REPORT

Development of new methods to study the amino terminal peptide of proteins and their applications in the biotechnological industry

✉ Aniel Sánchez, Lázaro Betancourt, Luis J González, Yassel Ramos, Jeovanis Gil, Yanni Solano, Vivian Morera, Yairet García, Félix Álvarez, Galina Moya, Jorge Fernández de Cossio, Gabriel Padrón, Vladimir Besada

Physical-Chemistry Division. Center for Genetic Engineering and Biotechnology, CIGB
Ave. 31 e/ 158 y 190, AP 6162, CP 10600, Cubanacán, Playa, Havana, Cuba
E-mail:aniel.sanchez@cigb.edu.cu

Introduction

The fast development of genetic engineering and biotechnology has made possible to obtain proteins naturally expressed at so low concentrations that it makes hard to be isolated. The recombinant DNA technology allows inserting a gene coding for a protein of interest into the genome of a host microorganism by means of an appropriate genetic construct. Then, the recombinant protein is obtained in sufficient amounts by growing the host microorganism, and its subcellular location can be predetermined to implement a successful purification strategy.

However, during translation, expression, purification or storage processes, proteins can be chemically modified, such modifications altering their primary structure and their subsequent biological activities. Additionally, degradation processes mediated by exopeptidases can frequently modify their amino (N-term) and carboxyl ends.

One of the most common post-translational modifications is the blocking of the N-term group. It is estimated that over 80% of eukaryotic proteins are blocked in the N-term [1]. Such modification makes sequencing impossible by the Edman's method. Therefore, the international regulatory agencies request the information of that end on the proteins to control the quality of recombinant proteins.

Most of the methods alternatives to the Edman's degradation method are based on mass spectrometry (MS) as analytical tool to study the proteins primary structure. They use a combination of enzyme digestion and/or chemical modification of primary amino groups preceded by a chromatographic step.

Our work comprises three new methods to study the N-term peptide of proteins, whether blocked or not. One of them (named method 1) selectively isolates the N-term peptide, while the other two (methods 2 and 3, respectively) allow its identification in a MS spectrum, starting with the protein isolation from a gel band after fractioning and direct identification.

Method 1 is optimized for proteins in solution and useful to study N-term-blocked (N-blocked) proteins, easily isolating the N-blocked peptide in a fast and effectively by successive digestions and a cationic exchange step (Figure 1).

Nevertheless, this method is not so efficacious to study neither proteins expressed in low amounts nor to analyze proteins with free N-terms or in stability studies to identify degradation products. Thus, the other two methods (methods 2 and 3) were optimized to study proteins isolated from polyacrylamide gels, and therefore, in very low amounts. Both allow the identification of the N-term peptide among the other peptides in the MS spectrum following enzyme digestion, a significant difference with previously reported methods.

This identification is made by means of the ESI-MS spectrum analysis, identifying the N-term peptide according to the isotopic distribution. Method 2 makes the study of both N-blocked and N-free proteins possible, and together with method 3 they can be used to study protein degradation products, one of the main goals of stability studies (Figure 2).

Method 3 combines the identification of the N-term peptide in an ESI-MS spectrum with the ability to differentiate the fragmentation series in the ESI-MS spectrum to achieve an casier and more reliable sequencing which is particularly advantageous when proteins are derived from organisms with unknown genome.

Results and discussion

Method design

The design of the methods proposed is based on using simple chemical reactions, combined to the use of enzymes common in protein chemistry together with conventional chromatography.

The first method allows the selective isolation of N-blocked peptide by simultaneous digestion of the protein with the trypsin and carboxypeptidase B enzymes. The released internal peptides of the protein are retained in a strong cationic exchanger, but not N-blocked peptide (Figure 1).

The other two methods use modifications commonly used in the chemistry of primary amino groups, as acylation and guanidination. All the analyses are carried out by MS (Figures 2 and 3).

Optimization

The optimization of specific conditions, the different steps and reactions of the methods were initially established by using standards, such as casomorphin-7, neuromedkin K and equine myoglobin proteins.

Analysis of the N-term peptide in different protein samples

The methods designed were successfully applied to characterize and identify the N-terminal sequences in samples of different proteins. Method 1, for selective isolation of the N-term peptide in

1. Brown JL, Roberts WK. Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are Nalpha-acetylated. Biol Chem 1976; 251:1009-14.

✉ Author of correspondence

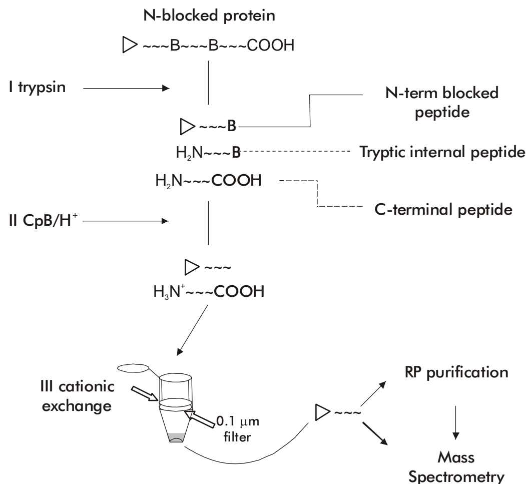

Figure 1. General diagram of the strategy proposed to identify the N-ter blocked peptide (method 1). The tryptic digestion is treated with carboxypeptidase B (CpB/H$^+$). After the cationic exchange, the N-term peptide is the only one not retained in the matrix. RP- reverse phase chromatography.

N-blocked proteins, was used to study the following samples [2].

Cytochrome C

Cytochrome C is a protein frequently used as a model to establish new selective isolation methods of N-blocked peptides. Its N-term peptide was successfully isolated.

Parvalbumin

The isolation of parvalbumin N-blocked peptide allowed us to establish and study the behavior of this type of peptides when analyzed by MALDI-MS. The peptide molecular mass was unequivocally assigned to the isolated peptide with spectra from different ionizing sources.

Human interferon α-2b

The partially blocked N-term peptide of the human interferon α-2b was successfully isolated. This result indicated the potentialities of the method to characterize proteins with heterogeneous N-terms.

Troponin C

The N-term acetylated peptide of troponin C was selectively isolated and identified, demonstrating that the proposed procedure can be effectively applied to detect species of blocked proteins from mixtures of proteins.

Levansucrase from *Gluconacetobacter diazotrophicus*

The N-term sequence of the mature Levansucrase from *Gluconacetobacter diazotrophicus* was determined, detecting a blocking pyroglutamic group that resulted from the hydrolysis of the signal peptide prior to the secretion of the protein to the culture media.

Hepatitis B surface antigen

The surface antigen of the hepatitis B virus is the active pharmaceutical ingredient of the vaccine de-

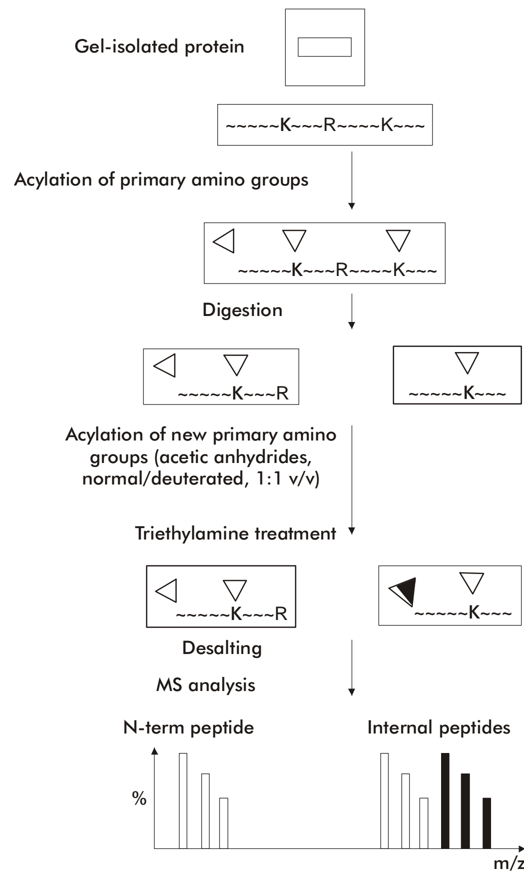

Figure 2. General diagram of the proposed method. The protein extracted from the polyacrylamide gel is acylated with succinic or acetic anhydride. The mixture of tryptic peptides is re-acylated with an equimolar mixture of acetic normal and deuterated anhydrides. The N-term peptide is recognized in the MS spectrum because it is the only specie showing a normal isotopic distribution.

veloped against this pathogen. The identification and sequencing of the N-blocked peptide allowed to determine the chemical nature of the blocking group and ii) to locate the fraction containing this peptide in peptide maps obtained from periodical quality controls of the vaccine preparation for the first time.

Methods 2 and 3, based on recognizing the N-term peptide signal in MS spectra [3, 4], can be applied to study proteins with N-blocked or N-free ends. In these methods, the N-term peptide acquires an isotopic distribution distinctive from those of the peptides internal to the protein. Both methods were applied to the following proteins:

Lysozyme C (14 kDa), trypsin inhibitor (24 kDA), p64k (62 kDa), phosphorylase B (N-term acetylated, 97 kDa) and a mix of proteins from *Escherichia coli*.

These are protein samples of different molecular weights and compositions, with increasingly complex mixes of peptides arising from their respective enzyme hydrolysis. Nevertheless, their respective N-blocked peptide sequences were unequivocally identified.

2. Betancourt L, Besada V, González LJ, Morera V, Padrón G, Takao T, *et al*. Selective isolation and identification of N-terminal blocked peptides from tryptic protein digests. J Pept Res 2001;57:345-53.

3. Sánchez A, Ramos Y, Solano Y, González LJ, Besada V, Betancourt L, *et al*. Double acylation for identification of amino-terminal peptides of proteins isolated by polyacrylamide gel electrophoresis. Rapid Commun Mass Spectrom 2007;21:2237-44.

4. Sánchez A, Ramos Y, Solano Y, González LJ, Betancourt L, Gil J, *et al*. Specific isotope labeling for the identification of free N-terminal peptide of proteins separated by polyacrylamide gel electrophoresis. Eur J Mass Spectrom (Chichester, Eng) 2007; 13:307-9.

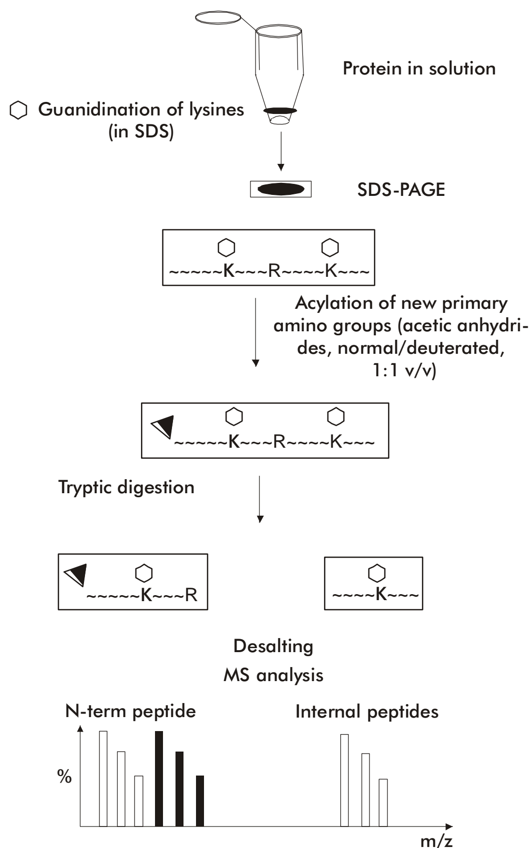

Figure 3. General diagram of the strategy proposed to identify the N-term peptide. The intact protein is guanidinated , following a tryptic digestion, and the peptides are acylated with an equimolar mixture of normal and deuterated acetic anhydrides. The N-term peptide is recognized in the MS spectrum because it appears as the only one with a complex isotopic distribution, which is not the case for internal peptides.

scFv multivalent antibody fragment against the carcinoembryonic antigen

This protein was obtained recombinant, to be used in immune radiodiagnosis or for immunoradiotherapy of different tumors [5]. One of the isolation strategies was applied to two gel degradation bands of the fragment. This allowed us to define from which end the fragment was degraded, also verifying the protein sequence.

phR3 antibody light chain

The R-3 antibody, commercially known as TheraCIM®, is used to treat head and neck tumors [6]. This molecule was transiently expressed in the leaves of tobacco plants, preserving its biological activity. The sweet potato (*Ipomoea batatas* L.) sporamine signal peptide was used, to favor the biological stability of the antibody in the genetic construct, bearing both, its light and heavy chains. Due to the outstanding relevance of this product and the differences between the native signal peptide of the antibody expressed in mice ascytes and that of the sporamine protein in the plantibody, both aminoacid sequences of the antibody were verified. The application of one of the methods allowed to identify the N-term peptide and to verify the primary sequence of the protein.

Relevance of the study

The methods described above are applicable to study the primary structure of proteins. One of the main objectives of our team work is to provide high quality services of protein primary structure analyses in Cuba, specifically of proteins produced for therapeutic purposes. Most of the products of the Center for Genetic Engineering and Biotechnology (CIGB) are rich in protein content, as well as products under research and development.

The development of methods to study the N-term peptide has allowed the analysis of the N-term or highly relevant proteins. The study of the Levansucrase from *Gluconacetobacter diazotrophicus* helped to understand the secretion mechanism of this enzyme. The study of the primary structure of the hepatitis B virus surface antigen (component of the main product of the CIGB) allowed complying with quality specifications for its use, as required by national and international regulatory authorities.

Moreover, these tools were used to characterize proteins in low amounts, isolated from polyacrylamide gels, such as the fragment of a specific antibody against the carcinoembryonic antigen, corroborating that the degradation products comprised the N-term. It was also possible to verify and analyze the N-term of the plantibody phR-3 transiently expressed in tobacco leaves. All these methods, while contributing to develop the analytical services offered by the CIGB, are scientifically novel and were included in the list of methods reported in international journals.

Acknowledgements

We want to thank Meylin Rodríguez, Merardo Pujol and Lincidio Pérez, from the CIGB, and Toshifumi Takao and Yasutsugu Shimonishi from the University of Osaka, Japan, for their contributions to this work.

5. Tormo BR, Gavilondo JV, Domínguez C, Freyre M, Rodriguez T, Biberfeld P. CEA in colonic adenocarcinomas and precancerous lesions. An immunohistochemical study with a novel monoclonal antibody. APMIS 1989 Dec;97(12):1073-80.

6. Mateo C, Moreno E, Amour K, Lombardero J, Harris W, Pérez R. Humanization of a mouse monoclonal antibody that blocks the epidermal growth factor receptor: recovery of antagonistic activity Immunotechnology 1997;3(1):71-81.