

# A novel method for proteomic analysis of immunoglobulin light chains and attribution to a germline gene-based family

Alessandra Tiengo

Department of Computer Engineering and  
Systems Science



BIO-MEDICAL INFORMATICS  
*"Mario Stefanelli"*

UNIVERSITÀ DI PAVIA



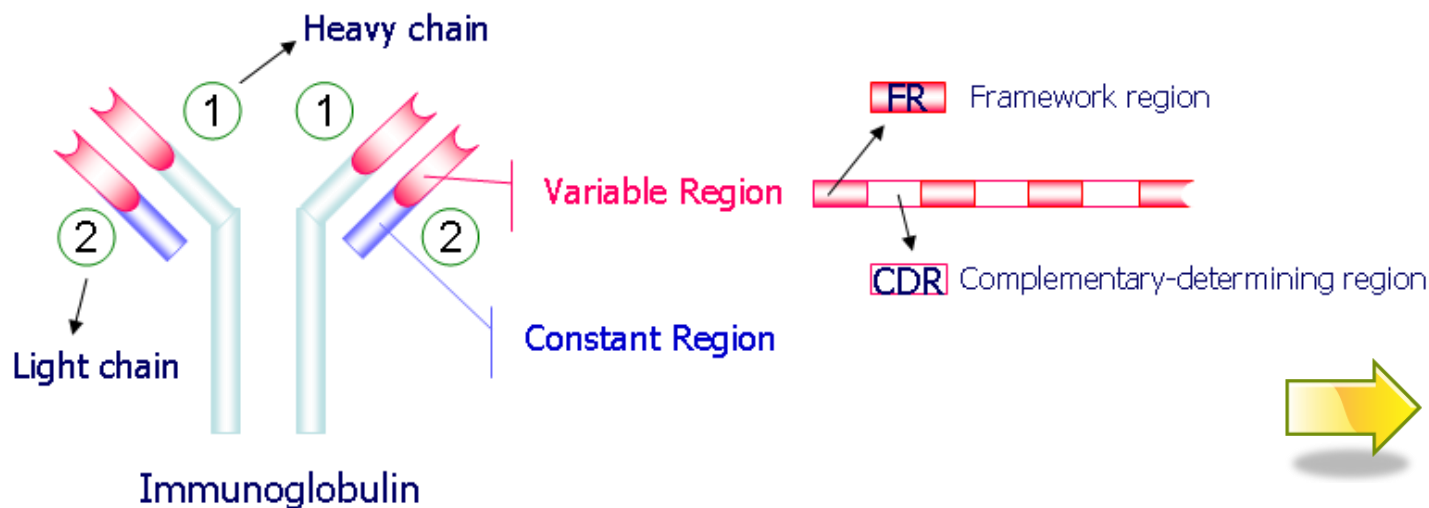
# Amyloidosis

Amyloidosis

Tissue deposition of misfolded proteins as amyloid fibrils

AL Amyloidosis

Deposition of monoclonal free immunoglobulin light chains (LC)



Subgroups  
or families

# Subgroup assignment

## Gene-based approach

**1. Coding mRNA  
(bone marrow plasma cells)**

**2. cDNA sequencing**



**Requires bone marrow  
(Invasive and  
long procedure)**

## Proteomic approach (PMF)

**1. MS experiments**

**2. Database of theoretical  
peptides  
(e.g. Swiss-Prot, IMGT)**

**3. Comparison between  
theoretical and experimental  
spectra**



**Variable Regions**

# Subgroup assignment

**1. Database  
processing**

**2. Peptide  
analysis**

**3. Subgroup  
identification**

# Database processing

## 1. Database processing

## 2. Peptide analysis

## 3. Subgroup identification

- @ IMGT database
- @ In silico digestion
- @ MCs and PTMs
- @ FR coverage
- @ LC subgroup



# Peptide analysis

**1. Database processing**

**2. Peptide analysis**

**3. Subgroup identification**

- Ⓢ Peptide frequency in the subgroup
- Ⓢ Median FR coverage of the peptide
- Ⓢ Output: ASCII file containing the peptide mass, the frequency, the median FR coverage and the subgroup

ASCII file



m/z	Freq.	FR cov.	Class
801	0.1	0.3	LC1
1100	0	0.8	LC10
1398	0.5	0	LC1
2100	0	0	LC3
2168	0.2	0.1	LC1
...	...	...	...

# Subgroup identification

1. Database processing

2. Peptide analysis

3. Subgroup identification

Ⓞ Comparison between MS spectra and peptides in the ASCII file (tolerance window in Da or ppm)

Ⓞ Subgroup identification

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

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

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

Peak list

m/z	Int
801	25.0
1100	39.0
1400	15.0
1406	100.0
1512	24.0
...	...

ASCII file

m/z	Freq.	FR cov.	Class
801	0.1	0.3	LC1
1100	0	0.8	LC10
1398	0.5	0	LC1
2100	0	0	LC3
2168	0.2	0.1	LC1
...	...	...	...

N matching peaks

# Results

- @ 20 MS spectra (10 patients)
- @ IMGT database (3178 human proteins with annotated LC subgroup)
  - @ 1 MC
  - @ PTMs: CAM (C), oxydation (M), pyro-Glu (Q), deletion of first amino acid
- @ Tolerance mass: 0.3 Da
- @ Comparison between the scoring functions and MsPI (PMF approach)
- @ cDNA sequencing assignment considered as correct

Patient ID		1	2	3	4	5	6	7	8	9	10	
MS spectrum		1 2 3 4	1 2	1 2	1 1	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	3	10	Correctly assigned
Score 1	$\Sigma \text{Int} * \text{Freq}$	1 1 1 1	1 5	6 6	4 1	3 8	1	3 5	3 3	3 6 3	9 (45%)	
Score 2	$\Sigma \text{Int} * \text{Freq} * \text{FR}$	1 1 1 1	1 1	1 1	4 1	3 3	1	3 1	3 3	3 6 3	13 (65%)	
Score 3	$\Sigma \text{Int} * \text{FR}$	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%)	



# Thanks!



BIO-MEDICAL INFORMATICS

*"Mario Stefanelli"*

