Marquette University
e-Publications@Marquette

Chemistry Faculty Research and Publications

Chemistry, Department of

12-1-2006

The Proteomics of N-terminal Methionine Cleavage

Frédéric Frottin
Institut des Sciences du Végétal

Aude Martinez
Institut des Sciences du Végétal

Philippe Peynot
Institut des Sciences du Végétal

Sanghamitra Mitra
Loyola University Chicago

Richard C. Holz
Marquette University, richard.holz@marquette.edu

See next page for additional authors

Authors

Frédéric Frottin, Aude Martinez, Philippe Peynot, Sanghamitra Mitra, Richard C. Holz, Carmela Giglione, and Thierry Meinnel

The Proteomics of N-terminal Methionine Cleavage*⑤

Frédéric Frottin‡§, Aude Martinez‡§, Philippe Peynot‡, Sanghamitra Mitra¶‖, Richard C. Holz¶‖, Carmela Giglione‡, and Thierry Meinnel‡**

**Methionine aminopeptidase (MAP) is a ubiquitous, essential enzyme involved in protein N-terminal methionine excision. According to the generally accepted cleavage rules for MAP, this enzyme cleaves all proteins with small side chains on the residue in the second position (P1′), but many exceptions are known. The substrate specificity of *Escherichia coli* MAP1 was studied *in vitro* with a large (>120) coherent array of peptides mimicking the natural substrates and kinetically analyzed in detail. Peptides with Val or Thr at P1′ were much less efficiently cleaved than those with Ala, Cys, Gly, Pro, or Ser in this position. Certain residues at P2′, P3′, and P4′ strongly slowed the reaction, and some proteins with Val and Thr at P1′ could not undergo Met cleavage. These *in vitro* data were fully consistent with data for 862 *E. coli* proteins with known N-terminal sequences *in vivo*. The specificity sites were found to be identical to those for the other type of MAPs, MAP2s, and a dedicated prediction tool for Met cleavage is now available. Taking into account the rules of MAP cleavage and leader peptide removal, the N termini of all proteins were predicted from the annotated genome and compared with data obtained *in vivo*. This analysis showed that proteins displaying N-Met cleavage are over-represented *in vivo*. We conclude that protein secretion involving leader peptide cleavage is more frequent than generally thought. *Molecular & Cellular Proteomics 5: 2336–2349, 2006.*

Protein N-terminal methionine excision (NME)[1] is an essential cotranslational process that occurs in the cytoplasm of all organisms and in the two organelles (*i.e.* mitochondria and plastids) displaying protein synthesis (for reviews, see Refs. 1 and 2). NME involves two types of methionine aminopeptidase (MAP), MAP1 (type-I) and MAP2 (type-II), which have similar three-dimensional structures despite having only low

levels of sequence identity (for a review, see Ref. 3). Higher eukaryotes have at least one MAP1 (MAP1A, also known as MetAP1b; Ref. 4) and one MAP2 in the cytoplasm and one MAP1 (MAP1D) in the organelles. Archaea, such as *Pyrococcus furiosus*, have one MAP2, and eubacteria, such as *Escherichia coli*, have only MAP1. In eubacteria and the organelles, a peptide deformylase systematically removes the *N*-formyl group, leading to the activation of MAP cleavage (5–7). Bacterial MAP1 cannot cleave *N*-blocked polypeptides (6).

The role of NME remains poorly understood (2), but this process is recognized to be the major source of N-terminal amino acid diversity. It is thought that up to 80% of the proteins of any given proteome undergo this modification (8–10). Early biochemical and genetic studies indicated that this activity was very specific for peptides with an N-terminal Met residue (P1 position according to Schechter's nomenclature, Ref. 11), and the penultimate position (P1′) was identified as the major determinant for cleavage (for reviews, see Refs. 12 and 13). The rule of thumb is that cleavage occurs if the side chain is small enough as is the case for Ala, Cys, Pro, Ser, Thr, and Val. According to this rule, cleavage is not possible for larger side chains. This stochastic rule, which emerged from early bioinformatics analysis based on compilation of the few protein sequences available at the time (14), was confirmed by biochemical analysis of MAP activity *in vitro* with about a dozen model tri- and tetrapeptides (6, 15, 16). Edman degradation sequencing of two reporter proteins in *E. coli* was used to further the analysis (17, 18). The authors of these studies suggested that the process was statistical rather than stochastic and that cleavage efficiency was correlated with side-chain length or gyration radius as defined by Levitt (19) at P1′ (20). This is fully confirmed by the structural analysis of many MAPs (3, 4). Cleavage probability was found to be highest for Gly (97%) followed by Ala, Thr (90%), Pro, Ser, and Val (84%). Cleavage was less probable for the substrates Cys (71%), Ile (18%), Asp, Leu, and Asn (16%). The underlying idea is that the S1′ binding pocket, into which the P1′ side chain must fit, is small and tolerates smaller side chains with Gly being the optimal residue. However, these two analyses were based on only two reporter proteins, although the comparison was straightforward as the protein sequences differed only at position P1′. These findings form the basis of our current understanding of the process (for a review, see Ref. 1; see also Fig. 1 in Ref. 21), which is used for bioinformatics

[1] The abbreviations used are: NME, N-terminal Met excision; LPR, leader peptide removal; MAP, methionine aminopeptidase; Nva, norvaline; Aba, α-aminobutyrate.

analysis of the process in genomes (see Scheme 1 and Fig. 4 in Ref. 22). More recent proteomics analysis in *E. coli* (23, 24) has generated a comprehensive overview, and the N terminus has now been clearly determined for the products of a total of 862 open reading frames (of the 4071 proteins in the *E. coli* proteome). However, the conclusions of this analysis (23) conflicted with the deduced rules of NME. Unlike proteins with Ala and Ser at P1′, the authors concluded that those with Gly, Pro, and Thr in this position displayed fastidious, "variable cleavage." Finally proteins with Val at P1′ were found to resist NME.

It is therefore difficult to predict the rules of N-terminal cleavage reliably based on a compilation of all these data. For example, it is unclear whether proteins with Gly, Pro, Thr, Val, Cys, Ile, Asp, Asn, or Leu at P1′ are cleaved and, if so, whether some of the rules for MAP cleavage have yet to be determined or whether other competing mechanisms, such as protein exportation, secretion, or insolubility due to membrane localization, may have biased analysis *in vivo*. This ambiguity makes reliable proteome annotations difficult for 27% of the proteins in the bacterial proteome and renders the production of recombinant proteins of therapeutic interest risky given the high antigenicity of the N terminus if incorrectly processed (25). This problem was initially encountered in the production of human hemoglobin (26). Closer examination of published biochemical analyses (6, 15, 16, 27) showed that these analyses were not really systematic as they compared peptides of different lengths with different residues in positions 3 (P2′) and 4 (P3′). These differences may have greatly influenced data interpretation. In addition, by analogy with the cleavage rules of similar peptidases with extended recognition regions around the cleavage site, it is unclear whether MAP enzymes have a P2′ or a P3′ recognition site (S2′ and S3′) and whether these two sites influence substrate specificity *in vivo*. Furthermore we noted that, in analyses *in vitro*, data were systematically compared at fixed peptide concentrations (usually 4 or 20 mM), although K_m values are known to range from 1 to 5 mM, and the steady-state concentration of the nascent peptide is of the order of 0.1 mM (28). Thus, k_{cat}/K_m measurements (*i.e.* catalytic efficiency) were required for the modeling of MAP activity from *in vitro* studies, facilitating the comparison of coherent peptide series essential for the drawing of definitive conclusions.

The aim of this study was to reconcile the data from *in vivo* and *in vitro* analysis using combined biochemical and bioinformatics analysis to draw a definitive and coherent picture of NME and its cleavage rules *in vivo*. We first used *E. coli* as a model system, making use of the large body of *in vivo* sequence determination data for the direct comparison of *in vitro* and *in vivo* data. We found that proteins with Thr and Val at P1′ were poor substrates of MAP1 and MAP2, frequently resisting NME cleavage, and that the P2′ and P3′ positions had a strong effect on specificity for both MAP types. These elements would 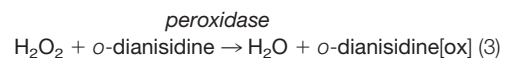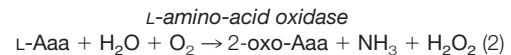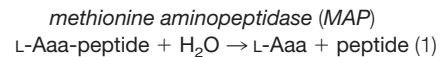be expected to have different effects on intrinsic protein processing in the natural proteome or the overproduction of foreign, recombinant proteins in bacteria.

EXPERIMENTAL PROCEDURES

Chemicals and Peptides—All chemicals were purchased from Sigma. Most peptides were synthesized as custom products at Genscript Corp. (Piscataway, NJ). Others were purchased from Bachem (Well-am-Rhein, Germany) or Sigma. All peptides were >95% pure and were dissolved in water at a final concentration of 50–150 mM. Heating at 50 °C was required to achieve dissolution in a few cases. Four peptides (MGF, MFG, MWG, MGW, and MPG) had to be dissolved in water plus dimethyl sulfoxide (5–10% final concentration) at a concentration of 3–10 mM. Each addition of 1% dimethyl sulfoxide to the MAP-coupled assay was found to decrease MAP activity by only 10%.

Methionine Aminopeptidase Purification—Native *E. coli* methionine aminopeptidase (EcMAP1) was overproduced from plasmid pXL1071 (29). JM83 cells expressing the plasmid were grown at 37 °C for 8–12 h in 2× TY medium (16 g/liter Bacto-tryptone, 10 g/liter Bacto-yeast extract, 5 g/liter NaCl and adjusted to pH 7.0 with NaOH) supplemented with 50 μg/ml ampicillin to an A_{600} of ~ 0.9. Cells were induced with 0.3 mM isopropyl 1-thio-β-D-galactopyranoside and incubated for a further 12 h with shaking. The cells were harvested by centrifugation and resuspended in 10–20 ml of buffer A consisting of 50 mM KHPO₄ (pH 7.5) and 0.2 mM CoCl₂. The samples were sonicated, and cell debris were removed by centrifugation. The supernatant was subjected to 0–80% ammonium sulfate precipitation and centrifuged for 30 min at 4 °C. The pellet was resuspended in 5 ml of buffer A, applied to a Superose-6 column (1.6 × 60 cm; GE Healthcare) and eluted at a flow rate of 0.5 ml/min in buffer A. The pool with MAP activity (30 ml) was loaded on a Q-Sepharose (1.6 × 10 cm; GE Healthcare) anion-exchange column equilibrated in buffer A, and the sample was eluted with a 0.2 M/h linear NaCl gradient (2.5 ml/min). The proteins recovered were homogeneous and were stored at −30 °C in buffer A plus 55% glycerol. *P. furiosus* MAP2 (PfMAP2) was purified as described elsewhere (30).

MAP Activity Measurements—MAP activity was assayed at 30 °C by continuously monitoring the absorbance of oxidized *o*-dianisidine at 440 nm, coupling MAP activity to both L-amino-acid oxidase and peroxidase activities, according to Scheme 1 (where Aaa is any α-amino acid).

$$\text{methionine aminopeptidase (MAP)}$$
$$\text{L-Aaa-peptide} + H_2O \rightarrow \text{L-Aaa} + \text{peptide} \quad (1)$$

$$\text{L-amino-acid oxidase}$$
$$\text{L-Aaa} + H_2O + O_2 \rightarrow \text{2-oxo-Aaa} + NH_3 + H_2O_2 \quad (2)$$

$$\text{peroxidase}$$
$$H_2O_2 + o\text{-dianisidine} \rightarrow H_2O + o\text{-dianisidine[ox]} \quad (3)$$

SCHEME 1

The conditions of this assay were set according to published observations (16, 31–33). The standard assay was performed in a final volume of 100 μl in plastic cuvettes with a 1-cm optical path (UV-ettes; Eppendorf). Changes in absorbance over time were followed using an Ultrospec-4000 spectrophotometer (GE Healthcare) equipped with a thermostat and a six-position Peltier heated cell changer. The reaction mixtures (95 μl) contained (final concentrations in 100 μl) 45 mM Hepes, pH 7.4, 0.2 mM CoCl₂ (Sigma; catalog number C8661) unless otherwise stated (Fig. 1*A*), 0.1 mg/ml *o*-dianisidine (Sigma; catalog number D9154, solution prepared from one tablet every other day and stored at 4 °C in the dark), 3 units of horseradish peroxidase (2000 units/ml, Sigma; catalog number P8415, stored for months at −20 °C), 0.5 units of L-amino-acid oxi-

dase (63 units/ml, Sigma; catalog number A9378, stored for months at 4 °C in the dark), and 0.01–20 mM peptide (see above). This premixture was incubated for 4–15 min in the spectrophotometer at 30 °C until the base line at 440 nm was stable. At this point, the spectrophotometer was set to zero. The reaction was started by adding 5 μl of purified 0.1–20 μM (final concentration) MAP in 50 mM Hepes, pH 7.4, 0.2 mM CoCl$_2$, and 150 mM NaCl, and the reaction was followed for 2–15 min. The initial velocity could generally be calculated after the first 2 min for MAP1 and the first 5 min for MAP2. This assay can measure six velocities of 0.001–0.5 A_{440}/min in parallel; this is generally sufficient to determine the catalytic constants. Above 0.5 A_{440}/min, MAP must be diluted. The measured velocities were transformed into s^{-1} by dividing the values by the enzyme concentration used, the molar extinction coefficient of oxidized o-dianisidine (10,580 M^{-1}·cm^{-1} as determined experimentally by incubating L-Met in the reaction mixture and allowing the reaction to continue to completion), and the length of the optical path (1 cm). Amino-acid oxidase has a broad enough substrate specificity for the efficient oxidation of many amino acids (31). Cysteine-containing peptides were systematically avoided as they are not compatible with the assay. This incompatibility is probably due to reduction, by the thiol group of Cys, of the H$_2$O$_2$ produced by amino-acid oxidase and used as the substrate of peroxidase, resulting in assay inhibition. We confirmed that the coupled assay was indeed inhibited by the Cys-Gly dipeptide. In a few cases, at concentrations higher than the K_m value, we noted that establishment of the stationary phase was delayed, although the associated reason was unclear. This problem was easily avoided provided kinetics was followed for a longer time.

Interpretation of Kinetic Data—The kinetic parameters k_{cat} and K_m were obtained using Enzyme Kinetics module 1.1 of Sigma plot (version 8.0) by non-linear Michaelis-Menten equation fitting. The confidence limits given are those associated with the data set. The kinetic parameter k_{cat}/K_m was derived from iterative non-linear least square fits of the Michaelis-Menten equation using the experimental data (34). Confidence limits for the fitted k_{cat}/K_m values were determined by 100 Monte Carlo iterations using the experimental standard deviations on individual measurements. We obtained similar data with the two peptides of the same sequence obtained from (i) the same supplier and (ii) two different suppliers.

Sequence Databases and Protein Pattern Syntax—The 4071 ORF sequences of *E. coli* were retrieved from the EcoGene website (ecogene.org/VerifiedInfo.php?download=true). The compiled N-terminal protein sequence data were retrieved from bmb.med. miami.edu/EcoGene/EcoWeb/CeSSPages/VerifiedProts.htm (see also Ref. 35). The data were ordered and classified by categories for statistical analysis. The complete proteomes of *Bacillus subtilis* and *P. furiosus* were retrieved at www.ebi.ac.uk/integr8/FtpSearch.do; jsessionid=98818969299187E6D1B9222E3B694316?orgProteomeId =6 and www.ebi.ac.uk/integr8/OrganismHomeAction.do?orgProteomeID=77, respectively.

We searched for protein patterns (36) at www.infobiogen.fr/ services/analyseq/cgi-bin/patternp_in.pl. The pattern syntax used in this text was: "<M" constrains the pattern to the N-terminal residue (*i.e.* an initiator Met), "(^XY)" means that residues *X* and *Y* (or a subset list) are excluded, and "(Y/Z)" means that residue *Y* or *Z* (or a subset list) is included. *X*, *Y*, or *Z* correspond to any normal amino acid.

RESULTS

Validation of the Experimental in Vitro Model, E. coli Methionine Aminopeptidase (EcMAP1)—A high throughput, substrate-independent assay was required to investigate EcMAP1 substrate specificity. Most of the available assays were not suitable (37–39). The only available test that appeared fully

independent of the peptide sequence had only been used in a discontinuous manner. This assay involved coupling with two other activities, those of amino-acid oxidase and peroxidase (16, 31–33). We set up this assay for use in continuous conditions. This continuous version of the assay proved extremely rapid, reliable, and cheap, and all the peptides used were soluble enough for kinetic measurements. That Cys-containing peptides inhibited the assay as a result of the reducing effect of the thiol group of Cys made it necessary to ensure that none of the peptides tested included a Cys. Other MAP activity assays are also subject to interference from Cys-containing peptides (40). However, this problem is not particularly important as Cys is the rarest amino acid at the N termini of proteins. For example, only six (0.17%) of the 4071 ORFs of *E. coli* are predicted to have a Cys as the second residue, 14 (0.34%) are predicted to have a Cys as the third residue, and 26 are predicted to have a Cys as the fourth residue (0.64%). Thus, our study covers 99% of the proteins of *E. coli*.

We first assessed the relevance of our assay with our purified EcMAP1. We evaluated the impact of the nature of the cocatalytic metal cation (see references compiled in Ref. 2) (Fig. 1*A*). Cobalt cations appeared to be the most efficient and were used in all MAP assays. In this study, our final goal was to measure *in vitro* the values of catalytic cleavage of Met cleavage (*i.e.* k_{cat}/K_m) and to compare these data with those derived from N-terminal sequence analysis of proteins expressed *in vivo*. We therefore systematically measured the k_{cat} and K_m values associated with a given peptide. Fig. 1*B* provides an example of fit quality and shows the relevance of fitting the Michaelis-Menten model to the data. With peptide Met-Ala-Met-Lys-Ser, the substrate most efficiently processed in this study, a catalytic efficiency value of 81,700 ± 8,000 M^{-1}·s^{-1}, associated with K_m = 0.05 ± 0.01 mM and k_{cat} = 3.9 ± 0.2 s^{-1}, was measured. These data, obtained for more than 120 different peptides (compiled data are shown in Supplemental Table S1), showed that both k_{cat} and K_m values were extremely variable with ranges covering more than 2 orders of magnitude (0.01–4 s^{-1} and 0.05–15 mM, respectively). It would therefore not be possible to make reliable comparisons between the data obtained for peptide concentrations of 4 and 20 mM in previous publications. These findings also suggest that the enzyme worked below the k_{cat}/K_m rate of cleavage as the steady-state peptide concentration is about 0.1 mM.

In Vitro Characterization of E. coli Methionine Aminopeptidase: the S1 Site and the Influence of Peptide Length—We assessed whether EcMAP1 could be used as a broader specificity aminopeptidase by varying the nature of the P1 side chain (Fig. 2*A*). All experiments were carried out with tripeptides containing Ala and Ser in the second and third positions, respectively. Met-Ala-Ser proved to be one of the most efficient substrates in our analysis (see below and Supplemental Table S1). We found that Met and its unnatural norleucine
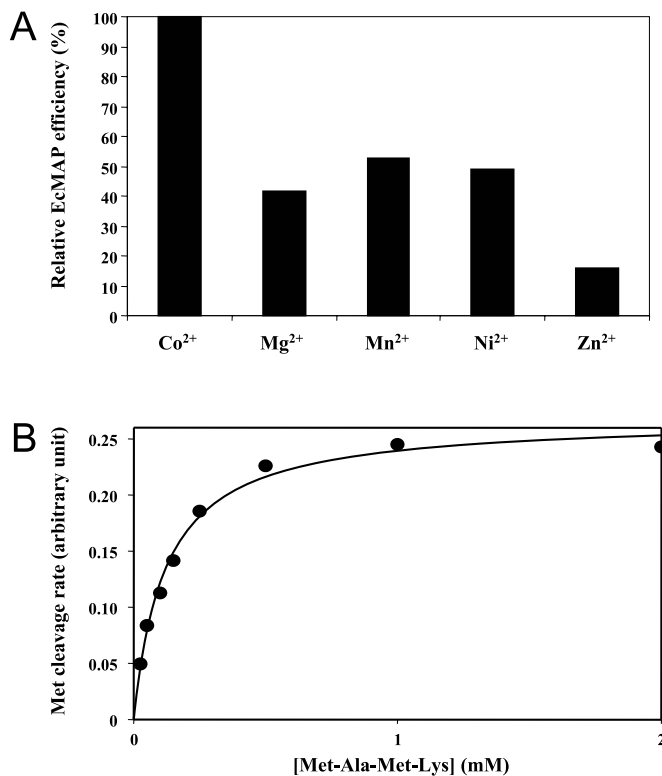
FIG. 1. **Validation of the continuous enzymatic assay of MAP for the medium throughput determination of catalytic efficiency.** *A*, influence of the metal cation on EcMAP cleavage efficiency. The assay was performed in the presence of 0.2 mM of the indicated metal cation, provided as a dichloride derivative in the solutions. The enzyme was also diluted in the presence of the concentration of the metal salt. *B*, example of the determination of kinetic parameters and fitting to a data set of the theoretical Michaelis-Menten equation. The peptide used was Met-Ala-Met-Lys.

(Nle) classic mimic were the only residues at position P1 that gave efficient cleavage. This finding is consistent with the tapered shape of the S1 pocket (41). The natural amino acids Leu and Phe could be processed *in vitro* but with a catalytic efficiency more than 3 orders of magnitudes lower. Similar results were obtained with methionine sulfoxide (Mox) or norvaline (Nva), which features an *n*-propyl side chain. Decreasing the length of the side chain to α-aminobutyrate (Aba), an unusual amino acid with a two-carbon linear side chain mimicking Cys and Ser, or Ala led to resistance to enzyme cleavage. We concluded that EcMAP1 could not further process its natural substrates and that the kinetic data were representative of a single cleavage site between P1 and P1′ provided that Met was the first amino acid of the peptide.

We investigated the role of the interface between the P1 and P1′ sites by synthesizing three peptides derived from Met-Ala-Ser: one with 2-methyl alanine (2mA), the second with oxamic acid (no side chain but with a keto group instead, Oxa), and the third with a D-Ala in position 2 replacing L-Ala. The first two peptides were not cleaved. This was in contrast to the D-Ala variant at P1′ that was 2 times more efficiently
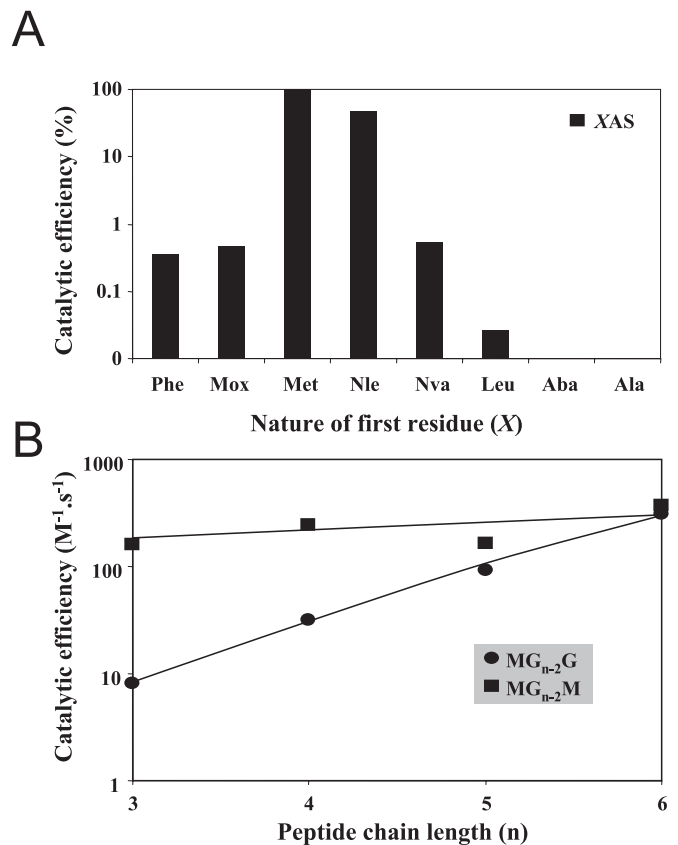

FIG. 2. **Impact of P1 and peptide chain length on cleavage by MAP.** *A*, influence of P1 on MAP cleavage efficiency *in vitro*. The peptide series was *X*-Ala-Ser where *X* is the indicated amino acid. A value of 100 was assigned to the k_{cat}/K_m value of Met-Ala-Ser. *B*, influence of peptide chain length. Two series of peptides were assayed, Met-(Gly)$_{n-2}$-Gly and Met-(Gly)$_{n-2}$-Met where $n = 3$–6 corresponds to the length of the peptide. Differences in length between peptides were ensured by a poly-Gly linker. The *x* axis corresponds to the full-length sequence. The shortest peptide tested was three amino acids long. The k_{cat}/K_m value is plotted as a function of peptide chain length. *Mox*, methionine sulfoxide; *Nle*, norleucine.

cleaved (Table S1). Thus, both geometry and the hydrogen environment around the α carbon of the second residue had a critical influence on cleavage by EcMAP1.

We also investigated the impact of the length of the polypeptide chain. We first confirmed that EcMAP1 cleaved dipeptides extremely inefficiently: Met-Ala or Met-Gly were cleaved with a k_{cat}/K_m value 2 orders of magnitude lower than that for the reference tripeptide with a serine in position P2′ (Supplemental Table S1). In contrast, tripeptides and larger peptides starting with Met-Gly proved to be efficient substrates (Fig. 2*B*). We assessed the impact of side chain and length by comparing two peptides series. The last amino acid in these two series of peptides was Gly or Met, and the linker peptide consisted of Gly. The k_{cat}/K_m value was found to be maximal for peptides more than five residues long (Fig. 2*B*). Comparison of the data obtained for peptides of a given peptide length between the two series indicated that the
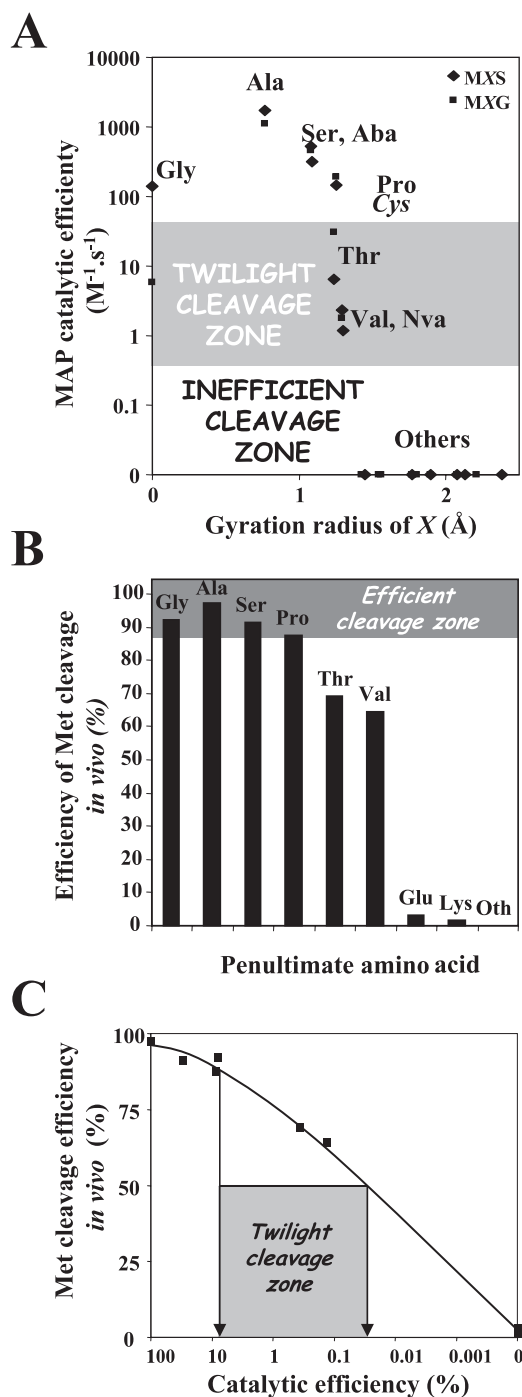
A

B

C

FIG. 3. **Tight correlation between *in vitro* and *in vivo* cleavage efficiency with variation of the penultimate (P1′) residue.** *A*, influence of P1′ *in vivo* and correlation with side-chain radius gyration (19). Two series were tested, Met-*X*-Gly and Met-*X*-Ser, where *X* is any amino acid. Cys could not be tested directly (it is therefore shown in *italics* here; see text); two unnatural amino acids, Aba and Nva, were included. The catalytic efficiency (k_{cat}/K_m) is plotted as a function of the side-chain radius gyration (data from Ref. 19). The nature of the corresponding amino acid is given. "Others" corresponds to all natural amino acids not indicated in the figure. *B*, influence of P1′ on *in vivo* cleavage efficiency. Data were taken from the 862-sequence database of proteins with determined N-terminal sequences. "Oth"

nature of the side chain at position 3 (P2′ site) or 4 (P3′ site) had a strong effect and that position 5 (P4′ site) also had a significant effect. A thorough analysis of the impact of the amino acids at these positions is therefore required.

The S1′ Subsite of EcMAP1: Difficult Cleavage of the Thr and Val Residues—MAPs are known to cleave peptides selectively between the terminal Met and the penultimate or P1′ residue. In two tripeptide series, one with Ser and the other with Gly at P2′, we measured the catalytic efficiency of Met cleavage for the complete set of amino acids with the exception of Cys (Fig. 3*A*). Ala was cleaved most efficiently followed by Ser, Gly, and Pro. Thr and Val were cleaved less efficiently with k_{cat}/K_m values 2 orders of magnitude lower in both series. We could not determine k_{cat}/K_m values for other amino acids (at least 5 orders of magnitude lower). In particular, peptides with Ile, Asn, Asp, Met, or Leu at P1′ were not cleaved. These findings contrast with two reports using reporter proteins *in vivo* (17, 18). At this stage, we do not know whether our *in vitro* assay was representative of *in vivo* conditions or whether it was not sensitive enough. Nevertheless there was a clear relationship between gyration radius (as defined by Levitt (19)) of the side chain at P1′ and catalytic efficiency as reported previously (Fig. 3*A*). This relationship was fully confirmed in the context of the most efficient series, the one derived from tetrapeptide M*X*MK (with *X* = Ala, Gly, Pro, Ser, Thr, Val, Asn, Ile, Leu, Asp, Glu, Phe, Gln, Lys, or Arg; data reported in Supplemental Table S1). In contrast to previous reports suggesting that Gly in the penultimate position gave the most efficient processing, our data clearly show that the side chain of Ala is optimal at P1′. We analyzed two tripeptides with unusual P1′ side chains. One contained Nva, and the second contained Aba, mimicking Cys and Ser, respectively. The deduced k_{cat}/K_m values for these peptides were entirely consistent with the model. This strongly suggests that the values for Cys would be similar to that of Pro, which has the most similar gyration radius (19).

In Vivo N-terminal Met Cleavage Is a Statistical Process Mostly Dependent on the Catalytic Efficiency of MAP—The data described above indicated that only peptides with Ala, Gly, Ser, Cys, or Pro at P1′ are substrates of EcMAP1 and that the cleavage of peptides with Thr or Val at this position is less predictable, possibly depending on other features. We compared this *in vitro* analysis with *in vivo* results using compiled data from N-terminal protein sequencing (data available at ecogene.org/VerifiedInfo.php?download=true). This analysis

corresponds to all natural amino acids not indicated in the figure. *C*, Met cleavage efficiency at P1′ *in vivo* is directly correlated with MAP cleavage efficiency assessed *in vitro*. The data set in Fig. 1 was associated with that of *A* for a given amino acid at position P1′. The panel shows a plot of the cleavage efficiency *in vivo* for a given P1′ position *versus* the activation energy (a value proportional to the logarithm of the k_{cat}/K_m value; see Ref. 79) of the reaction *in vitro* measured with a model substrate with the same P1′.

FIG. 4. **Comparison of the impact of the nature of the P2′ position on NME *in vitro* and *in vivo*.** *A*, influence of the P2′ position *in vitro*. The data came from two data sets, one with Met-Ala-*X* (*black bars*) and the second with Met-Gly-*X* (*gray bars*), where *X* is any natural amino acid. The relative catalytic efficiency (k_{cat}/K_m) is plotted as a function of *X*. A value of 100 was assigned to *X* = Ser in both series. A *black* or *gray asterisk* (*) indicates that the corresponding peptide in either series was not characterized. *B*, influence of the P2′ position *in vivo*. Data were taken from the 862-sequence database of proteins with determined N-terminal sequences. The efficiency of Met cleavage was defined as the percentage of proteins with a <M(A/C/G/P/S/T/V) pattern resistant to Met cleavage (385 instances). The general mean value of 87% was used to display differences optimally. Given the small number of instances, an *error bar* corresponding to one more or less uncleaved example was introduced. *Arrows* indicate the amino acids leading to significant decrease of cleavage efficiency in the series.
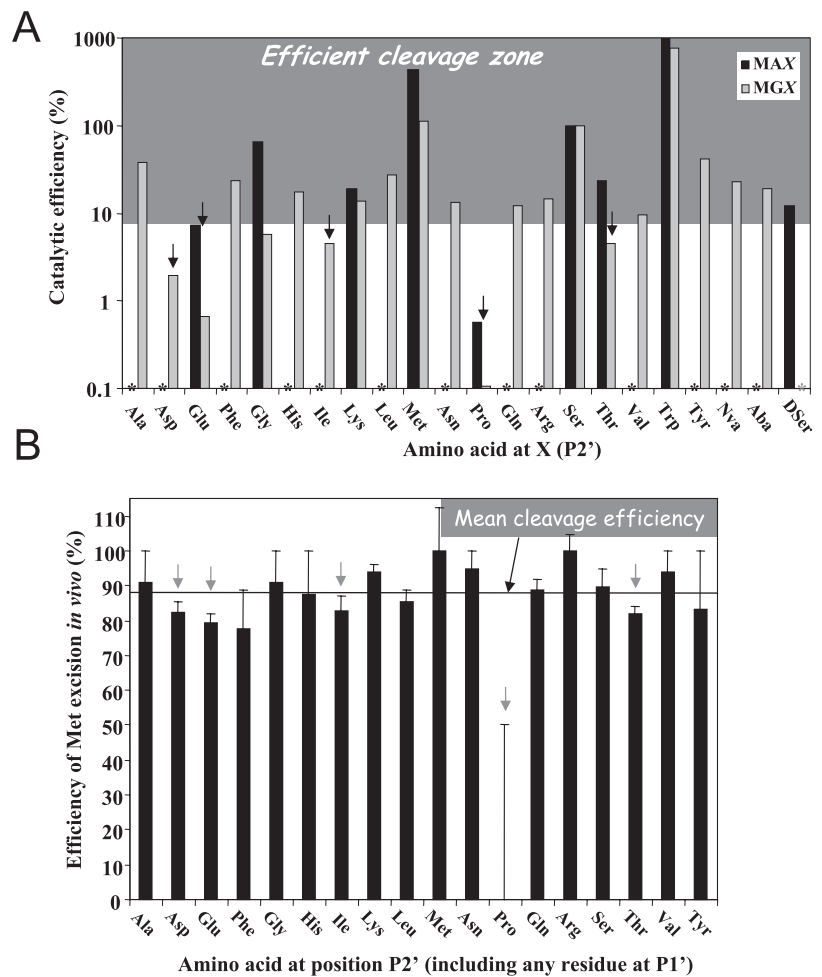
showed that 14.4% of the proteins are cleaved by a signal peptidase. Cleavage usually occurs between two alanine residues located at positions 23–27 of the polypeptide sequence. The remaining proteins (85.6%) either retain or lose their N-terminal Met. We analyzed this pool of proteins (738 sequences) to draw up the rules of NME *in vivo* and compared the data obtained with our *in vitro* analysis. Cleavage was not reported with Ile, Asp, Asn, Leu, Met, or Gln at P1′. This fully confirmed the results of our *in vitro* analysis, conflicting with previous analyses based on reporter proteins (17, 18). No cleavage was reported if the P1′ residue was Arg, Phe, His, Tyr, or Trp as suggested *in vitro*. Finally one cleavage with Glu (42) and two with Lys (43, 44) at P1′ were reported, but these cleavages concerned only 3.0 and 1.5%, respectively, of all the sequences containing these amino acids at this position. These exceptional cases of cleavage may result from the action of dedicated acylaminopeptidases rather than MAP as shown previously for actin in animals (45, 46).

Met excision was not systematic with Ala, Cys, Pro, Ser, Thr, and Val at P1′ as 13% of such proteins escaped Met cleavage (NME). Cleavage efficiency was maximal with Ala (97%) and minimal with Thr (69%) and Val (64%) (Fig. 3*B*).

This efficiency was similar to that found in the kinetic analysis shown in Fig. 3*A*. For each type of residue found at P1′, we therefore plotted *in vivo* cleavage efficiency as a function of EcMAP1-mediated cleavage efficiency measured *in vitro*. There was an extremely strong correlation between the two data sets (Fig. 3*C*). Thus, NME or resistance to NME was limited essentially by MAP catalytic efficiency *in vivo*, and the data obtained *in vitro* were reliable for the modeling of NME *in vivo*. We defined a "twilight cleavage zone" as the catalytic cleavage efficiency leading to partial cleavage efficiency *in vivo* (Fig. 3). Below this zone, cleavage is considered as inefficient.

The S2′ Subsite of EcMAP1: Pro and Glu Are Inefficiently Cleaved Residues—We first investigated the impact of the P2′ amino acid in the context of two series of tripeptides beginning with Met-Ala, the most efficient substrate of EcMAP1, and Met-Gly, an intermediate substrate for EcMAP1. The P2′ side chain clearly had a strong effect on the catalytic efficiency of Met excision catalyzed by EcMAP1. Some residues, such as Pro and Glu, decreased the k_{cat}/K_m ratio by more than an order of magnitude in both series with others, such as Asp, Ile, and Thr, significantly decreasing this value by at least 1

FIG. 5. **Impact of the nature of P2′ on NME when P1′ is a residue processed inefficiently (*i.e.* Thr or Val) for MAP.** *A*, influence of the P2′ position *in vitro*. The data were obtained from two data sets, one with Met-Val-*X* (*black bars*) and the second with Met-Thr-*X* (*gray bars*), where *X* is the indicated natural amino acid. The relative catalytic efficiency (k_{cat}/K_m) is plotted as a function of *X*. A value of 100 was assigned to Met-Ala-Ser. A *black* or *gray asterisk* (*) indicates that the corresponding peptide in either series was not characterized. *B*, influence of the P2′ position *in vivo*. Data were taken from the 862-sequence database of proteins with determined N-terminal sequences. The efficiency of Met cleavage was defined as the percentage of proteins with a <M(A/C/G/P/S/T/V) pattern resistant to Met cleavage. *Arrows* indicate the amino acids leading to significant decrease of cleavage efficiency in the series.
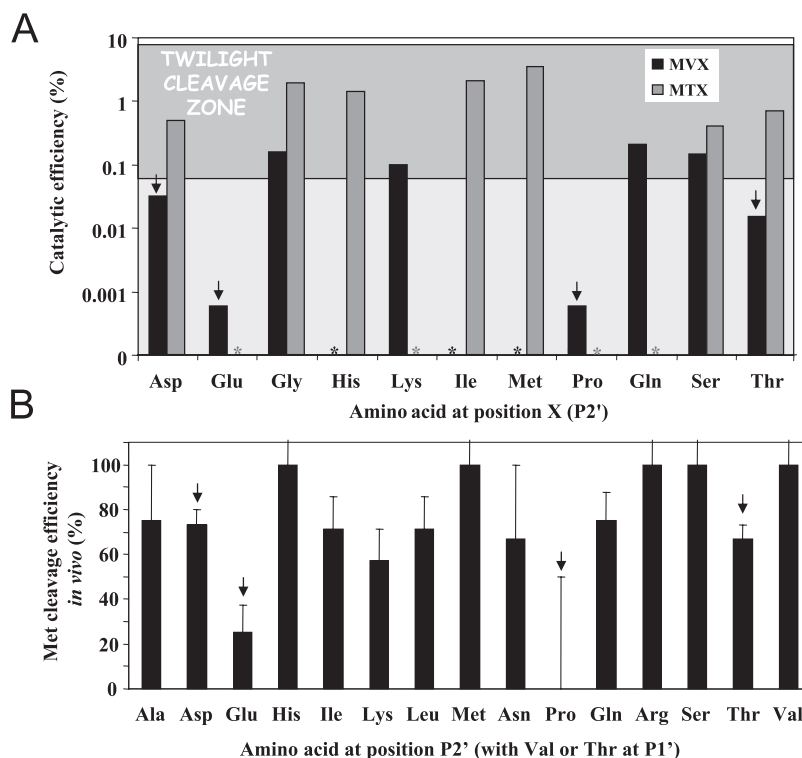

order of magnitude (Fig. 4*A*). Efficiency was optimal for peptides with Trp, Met, or Ser at P2′ as these residues increased the k_{cat}/K_m ratio. Other residues gave intermediate values. We next analyzed the impact of the P2′ position on the efficiency of cleavage *in vivo* using cleavage data for this position in the 385-sequence data library. The data are displayed in Fig. 4*B*. They show a clear trend toward difficult cleavage when the P2′ residue is Asp, Glu, Ile, Pro, or Thr. Phe and Tyr were also suggested to decrease cleavage efficiency *in vivo*, although this prediction was not consistent with the data for catalysis. Nevertheless the small number of peptides containing these residues resulted in the data not being robust enough for confident prediction of the *in vivo* behavior linked to the presence of these two aromatic amino acids at position P2′ (Fig. 4*B*). We concluded that the kinetic data were representative of the *in vivo* efficiency of cleavage.

In a second series of investigations of the impact of position P2′ in EcMAP1 cleavage, we assessed tripeptides with Thr or Val at position P1′, corresponding to substrates for fastidious cleavage by EcMAP1 (*i.e.* zone leading to the twilight cleavage zone). A similar trend was observed for this series (Fig. 5*A*): Pro and Glu and, to a lesser extent, Asp and Thr at P2′ were less likely to result in cleavage than other residues. If we considered only those proteins starting with Val or Thr at P1′, the number of examples was significantly reduced, and the statistics were therefore not robust enough for significant conclusions to be drawn, but Glu and Pro seemed to reduce cleavage efficiency (Fig. 5*B*). We concluded that the cleavage of proteins with <M(V/T)(E/P) patterns was virtually impossi-

ble *in vivo* and that the cleavage of proteins with <M(V/T)(D/T) patterns was difficult (*i.e.* "inefficient"), resulting in partial Met retention in some cases.

The S3′ Subsite of EcMAP1 Shows Negative Discrimination Similar to That for the S2′ Subsite—In a final systematic search for EcMAP1 cleavage criteria, we investigated the role of position P3′ and P4′. Data were obtained with a tetrapeptide, Met-Gly-Met-*X*, in which position *X* (P3′) was varied with all natural amino acids (except Cys). Significant differences were observed, but only Asp, Glu, Pro, and Thr, *i.e.* the same residues as for P2′, resulted in a significantly lower cleavage efficiency (Fig. 6*A*). Another set of data were obtained with a pentapeptide, Met-Gly-Gly-Gly-*X*, in which position *X* (P4′) was varied with a reduced set of 10 amino acids (Fig. 6*B*). Phe was the most efficient, and Glu (again) was the least efficient residue. Cleavage efficiency was only modestly increased when peptides Met-Ala-Met-Lys and Met-Ala-Met-Lys-*Ser* were compared (Supplemental Table S1). The impact of P3′ or P4′ *in vivo* was investigated (Supplemental Fig. S1). Concerning position P4′, a negative influence of Cys and to a lesser extent Glu and His emerged (Supplemental Fig. S1B). The set of fastidious cleavage substrates with Val and Thr at position P1′ was next investigated. The very small size and the composition of the sample analyzed with Val and Thr (only 91 cases covering 18 different residues) resulted in less conclusive data than for position P2′ (Supplemental Fig. S1AB). Nevertheless the negative impact of Cys, Glu, His, and Pro at P4′ appeared clearly (Supplemental Fig. S1C).

A

B

FIG. 6. **Comparison of the influence of the nature of the P3′ and P4′ positions on Met cleavage *in vitro*.** *A*, influence of the P3′ position *in vitro*. The data were obtained with the Met-Gly-Met-*X* peptide where *X* is any natural amino acid. The relative catalytic efficiency (k_{cat}/K_m) is plotted as a function of *X*. A value of 100 was assigned to *X* = Gly. *B*, influence of the P4′ position *in vitro*. The data were obtained with the Met-Gly-Gly-Gly-*X* peptide where *X* is any natural amino acid. The relative catalytic efficiency (k_{cat}/K_m) is plotted as a function of *X*. A value of 100 was assigned to *X* = Ser. *Arrows* indicate the amino acids leading to significant decrease of cleavage efficiency in the series.
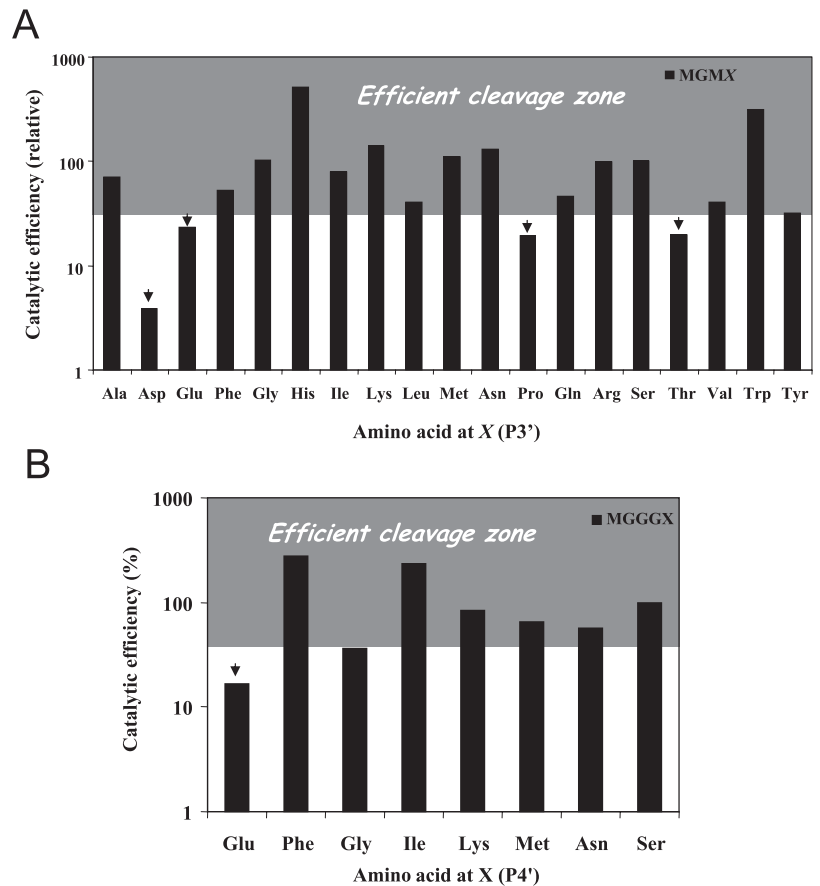
Cleavage Specificity Is Very Similar for Type-1 and Type-2 MAPs—We extended our analysis to type-2 MAP, performing the same analysis with *P. furiosus* MAP (PfMAP2), an archaeal enzyme. Little is known about the cleavage capacity of Pf-MAP2. Cleavage data are reported in the Takara catalog (catalog number 7335), but no detailed protocol is described. We extended our study aiming to confirm the validity of our peptide array by studying PfMAP2-mediated cleavage in detail. This enzyme was characterized under the same conditions as EcMAP1. A catalytic efficiency of 44,275 ± 10,000 $M^{-1} \cdot s^{-1}$, associated with $K_m = 0.9 \pm 0.2$ mM and $k_{cat} = 38 \pm 3$ s^{-1}, was measured for this enzyme with Met-Gly-Met-Met, the most efficiently processed substrate studied. This catalytic efficiency is similar to that reported for EcMAP1 (see above).

We then studied a set of substrates with various residues in the P1′ and P2′ positions (the complete data are reported in Supplemental Table S2 and displayed in Fig. 7). For P1′, the data obtained were similar to those obtained with EcMAP1 (Fig. 7A) and those available from Takara with peptide series M*X*AAA (where *X* is any natural amino acid except Cys). Thus, although MAP1 and MAP2 differ in terms of their amino acid sequences, the S1′ binding pockets of these two types of enzyme have similar sieve capacities. Cleavage was optimal for Ala and Pro and minimal for Thr and Val. The enzyme

showed a strong preference for large hydrophobic side chains, such as those of Met and Trp, in the P2′ position with strong hindrance observed for residues such as Glu and Pro and, to a lesser extent, Thr, Ile, and Asp (Fig. 7B). Some differences were observed between MAP1 and MAP2. For example, Ser and, to a lesser extent, Lys in P2′ were less optimal for MAP2 cleavage than for MAP1 cleavage, but the presence of these residues is unlikely to have a major effect on cleavage efficiency.

Thus, the substrate specificities of PfMAP2 and EcMAP1 seem to be similar, including for instance difficult cleavage if Thr or Val is at P1′ and if Glu or Pro is at P2′. This similarity was unexpected given the differences in the sequences of the two enzymes. To give access to these data for the whole community, we created a prediction tool, TermiNator2. This predictor was created as a web tool by combining php/HTML/MySQL/javascript languages to predict NME in prokaryotes (*i.e.* both eubacteria and Archaea) and other associated N-terminal modifications in eukaryotes. It is available on line at www.isv.cnrs-gif.fr/terminator2/index.html. This tool allows also complete proteomes to be analyzed (see below).

The Predicted N-terminal Proteome Shows a Strong Bias toward MAP Substrates with Respect to the Proteome Deduced from the Most Abundant Proteins—The *E. coli* data sets studied included both *in vivo* and *in vitro* studies, making
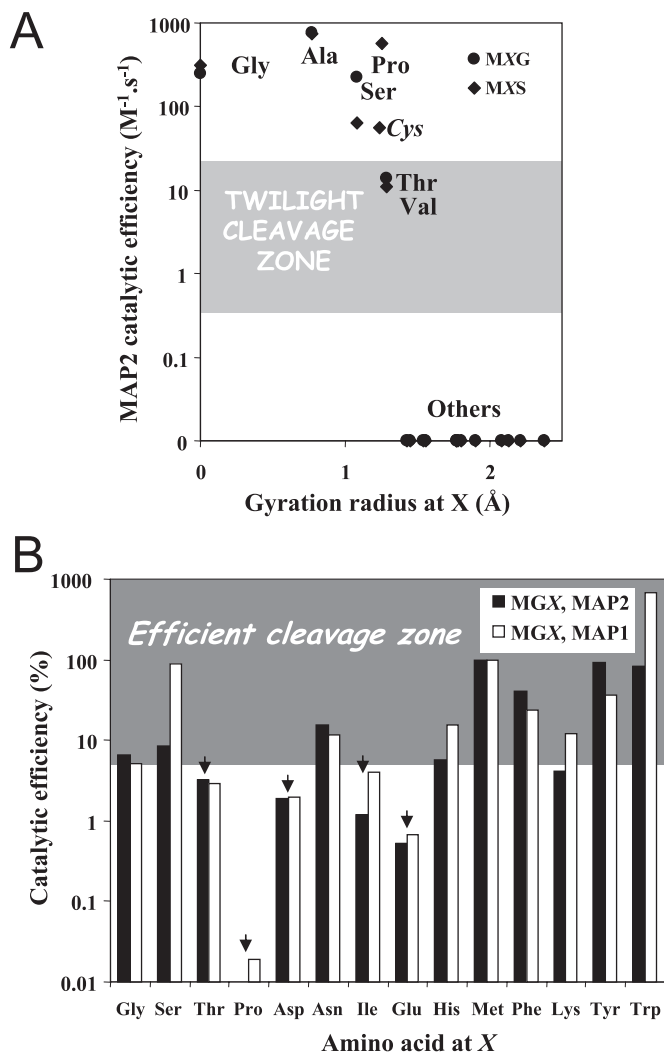
FIG. 7. **Impact of P1′ and P2′ positions on MAP2 cleavage efficiency *in vitro*.** *A*, influence of P1′ *in vivo* and correlation with side-chain radius gyration (19). Two series were tested, Met-*X*-Gly and Met-*X*-Ser, where *X* is any amino acid. "*Oth*" corresponds to all natural amino acids not indicated in the figure. *B*, influence of the P2′ position *in vitro* (*black bars*) and comparison with EcMAP (*white bars*). The data were obtained from the Met-Gly-*X* series where *X* is any natural amino acid. The relative catalytic efficiency (k_{cat}/K_m) is plotted as a function of *X*. A value of 100 was assigned to *X* = Met in both series. *Arrows* indicate the amino acids leading to significant decrease of cleavage efficiency in the series.

this a unique system for further functional analysis of the impact of the N-terminal residue. In a previous analysis of the *E. coli* NME machinery, we showed that MAP was the primary determinant of the nature of the N terminus of a protein and that the most sensitive position for EcMAP1 action was position P1′ (Fig. 3). We investigated the frequency and nature of particular N-terminal residues in *E. coli*. We extracted all proteins with the same penultimate (P1′) residue from the complete proteome of *E. coli* (4071 entries) and calculated the percentage of proteins with each of the amino acids in the penultimate position. We compared this data set with that

calculated only for proteins with N-terminal sequences that had been determined and shown to follow the rules of NME (738 entries; Fig. 8*A*). All residues predicted to undergo NME, except Val, the least efficiently processed of these substrates, were underrepresented by a factor of 2–3 in the complete proteome with respect to the restricted proteome. These data suggest that the accumulation of proteins sensitive to NME is favored in bacteria. Using the MAP cleavage prediction tool, TermiNator2, and the complete proteomes of *E. coli* (representative of Gram-negative bacteria), *B. subtilis* (representative of Gram-positive bacteria), and *P. furiosus* (representative of Archaea), we predicted the N-terminal amino acid distribution. In each of these cell types, the N-terminal amino acid distribution was very similar, featuring Met as the major N-terminal residue (Fig. 8*B*). Residues appearing as a result of NME corresponded to less than 30% of the proteome. These data indicate that (i) *E. coli* is a good model for NME studies in prokaryotes and that (ii) NME involves only a minority of the proteins in a given proteome in contrast to the general belief.

In addition, we investigated whether these findings were biased by the lack of consideration of leader peptide removal (LPR). LPR may play an important role, in addition to MAP cleavage, in influencing the exposure of specific and functionally important N-terminal residues. Another possible bias is the dependence of NME on the residue in the P1′ position and also on the residues in the P2′ and P3′ positions (Fig. 3*B*). Therefore, a second series of predictive analyses were performed based on the 4071 entries of the compete proteome of *E. coli*, taking into account both (i) the extent of LPR and (ii) Met cleavage efficiency (data shown in Fig. 3*B*). Statistical analysis of the restricted proteome (862 entries) was used (i) to predict the extent of LPR (14.4%) in a robust manner and (ii) to determine the nature and probability of the appearance of a new N-terminal amino acid as a result of LPR (*i.e.* Ala ≫ Asp, Glu, Lys > Val > Gln, His, Tyr; values indicated in the legend to Fig. 8). Our data are consistent with published results (47), which reveal a clear bias in cleavage site selection and in the nature of the amino acid unmasked at the N terminus. We used these data and the associated probabilities to calculate the mean percentage of mature proteins from the complete proteome with a given amino acid at their N terminus (Fig. 8*C*, *black bars*). As many as 60% of the proteins considered were predicted to retain their N-terminal Met residue. These data were compared with experimental data obtained by Edman N-terminal sequence determination of crude *E. coli* extracts (*white bars*; average of data from Refs. 8–10 and 23). The experimental data set differs from the calculated data in that only 35 ± 5% of the proteins were found to retain their N-terminal Met residue. This value is half that obtained theoretically. In contrast, both Ala and Ser appeared to be overrepresented as N-terminal residues in the experimental data sets. These data indicate that taking into account LPR is not sufficient for significant modification of the data shown in Fig. 8, *A* and *B*, and that there is indeed (i) a strong bias
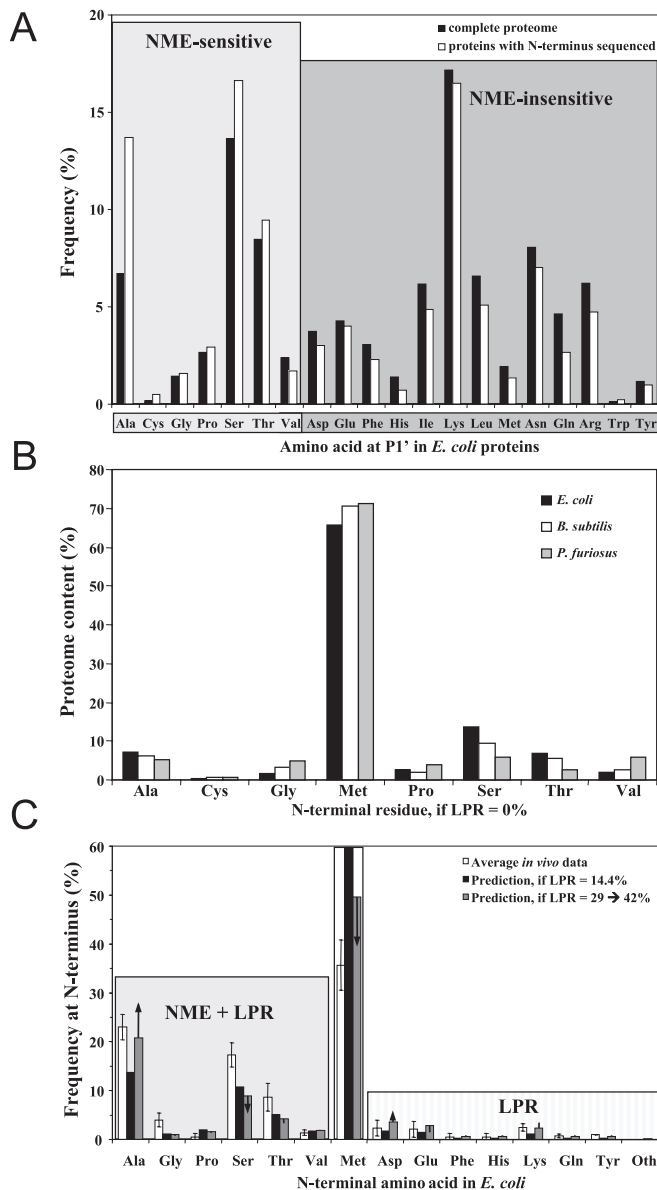
FIG. 8. **Existence of a strong positive bias between proteins with an N terminus identical to that revealed by NME and proteins accumulating *in vivo*.** *A*, comparison between the penultimate (P1′) residues predicted in the 4071 ORFs annotated in the *E. coli* genome (*black bars*) and those of the 862-protein set for which the N-terminal residue has been determined by sequencing (*white bars*). The proteins undergoing LPR (14.4%) in this second set were excluded from the analysis. *B*, amino acids unmasked by NME in various prokaryotic proteomes. The TermiNator2 prediction tool was used to analyze the three complete proteomes from *E. coli*, *B. subtilis*, and *P. furiosus*. LPR was not considered and assumed to be 0 (see *C*). *C*, comparison of a model of N-terminal protein maturation *in vivo*, predicting the nature of the N-terminal residue (calculation involved the 4071 ORFs; *black bars*), with the average of N termini (*white bars*) determined from two-dimensional gel electrophoresis (data from Ref. 23) and crude extracts (data from Refs. 8–10). The bioinformatic model took into account the efficiency of NME *in vivo* as a function of the nature of the residue in the P1′ position, as shown in Fig. 3*B*, and the LPR of 14.4% of the proteins. In this subset (data available from Ref. 23), we assumed that 57% of the proteins undergoing LPR eventually start with

toward the *in vivo* accumulation of proteins losing their N-terminal Met and/or (ii) a negative bias against proteins resistant to NME.

DISCUSSION

We report here the first extensive, coherent, and comprehensive analysis of the substrate specificity of MAP *in vitro*. Our data indicate that the P1′, P2′, and P3′ positions of the protein substrate (*i.e.* the three residues following the first Met (P1 position)) have a major effect on cleavage efficiency. Our data were obtained with MAPs of both types: type-1, exemplified by *E. coli* MAP, and type-2, exemplified by *P. furiosus* MAP. The *E. coli* system, with a large number of available protein sequences, has provided unique information, making it possible to demonstrate a direct correlation between MAP activity *in vitro* and NME *in vivo*. Together these data indicate that the assessment of MAP catalytic efficiency *in vitro* is the best way to assess MAP cleavage efficiency *in vivo*. Based on our data, we also developed a new bioinformatics tool for improving Met cleavage prediction for prokaryotic, eubacterial and archaeal, MAPs that is available from www.isv.cnrs-gif.fr/terminator2/index.html. The prediction of Met cleavage for organellar and nucleus-encoded eukaryotic proteins is also possible via the same site using data for the prokaryotic enzyme (this work and Ref. 48).

NME Prediction of Natural Bacterial Substrates and Recombinant Proteins Overproduced in E. coli—The nature of position P1′ in a polypeptide chain affects cleavage efficiency, depending on the nature of the amino acid, from highly efficient cleavage, as observed with Ala, to fastidious cleavage, as seen with Thr and Val. MAP k_{cat}/K_m (catalytic efficiency) and *in vivo* cleavage efficiency were found to be directly correlated (Fig. 3*C*). This explains why most proteins resistant to NME according to the usual rules of thumb have Val or Thr in the penultimate residue. Positions P2′ and, to a lesser extent, P3′ and P4′ also affected cleavage probability with the same series of amino acids, Asp, Glu, Pro, and Thr, having a negative effect. A Glu or a Pro at P2′ appeared to have the most negative effect. For the prediction of MAP cleavage *in vivo* directly from analysis of the N terminus of a protein, it is important to take into account whether the protein is (i) a natural protein substrate or (ii) an overproduced recombinant protein. For natural substrates, we suggest that proteins beginning with a <M(^ACGPSTV) or <M(V/T)(E/P) (note that "(^*XY*)" means that residues *X* and *Y* or a subset list is ex-

Ala, 12.2% with Asp, 10.2% with Glu, 8.2% with Lys, 4.1% with Val, and 2% with each of Gln, His, Phe, and Tyr, respectively. "*Oth*" corresponds to the sum of all the other amino acids (0.1%). This assessment was highly consistent with the data set shown in Fig. 1 (for Gram-negative bacteria) of Ref. 47. Data were calculated from the same set but with a probability of LPR of 29% (*gray bars*, tendency up to 42% is schematized with a *black arrow* on the *bar*) rather than 14.4% (*black bars*).

cluded, and "(*Y/Z*)" means that residue *Y* or *Z* or a subset list is included; *X*, *Y*, or *Z* correspond to any normal amino acid) are unlikely to be cleaved. There is a strong risk of lack of cleavage for proteins beginning with <M(G/P/V/T)(D/T)(D/E/P/T), <M(G/P/V/T)(D/T)(^DEPT)(C/E/H/P), and <M(G/P/V/T)-(^DEPT)(D/E/P/T)(C/E/H/P). In this last situation, incomplete maturation *in vivo* is most likely to occur, resulting in two types of the protein: one, major with an N-terminal Met, and the other, minor without an N-terminal Met. There are already a few well documented examples of such behavior (23). These simple rules account for 60% of the proteins displaying as yet unexplained Met retention in the proteome of *E. coli*.

E. coli is widely used for the overproduction of recombinant proteins. In the last 20 years, many instances of improperly processed N termini have been reported. This recurrent issue, known as the "methionine problem" (49), has not yet been resolved. MAP cleavage is thought to become limiting because the amount of overproduced protein exceeds the capacity for MAP activity. Our findings are entirely consistent with MAP activity being the limiting factor for NME *in vivo* (Fig. 3*C*). They also suggest that the conclusions drawn from studies based on overproduced reporter proteins (17, 18) should be interpreted with caution. Unexpected complete or partial Met retention is known to lead to many defects, including an increase in antigenicity or a decrease in biological activity. This is true for the N-terminal nucleophile hydrolase (Ntn) family of enzymes, which use both the free N terminus and the corresponding Thr, Ser, and Cys side chains in the catalytic mechanism (50) as reviewed in Ref. 51. Before envisaging the production of a protein in *E. coli*, particularly for heterologous proteins, the NME behavior of the protein must be correctly predicted to ensure that the expected protein, with the expected N terminus, is produced. The data reported herein is particularly useful in this respect as illustrated by the following examples concerning the production of human proteins of therapeutic interest by recombinant DNA methods in *E. coli*.

"Unexpected" blockage of Met cleavage has been reported for granulocyte colony-stimulating factor (Met-*Thr-Pro*-Leu; see Ref. 52) and α- or β-hemoglobins (Met-*Val*-His-Leu; see Ref. 53). In all these cases, the overproduced protein contains a fastidious P1′ (Val or Thr) residue as revealed in our study (Fig. 3), and one of these proteins also contains the most negative P2′ residue, Pro. Other well documented examples of Met retention include the cytokine RANTES (regulated on activation normal T cell expressed and secreted) (Met-Ser-*Pro*-Tyr) and the interleukins 1β and 2 (Met-Ala-*Pro-Thr/V*al; see Refs. 54 and 55). For the interleukins, Met cleavage is crucial as the N-terminal Ala must interact with the C-terminal Thr-133 (56). Our study provides a rationale for the retention of Met as both Pro and Thr at positions P2′ and P3′ tend to inhibit Met cleavage. Another example is provided by interleukin-6, which begins with Met-Pro-Val-*Pro* (57). We show here that a Pro at position P3′ significantly decreases the efficiency of the process. We conclude from the aforemen-

tioned examples that, when a given protein is overproduced, substrates displaying fastidious cleavage by MAP may not be completely processed by this enzyme. The prediction of MAP cleavage must therefore be more stringent than for intrinsic proteins. We suggest that processing is highly unlikely to occur for overproduced proteins beginning with <M(A/G/P/S/T)(D/E/P/T)(D/E/P/T), <M(A/G/P/S/T)(^DEPT)*X*(C/E/H/P), <M(A/G/P/S/T)*X*(^DEPT)(C/E/H/P), or <MV(D/T) and that there is a risk of non-processing for proteins beginning with <M(A/G/P/S/T)(D/E/P/T) or <MV(^DEPT), leading to partial Met retention. This rule should be useful for predicting the likelihood of cleavage for recombinant proteins. To this aim, we therefore added an option relative to the overexpression of any favorite protein in the TermiNator2 prediction tool. If the protein has an N terminus associated with a risk of non-processing, various methods may be used to prevent Met retention: (i) adjusting expression conditions to optimize MAP activity according to the rate of protein synthesis *in vivo* (58), (ii) *in vitro* processing with purified wild-type or engineered MAP (59–62), (iii) dual co-overproduction of MAP and the protein of interest *in vivo* (25, 63, 64), (iv) processing *in vitro* or *in vivo* of the N terminus with another amino- or endoprotease (65–67), or (v) fusion of the ORF to a propeptide or ubiquitin sequence and use of endogenous signal peptidase specificity (68, 69). This last strategy has also proved successful for residues that cannot be unmasked by NME (this study, Fig. 3), such as the thrombin inhibitor hirudin, which must have Ile at its N terminus for biological activity, or human growth hormone, which must have Phe at its N terminus (70–73).

Comparison of NME Rules between Eubacteria and Archaea—In Archaea, most mature proteins have Ala, Met, Gly, Ser, Thr, or Pro at their N terminus (10). The N-terminal Met of proteins is believed to be frequently cleaved by MAP2 in Archaea, although few data are available to confirm this. This assumption is consistent with the *in vitro* data presented here (Fig. 7*A*). Specifically we have demonstrated that the rules governing cleavage are similar for MAP2 and MAP1, involving not only the P1′ site but also the true P2′ and P3′ sites (Fig. 7*B*).

Archaea display 8–9% signal peptide cleavage, a frequency significantly lower than that in eubacteria (15–37%; Ref. 74). NME is therefore of greater importance for exposing the N termini of most proteins. The consequences of the observed differences in MAP substrate specificity for Met cleavage *in vivo* should be taken into account. In Archaea, the rules governing NME do not differ significantly from those of eubacteria with similar trends observed at P1′ and P2′. We conclude that the use of identical cleavage rules should make it possible to predict NME accurately in Archaea.

Physiological Relevance of Improved NME Prediction for the E. coli Proteome—By improving NME prediction for natural proteins of the bacterial proteome, we were able to model the effect of this process on the whole proteome (Fig. 8). We observed an overrepresentation of proteins undergoing NME in bacteria: those that resisted the process and had se-

quences starting with a Met were underrepresented (Fig. 8*C*).

We tried to account for this discrepancy by modeling the impact of inadequate LPR assessment. The frequency of LPR used (14.4%) was directly deduced from data obtained from the 862-protein database. Interestingly this value is identical to that obtained from bioinformatics analysis using a maximal stringency cutoff applied to the complete proteome of *Haemophilus influenzae* (14% in Ref. 47). *H. influenzae* is a Gram-negative bacterium closely related to *E. coli* but with a significantly lower (by a factor of 2.5) number of associated ORFs (75). Doolittle (75) estimated that the signal peptide cleavage frequency might reach 15–20%, which is slightly higher than the frequency used here. A maximal value of 28% was proposed with lower stringency scoring. This value is consistent with other predictions of 25% in *H. influenzae* and 29% in *E. coli* (74). A higher frequency of signal peptide cleavage might therefore account for the bias observed in the distribution of N-terminal amino acids. It should be borne in mind that the proteins analyzed correspond to a class of proteins with a compartmentation bias as these proteins are (i) soluble (as opposed to membrane proteins) and (ii) not secreted outside the bacterium as the only secreted proteins that can be retrieved by the fractionation strategies used are those exported to the periplasmic space. LPR is thought to be frequent in membrane and secreted proteins. We therefore modeled the impact of increasing LPR frequency. Only a huge increase in the proportion of these proteins to 42% (3 times experimental values) gave a satisfactory fit of N-terminal Met frequencies to the experimental values (Fig. 8*C*, *end* of *arrows* in *gray bars*). However, the correlation with Ala, Ser, and Thr was very weak with this value. Finally this value of 42% is much higher than the upper value of 29% proposed previously (74). A value of 29% gave a much better fit to the data (Fig. 8*C*, *gray bars*). We conclude that the protein pool undergoing signal peptide cleavage is probably underestimated in our study but that this underestimation alone is unlikely to account for the trend observed toward a lower content of proteins retaining their N-terminal Met.

Most of the data for the 862-sequence compilation concerned proteins extracted by two-dimensional gel electrophoresis (23). As a result of this, these proteins were probably the most abundant proteins in the bacterium. Another possible reason for the poor correlation between the data sets in Fig. 8*C* is likely due to the assumption that the genes for these abundant proteins may have been selected to encode residues such as Ala or Ser at P1′ so that their N-terminal Met is efficiently cleaved and recycled. Met is indeed an "expensive" amino acid to biosynthesize. The capacity to recover a majority of the N-terminal Met (*i.e.* especially those corresponding to the most abundant proteins) could have been favored throughout bacterial evolution (see also concluding remarks in Ref. 21 for further discussion). One could also suggest that all or some of the proteins retaining their N-terminal Met are likely to be less stable. Recent data obtained in the search for proteins cleaved by the ClpAXP proteolytic machinery of *E. coli* have indeed identified an N-terminal motif including an N-terminal Met. The recurring N-terminal motif $<$MK$\Phi\Phi X\Phi$ (where Φ is any hydrophobic amino acid and X is any amino acid) is involved in cleavage by the bacterial ClpAXP protease (76). This hypothesis seems likely given that Lys is the most frequent amino acid in position 2 (Fig. 8*A*) and prevents NME. However, a less stringent motif would be required to obtain a better fit to the experimental values. It should be noted that a single proteolytic cleavage of a mature protein by cellular machinery other than a leader peptidase and ClpAXP may lead to enhanced protein degradation. This hypothesis, in which protein half-life and signal peptide prediction accounts for the amino acid bias observed in experimentally determined protein N termini with respect to proteomic predictions, may also account for the accumulation of fewer proteins with an N-Met than predicted on the basis of the genome sequence (Fig. 8*B*).

Generalization, Introducing a Bioinformatics Tool to Predict NME—We show here that detailed characterization of the catalytic efficiency of MAP can be used to predict the *in vivo* Met cleavage of any protein of a given proteome. NME prediction within the proteome is indeed crucial as a database of protein N termini as representative (*i.e.* about 20%) as that available for *E. coli* will probably never be available for any other organism. However, published data for the two cytosolic *Saccharomyces cerevisiae* MAPs (ScMAPs) suggest that the substrate specificity of these enzymes may differ from that of prokaryotic MAPs. For example, studies of ScMAP1 have shown fastidious processing of peptides with Pro at P1′ or with Val or Thr in this position (Fig. 4A in Ref. 40). Studies with ScMAP2 and low concentrations of peptide (*i.e.* mimicking k_{cat}/K_m conditions) have suggested fastidious processing for peptides with Pro at P1′ but not with Val in this position (Table V in Ref. 77). Thus, in higher eukaryotes, prediction is complicated by the occurrence of at least one MAP1 and one MAP2, which do not necessarily have identical specificities. If the two enzymes have similar specific activities *in vivo*, cleavage efficiency is likely to be increased. However, this does not appear to be the case in yeast in which MAP1 behaves as the most active, major MAP (78). Data on relative MAP expression and regulation are therefore required to complement prediction algorithms.

The associated crucial physiological consequences, such as protein stabilization associated with Met removal, highlight the importance of correctly predicting the N termini of proteins as recently reported for higher eukaryotes (48). Further detailed kinetic characterization of the actual substrate specificity of each eukaryotic MAP is required to increase the confidence of predictions made with this tool.

REFERENCES

1. Bradshaw, R. A., Brickey, W. W., and Walker, K. W. (1998) N-terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends Biochem. Sci.* **23,** 263–267
2. Giglione, C., Boularot, A., and Meinnel, T. (2004) Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61,** 1455–1474
3. Lowther, W. T., and Matthews, B. W. (2000) Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta* **1477,** 157–167
4. Addlagatta, A., Hu, X., Liu, J. O., and Matthews, B. W. (2005) Structural basis for the functional differences between type I and type II human methionine aminopeptidases. *Biochemistry* **44,** 14741–14749
5. Giglione, C., Pierre, M., and Meinnel, T. (2000) Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. *Mol. Microbiol.* **36,** 1197–1205
6. Solbiati, J., Chapman-Smith, A., Miller, J. L., Miller, C. G., and Cronan, J. E., Jr. (1999) Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. *J. Mol. Biol.* **290,** 607–614
7. Giglione, C., and Meinnel, T. (2001) Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci.* **6,** 566–572
8. Waller, J.-P. (1963) The NH₂-terminal residue of the proteins from cell-free extract of *E. coli. J. Mol. Biol.* **7,** 483–496
9. Brown, J. L. (1970) The N-terminal region of soluble proteins from procaryotes and eucaryotes. *Biochim. Biophys. Acta* **221,** 480–488
10. Matheson, A. T., Yaguchi, M., and Visentin, L. P. (1975) The conservation of amino acids in the N-terminal position of ribosomal and cytosol proteins from *Escherichia coli*, *Bacillus stearothermophilus*, and *Halobacterium cutirubrum. Can. J. Biochem.* **53,** 1323–1327
11. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162
12. Miller, C. G. (1975) Peptidases and proteases of *Escherichia coli* and *Salmonella typhimurium. Annu. Rev. Microbiol.* **29,** 485–504
13. Miller, C. G. (1987) Protein degradation and proteolytic modification, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E., eds) pp. 680–691, American Society for Microbiology, Washington, D. C.
14. Flinta, C., Persson, B., Jornvall, H., and von Heijne, G. (1986) Sequence determinants of cytosolic N-terminal protein processing. *Eur. J. Biochem.* **154,** 193–196
15. Miller, C. G., Strauch, K. L., Kukral, A. M., Miller, J. L., Wingfield, P. T., Mazzei, G. J., Werlen, R. C., Graber, P., and Movva, N. R. (1987) N-terminal methionine-specific peptidase in *Salmonella typhimurium. Proc. Natl. Acad. Sci. U. S. A.* **84,** 2718–2722
16. Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A., and Chang, S. (1987) Processing of the initiation methionine from proteins: Properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.* **169,** 751–757
17. Dalboge, H., Bayne, S., and Pedersen, J. (1990) *In vivo* processing of N-terminal methionine in *E. coli. FEBS Lett.* **266,** 1–3
18. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. (1989) Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8247–8251
19. Levitt, M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104,** 59–107
20. Sherman, F., Stewart, J. W., and Tsunasawa, S. (1985) Methionine or not methionine at the beginning of a protein. *BioEssays* **3,** 27–31
21. Meinnel, T., Mechulam, Y., and Blanquet, S. (1993) Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli. Biochimie* (*Paris*) **75,** 1061–1075
22. Demirev, P. A., Lin, J. S., Pineda, F. J., and Fenselaut, C. (2001) Bioinformatics and mass spectrometry for microorganism identification: proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori. Anal. Chem.* **73,** 4566–4573
23. Link, A. J., Robison, K., and Church, G. M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* **18,** 1259–1313
24. Wasinger, V. C., and Humphery-Smith, I. (1998) Small genes/gene-products in *Escherichia coli* K-12. *FEMS Microbiol. Lett.* **169,** 375–382
25. Ben-Bassat, A., and Bauer, K. (1987) Amino-terminal processing of proteins. *Nature* **326,** 315
26. Olson, K. C., Fenno, J., Lin, N., Harkins, R. N., Snider, C., Kohr, W. H., Ross, M. J., Fodge, D., Prender, G., and Stebbing, N. (1981) Purified human growth hormone from *E. coli* is biologically active. *Nature* **293,** 408–411
27. Boufous, H., and Vadeboncoeur, C. (2003) Purification and characterization of the *Streptococcus salivarius* methionine aminopeptidase (MetAP). *Biochimie* (*Paris*) **85,** 993–997
28. Ragusa, S., Blanquet, S., and Meinnel, T. (1998) Control of peptide deformylase activity by metal cations. *J. Mol. Biol.* **280,** 515–523
29. Mayaux, J. F., Soubrier, F., and Latta, M. (1987) Cloning *E. coli* genes by oligonucleotide hybridization. *Nucleic Acids Res.* **15,** 10593–10594
30. Meng, L., Ruebush, S., D'Souza, V. M., Copik, A. J., Tsunasawa, S., and Holz, R. C. (2002) Overexpression and divalent metal binding properties of the methionyl aminopeptidase from *Pyrococcus furiosus. Biochemistry* **41,** 7199–7208
31. Massey, V., and Curti, B. (1967) On the reaction mechanism of *Crotalus adamanteus* L-amino acid oxidase. *J. Biol. Chem.* **242,** 1259–1264
32. Lewis, W. H., and Harris, H. (1967) Human red cell peptidases. *Nature* **215,** 351–355
33. Carter, T. H., and Miller, C. G. (1984) Aspartate-specific peptidases in *Salmonella typhimurium*: mutants deficient in peptidase E. *J. Bacteriol.* **159,** 453–459
34. Dardel, F. (1994) MC-Fit: using Monte-Carlo methods to get accurate confidence limits on enzyme parameters. *Comput. Appl. Biosci.* **10,** 273–275
35. Rudd, K. E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.* **28,** 60–64
36. Cockwell, K. Y., and Giles, I. G. (1989) Software tools for motif and pattern scanning: program descriptions including a universal sequence reading algorithm. *Comput. Appl. Biosci.* **5,** 227–232
37. Walker, K. W., Yi, E., and Bradshaw, R. A. (1999) Yeast (*Saccharomyces cerevisiae*) methionine aminopeptidase I: rapid purification and improved activity assay. *Biotechnol. Appl. Biochem.* **29,** 157–163
38. Zhou, Y., Guo, X. C., Yi, T., Yoshimoto, T., and Pei, D. (2000) Two continuous spectrophotometric assays for methionine aminopeptidase. *Anal. Biochem.* **280,** 159–165
39. Larrabee, J. A., Thamrong-Nawasawat, T., and Mon, S. Y. (1999) High-pressure liquid chromatographic method for the assay of methionine aminopeptidase activity: application to the study of enzymatic inactivation. *Anal. Biochem.* **269,** 194–198
40. Walker, K. W., and Bradshaw, R. A. (1999) Yeast methionine aminopeptidase I. Alteration of substrate specificity by site-directed mutagenesis. *J. Biol. Chem.* **274,** 13403–13409
41. Lowther, W. T., Orville, A. M., Madden, D. T., Lim, S., Rich, D. H., and Matthews, B. W. (1999) *Escherichia coli* methionine aminopeptidase: implications of crystallographic analyses of the native, mutant, and inhibited enzymes for the mechanism of catalysis. *Biochemistry* **38,** 7678–7688
42. Schurmann, M., and Sprenger, G. A. (2001) Fructose-6-phosphate aldolase is a novel class I aldolase from *Escherichia coli* and is related to a novel group of bacterial transaldolases. *J. Biol. Chem.* **276,** 11055–11061
43. Zavitz, K. H., DiGate, R. J., and Marians, K. J. (1991) The *priB* and *priC* replication proteins of *Escherichia coli*. Genes, DNA sequence, overexpression, and purification. *J. Biol. Chem.* **266,** 13988–13995
44. Poulsen, P., Jensen, K. F., Valentin-Hansen, P., Carlsson, P., and Lundberg, L. G. (1983) Nucleotide sequence of the *Escherichia coli pyrE* gene and of the DNA in front of the protein-coding region. *Eur. J. Biochem.*

     135, 223–229

45. Sheff, D. R., and Rubenstein, P. A. (1992) Isolation and characterization of the rat liver actin N-acetylaminopeptidase. *J. Biol. Chem.* **267,** 20217–20224

46. Polevoda, B., and Sherman, F. (2000) Nα-terminal acetylation of eukaryotic proteins. *J. Biol. Chem.* **275,** 36479–36482

47. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6

48. Meinnel, T., Peynot, P., and Giglione, C. (2005) Processed N-termini of mature proteins in higher eukaryotes and their major contribution to dynamic proteomics. *Biochimie (Paris)* **87,** 701–712

49. Bialy, H. (1989) The methionine problem solved, ubiquitously. *Bio/Technology* **7,** 649

50. Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C., Smith, J. L., Tomchick, D. R., and Murzin, A. G. (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378,** 416–419

51. Meinnel, T., Serero, A., and Giglione, C. (2006) Impact of the N-terminal amino acid on targeted protein degradation. *Biol. Chem.* **87,** 839–851

52. Devlin, P. E., Drummond, R. J., Toy, P., Mark, D. F., Watt, K. W., and Devlin, J. J. (1988) Alteration of amino-terminal codons of human granulocyte-colony-stimulating factor increases expression levels and allows efficient processing by methionine aminopeptidase in *Escherichia coli. Gene (Amst.)* **65,** 13–22

53. Shen, T. J., Ho, N. T., Simplaceanu, V., Zou, M., Green, B. N., Tam, M. F., and Ho, C. (1993) Production of unmodified human adult hemoglobin in *Escherichia coli. Proc. Natl. Acad. Sci. U. S. A.* **90,** 8108–8112

54. Liang, S. M., Allet, B., Rose, K., Hirschi, M., Liang, C. M., and Thatcher, D. R. (1985) Characterization of human interleukin 2 derived from *Escherichia coli. Biochem. J.* **229,** 429–439

55. Kikumoto, Y., Hong, Y. M., Nishida, T., Nakai, S., Masui, Y., and Hirai, Y. (1987) Purification and characterization of recombinant human interleukin-1 beta produced in *Escherichia coli. Biochem. Biophys. Res. Commun.* **147,** 315–321

56. Endo, S., Yamamoto, Y., Sugawara, T., Nishimura, O., and Fujino, M. (2001) The additional methionine residue at the N-terminus of bacterially expressed human interleukin-2 affects the interaction between the N- and C-termini. *Biochemistry* **40,** 914–919

57. Yasueda, H., Nagase, K., Hosoda, A., Akiyama, Y., and Yamada, K. (1990) High-level direct expression of semi-synthetic human interleukin-6 in *Escherichia coli* and production of N-terminus met-free product. *Biotechnology (N. Y.)* **8,** 1036–1040

58. Covalt, J. C., Jr., Cao, T. B., Magdaroag, J. R., Gross, L. A., and Jennings, P. A. (2005) Temperature, media, and point of induction affect the N-terminal processing of interleukin-1β. *Protein Expr. Purif.* **41,** 45–52

59. Nakagawa, S., Yamada, T., Kato, K., and Nishimura, O. (1987) Enzymatic cleavage of amino terminal methionine from recombinant human interleukin 2 and growth hormone by aminopeptidase M. *Bio/Technology* **5,** 824–827

60. Ben-Bassat, A., Bauer, K., Chang, S., and Chang, S.-Y. (May 7, 1991) Bacterial methionine N-terminal peptidase. U. S. Patent 5,013,662

61. Liao, Y. D., Jeng, J. C., Wang, C. F., Wang, S. C., and Chang, S. T. (2004) Removal of N-terminal methionine from recombinant proteins by engineered *E. coli* methionine aminopeptidase. *Protein Sci.* **13,** 1802–1810

62. Liu, L. F., Hsieh, C. H., Liu, P. F., Tsai, S. P., and Tam, M. F. (2004) Escherichia coli methionine aminopeptidase with Tyr168 to alanine substitution can improve the N-terminal processing of recombinant proteins with valine at the penultimate position. *Anal. Biochem.* **329,** 345–347

63. Shen, T. J., Ho, N. T., Zou, M., Sun, D. P., Cottam, P. F., Simplaceanu, V., Tam, M. F., Bell, D. A., Jr., and Ho, C. (1997) Production of human normal adult and fetal hemoglobins in *Escherichia coli. Protein Eng.* **10,** 1085–1097

64. Hwang, D. D., Liu, L. F., Kuan, I. C., Lin, L. Y., Tam, T. C., and Tam, M. F. (1999) Co-expression of glutathione S-transferase with methionine aminopeptidase: a system of producing enriched N-terminal processed proteins in *Escherichia coli. Biochem. J.* **338,** 335–342

65. Yasueda, H., Kikuchi, Y., Kojima, H., and Nagase, K. (1991) In-vivo processing of the initiator methionine from recombinant methionyl human interleukin-6 synthesized in *Escherichia coli* overproducing aminopeptidase-P. *Appl. Microbiol. Biotechnol.* **36,** 211–215

66. Latta, M., Knapp, M., Sarmientos, P., Bréfort, G., Becquart, J., Guerrier, L., Jung, G., and Mayaux, J.-F. (1987) Synthesis and purification of mature human serum albumin from *E. coli. Bio/Technology* **5,** 1309–1314

67. Dalboge, H., Dahl, H.-H.M., Pedersen, J., Hansen, J., and Christensen, T. (1987) A novel enzymatic method for production of authentic hGH from an *Escherichia coli* produced hGH precursor. *Bio/Technology* **5,** 161–164

68. Miller, H. I., Henzel, W. J., Ridgway, J. B., Kuang, W. J., Chisholm, V., and Liu, C. C. (1989) Cloning and expression of a yeast ubiquitin-protein cleaving activity in *Escherichia coli. Bio/Technology* **7,** 698–704

69. Soares, C. R., Gomide, F. I., Ueda, E. K., and Bartolini, P. (2003) Periplasmic expression of human growth hormone via plasmid vectors containing the λPL promoter: use of HPLC for product quantification. *Protein Eng.* **16,** 1131–1138

70. Kohr, W. J., Keck, R., and Harkins, R. N. (1982) Characterization of intact and trypsin-digested biosynthetic human growth hormone by high-pressure liquid chromatography. *Anal. Biochem.* **122,** 348–359

71. Becker, G. W., and Hsiung, H. M. (1986) Expression, secretion and folding of human growth hormone in *Escherichia coli.* Purification and characterization. *FEBS Lett.* **204,** 145–150

72. Loison, G., Findelli, A., Bernard, S., Nguyen-Juilleret, M., Marquet, M., Riehl-Bellon, N., Carvallo, D., Guerra-Santos, L., Brown, S., Courtney, M., Roitsch, C., and Lemoine, Y. (1988) Expression and secretion of biologically active leech hirudin. *Bio/Technology* **6,** 72–77

73. Sabin, E. A., Lee-Ng, C. T., Shuster, J. R., and Barr, P. J. (1989) High-level expression and *in vivo* processing of chimeric ubiquitin fusion proteins in *Saccharomyces cerevisiae. Bio/Technology* **7,** 705–709

74. Saleh, M. T., Fillon, M., Brennan, P. J., and Belisle, J. T. (2001) Identification of putative exported/secreted proteins in prokaryotic proteomes. *Gene (Amst.)* **269,** 195–204

75. Doolittle, R. F. (1997) Microbial genomes opened up. *Nature* **392,** 339–342

76. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11,** 671–683

77. Kendall, R. L., and Bradshaw, R. A. (1992) Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the co-translational processing of proteins. *J. Biol. Chem.* **267,** 20667–20673

78. Chen, S., Vetro, J. A., and Chang, Y. H. (2002) The specificity *in vivo* of two distinct methionine aminopeptidases in *Saccharomyces cerevisiae. Arch. Biochem. Biophys.* **398,** 87–93

79. Fersht, A. R. (1998) *Structure and Mechanism in Protein Science*, pp. 103–129, W. H. Freeman & Co., New York