Arai, K., Clark, B.F., Duffy, L., Jones, M.D., Kaziro, Y., Laursen, R.A., L'Italien, J., Miller, D.L., Nagarkatti, S., Nakamura, S., Nielsen, K.M., Petersen, T.E., Takahashi, K. and Wade, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1326–1330.
- [12] Smith, V.F., Schwartz, B.L., Randall, L.L. and Smith, R.D. (1996) Protein Sci. 5, 488–494.
- [13] Charbaut, E., Curmi, P.A., Ozon, S., Lachkar, S., Redeker, V. and Sobel, A. (2001) J. Biol. Chem. 276, 16146–16154.
- [14] Curmi, P., Maucuer, A., Asselin, S., Lecourtois, M., Chaffotte, A., Schmitter, J.M. and Sobel, A. (1994) Biochem. J. 300, 331–338.
- [15] Honda, S., Asano, T., Kajio, T. and Nishimura, O. (1989) Arch. Biochem. Biophys. 269, 612–622.
- [16] Takao, T., Kobayashi, M., Nishimura, O. and Shimonishi, Y. (1987) J. Biol. Chem. 262, 3541–3547.
- [17] Persson, B., Flinta, C., von Heijne, G. and Jörnval, H. (1985) Eur. J. Biochem. 152, 523–527.
- [18] Polevoda, B. and Sherman, F. (2000) J. Biol. Chem. 275, 36479–36482.
- [19] Urbancikova, M. and Hitchcock-DeGregori, S.E. (1994) J. Biol. Chem. 269, 24130–24135.
- [20] Ogawa, H., Gomi, T., Takata, Y., Date, T. and Fujioka, M. (1997) Biochem. J. 327, 407–412.