

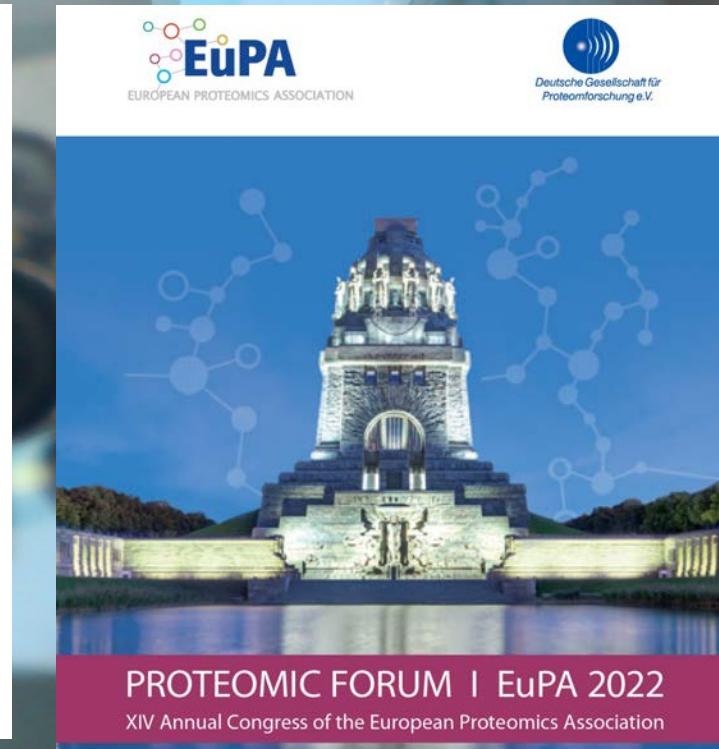
# Expanding the role of proteoforms in the field of infectious disease

---

**Julia Chamot-Rooke**

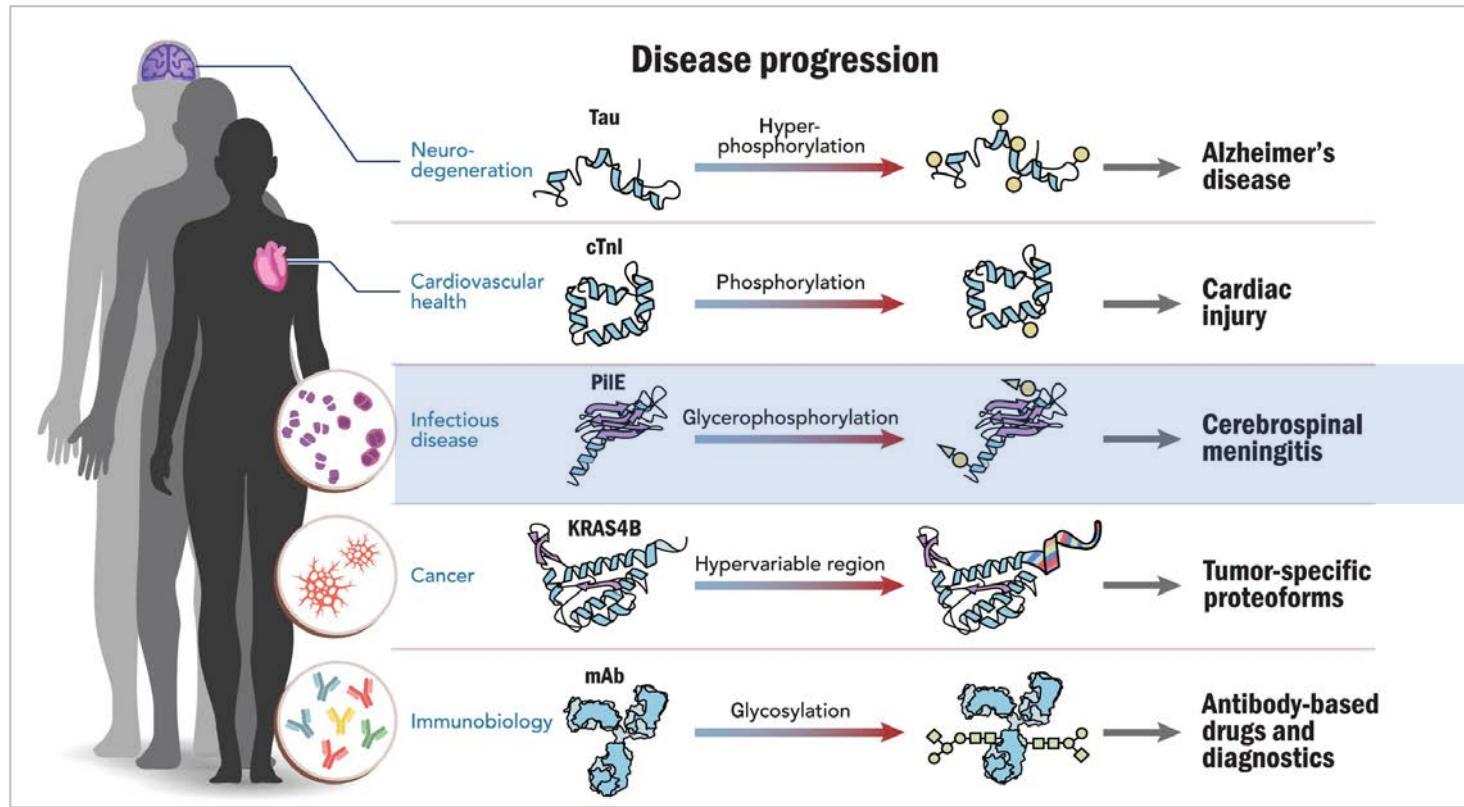
*Institut Pasteur, Paris, France*

04/04/2022



# Proteoforms & infectious disease: two main axes

- Characterization of proteoforms involved in infection (*understanding of the pathogenicity*)

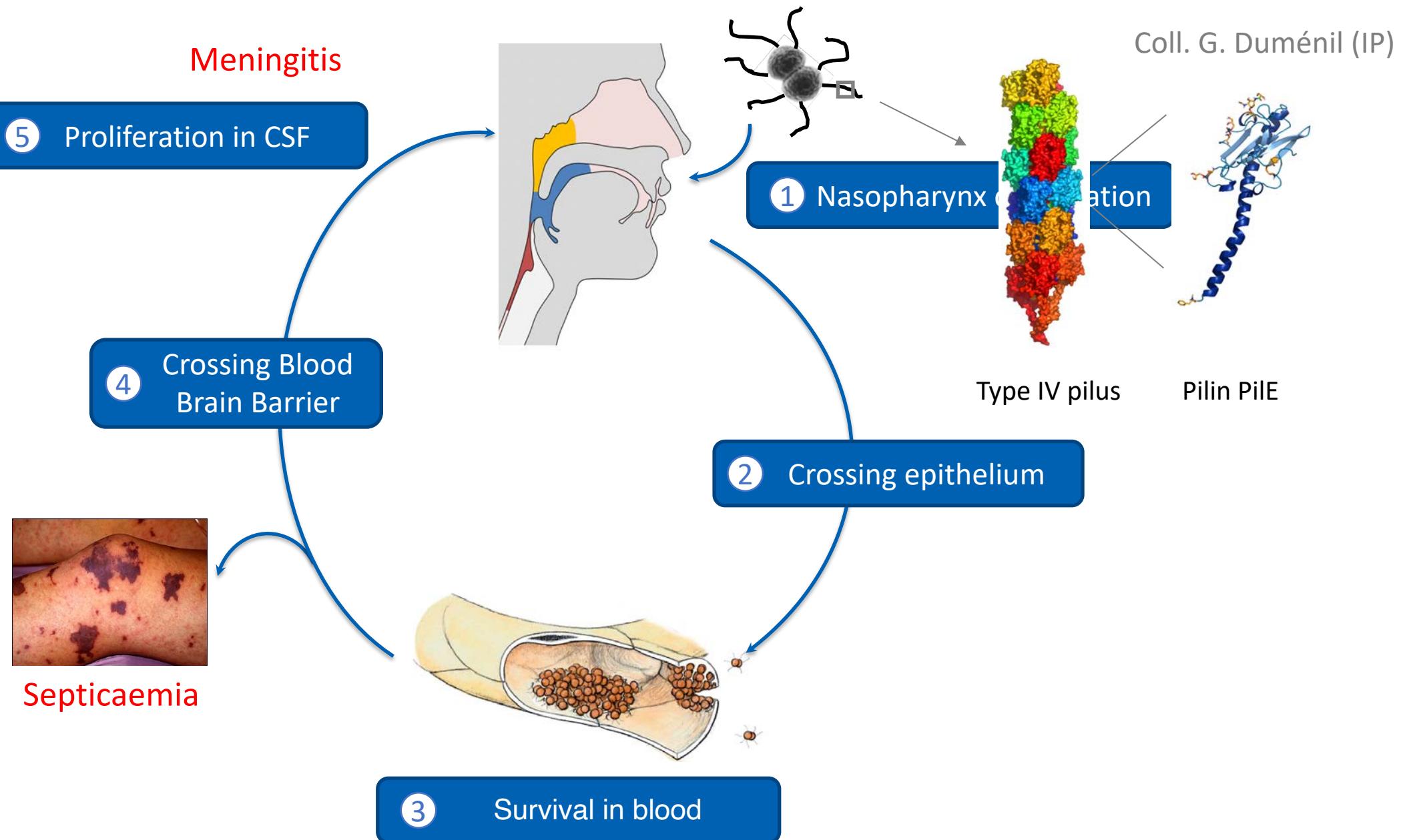


Smith *et al.*, *Science Advances* 7, eabk0734 (2021)

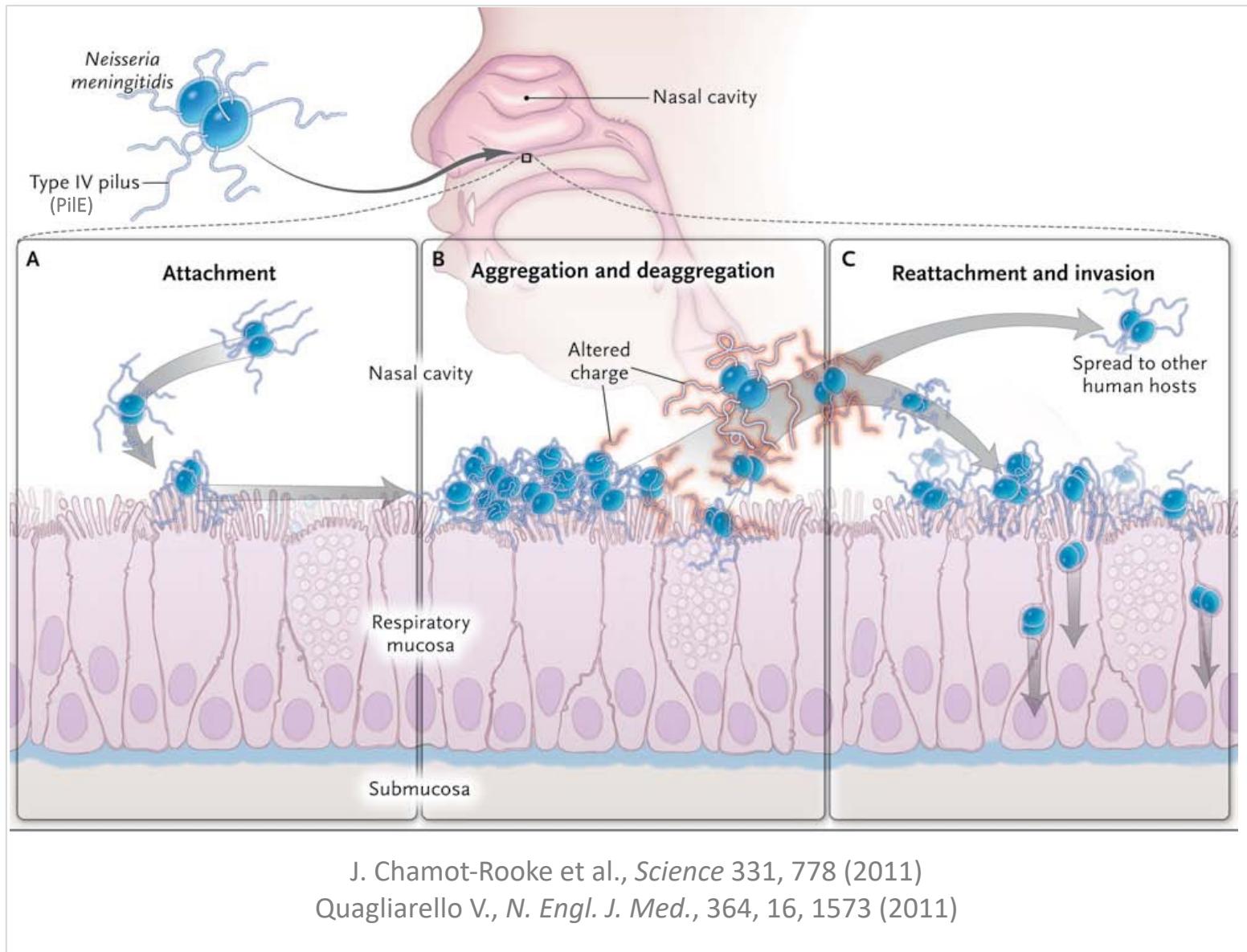


- Proteoform-based discrimination of bacterial pathogens (*diagnosis in hospital settings*)

# *Neisseria meningitidis*



# PilE proteoform analysis (reference strain)



Coll. G. Duménil (IP)

Pili modification with PG facilitates systemic invasion and the spread of colonization to close contacts

# PilE proteoform analysis (hypervirulent clinical strains)

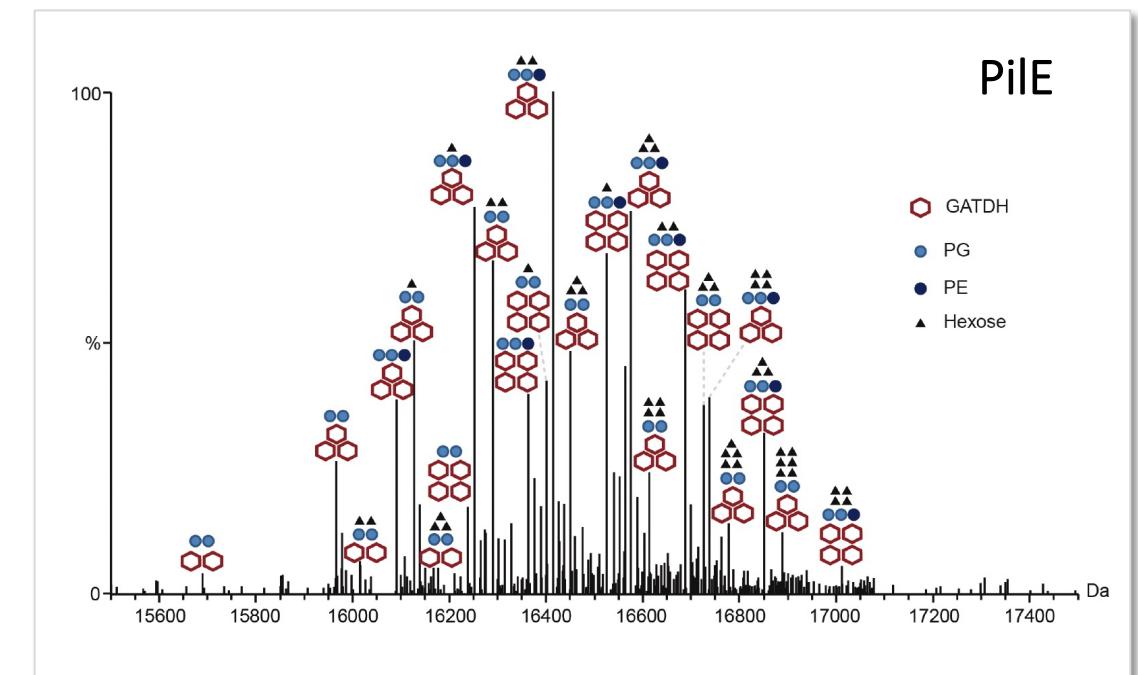
- Hypervirulent clinical isolates collected from patients with evidence of meningitis
- Selection of several isolates from variable location and serogroup

Coll. M-C. Ploy (Limoges)



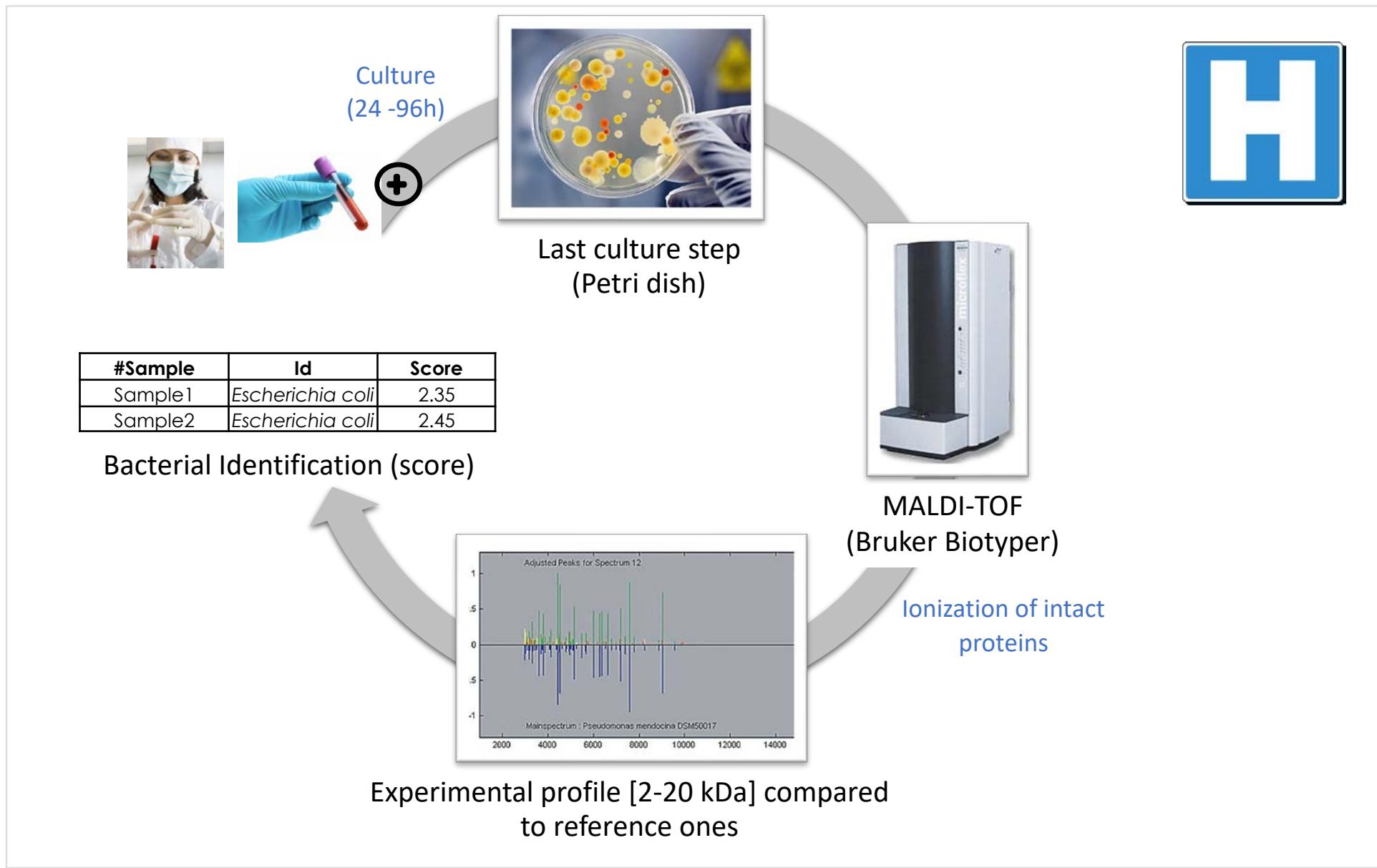
CSF  
Throat

Blood



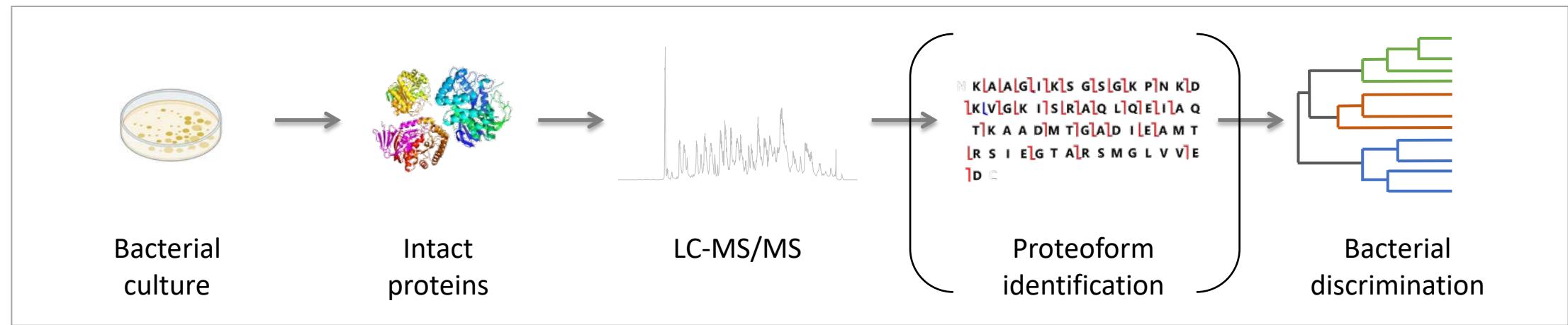
Unexpectedly high level of glycosylation associated with immune escape

# Bacterial identification in hospitals (MALDI-TOF)

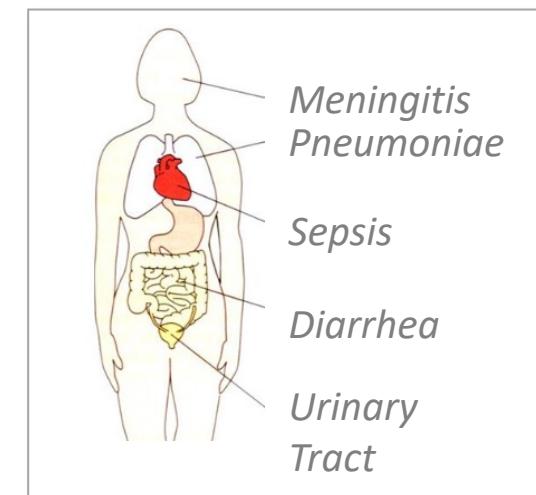


# Proteoform-based discrimination of bacterial pathogens

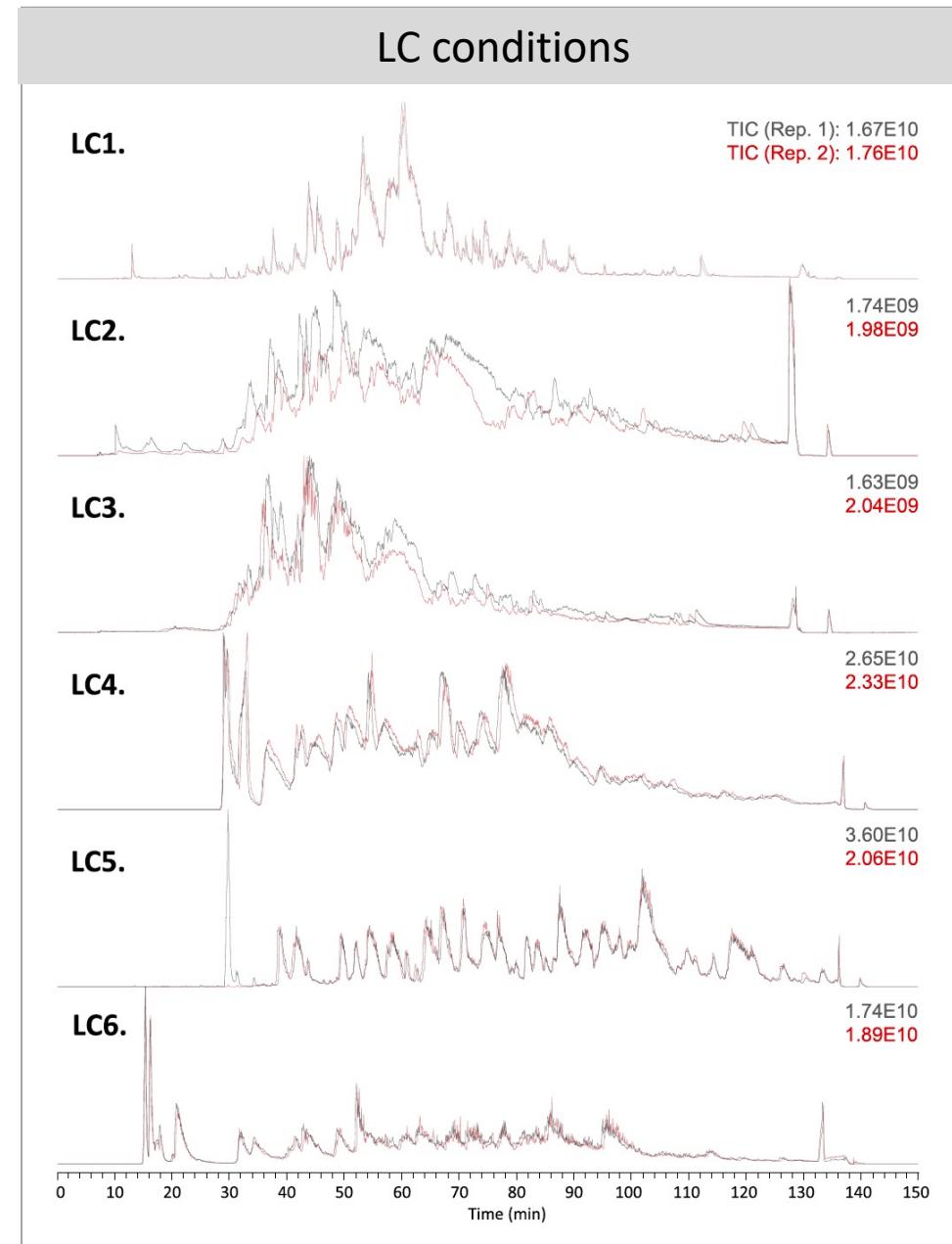
- Objective: Use top-down proteomics to characterize bacterial pathogens in a clinical context  
(Collaboration with 3 French hospitals)



- Differentiation of enterobacterial pathogens (12 strains, 3 replicates)



# Optimization of the top-down proteomics workflow



### MS parameters (Orbitrap Lumos)

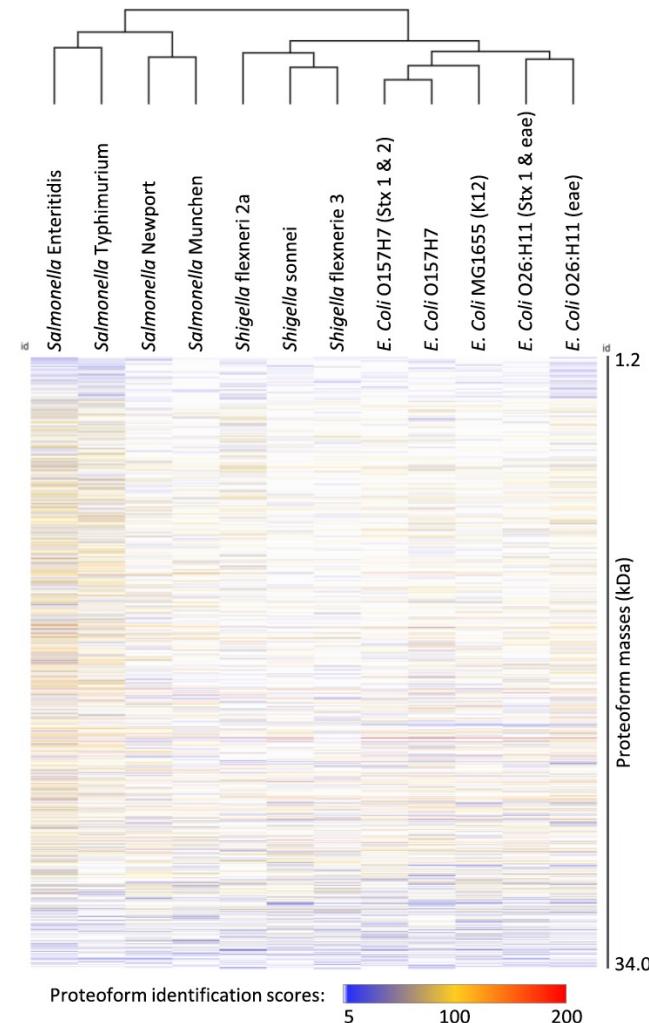
AGC target (ms)	100; 250; 500
charge-state exclusion	yes; no
DDA mode	top $N$ : 2, 4, 6; top speed: 5 s
$\mu$ scans (MS) <sup>a</sup>	1; 2; 3; 4; 6; 12; 24
MS resolution	15k; 30k; 60k; 120k
$\mu$ scans (MS/MS) <sup>a</sup>	2; 3; 4; 6; 12
MS/MS resolution	30k; 60k; 120k
MS/MS mode	ETD (@5 ms; @10 ms); HCD (@15NCE; @20NCE; @25NCE); EThcD (@10 ms@SNCE; @10 ms@10NCE)

### Optimized conditions

- PBS buffer, C4 nanoLC column
- 60k MS & MS/MS, 2  $\mu$ scans, Top4, EThcD
- ~ 500 proteoforms (PrositePD) in a single run

# Proteoform-based differentiation

Phylogenetic Tree for *E. coli*, *Shigella*, *Salmonella*  
(3,500 proteoforms, ProsightPD)



YegP Discriminating Proteoforms

**A.**

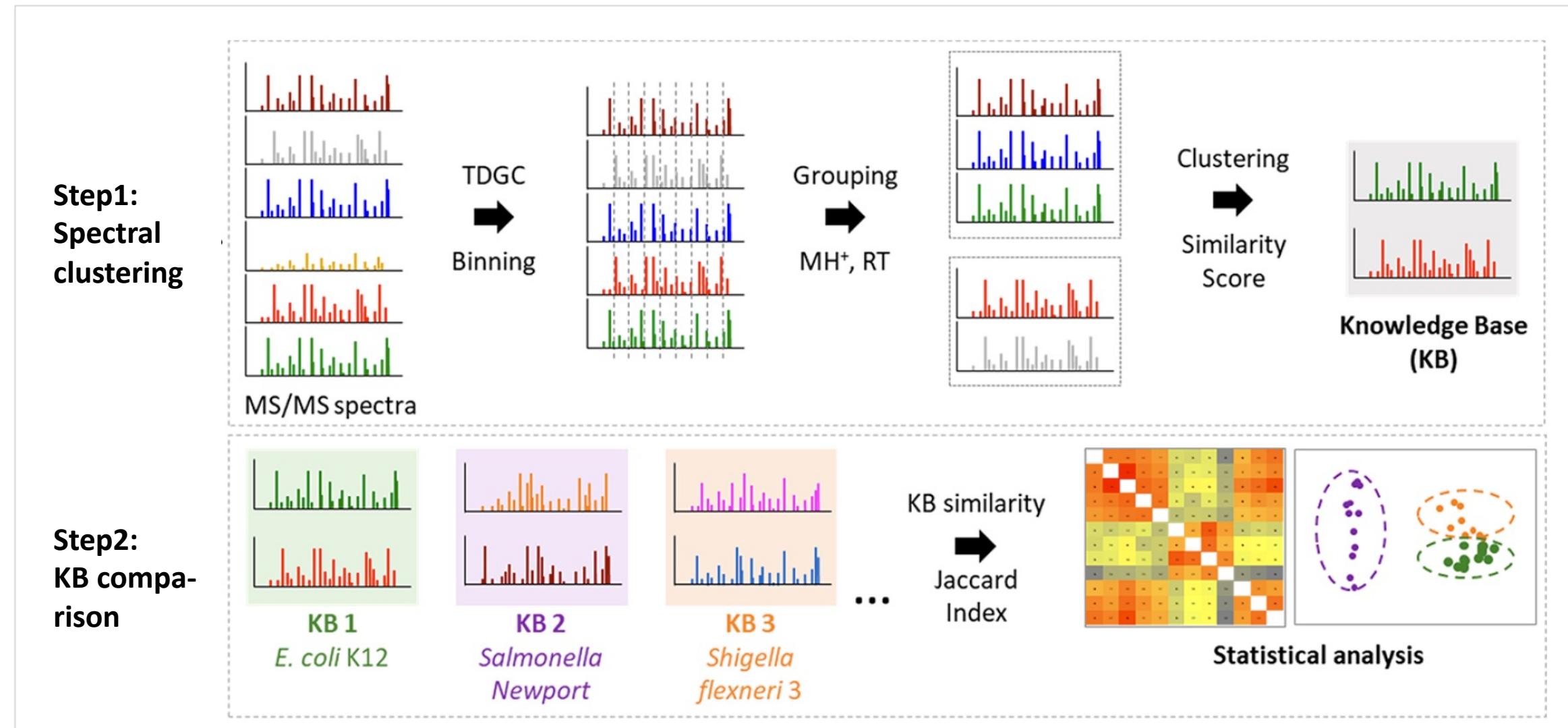
YegP proteoforms	<i>Shigella</i>			<i>Escherichia coli</i>				
	<i>sonnei</i>	<i>flexneri</i> 2a	<i>flexneri</i> 3	O157:H7 (Stx 1 & 2)	O157:H7	O26:H11 (Stx 1 & eae)	O26:H11 (eae)	MG1655 (K12)
P1	11885.93	x	x	x	x	x	x	
P2	11855.92	x	x	x	x	x	x	
P3	11828.91		x	x				

**B.**

P1	M <sub>measured</sub> = 11 885.94 Da M <sub>theoretical</sub> = 11 885.93 Da Δmass = 0.003 Da -log P-score = 137.2 % residue cleavage = 58 Matching fragments = 92	N A G W F E L S K S S D N Q F R F V L K A G N G E T 25 26 I L T S E L Y T S K L S A E K G I A S V R S N S P 50 51 Q E E R Y E K K T A S N G K F Y F N L K A A N H Q 75 76 I I G S S Q M Y A T A Q S R E T G I A S V K A N G 100 101 T S Q T V K D N T C
P2	M <sub>measured</sub> = 11 855.91 Da M <sub>theoretical</sub> = 11 855.92 Da Δmass = -0.010 Da -log P-score = 132.5 % residue cleavage = 62 Matching fragments = 99	N A G W F E L S K S S D N Q F R F V L K A G N G E T 25 26 I L T S E L Y T S K L S A E K G I A S V R S N S P 50 51 Q E E R Y E K K T A S N G K F Y F N L K A A N H Q 75 76 I I G S S Q M Y A T A Q S R E T G I A S V K A N G 100 101 T S Q T V K D N T C
P3	M <sub>measured</sub> = 11 828.99 Da M <sub>theoretical</sub> = 11 885.91 Da Δmass = 0.076 Da -log P-score = 135.1 % residue cleavage = 59 Matching fragments = 93	N A G W F E L S K S S D N Q F R F V L K A G N G E T 25 26 I L T S E L Y T S K L S A E K G I A S V R S N S P 50 51 Q E E R Y E K K T A S N G K F Y F N L K A A N H Q 75 76 I I G S S Q M Y A T A Q S R E T G I A S V K A N G 100 101 T S Q T V K D N T C

# DiagoTop: a new computational pipeline

- Compares TDP datasets and extract discriminative MS/MS spectra without database search



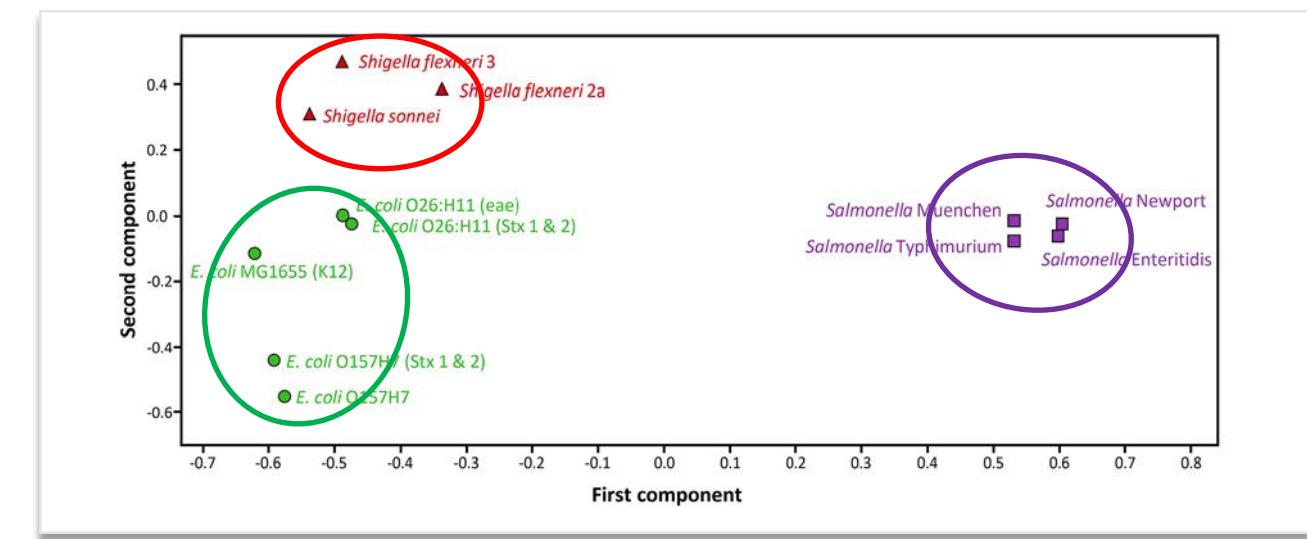
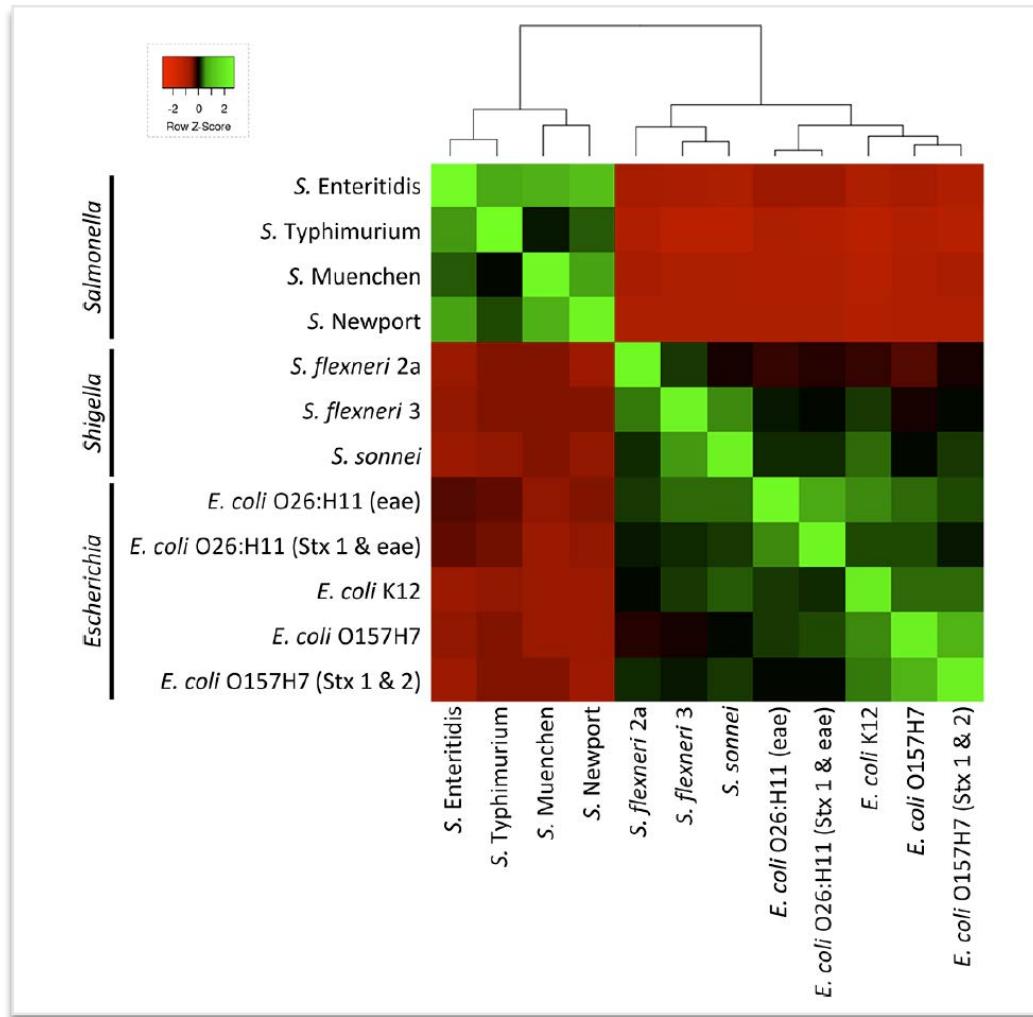
# Results for the 12 enterobacterial strains

Strains	DiagnoTop		ProsightPD	
	Total clusters	Exclusive clusters	Total proteoforms	Exclusive proteoforms
<i>Salmonella</i> Enteritidis	1750	1156	1190	521
<i>Salmonella</i> Typhimurium	1732	1716	863	375
<i>Salmonella</i> Newport	1342	654	578	139
<i>Salmonella</i> Muenchen	1222	706	473	147
<i>Shigella</i> sonnei	1319	560	432	84
<i>Shigella</i> flexneri 2a	1248	759	532	189
<i>Shigella</i> flexneri 3	1277	618	430	77
<i>E. coli</i> O157:H7 (Stx 1 & 2)	1335	537	454	58
<i>E. coli</i> O157:H7	1366	475	604	105
<i>E. coli</i> O26:H11 (Stx 1 & eae)	1443	659	499	97
<i>E. coli</i> O26:H11 (eae)	1617	782	644	175
<i>E. coli</i> K12	1220	333	478	82

Higher number of  
discriminative features  
with DiagnoTop

Reduced computing time  
for DiagnoTop  
(6h30 vs 10 days)

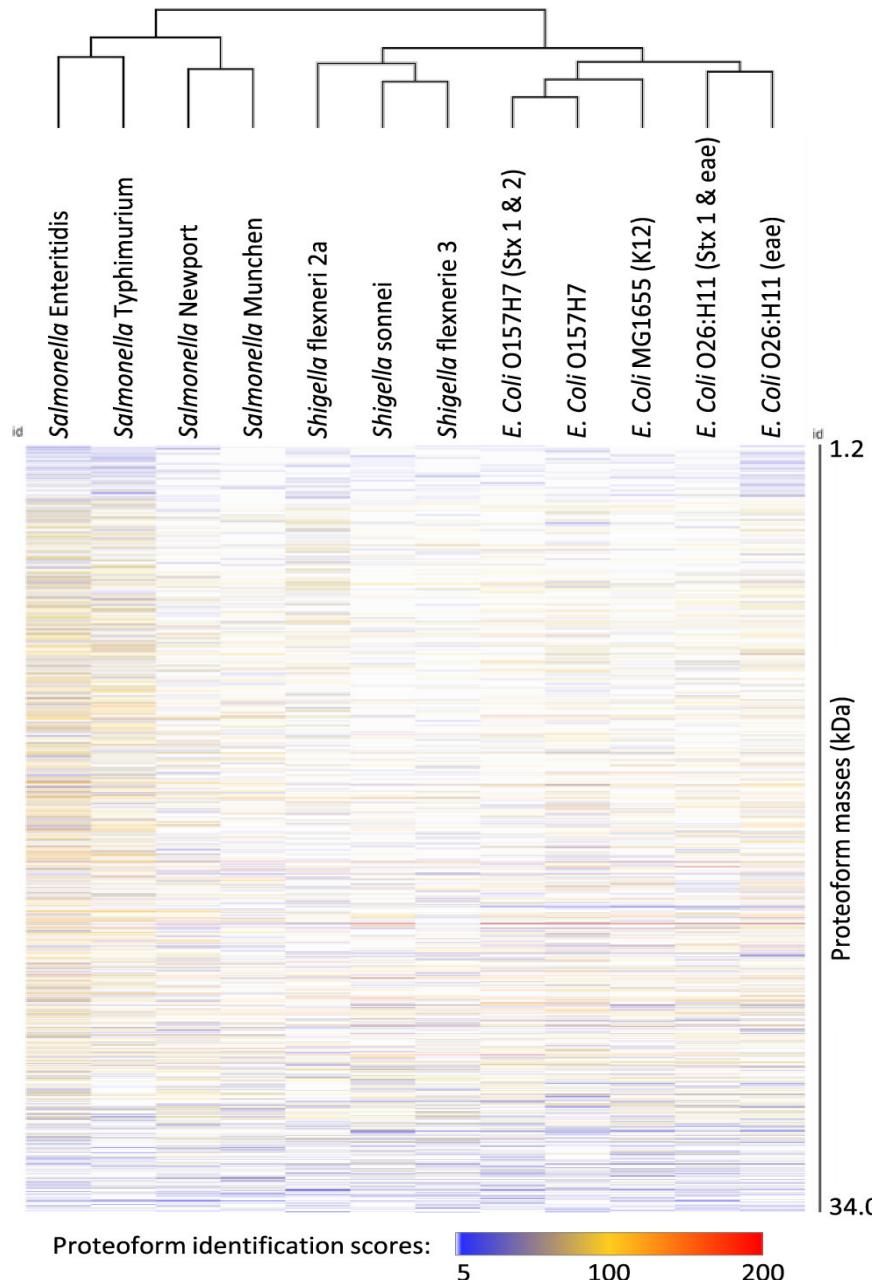
# Data visualization



Discrimination of  
enterobacterial pathogens  
without database search

# Most recent developments: DiagoMass

P. Carvalho, M. Santos



Phylogenetic tree built  
in 12 minutes!

# Future direction

---

## Improved Instrumentation (higher mass)



**Orbitrap Tribrid Eclipse: PTCR  
(2021)**



**Modification of a Q-  
Exactive to install an  
Omnitrap**

## Improved Data Analysis



**Dr. David Tabb**



# Acknowledgements

---



## Mass Spectrometry for Biology Unit Institut Pasteur

D. Borges-Lima	M. Rey
J. Dhenin	D. Tabb
M. Duchateau	<b>J. Gault</b>
<b>M. Dupré</b>	M. Gant
C. Malosse	K. Druart

**Pathogenesis of Vascular Infections Unit**  
**Institut Pasteur**  
**G. Duménil**  
**Y. Kong**

**Computational Proteomics**  
**Carlos Chagas Institute, Fiocruz, Brazil**  
**P. Carvalho**  
**M. Santos**

