Rendiconti Accademia Nazionale delle Scienze detta dei XL
Memorie di Scienze Fisiche e Naturali
ISBN 987-88-548-7171-7
DOI 10.4399/97888548717174
pag. 77–86

Proteomics in Forensic Science

Angela Amoresano, Angela Flagiello,
Michela Balsamo, Andrea Carpentieri, Piero Pucci

1. Abstract

In forensic investigations on crime scenes, the identification of a biological matrix, even if present in mixtures, and the species of origin of a forensic trace without destruction of the DNA is of utmost importance. A new Proteomic–based approach was developed to identify unambiguous protein biomarkers for each individual matrix with the aim to constitute a specific database for matrix characterization. This approach fulfills all the requirements needed by forensic science for the rapid determination of biological matrices. In fact, MS–based proteomics procedures can be used to identify the most prominent proteins specifically present in each biological matrix found at a crime scene (blood, saliva, semen, vaginal fluid, nasal secretion, and urine). One single test is sufficient to unambiguously determine the identity of the biological matrices, even in complex mixtures. In addition, species identification is also possible and the analysis is performed on the first "washing" step of DNA extraction, thus saving most of the sample for subsequent DNA analysis. The protein biomarkers for the most frequently encountered biological matrices (saliva, blood, semen and lacrimal fluid) were clearly identified on real forensic samples. Moreover, a peptide mass fingerprinting procedure combined with tandem MS experiments was used to unambiguously assess the species of origins of human and animal hair.

These methodologies can be considered as useful additional tools for scientific investigations of biological traces in forensic science.

2. Introduction

In the last three decades the link between forensic investigations and scientific disciplines, such as Chemistry and Biology, has enormously strengthen. Nowadays, all the scientific contributions available to the investigators allow the detailed characterization of the crime scene up to the identification of a specific subject starting from very tiny traces that once might have been considered as science fiction.

During the last 15 years, forensic genetics strongly developed due to new discoveries and technologies in molecular biology. Genetic typization starting from only few cells, such as those found on "solely touched" objects, as well as the identification of many biological samples performed in a very short time are currently possible. Genetic analyses aimed at identifying people present on a crime scene are performed on DNA samples extracted from biological sources such as blood, saliva, semen, bones, hairs, urine, etc.). A specific set of 10–15 polymorfic genetic markers named STR (short tandem repeat), localized on specific chromosomes (autosomes), are characterised. The simultaneous analysis of this set of polymorphisms allows the investigators to define a genetic profile that is unique to each individual and virtually unique in the reference population. DNA analysis for forensic purposes is basically a comparative assessment between the genetic profiles obtained from biological traces found on a crime scene and the profile of a suspect. When the genetic profiles are different (incompatibility), the particular subject is not responsible for the biological traces, while in case of perfect overlap of the genetic profiles (compatibility) the specific subject is unambiguously located on the crime scene. In fact, the probability of finding another person in the population with the same genetic profile is lower than 1.7 in 10^{-16} (1 in 594 trillions) almost equivalent to uniqueness and, consequently, to the certain identification of the subject. Further investigations may be conducted by characterizing polymorphisms from the Y chromosome and from mitochondrial DNA; this information represents a useful marker in the reconstruction of parental relationships for the paternal line and maternal line, respectively (1).

Contrary to DNA, proteins were rarely considered as sources of useful biological traces and were scarcely used in crime scene investigations for the following reasons. Proteins tend to be less stable

than DNA and are easily degraded making their identification by immunological methods unfeasible. Biochemical tests have the main disadvantage to be destructive, resulting in sample loss for subsequent DNA analysis (2), while destruction of DNA should be avoided since forensic samples are often found in low quantity. Moreover, individual diagnosis by protein identification is impractical because protein variability is low and the amount of samples available cannot be amplified by "PCR–like" procedures thus requiring the use of analytical techniques with extremely high sensitivity.

All these negative features were rapidly overcome with the introduction of proteomics–based protein identification procedures as described below.

3. Traces analysis and diagnosis

Scientific investigations on biological traces occurring on a crime scene are carried out through a series of sequential steps involving generic diagnosis, species diagnosis, regional diagnosis and individual diagnosis. All together these steps allow the investigators to reconstruct the dynamics of the criminal event and to check the reliability of assertions by putative suspect people.

3.1. *Generic diagnosis*

The first step in the investigation of biological traces on a crime scene consists in the identification of the real nature of a given trace be it blood, semen, saliva, urine, etc. by using preliminary and confirmatory tests. For example, blood is quickly detected by means of luminol or benzidine (3), whereas traces of semen, saliva, sweat, urine and other biological fluids are monitored by ultraviolet light or by other light sources at specific wavelengths. Confirmatory tests are essentially based on immunochemical techniques or ELISA methods. Forensic tests currently focus on the prostate–specific antigen (PSA) for semen (4) while alpha–amylase tests are employed for the detection of saliva. However, the alpha–amylase assay is an indirect test that measures the activity of amylase. A further drawback is the lack of specificity as it cannot differentiate between alpha–amylase

1, present in saliva, and alpha–amylase 2, occurring in semen and vaginal secretion (5). In general terms, color–based presumptive tests, like those just mentioned, can be challenging to interpret for *weak* to *trace* positives or for mixtures with blood (6) and in addition are specific for only one biological matrix. It follows that several cascade tests might be needed before the biologic nature of a certain sample is uncovered. Moreover, when the identity of one biological matrix is determined, no or little effort is made to find out whether the sample is a mixture of different biological matrices due to the high cost of the tests, the long–lasting procedure of the cascade testing and the increased loss of sample when multiple tests are performed. However, the main disadvantage of these biochemical tests is their destructive nature, resulting in sample loss for subsequent DNA analysis.

3.2. *Species diagnosis*

Besides the identification of a certain biological matrix, determining the species of the donor of the sample can be fundamental. For example, when the presence of blood is ascertained by preliminary methods, other tests based on the use of highly specific antibodies are needed to distinguish human from animal blood. However, these tests can only indicate the presence or absence of blood from one specific species. If negative, another test has to be performed in order to potentially identify the animal species wherefrom the blood originated.

The diagnosis of species on hair is usually done by microscopic examination which again is able to discriminate between human and animal species with few chances to properly identify the specific species.

These considerations show that forensic science is needing a universal, specific, and unbiased method to unambiguously identify biological matrices of different species origin, even in mixture, without destruction of the DNA and preferably without any additional sample consumption, so that the full trace is available for DNA extraction.

4. Proteomics procedures for protein identification

Following the conclusion of the Human Genome Project, it appeared immediately clear that the knowledge of the entire DNA sequence

of an organism provided a wealth of information, but constituted more a starting point than the "end of the story" in understanding the function of living cells at the molecular level. The challenge shifted again from genes to proteins, giving rise to what has been termed the "proteomic era". The aim of this new era is to identify and characterize proteins located within a given organelle, cell or organism, as well as to unravel the in vivo protein pathways. Proteome analysis is essentially accomplished by three consecutive steps:

a) fractionation of complex protein mixtures (mainly performed by bi–dimensional electrophoretic and/or chromatographic techniques);
b) identification of individual proteins by mass spectrometry (MS)–based methodologies;
c) screening protein databases using mass spectrometry data (7, 8).

Because of its unsurpassed sensitivity and dynamic range of analysis, mass spectrometry is the method of choice in proteomics procedures for protein identification, nowadays routinely applied in proteomics laboratories (9). The protein in solution, or spotted on any solid surface, is digested enzymatically *in situ* and the resulting peptide mixtures analysed directly by capillary Liquid Chromatography–Mass Spectrometry techniques, LCMS/MS. The fractions eluted from the column are directly inserted into the Electro Spray (ES) mass spectrometry source and their mass values determined. Peptide ions will be isolated simultaneously and fragmented within the mass spectrometer, producing daughter ion spectra from which sequence information on individual peptides can be obtained. This information is then used together with the peptide mass values to search protein databases, leading to the unambiguous identification of the protein component (Fig. 1).

The approach delineated fulfills all the requirements needed by forensic science for the rapid determination of biological matrices. MS–based proteomics procedures, in fact, can be used to identify the most prominent proteins specifically present in each biological matrix that might be found at a crime scene (blood, saliva, semen, vaginal fluid, nasal secretion, and urine). One single test is sufficient to unambiguously determine the identity of the biological matri-

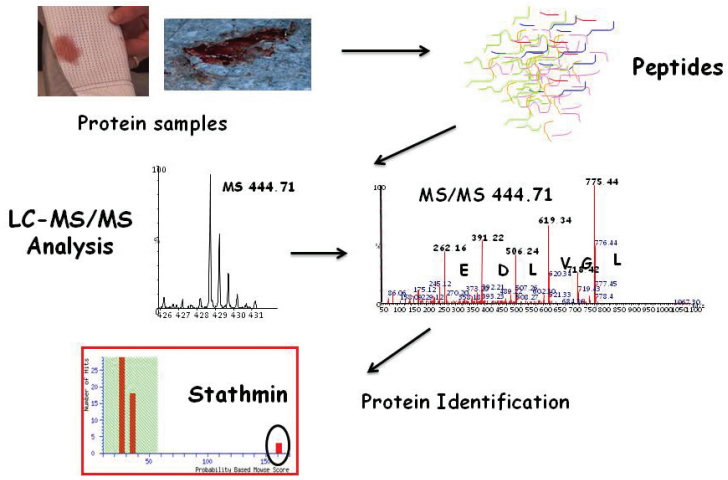

Figure 1. Proteomic strategy for protein identification.

ces, even in complex mixtures, avoiding long lasting and expensive cascade tests. Additionally, by means of this proteomic approach, species identification is also possible and the analysis is performed on the first "washing" step of DNA extraction, a solution which is normally discarded, saving most of the sample for subsequent DNA analysis. In this respect, degradation of proteins is no longer a limitation since in proteomics procedures proteins are identified from their fragments. Moreover, modern mass spectrometric techniques are sensitive enough to identify proteins from a very minute amount of material.

5. Identification of protein biomarkers of biological matrices

The proteomic approach was applied to different biological matrices to identify particular protein biomarkers for each individual matrix with the aim to constitute a specific database for matrix characterization. Four different human biological matrices, namely blood, saliva, semen and lacrimal fluid, found at real crime scenes were collected from different solid surfaces, i.e. cotton, wool, clothes, bristle broom, etc. The real forensic samples were cut in small pieces and extracted with suitable aqueous solutions. Thereafter, the extracted proteins

were directly digested with trypsin and the resulting peptide mixtures analysed by capillary liquid chromatography coupled to tandem mass spectrometry (nanoLC–MS/MS). The accurate peptide mass values determined by MS and the sequence information on individual peptides obtained by the daughter ion spectra were then used to search protein databases using the Mascot software, leading to the identification of the protein components. As expected, different proteins were identified in different matrices. Since the amount of biological matrix in forensic samples can be very limited, the selection criteria for biomarkers were based on the specificity and the abundance of the proteins.

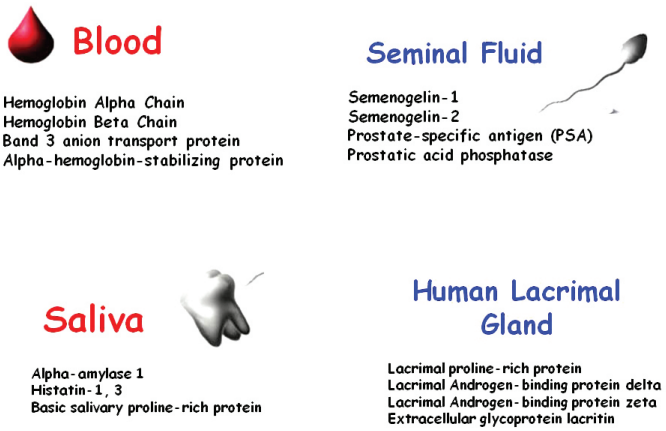

Figure 2. Protein markers of individual biological matrices.

Fig. 2 shows the protein biomarkers identified for each individual biological matrix. For blood, the alpha and beta subunits of hemoglobin (Hb) were chosen as biomarkers since Hb is specific for blood and present in large amounts. However, besides the Hb subunits, other blood specific proteins, the Band 3 anion transport protein and the Alpha–Hemoglobin–Stabilizing Protein (AHSP), could also be identified. Additionally, since the proteomic approach is also able to assess the origin of the identified proteins, the unambiguous identification of human hemoglobin allows the investigator to distinguish blood from different species, thus providing both generic and species diagnoses at the same time. The most commonly used conventional tests can differentiate between human and non–human blood, whereas a

proteomic investigation can easily pinpoint the species as well, again without the need of cascade analysis using several antibodies and without any loss of trace or cellular material needed for subsequent DNA analysis.

Semenogelin 1 and 2, but also other semen–specific proteins such as prostatic acid phosphatase and the Prostate–Specific Antigen (PSA) were found by proteomic investigation and used to specifically characterise the matrix. Detection of all these different proteins in one single test confirmed the origin of the matrix unambiguously.

In the same way, alpha–amylase 1, the highly abundant and specific protein for saliva, together with Histatin–1 and 3 and the Basic salivary proline–rich protein were identified by mass spectrometry as specific biomarkers for saliva whereas Lacrimal proline–rich protein, Lacrimal Androgen–binding protein delta and zeta and the Extracellular glycoprotein lacritin were used to characterise the lacrimal fluid. It should be underlined that mass spectrometry analysis was able to distinguish between the two forms of alpha–amylase, alpha amylase 1, present in saliva, and alpha–amylase 2, occurring in semen and vaginal secretion, thus increasing specificity.

6. Assessment of hair species by mass spectrometry

As stated above, microscopic examination is the most common method for the diagnosis of species on hair. However, even though this approach can easily discriminate only between human and animal species, in few cases it is able to properly identify the specific non–human species. Human and non human hair are principally composed of keratins, a family of conserved proteins that form a complex network giving the hair its rigidity and mechanical properties. Although keratins share a good degree of homology, their amino acid sequences are species–specific and can then be used to distinguish among non human hair. In fact, following enzymatic hydrolysis species–specific keratins give origin to a specific set of peptides characterised by a unique molecular mass and amino acid sequence.

Based on these considerations, a fast and reliable proteomic mass spectrometry (MS)–based strategy for the identification of the origin of native hair was developed. The method is exclusively based on

the analysis of proteins and uses minute amounts of peptides directly derived from tryptic hair digests without any separation or enrichment steps. However, since animal keratin sequences from many species are still unknown and are absent in databases, identification of the specific origin of hair proteins using classical proteomic approaches was not trivial.

Preliminary experiments were designed to constitute a reference databank of mass spectrometry data to be used for unknown samples investigation. Without any prior cleaning or isolation of single proteins, hair from human and other nine animal species was enzymatically digested with trypsin. Fragments generated were analyzed by MALDI–TOF mass spectrometry and peak groups of different selectivity were established for every animal species. For the identification of individual animal species, only unique species–specific peaks were selected as reported in Table 1. Precursor ions identified as species–specific diagnostic peptide ions were also selected for MS/MS analysis to provide sequence information. Peptide mass fingerprinting data combined with those from tandem MS experiments were then used to generate the reference database. Despite the high degree of homology of the samples, reliable identification of the species of origins of human and all the animal species was possible.

Table 1. Hair analysis by Mass Spectometry Peptide Ions specific for each species

Human	Cat	Fox	Sheep	Rabbit	Wolf	Dog	Hare	Raccoon	ALPACA
2975.4	1976.8	1759.8	2515.1	2549.3	1520.8	2135.1	2152.9	2705.3	2561.4
3000.6	1991.7	1763.9	2519.3	2550.3	1596.8	2136.2	2231.1	2717.1	2573.4
3022.4	1998.7	2068.1	2536.3	2681.5	1657.9	2167.1	2232.2	3223.5	2615.4
3074.6	2007.9	2094.1	2581.3	2682.5	1996.1	2168.2	2293.1	3265.7	2746.3
3193.5	2170.1	2095.1	2608.3	2874.4	2249.3	3281.7	2294.1	3297.5	3083.6
3259.6	2172.8	2496.2	2623.3	2875.4	2250.3	3282.7	2381.1	3298.5	3101.7
3342.7	2337.0	2546.4	2776.6	3089.5	2660.3	3459.0	2414.2	3312.5	3111.7
3410.7	2410.3	2597.3	2804.3	3090.5	3229.5	3459.9	2481.2	3313.5	3272.7
3544.7	2769.3		2866.2	3270.4	3230.6		2482.2	3392.7	3284.7
3752.8				3271.4			2571.4	3393.7	3326.7
							2579.3		

Unknown species can thus be identified by comparison of their peptide ion patterns with known mass values stored in the reference database. If more than one potential hit remains, specific diagnostic peptide ions are used to stepwise exclude incorrect matches.

References

[1] Roewer L., *DNA fingerprinting in forensics: past, present, future*. Investig Genet. 4 (2013) 22.

[2] Virkler K, Lednev IK., *Analysis of body fluids for forensic purposes: from laboratory testing to non–destructive rapid confirmatory identification at a crime scene*. Forensic Sci Int 188 (2009) 1.

[3] Tobe SS, Watson N, Daeid NN., *Evaluation of six presumptive tests for blood, their specificity, sensitivity, and effect on high molecular–weight DNA*. J Forensic Sci. 52 (2007) 102.

[4] Pang BCM, Cheung BKK., *Identification of human semenogelin in membrane strip test as an alternative method for the detection of semen*. Forensic Sci Int 169 (2007) 27.

[5] Quarino L, Dang Q, Hartmann J, Moynihan N., *An ELISA method for the identification of salivary amylase*. J Forensic Sci 50 (2005) 873.

[6] Myers JR, Adkins WK., *Comparison of modern techniques for saliva screening*. J Forensic Sci 53 (2008) 862.

[7] Godovac–Zimmermann J, Brown LR., *Perspectives for mass spectrometry and functional proteomics*. Mass Spectrom Rev 20 (2001) 1.

[8] Patterson SD, Aebersold RH., *Proteomics: the first decade and beyond*. Nat Genet. 33 (2003) 311.

[9] Van Riper SK, de Jong EP, Carlis JV, Griffin TJ., *Mass spectrometry–based proteomics: basic principles and emerging technologies and directions*. Adv Exp Med Biol. 990 (2013) 1.

Angela Amoresano, Angela Flagiello, Michela Balsamo,
Andrea Carpentieri, Piero Pucci
Dipartimento di Scienze Chimiche
Università "Federico II", Napoli
pucci@unina.it