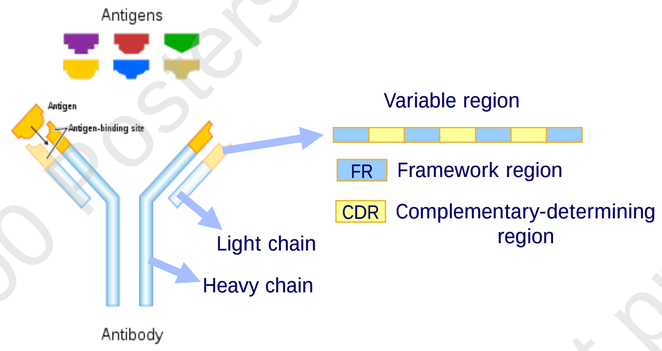
A NOVEL METHOD FOR PROTEOMIC ANALYSIS OF IMMUNOGLOBULIN LIGHT CHAINS AND ATTRIBUTION TO A GERMLINE GENE-BASED FAMILY

Alessandra Tiengo[1], Francesca Lavatelli[2], Nicola Barbarini[1], Giampaolo Merlini[2] and Paolo Magni[1]

[1]Dipartimento di Informatica e Sistemistica, Università degli Studi di Pavia, Pavia, Italy
[2]Amyloid Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

MOTIVATION. Amyloidoses are a group of diseases caused by tissue deposition of misfolded proteins as amyloid fibrils. In the systemic forms, amyloid deposition is widespread and associates with severe dysfunction of the involved organs [1]. Precise typing of amyloidosis, *i.e.* identification of the protein constituting the deposits, is the key for establishing a proper treatment, a correct clinical management and prognosis assessment. AL amyloidosis is the most common systemic form in Western countries and it is caused by deposition of monoclonal free immunoglobulin light chains (LC), produced in excess by bone marrow plasma cells. Immunoglobulin light chains are ~22-25 kDa proteins that, under normal circumstances, associate to heavy chains to form intact antibodies.

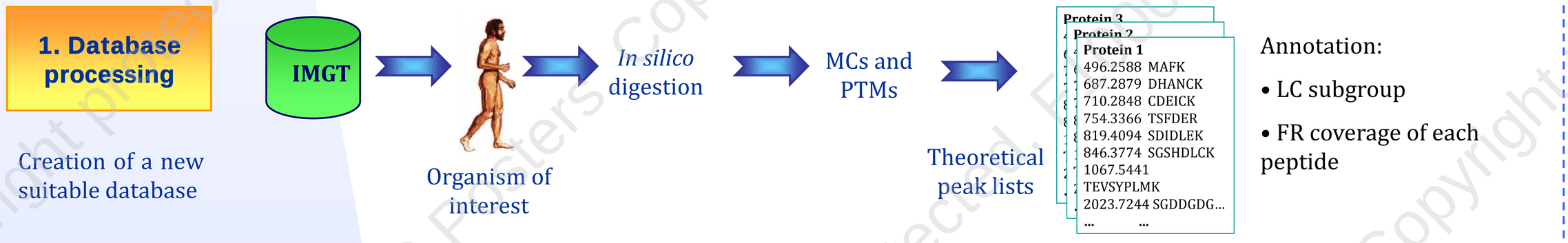


Light chains are constituted by two structurally and biologically distinct domains: a constant region (CR), which distinguishes the LC in two main groups or isotypes (κ and λ) and shows little or no amino acid variation within the same isotype, and a variable region (VR), which, instead, profoundly differs also within the same LC and determines the antigen binding and the biochemical properties [2]. This translates into the fact that thousands of different light chains exist.

The existence of homologies in sequence of VR, especially in FR1, provides the basis for classification of LC into multiple subgroups or families [2]. Assignment of the LC gene family is thus obtained by alignment of the full length nucleotidic sequence with germline sequences deposited in databases, but it is both time consuming and rather invasive.

In contrast, proteomic analysis allows the direct investigation of small amounts of a given protein, based on mass spectrometry (MS). The presence of variations in sequence of VR makes the search of the MS data in protein databases very difficult, because the primary structure of the light chain obtained from a patient rarely overlaps properly with the annotated sequences [3, 4].

METHODS. The proposed method consists of the following three steps: the database processing, the peptides analysis and the subgroup identification.



2. Peptides analysis

For each peptide in the new DB:
- occurrence in every LC subgroup
- frequency in every LC subgroup
- median FR coverage

New ASCII file (output):
- peptide mass
- max frequency
- median FR coverage
- LC subgroup

High FR coverage ? LC subgroup

Pearson correlation

Occurrence – Median FR cov.: r=0.09 p-value=5.91*10-11

Frequency – Median FR cov.: r=0.39 p-value=0

3. Subgroup identification

- Comparison between MS spectrum and ASCII file
- Mass tolerance in Da or ppm
- Three scoring functions tested

$$Score_1(S) = \sum_{n=1}^{N} Int_n * Freq_{m(S)}$$

$$Score_2(S) = \sum_{n=1}^{N} Int_n * Freq_{m(S)} * FR_{m(S)}$$

$$Score_3(S) = \sum_{n=1}^{N} Int_n * FR_{m(S)}$$

N: number of matches
S: LC subgroup
m(S): mass for the subgroup S
Int_n: intensity of the *n-th* peak
$Freq_{m(S)}$: frequency of m(S)
$FR_{m(S)}$: FR coverage of m(S)

RESULTS. The proposed method was tested on a dataset of 20 MS spectra obtained from 10 patients. The results achieved using the proposed method were compared with the identifications obtained by a standard method such as that implemented in MsPI, considering the same protein database. MsPI is a Perl software tool for protein identification through PMF approach [5]. cDNA sequencing has been considered as correct.

Compared with gene-based methods, the use of proteomics to assign each LC to its subgroup allows a significantly easier biological sample acquisition, wider applicability of the analysis and more direct investigation of the involved species. Considering that proteomic analysis of tissues is becoming the new standard in diagnosis of amyloidoses [6], this method could find wide clinical applicability.

Patient ID	1	2	3	4	5	6	7	8	9	10	Correctly assigned
MS spectrum	1 2 3 4	1 2	1 2	1	1	1 2	1	1 2	1 2	1 2 3	
Subgroup assigned through cDNA sequencing	1	1	1	2	2	3	3	3	3	10	
Score 1	**1 1 1 1**	**1** 5	6 6	4	1	**3** 8	1	**3** 5	**3 3**	3 6 3	9 (45%)
Score 2	**1 1 1 1**	**1 1**	**1 1**	4	1	**3 3**	1	**3** 1	**3 3**	3 6 3	13 (65%)
Score 3	**1 1 1 1**	**1 1**	**1 1**	1	1	**3 3**	1	**3** 1	3 3	3 6 3	13 (65%)
MsPI	**1 1 1 1**	5 **1**	3 **1**	1	5	5 5	1	**3** 5	1 2	**10** 8 3	8 (40%)

IMGT database (3178 human proteins)
1 MC
PTMs: CAM of cysteine (+57.02),
oxidation of methionine (+15.99)
pyro-Glu from glutamine (-17.03)
deletion of the first amino acid
Mass tolerance: 0.3 Da

REFERENCES

[1] Merlini G and Bellotti V. Molecular mechanisms of amyloidosis. N Engl J Med 2003;349:583-596.
[2] Solomon A and Weiss DT. Structural and functional properties of human lambda light chain variable-region subgroups. Clin Diagn Lab Immunol. 1995;2:387-394.
[3] Lavatelli F et al. Amyloidogenic and associated proteins in systemic amyloidosis proteome of adipose tissue. Mol Cell Proteomics. 2008;7:1570-1583.
[4] Lavatelli F et al. A novel approach for the purification and proteomic analysis of pathogenic immunoglobulin free light chains from serum. Biochim Biophys Acta. 2011;1814:409-419.
[5] Tiengo A et al. A Perl procedure for protein identification by Peptide Mass Fingerprinting. BMC Bioinformatics. 2009;10:S11.
[6] Vrana et al. Classification of amyloidosis by laser microdissection and mass spectrometry-based proteomic analysis in clinical biopsy specimens. Blood. 2009;114:4957-9.