

Functional analysis of natural microbial consortia
using community proteomics
利用群体蛋白质组学对天然微生物类群进行的功能分析

Contents



Part one 研究背景



Part two 微生物蛋白质组学



Part three 微生物群落的蛋白质组学



Part four Strain-resolved蛋白质组学



Part five 复杂生物系统中的蛋白质组学

The slide features a solid blue background. In the top-left and bottom-right corners, there are decorative geometric shapes consisting of overlapping triangles in white and yellow. The main title is centered in the middle of the slide.

研究背景

Part one

如今，我们对微生物群落的代谢功能和进化动力学知之甚少，然而，随着近段时间全面的测序技术的发展，我们能够以全新的视角在分子水平上对微生物群落如何塑造生物圈有一个深入的认识。我们在此探索了微生物生态的趋同性，基因组学，生物质谱，和微生物群体学的新领域——蛋白基因组学。我们探讨了蛋白基因组学的首次应用，以及将其应用在生理学、生态学、和群体微生物进化的研究上的潜能。

微生物分布于：

人体、土壤、植物根系、海洋和极端环境（例如酸性矿物废水——ADM和高温系统）中

存在的问题：

不能人工培育、

分离菌株与自然环境的行不一致

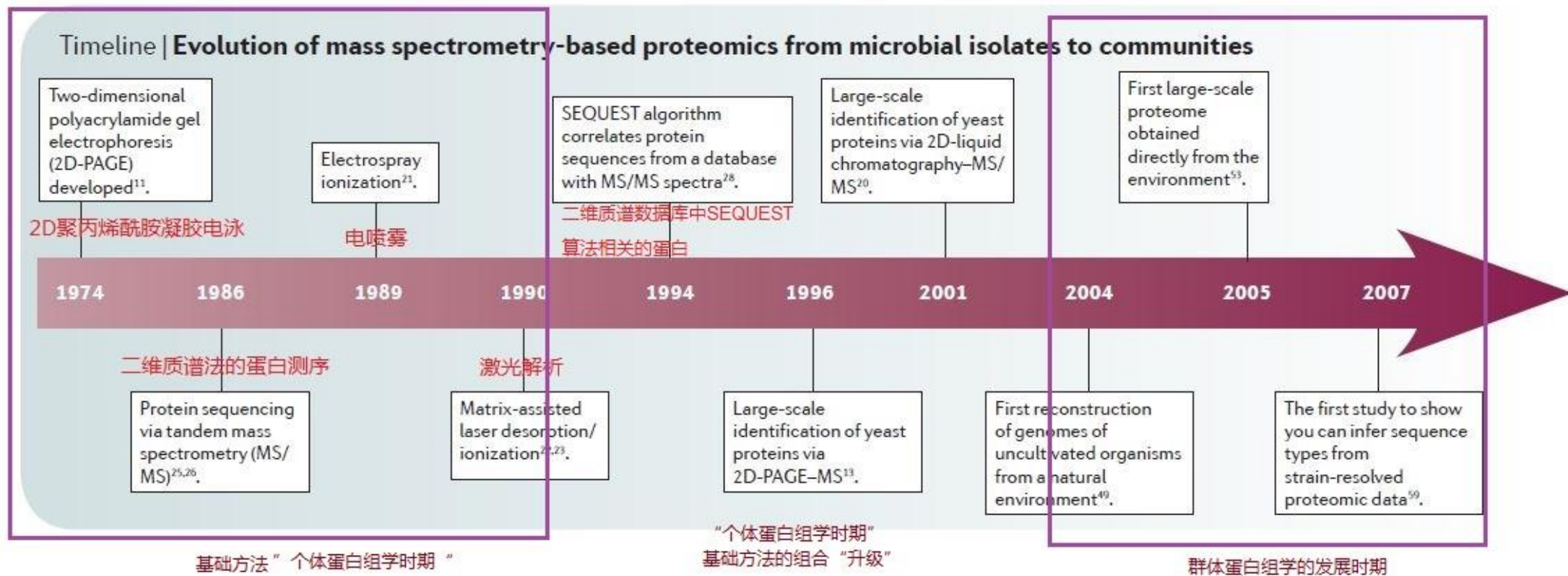
组学：

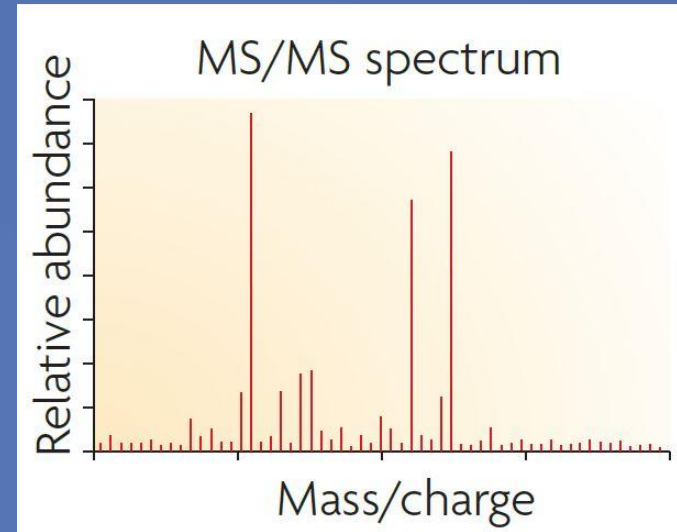
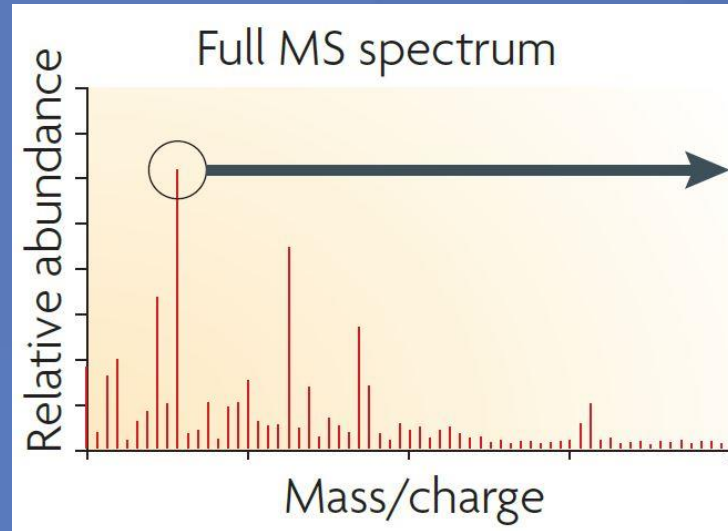
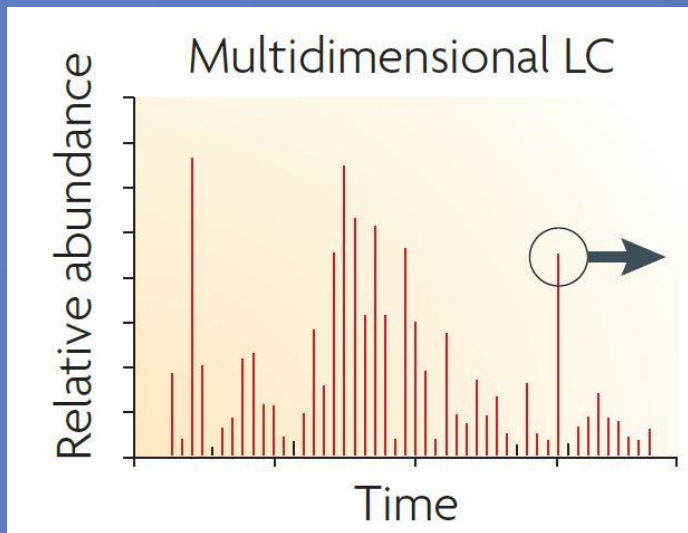
代谢组学、群体微生物蛋白组学、
群体微生物蛋白基因组学

微生物蛋白质组学

Part two

基于质谱分析的蛋白组学演变过程





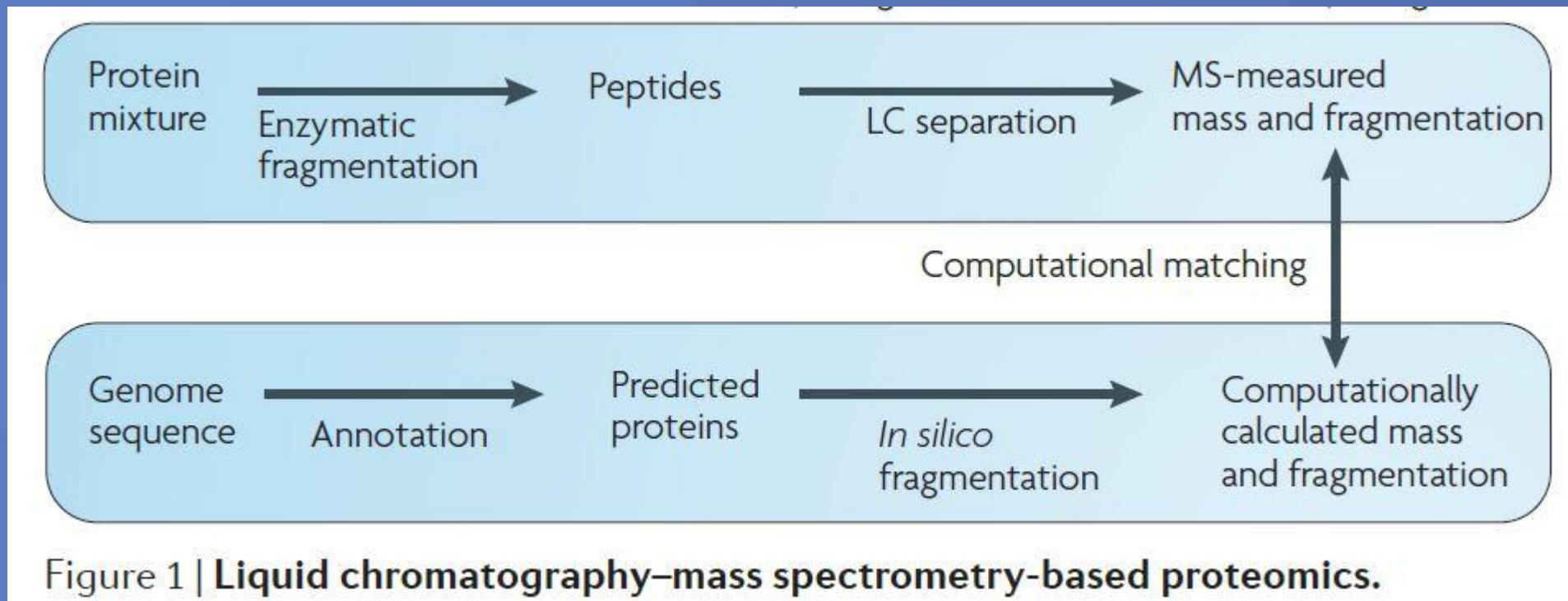


Figure 1 | **Liquid chromatography-mass spectrometry-based proteomics.**

液相色谱法——以质谱为基础的蛋白质组学

蛋白质混合物被酶促分解(例如,使用胰蛋白酶)——生成肽混合物——
多维液相色谱分离(LC)——通过使用质谱(MS)评估肽,之后通过二维
质谱(MS / MS), 打碎分散获得序列信息。

基因组序列数据的注释——计算机预测的蛋白数据库——计算机模
拟分解蛋白, 形成肽段

通过电脑分析质谱所生成的序列与计算机生成序列, 进行匹配

微生物群落的蛋白组学

Part three

对应的基因组序列数据的可靠性是蛋白组学从独立过渡到群体的关键。

从2004年开始，在许多群落基因组都被完整测序后，群体蛋白组学得以发展。

AMD生物膜原理

第一次大规模蛋白质组学实验

EPBR沉淀物

AMD生物膜是第一个因为群体中的优势物种而保证基因序列信息有效性的生态系统（BOX 1a, top line）。AMD是一种化能自养的生物膜，只受到少数类群的控制。它可以使用较低的能源（需氧型铁氧化）保证微生物群落的数量与丰度，因此成为理想的微生物群落培养方法，群体蛋白组学实验的首选。

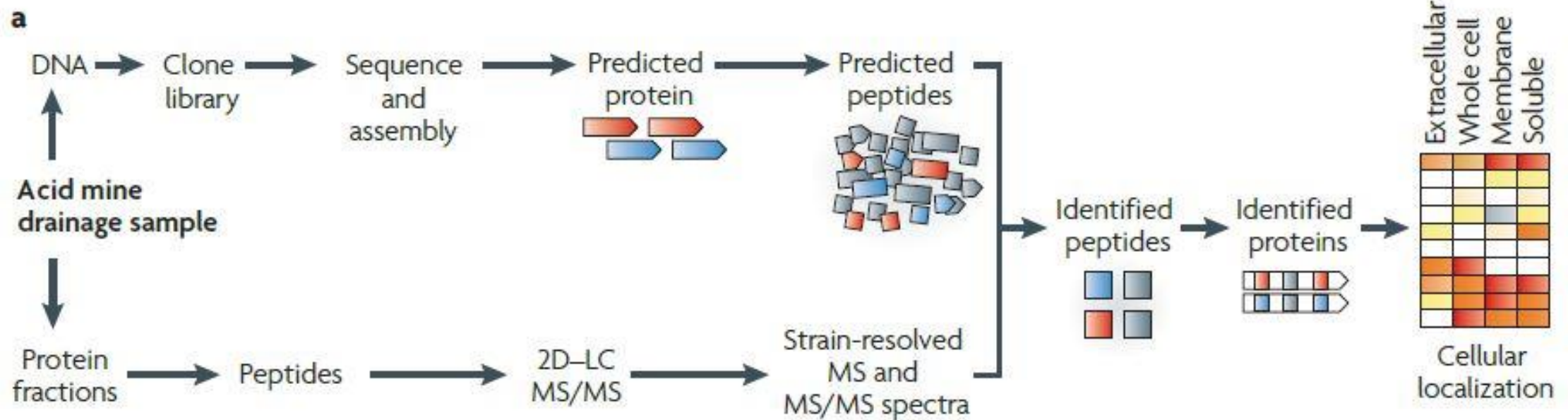
第一次大规模蛋白质组学实验，利用AMD生物膜欲得到系统水平上的蛋白质丰度，但当时没有对应的群体基因组学数据。多肽和蛋白质的识别（BOX 1a, bottom line）反而依赖于基因组序列。尽管数据库中与实际样本中的预测蛋白序列存在差异，但多肽还是可以与MS / MS谱匹配，从而识别了5个丰度群体中的2000多个蛋白质。样本中丰度最高的生物体——Leptospirillum group II，有50%的蛋白质被识别。实验证实了570种预测蛋白的存在，揭示了群体中不同生物体的不同功能，和不同基因区域的多种蛋白表达。

蛋白组学也被应用于研究废水污泥微生物群落，用于增强生物除磷(EPBR)。与AMD相比，EPBR具有物种丰度分布不均的特征。用lc-MS/MS和AMD改善后的方法进行后续分析，识别了2300多个蛋白质，使以EPBR为中心的代谢途径的扩展分析成为可能。

Strain-resolved 蛋白质基因组学

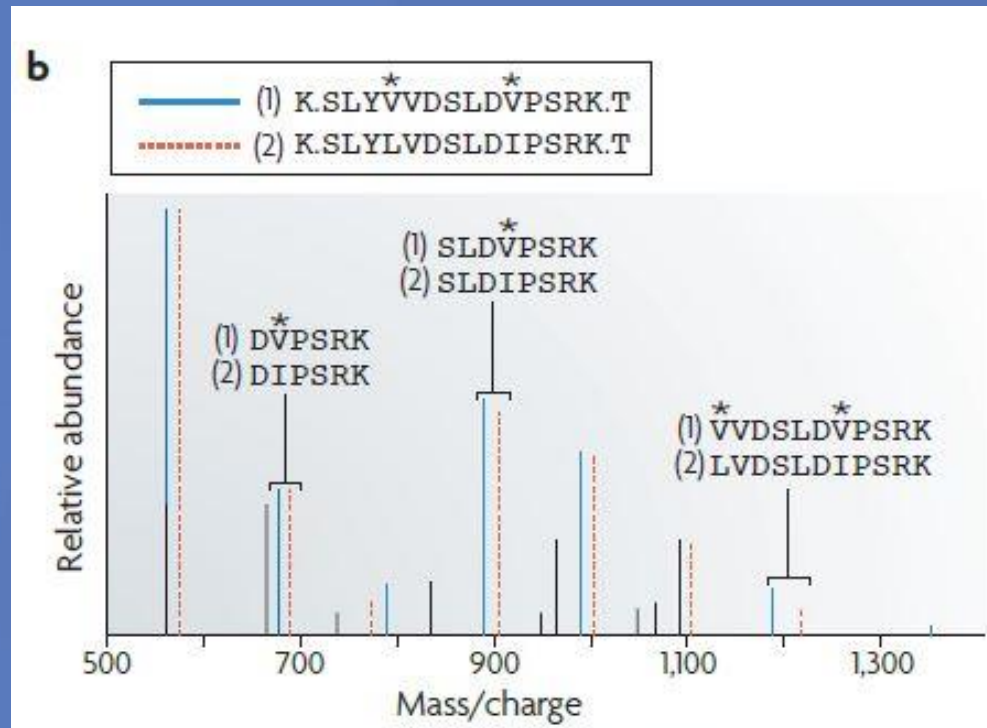
Part four

Box 1 | Strain-resolved proteomics from environmental samples



蛋白基因组学方法依赖于同一样本的基因组学和蛋白质组学数据(如上图)。DNA是从生物样品中提取,破碎,克隆和测序,由此产生的序列被读取装配/或被抛弃。基因注释后,建立蛋白质序列数据库,同时在一个在电脑上模拟胰蛋白酶分解预测的蛋白质,生成肽数据库(图片上部)。从相同或相似的生物样本,提取群体总蛋白,然后用胰蛋白酶消化。通过二维nano—液相色谱法(LC)和二维质谱分析(MS / MS)分离肽(见图1底部),光谱与数据库中的肽匹配,过滤后得到特定多肽的列表。

某些肽(唯一的肽,图中红色和蓝色的)在整个数据库中只出现在唯一的蛋白质中,据此,可以明确地追踪回相应的蛋白质,因而可以成为可靠的蛋白质鉴定。非唯一肽(灰色)不能用于唯一地标识一个蛋白质,但是这些数据用于计算蛋白质覆盖率和丰富措施。确定的蛋白质被放置回它们来源的生物的基因组,从而进行生物数据的挖掘。而且因为最初蛋白质混合物是分馏(胞外、胞内和膜上)提取的,分离的数据可以提供蛋白质定位的信息。



*代表氨基酸被替换了

通过质谱分析，由于其高质量准确性证明了覆盖两个光谱密切相关肽(见图b)通过强调主峰和连接他们相应的氨基酸序列(蓝色为一个肽转化谱，红色是另外一个的)。在这个例子中,约14道尔顿转变由于异亮氨酸和缬氨酸的替换第二个转变的另外14个道尔顿的替换缬氨酸,亮氨酸。由于质量精度高父离子测量和MS / MS的测量,一个氨基酸的变化在两个同源多肽可以很容易地分辨,从而使strain-resolved蛋白质组学成为可能。

复杂生物系统中的蛋白质组学

Part five

挑战：
特征群落产生高度
可变的丰度水平
(动态范围) 的特
征蛋白。

复杂群落中深度蛋白取样

样本较大时现有技术提供的信息可靠性

当蛋白质丰度、生物丰度和基因型不同时，定量比较复杂环境样品中蛋白质的丰度

代谢网络的动力学和分子水平上的群落功能

所测多肽的蛋白质来源和生物体来源

Table 1 | Overview of meta- and community proteomic studies

Microbiome	Number of peptides/ proteins identified*	Protein/peptide separation method	MS platform	Peptide identification method	Ref.
Ocean	184/NA [†]	2D-PAGE, 2D nano-LC	LCQ, MS/MS	Spectral matching, de novo	76
Acid mine drainage	6,188 [‡] /2,033 (2p) 6,931 [‡] /5,090 (1p)	2D nano-LC	LTQ MS/MS	Spectral matching	52
Lake and soil	NA/513 (1p) [‡]	2D nano-LC	Q-ToF, MS/MS	Spectral matching	65
Estuary	7/3 (2p)	2D-PAGE + LC	Q-ToF, MS/MS	De novo	77
Ocean	3/1 (2p) [‡]	1D-PAGE	MALDI-ToF, MS	Spectral matching	62
Riftia symbionts	NA/220 (2p) [‡]	2D-PAGE, 1D-PAGE + 2D nano-LC	MALDI-ToF MS, Q-ToF MS/MS	Spectral matching	78
Infant gastrointestinal tract	11/1 (1p)	2D-PAGE	MALDI-ToF MS	De novo	79
Acid mine drainage	8,137 [‡] /3,234 (2p)	2D nano-LC	LTQ MS/MS	Spectral matching	59
Waste water treatment reactor	NA/109 (2p) [‡]	2D-PAGE	MALDI-ToF, MS/MS	Spectral matching, de novo	80
Contaminated soil/groundwater	NA/59 (1p) [‡]	1D + 2D-PAGE + LC	MS/MS	Spectral matching	66
Sludge	NA/46 (2p) [‡]	2D-PAGE	MALDI-ToF MS, Q-ToF MS/MS	Spectral matching	56
Sludge	4,472 [‡] /2,378 (2p)	2D nano-LC	LTQ MS/MS, Orbitrap, MS/MS	Spectral matching	57
Sludge EPS	50/10 (1p) [‡]	1D-PAGE + LC	4000Qtrap, MS/MS	Spectral matching	81
Ocean	6,533/1,042 (1p-2p)	2D nano-LC	LTQ MS/MS	Spectral matching	63
Acid mine drainage	NA/2,752 ^{**} (2p)	2D nano-LC	Orbitrap, MS/MS	Spectral matching	60
Gut	NA/2,214 (2p)	2D nano-LC	Orbitrap, MS/MS	Spectral matching	64

对于群落或者宏蛋白质组学研究2D lc-MS/MS方法更好，有时SDSPAGE分离有优势。

生物多样性(物种丰富度和丰度)

- 可观的丰度水平的表现在于：
几个主要生物 or 所有生物？
独立的参考基因组 or 相关的宏基因组？
- 对低丰度物种的分析具有挑战性：
细胞富集技术，但存在缺陷；
大规模测序

土壤和沉积物样品

土壤中只有一小部分蛋白质已经被确定或被量化，因此限制了从生物角度对土壤系统的理解。

人类微生物组项目

例如,人类肠道所包含密集, 复杂, 多样的微生物群落, 称为肠道微生物组, 对健康和疾病至关重要。用蛋白组学的方法对人类双胞胎的粪便样本中的微生物群落进行蛋白质组的测量, 发现了一些未知的蛋白和微生物途径。

蛋白质提取中优化偏好性

例如, 由于溶菌作用, 低丰度群落物种的蛋白质组有时可以优先取样来降低偏好性。

基因组注释的质量

影响蛋白质识别的因素：

- 缺少或错框预测蛋白质——利用蛋白质数据库避免
- 基因起始和信号肽预测错误

小质量样本分析

例如，溶解在微型离心管和低流量柱色谱分析，在微型芯片平台直接连接MS系统对进行细胞溶解、蛋白质消化和肽分离。

动态范围

在过去的5年里lc-MS系统的标准蛋白质组学的动态范围增加1-2个数量级。这种改进的实现是通过色谱提纯的方法，将多肽更好的连接在新一代的MS仪器上，由此可以保留系统中大量的离子，产生更好的灵敏度。

准确定位

准确定位的核心问题是有效的蛋白质组学方法。2D-lc-MS/MS方法存在假阳性和假阴性识别从而影响准确性。现有的质量精度高的MS系统利用从头测序和序列标签两种方法保证了一定的准确率，但同样有待进一步改善

尽管群体蛋白质组学领域处于起步阶段，但人们越来越清楚地认识到微生物分离技术的发展可以被扩展到这些更复杂的系统。现有技术可以识别总体蛋白质包括至少1%的群体且与其紧密相关的基因组序列也可用的，但未来技术的进步应该使群体蛋白质组学可以应用于不那么丰富的总体。这将允许扩展现有对群体功能分区、资源竞争和strain-resolved适应的研究。以蛋白质为中心的宏蛋白质组学方法也将成为研究微生物生态系统的重要工具。然而我们需要聚焦于群体蛋白质组学，同时推测关于基因型和识别蛋白质的动态。只有动态整合基因组学和蛋白质组学数据，我们才能进行不同样本和生态系统间，群体蛋白质组学数据的有意义的定量比较。