

探测能量状态发生变化的蛋白或配体与蛋白亲和力的方法

技术领域

本发明属于（蛋白质组学方向）功能蛋白质组学的研究领域，具体涉及在细胞或组织裂解液等复杂蛋白质混合物中检测能量状态发生变化的蛋白或鉴定配体的结合蛋白或测定配体与蛋白质的局部亲和力。

背景技术

蛋白质是生命活动的主要承担者，传统的定量蛋白质组学多用于评价不同样品中蛋白质丰度的差异，而众多疾病的发生（例如帕金森综合征和阿尔兹海默症）或配体（包含药物，内源性代谢物等）与蛋白相互作用，特定蛋白质的丰度并没有发生改变，主要是蛋白的折叠状态，蛋白质的翻译后修饰或蛋白质的功能发生了改变 (De Souza N , et al, Curr Opin Struct Biol, 2020,60:57-65; Cappelletti V, et al,Cell,2021,184(2):545-559)。而这些丰度外的蛋白质性质的改变，则会影响蛋白质的热力学稳定性也即能量状态，因此发展能探测蛋白质能量状态变化的方法在疾病的诊断，代谢分子作用通路分析，药物靶标的发现等领域均有重要意义。

自上而下 (Top-down)的质谱分析手段可以对完整蛋白直接分析进而监测蛋白的能量状态的变化 (Sheng Y et al, International Journal of Mass Spectrometry, 2010, 300 (2011) 118-122)，但是该方法通量低，主要用于纯化的蛋白，较难在复杂的蛋白混合物体系中应用。目前也有一些非质谱的方法来探测蛋白质能量状态的变化，例如荧光共振能量转移 (Heyduk T, et al, Curr Opin Biotechnol, 2002, 13, 292-296)，核磁共振 (Sakakibara D, et al, Nature, 2009, 458, 102-105)等。这些技术适用于纯化的蛋白，因此也无法在蛋白质组学层次上应用。

自下而上 (Bottom-up) 的质谱技术是将大的蛋白片段酶解为小肽段再进入质谱进行质谱分析，该技术能够同时鉴定或定量上千种甚至上万种蛋白质。近年来一系列结合自下而上的质谱技术来高通量分析蛋白质能量状态变化的方法，受到了国内外的广泛关注。这类方法主要基于蛋白质能量状态变化后，对热以及变性剂的耐受能力发生改变，因此施加同样强度的热或变性剂刺激，能量状态低的蛋白质由于更容易保持其原有构象，因此更不容易沉淀或结构展开，通过分析两组或多组样品中沉淀或结构展开的蛋白的量的差异，可以确定能量状态发生变化的蛋白。这类技术包含：细胞热转移分析 (CETSA, Cellular Thermal Shift Assay, Martinez Molina, et al, Science, 2013, 341(6141):84-87)，热蛋白质组分析技术 (TPP, Thermal Proteome Profiling, Savitski M M, et al, Science, 2014, 346 (6205): 55-+), 脉冲酶解法 (PP, Pulse Proteolysis, Chiwook P, et al, Nature methods, 2005, 2(3):207-212) 等。该类探测蛋白质能量状态变化的技术已成功的实现了对药物作用靶蛋白的鉴定 (Savitski M M, et al, Science, 2014, 346(6205):55-+; Perrin J, et al, Nature Biotechnology, 2020, 38(3):303-308)，蛋白质与蛋白质相互作用的分析 (Tan C S, et al, Science, 2018, 359(6380):1170-1177) 以及功能性翻译后修饰组分析 (Huang J X, et al, Nature methods, 2019, 16, 894-901)。但是此类方法在蛋白质层次上分析蛋白质能量状态的变化，因此无法确定蛋白质能量状态发生变化的区域。另外对于多结构域的蛋白，蛋白局部稳定性的变化可能对蛋白整体稳定仅有微小的影响，因此在蛋白质层次上可能较难捕捉到其稳定性的变化 (Mateus A,

et al, Proteome Sci,2016,15:13)。

限制性酶切法 (LiP, Limited Proteolysis)很久之前被用来研究蛋白质构象, 该方法基于蛋白质处于不同构象时其耐受酶切的能力不同, 采用较少量的非特异性酶对蛋白质进行轻微酶切, 使得酶切位点主要发生在处于高能量状态的无序区域, 产生较大的蛋白片段; 当蛋白质构象发生改变时 (例如无序区域变得更加有序), 构象变化部分对酶切的耐受程度发生变化, 通过免疫印迹检测产生的蛋白大片段的丰度差异确定蛋白质构象变化的区域。由于蛋白质能量状态的变化通常伴随着蛋白质构象的变化, 因此该方法也能用来研究蛋白质能量状态的变化以及确定蛋白质能量状态变化的区域 (Antonio Aceto, et al, The International Journal of Biochemistry & Cell Biology,1995,27(10):1033-1041; Polverino de Laureto, et al, J Mol Biol,2003,334(1):129-141)。然而由于大的蛋白片段难以直接在质谱自下而上的分析模式下进行通量化分析, 因此该技术局限于纯蛋白质的表征。

Paola等人将限制性酶切技术与自下而上的质谱分析技术结合(Feng Y H, et al, Nature Biotechnology,2014,32(10):1036-+), 将大片段的限制性酶切产物 (即上述蛋白大片段) 进一步用变性剂变性, 进行第二步完全胰蛋白酶酶切, 最终产生含有一个非特异性酶切位点的半胰蛋白酶切肽段及不含有非特异性酶切位点的全胰蛋白酶切肽段, 在质谱自下而上的分析模式下通过定量蛋白质组学分析半/全胰蛋白酶切肽段在两组或多组样品中丰度的差异, 来判定构象发生变化的蛋白并通过分析肽段在蛋白上所处的位置确定构象发生变化的区域。该方法一定程度了实现了对蛋白质能量状态变化的规模分析, 并在相对简单的生物体系中捕捉到了代谢物与蛋白质的相互作用(Piazza I, et al, Cell, 2018,172(1-2):358-372)以及由于翻译后修饰, 热刺激以及低渗压力导致的蛋白质结构的变化(Cappeletti V, et al,Cell,2021,184(2):545-559)。但该方法中两步酶切大大增加了样品的复杂度, 容易产生干扰肽段, 因此该方法大多数应用在相对简单的生物体系。Paola等人 (2020) 又发展了LiP-Quant(Piazza I, et al, Nature Communication,2020,11(1):4200), 该方法在LiP-MS的基础上需要多个配体浓度来增加靶蛋白鉴定的可信度。LiP-Quant在一定程度上实现了在哺乳动物细胞裂解液中对配体的结合蛋白及结合区域的鉴定, 但是由于两步酶切导致含有一个非特异性酶切位点的半胰蛋白酶切肽段难以鉴定, 此外含有一个非特异性酶切位点的肽段对配体结合的响应程度较小, 因此其对配体结合靶蛋白鉴定的灵敏度非常有限, 例如在对广谱激酶抑制剂星孢菌素 (staurosporine) 靶蛋白的鉴定中, 如果要求在所有鉴定到的靶蛋白中激酶的比例在80%以上, LiP-Quant仅能在HeLa细胞裂解液中鉴定到9个激酶靶蛋白(Piazza I, et al, Nature Communication,2020,11(1):4200), 而本发明同样在HeLa细胞裂解液中能鉴定到111个激酶靶蛋白(见实施例3)。

上述基于酶切的蛋白质构象研究方法基于采用较少量的酶, 使得酶切位点局限于高能量状态的无序区域, 不同构象的蛋白质产生不同的蛋白质大片段, 通过免疫印迹或变性剂辅助的二次酶切检测蛋白大片段产生的差异从而来判断构象发生变化的蛋白。没有提示利用 (较) 大量的酶使得蛋白质发生破坏性酶切 (即蛋白的高能量状态的结构首先被酶切破坏, 进而诱导蛋白处于低能量状态的酶切位点暴露出来, 在大量酶的存在下, 低能量状态的结构也被酶切破坏), 直接产生分子量小于 5KDa 的具有两个能反映蛋白质区域稳定性酶切位点的肽段, 在质谱自下而上的分析模式下, 通过

定量蛋白质组学分析该部分肽段丰度的差异确定能量状态发生变化的蛋白质及区域、配体与蛋白结合的局部亲和力。针对上述问题，我们发明了一种能在复杂的蛋白混合物体系中高灵敏度确定能量状态发生变化的蛋白及区域、并且测定配体与蛋白局部亲和力的方法。该方法通过破坏性酶切直接产生拥有两个能反映蛋白质区域稳定性酶切位点并适合自下而上质谱分析的小肽段，与 LiP-MS 中使用两种酶进行两步酶切的产物相比，大大降低了样品复杂度，且产生该小肽段，蛋白经历了较为充分的级联酶切（即蛋白的高能量状态的结构首先被酶切破坏诱导低能量状态的结构进一步酶切），在此级联酶切过程中每一级酶切差异的累积将使得小肽段对蛋白能量状态的变化或者配体结合有更大幅度的响应，使检测到所属肽段的变化幅度大大增加，因此该方法在蛋白质能量状态变化的检测方面具有极高的灵敏度。例如与 LiP-Quant 相比，同样要求在所有鉴定到的 staurosporine 靶蛋白中激酶比例大于 80% 时，该方法能在 HeLa 细胞裂解液中鉴定到 111 个激酶，比 LiP-Quant（9 个激酶）提高了 12.3 倍（实施例 3）。该方法以肽段为单位输出定量结果，在使用多个配体剂量时还可以用于计算蛋白局部区域与配体结合的亲和力。

15 发明内容

本发明的目的在于发展一种能在复杂体系中高灵敏度探测能量状态发生变化的蛋白及蛋白区域或测定配体与蛋白结合局部亲和力的方法，即以肽段为中心的蛋白局部稳定性探测方法(PEptide-centric Local Stability Assay)，以下简称 PELSA。

其原理在于蛋白质的能量状态发生变化后，蛋白质的区域或全局稳定性发生变化，因此其抵抗外界破坏的能力发生了变化。本发明利用酶切是一个破坏性过程，蛋白质能量状态发生变化之后对酶切产生的破坏的抵抗程度发生变化。在高的蛋白酶浓度下，蛋白首先被切为较大的亚基，由于酶切是一个破坏性过程，处于高能量状态或无配体结合的该亚基的结构被破坏并迅速展开，诱导自身进一步酶切，导致蛋白质处于低能量状态的结构也被酶切破坏，最终产生拥有两个能反映蛋白区域稳定性酶切位点的，适合自下而上质谱分析的小肽段 (< 5 KDa)。而处于低能量状态或有配体结合的该蛋白区域仍然维持稳定结构，较难进入蛋白酶的活性口袋，因此该区域不容易产生小肽段。通过比较此类肽段丰度差异来确定能量状态有差异的蛋白；分析差异肽段在蛋白上的位置，能够确定能量状态有差异的蛋白区域。在该方法中蛋白在非变性的条件下直接产生含有两个能反映蛋白区域稳定性的酶切位点的小肽段，由于产生该小肽段，蛋白经历了较为充分的级联酶切，在此级联酶切过程中每一级酶切差异的累积将使得小肽段对蛋白能量状态的变化或者配体结合有更大幅度的响应。另外由于该方法对蛋白进行了较为彻底的破坏性酶切，酶切位点不局限于高能量状态的无序区域因此能够产生更多反映蛋白稳定性变化的肽段。再者，该过程仅使用一种蛋白酶，酶切肽段的复杂度大大降低，因此能覆盖到更多的蛋白，以上特点使得该方法具有很高的灵敏度。

在复杂蛋白质样品中加入配体时，配体与其目标蛋白结合会导致目标蛋白的配体结合区域的能量状态发生变化。通过该方法对整个蛋白质组能量状态的分析能确定配体的结合蛋白及配体的结合区域。

该技术用来测定配体与蛋白结合局部亲和力的原理在于未达到饱和浓度时，目标蛋白质的局部稳定性随着所加入配体的浓度变化呈剂量依赖性变化。通过检测不同配

体浓度处理的样品中酶切下来的小肽段的丰度,从而能计算出配体与蛋白区域结合的局部亲和力;

该技术检测的小肽段拥有两个反映蛋白质区域稳定性的酶切位点;该酶切位点是指蛋白质在其待测构象下,不同区域有不同的稳定性,因此有不同的酶切难易程度,在此情况下产生的酶切位点称为反映蛋白区域稳定性的酶切位点。

本发明采用如下的技术方案:

(a) 将处于不同能量状态的蛋白溶液或者与不同浓度配体孵育的蛋白溶液。按照酶和蛋白一定的比例进行破坏性酶切,使大多数蛋白能够产生分子量小于 5KDa 的含有两个能反映蛋白区域稳定性酶切位点的小肽段;

(b) 酶切下来的小肽段在质谱自下而上的分析模式下定量分析;

(c) 分析酶切下来小肽段丰度的差异确定能量状态发生变化的蛋白和蛋白区域以及配体与蛋白结合的局部亲和力。

具体包含以下步骤:

- 1) 步骤(a)中所述蛋白溶液必须保持蛋白的待测构象,若通过裂解获得组织或细胞裂解液,则应采取温和的方式裂解组织或细胞,包括液氮反复冻融法,液氮研磨法以及匀浆法等中的一种或二种以上,所采用的裂解液也应该能保持蛋白的待测构象。
- 2) 在研究配体的结合蛋白时,将含有蛋白的生物样品分为两组,其中一组加入一定浓度的配体分子作为实验组,另一组加入空白溶剂作为对照组;但不局限于两组,配体组也可以是多组含有不同浓度配体的样品。
- 3) 按照酶与蛋白的质量比例 1:1-1:50 加入酶,在一定温度下酶解 0.5 min-60 min,酶切在震荡器上进行,转速 1000 rpm-1500rpm。
- 4) 将步骤3)中酶切后的样品,在金属浴中 100℃,加热 5min,终止酶解。
- 5) 步骤4)加热后的样品有沉淀产生,将其用 6M 盐酸胍溶解,并加入终浓度为 10mM 的三(2-羧乙基)膦即 TCEP 和终浓度为 40mM 的氯代乙酰胺即 CAA 在 95℃,加热 5min,使酶切之后的样品烷基化。
- 6) 步骤5)烷基化后的样品转移到 10KDa 的超滤管中,14000g,离心 20min-1h,收集离心下来的肽段,再用 200μL pH8.2 缓冲液洗超滤膜,合并超滤下来的肽段。
- 7) 步骤6)产生的肽段进行质谱定量,定量方式包括标记定量以及无标记定量。
- 8) 在肽段层次上分析蛋白的定量结果,用差异显著性或差异倍数第一大、第二大、第三大或第四大的肽段的差异显著性或差异倍数和作为该蛋白的差异显著性或差异倍数,来筛选可信的差异蛋白。
- 9) 将有差异的肽段归属到对应的蛋白序列或结构上,从而确定能量状态发生变化的蛋白区域;将不同配体浓度下肽段的变化倍数以及对应的配体浓度拟合为四参数对数方程 $Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{-(\text{LogEC}_{50} - X) * \text{Hill Slope}})$,计算配体与蛋白之间的局部亲和力即该方程里面的 EC_{50} 。

与已有的方法相比,本发明的创新和优点体现在如下几个方面:

- 1) 本发明即 PELSA 利用酶切是一个破坏性过程,采用较大量的酶,使得蛋白在非变性条件下,利用酶切的破坏效应使得蛋白质暴露的处于高能量状态的结构首先被酶切破坏,进而诱导蛋白质处于低能量状态的酶切位点暴露出来,在大量酶的

- 存在下,低能量状态的结构也被酶切破坏,使大多数蛋白能直接产生分子量小于 5KDa 的小肽段,并通过质谱对该部分肽段进行鉴定和定量。PELSA 所鉴定的小肽段由上述类似于级联反应的破坏性酶切产生并能直接进行自下而上的质谱分析,若蛋白质能量状态发生变化或者结合配体时,每一级级联酶切速度差异的累积使该小肽段有更大幅度的响应。且该方法使用较大量的酶,因此能产生更多对配体结合有响应的肽段。例如在实施例 2 中 PELSA 鉴定到的对配体有响应的肽段的数目是对比方法 LiP-MS 的 2-5 倍;且该部分肽段在 PELSA 中测得的响应幅度是在 LiP-MS 中的 4-6 倍。
- 5
- 2) 适用范围广。与传统的基于亲和色谱的方法相比,本发明所述方法在确定配体结合蛋白时,无需对配体分子进行化学修饰,理论上可以用于任何配体。此外对配体结合靶蛋白的鉴定不依赖于两者之间结合的强弱,因此可适用于低亲和力配体分子结合靶蛋白的鉴定。例如在实施例 4-6 中成功实现了对代谢分子叶酸,亮氨酸以及二酮戊二酸 (α KG) 靶蛋白的鉴定。其中实施例 6 中对 α KG 靶蛋白的鉴定代表目前仅在一次实验中鉴定到的最多的已知 α KG 靶蛋白数目。
- 10
- 3) 能够确定能量状态不同的蛋白区域或配体的结合区域。与目前广泛使用的 TPP (Thermal Proteome Profiling) 方法相比,该方法能通过分析发生变化的肽段所处蛋白的位置,进而确定能量状态发生变化的区域或配体的结合区域。如在实施例 3 所示,该方法实现了激酶抑制剂 staurosporine 在 150 个激酶上的作用区域的鉴定。
- 15
- 4) 灵敏度高。在蛋白层次捕捉蛋白质能量状态或者配体结合导致的酶切差异时,局部差异被平均化到整个蛋白上;该方法直接在肽段层次上捕捉蛋白质能量状态或者配体结合诱导的局部差异因此该方法更能捕捉到微小的差异。此外,该方法对配体作用区域的确定不依赖特殊性质肽段的鉴定(比如在 LiP-MS 中必须鉴定到含有非特异性酶切位点的肽段),因此更容易鉴定到有差异的肽段。例如在实施例 3 中,PELSA 鉴定到的广谱激酶抑制剂 staurosporine 的激酶靶蛋白数目是 TPP 的 2.1 倍,是 LiP-Quant 的 12.3 倍。
- 20
- 5) 评估配体和蛋白结合的局部亲和力。将蛋白样品分为多组,并分别加入不同浓度的配体。在未达到饱和浓度时,靶蛋白对酶切破坏的耐受能力随着配体的浓度的改变而改变,以酶切下来肽段的丰度的变化倍数作为纵坐标,以配体的浓度作为横坐标拟合四参数对数方程,即能得到亲和力信息。一个蛋白可能鉴定到多条肽段,每个肽段的亲和力曲线反映的是配体对该区域的局部结合亲和力。例如在实施例 11 中,PELSA 测得配体与蛋白之间的局部亲和力与热泳动技术测得的亲和力非常相近。
- 25
- 6) 有广阔的应用前景。蛋白质与配体结合,蛋白翻译后修饰,蛋白结构的改变以及蛋白与蛋白相互作用等均会导致蛋白能量状态的改变。因此该方法可以应用在各种外界扰动导致蛋白能量状态改变的体系中,为解决多种生物学问题提供了一个强大的工具。实施例 7 展示了 PELSA 用于分析由于蛋白与蛋白相互作用导致蛋白能量状态的变化;实施例 8 展示了 PELSA 用于分析由于含有翻译后修饰的肽段与蛋白相互作用导致蛋白能量状态的变化;实施例 9 展示了 PELSA 用于分析由于金属离子与蛋白相互作用导致蛋白能量状态的变化。
- 30
- 40

附图说明

图 1: PELSA 实施流程图: 能量状态不同的蛋白 (如图所示配体结合状态下的某蛋白和无配体结合状态下的该蛋白), 按照酶与蛋白一定的比例进行破坏性酶切, 配体组由于配体与蛋白质的结合, 导致结合区域更加稳定, 因此结构更不容易舒展, 导致切下来比空白组少。酶切后对酶切产物进行变性以及烷基化处理, 超滤分离酶切下来的肽段, 进行质谱定量分析。

图 2: PELSA 应用在分析抗乳腺癌药 lapatinib 处理细胞裂解液后能量状态发生变化的蛋白质; (A) 分析 PELSA 酶切位点所处的蛋白质二级结构: PELSA 的酶切位点是指该实验鉴定到的所有肽段的 N 端和 C 端的残基, 蛋白质的二级结构信息来源于 AlphaFold, 二级结构的分类是依据于文献 (Bludau I, et al, PLoS Biol, 2022, 20(5): e3001636), 其中 HELIX 代表低能量状态的螺旋结构, STRAND 代表折叠结构, BEND 是弯曲结构, TURN 是转角结构, unstructured 是高能量状态的无序结构; (B) 100nM lapatinib 对应的蛋白层次火山图: 100nM lapatinib 或 DMSO 处理 BT474 细胞裂解液 (4 次重复) 得到的肽段的定量值, 使用 Empirical Bayes t-test 计算每个肽段的 fold change(FC)及 P value。每个蛋白中 Pvalue 最小的肽段来代表该蛋白, 以蛋白的 log2FC 及 -log10Pvalue 作散点图; (C) ERBB2 的二维稳定性变化图: 定量到的 ERBB2 的肽段的 log2FC 及其对应 ERBB2 二维序列上的位置; (D) ERBB2 的局部亲和力图: 不同 lapatinib 浓度下 ERBB2 肽段的 log2FC; (E) 1 μ M lapatinib 对应的蛋白层次上火山图; (F) 以激酶 CHEK2, SLK, RIPK2, YES1 为例展示所鉴定到的 lapatinib 的脱靶激酶的二维稳定性变化图; (G) Western blotting 的结果证明了加入 lapatinib 后 PTGES2 被稳定。

除非特别说明, 本说明书中出现的 fold change、FC (fold change 的缩写) 或 ratio 均代表实验组与对照组肽段丰度的比值。

图 3: 比较 PELSA 和 LiP-MS 对甲氨蝶呤 (MTX) 以及 SHP099 处理 HeLa 细胞裂解液后能量状态发生变化的蛋白的鉴定; (A) (左上) / (右上) 10 μ M MTX 处理 HeLa 细胞裂解液, LiP-MS/PELSA 所获得的肽段层次上的火山图; (左下) / (右下) 10 μ M SHP099 处理 HeLa 细胞裂解液, LiP-MS/PELSA 所获得的肽段层次上的火山图; (B) 在 MTX 以及 SHP099 的处理体系中, PELSA 鉴定到的对配体有响应的靶蛋白的肽段分别是 LiP-MS 的 2 和 5.25 倍; (C) PELSA 鉴定到的对配体有响应的靶蛋白 DHFR 及 PTPN11 肽段的变化倍数分别是 LiP-MS 的 4.3 和 6.4 倍; (D) 结合 MTX 的 DHFR 的 PDB 结构图, 箭头指示的肽段位于 MTX 的结合位点且处于低能量状态的螺旋区, 该肽段在 PELSA 中被检测到了显著变化, 而在 LiP-MS 中没有变化; (E) 结合 SHP099 的 PTPN11 的 PDB 结构图, 箭头指示的肽段位于 SHP099 的结合位点且位于蛋白内部, 同样, 该肽段在 PELSA 中被检测到了显著变化, 而在 LiP-MS 中没有变化。

图 4: PELSA 在鉴定能量状态发生变化的蛋白时具有极高的灵敏度 (以筛选广谱激酶抑制剂星孢菌素(staurosporine)的作用靶蛋白为例说明); (A) 20 μ M staurosporine 处理 K562 细胞裂解液 (左) 和 HeLa 细胞裂解液 (右) 对应的蛋白层次的火山图; (B) PELSA 在 K562 细胞裂解液, PELSA 在 HeLa 细胞裂解液, 已发表的 LiP-Quant 在 HeLa 细胞裂解液以及已发表的 TPP 在 K562 细胞裂解液 (以下分别简称 PELSA (K562), PELSA (HeLa), LiP-Quant (HeLa) 以及 TPP (K562)) 鉴定到的 staurosporine 的靶蛋白的

数目对比,其中横坐标代表鉴定到的总的靶蛋白数目,纵坐标代表所鉴定到的靶蛋白中激酶的数目;(C) LiP-Quant(HeLa), TPP (K562), PELSA (K562),PELSA (HeLa)鉴定到的蛋白数目,肽段数目及最终筛选出来的激酶数目;(D)PELSA (HeLa)与 LiP-Quant (HeLa)鉴定到的所有蛋白质的蛋白序列覆盖度对比图(左); PELSA (HeLa)与 LiP-Quant (HeLa) 筛选出来的激酶的蛋白序列覆盖度对比图(右边);该图提示 PELSA 能够在蛋白序列覆盖度较小时也能鉴定为靶蛋白;(E)分割小提琴图表示 PELSA (HeLa),PELSA (K562),TPP (K562)筛选出来的激酶及所有定量到的激酶的蛋白质的热熔点分布,该图说明 PELSA 对热稳定及热敏感靶蛋白也有很好的鉴定效果;(F)PELSA(K562)与 PELSA (HeLa)鉴定到的所有 staurosporine 靶蛋白的重叠情况;(G) (左) PELSA (K562)和(右) PELSA (HeLa)鉴定到激酶肽段中,位于激酶结构域的肽段的稳定性变化显著性以及其他肽段稳定性变化显著性的密度分布图(横坐标表示变化的显著性即 $-\log_{10}P$ value,该值是由 Empirical Bayes t-test 计算得到,纵坐标表示密度大小),圆环图表示在显著变化的肽段中,位于激酶结构域,非常临近激酶结构域(距离激酶结构域 ± 10 个氨基酸残基),该肽段所属蛋白没有激酶结构域以及位于激酶结构域之外的肽段所占的比例,可以看出绝大部分位于显著变化的肽段均来自于激酶结构域。

图 5: PELSA 用于分析代谢物处理 HeLa 细胞裂解液后能量状态发生变化的蛋白;(A) 50 μ M 叶酸(folate)处理 K562 细胞裂解液对应的蛋白层次的火山图;(B)DHFR 结合 folate 后 PELSA 测得的三维稳定性变化图,稳定性变化最大的蛋白区域用箭头指示;(C)加入 50 μ M folate 后 PELSA 测得的 ATIC 的三维稳定性变化图,箭头指示稳定性变化最大的蛋白片段;(D)加入 50 μ M folate 后 PELSA 测得的 MTHFR 的二维稳定性变化图;(E)加入 50 μ M folate 后 PELSA 测得的 GART 的三维稳定性变化图,箭头指示稳定性变化最大的蛋白片段;(F)加入 50 μ M folate 后 PELSA 测得的 P3H1 的二维稳定性变化图;(G)5mM 亮氨酸(leucine)处理 K562 细胞裂解液对应的蛋白层次火山图;(H) 加入 5 mM leucine 后 PELSA 测得的 LARS1 的二维稳定性变化图;(I)膜蛋白 SLC1A5 的拓扑结构图,190-212, 493-502 及 523-541 为三条定量到的肽段的始末位置,其对应的 $|\log_2FC|$ 也分别标注在对应肽段旁;(J)加入 5 mM leucine 后 PELSA 测得的 PPIP5K1 以及 PPIP5K2 的二维稳定性变化图;(K) 2 mM 二酮戊二酸 (α KG) 处理 HeLa 细胞裂解液对应的蛋白层次火山图,虚线分别表示 $-\log_{10}P$ value = 3.4 以及 $\log_2FC = -0.5$, 满足 $-\log_{10}P$ value > 3.4 且 $\log_2FC < -0.5$ 的 40 个蛋白中, 30 个是已知的 α KG 的作用蛋白;(L) 2 mM α KG 处理 HeLa 细胞裂解液后 PELSA 测得的 EGLN1, RSNB1L 以及 KDM3B 的二维稳定性变化图; PELSA 测得能量变化区域均为已知的 α KG 作用区域。

图 6: PELSA 在 HeLa 细胞裂解液中分析由于蛋白与蛋白相互作用导致蛋白局部能量状态的变化(以抗体和抗原结合为例):(A)PELSA 在细胞裂解液中鉴定抗体结合表位的示意图;(B)(左) DHFR 抗体处理 HeLa 细胞裂解液对应的蛋白质层次的火山图, (右) CDK9 抗体处理 HeLa 细胞裂解液对应的蛋白质层次的火山图;(C)DHFR 抗体处理后, PELSA 测得的 DHFR 的三维稳定性变化图(PDB:1BOZ),球状阴影部分表示抗体的结合表位,箭头指示稳定性变化最大的蛋白片段;(D)(上) DHFR 抗体处理后 PELSA 测得的 DHFR 的二维稳定性变化图,(下) CDK9 抗体处理后 PELSA 测得的 CDK9 的二维稳定性变化图;(E)CDK9 抗体处理后, CDK9 的两条肽段

NPATTNQTEFERVF 以及 NPATTNQTEFER 丰度值的变化情况。

图 7: PELSA 在 BT474 细胞裂解液中分析由于翻译后修饰与蛋白相互作用导致蛋白质局部能量状态的变化 (以鉴定酪氨酸磷酸化的识别结构域为例); (A)PELSA 在细胞裂解液中鉴定翻译后修饰识别结构域的示意图; (B)使用 PELSA 分别以 pSEEI 及 YEEI 为对照组, 计算得到每个肽段的 $-\log_{10}Pvalue(pYEEI/pSEEI)$ 及 $-\log_{10}Pvalue(pYEEI/YEEI)$, 以 $-\log_{10}Pvalue(pYEEI/pSEEI)$ 及 $-\log_{10}Pvalue(pYEEI/YEEI)$ 之和最大的肽段代表该蛋白, 做蛋白层次上的散点图; (C)使用 Pulldown 的方法分别以 pSEEI 及 YEEI 为对照组, 计算得到每个蛋白的 $-\log_{10}Pvalue(pYEEI/pSEEI)$ 及 $-\log_{10}Pvalue(pYEEI/YEEI)$, 做蛋白层次上的散点图; (D)小提琴图表示 PELSA 筛选出来的 9 个含有 SH2 结构域的蛋白, 位于 SH2 结构域的肽段及在 SH2 结构域之外的肽段的 $\log_2FC$ 的分布; (E)PELSA 筛选出来的含有 SH2 结构域的蛋白的二维稳定性变化图, 以 YES1,TNS3,PLCG1 及 GRB10 为例; (F) PELSA 筛选出来的钙离子结合蛋白的二维稳定性变化图, 以 EFHD1 和 CALM1 为例。

图 8: PELSA 用于分析金属离子 (以锌离子为例) 处理 HeLa 细胞裂解液后能量状态发生变化的蛋白; (A)30 μ M 锌离子处理 HeLa 细胞裂解液对应的蛋白层次火山图; (B) (左) 所有鉴定到的蛋白质中金属离子结合蛋白所占比例, (右) 显著稳定的 280 个蛋白中金属离子结合蛋白所占比例, 右侧的饼状图表示在稳定的金属离子结合蛋白中锌离子结合蛋白所占比例; (C)柱状图表示被显著稳定的金属离子结合蛋白中各种金属离子结合蛋白的数目。(D)PELSA 筛选到的 60 个含有锌指结构的蛋白, 位于锌指结构之外的肽段和位于锌指结构内的肽段的 $\log_2FC$ 的分布; (E) PELSA 筛选到的含有锌指结构的蛋白的二维稳定性变化图: 以 SQSTM1,TRAD1,CHAMP1 以及 YY1 为例; (F)不含有锌指结构的锌离子结合蛋白 LIMA1 在加入 30 μ M ZnCl₂ 后二维稳定性变化图, 其中三个位于 LIM 结构域的肽段分别标注为 1, 2, 3; (G)LIMA1 的部分三维结构图 (LIM 结构域), 球表示锌离子, 三个鉴定到的 LIM 结构域的肽段的标注顺序与(F)相同; (H)PELSA 筛选到的 20 个含有 EF-hand/EH motif 的蛋白, 位于 EF-hand/EHmotif 外的肽段和位于 EF-hand/EHmotif 内的肽段的 $\log_2FC$ 的分布; (I)PELSA 筛选到的 9 个含有 Fe²⁺-binding 结构域的蛋白, 位于 Fe²⁺-binding 结构域外的肽段和位于 Fe²⁺-binding 结构域内的肽段的 $\log_2FC$ 的分布; (J)局部放大的 30 μ M ZnCl₂ 处理 HeLa 细胞裂解液对应的蛋白层次火山图, 只保留加入 ZnCl₂ 后能量状态升高的蛋白 ($-\log_{10}Pvalue > 6, \log_2FC > 0$); (K) 含有 IQ-motif 的蛋白 UBE3C,MYO1B,MYO1C 及含有 B30.2SPRY 结构域的 HNRNPU 在加入 30 μ M ZnCl₂ 后的二维稳定性变化图; (L)加入 30 μ M ZnCl₂ 后 PSMC1-6 蛋白位于 P-loop-NTPase 结构域之外的肽段与位于 P-loop-NTPase 的肽段的 $\log_2FC$ 的分布。

图 9: 比较基于不同酶酶切的 PELSA 的蛋白质组覆盖及对能量状态发生变化的蛋白的鉴定:(A) 比较基于 trypsin 的 PELSA, 基于 chymotrypsin 的 PELSA, 基于 proteinase K 的 PELSA (以下分别简称为 Trypsin-PELSA, Chymotrypsin-PELSA 及 Proteinase K-PELSA) 的肽段鉴定数目; (B)Venn 图表示 Trypsin-PELSA, Chymotrypsin-PELSA 及 Proteinase K-PELSA 鉴定到的蛋白的重叠情况; (C)20 μ M 甲氨蝶呤(MTX) 处理 HeLa 细胞裂解液 Trypsin-PELSA(左), Chymotrypsin-PELSA(中), Proteinase K-PELSA (右) 得到的蛋白层次火山图。

- 图 10: PELSA 结合二甲基标记的定量方式对三个热休克蛋白抑制剂处理 HeLa 细胞裂解液后能量状态发生变化的蛋白的鉴定; (A)三个热休克蛋白抑制剂的结构图, geldanamycin (左), tanespimycin(中), ganetespib (右); (B)实验组分别为加入 100 μ Mgeldanamycin, 100 μ Mtanespimycin 以及 100 μ Mganetespib, 对照组为等体积的空白溶剂。使用第二大变化的肽段代表所属蛋白, 两次质谱重复得到的两个 $\log_2FC$ 分别为横坐标和纵坐标的散点图; (C)加入 geldanamycin 后, HSP90AA1, HSP90AB1, HSP90B1, HSP90AB2P 不同结构域肽段丰度的变化倍数。横坐标代表蛋白的序列, 纵坐标代表 $\log_2FC$; (D) 用热迁移法验证 MAT2A 和 ganetespib, AKR1C2 和 ganetespib 以及 AKR1C2 和 geldanamycin 之间的相互作用。
- 图 11: PELSA 评估配体与蛋白之间局部的亲和力; (A) 不同 ganetespib 浓度下, 热休克蛋白 HSP90AA1, HSP90AB1, HSP90B1 和 TRAP1 肽段丰度的变化倍数; (B)用 PELSA 计算得到的热休克蛋白家族的蛋白和三个抑制剂之间的亲和力, 横坐标代表不同的抑制剂浓度, 从左到右抑制剂浓度依次增加, 纵坐标代表肽段加药前后的丰度变化倍数, 变化倍数为 1 时表示没有变化, 所标注的数值代表 PELSA 计算得到的三个抑制剂对 HSP90AA1 的半数抑制浓度; (C)表示用微量热泳动仪(MST)表征三种热休克蛋白抑制剂与纯化的 HSP90AA1 之间的亲和力, 图中标注的数值为 MST 测得的 HSP90AA1 与三种热休克蛋白抑制剂之间的解离常数。

具体实施方式

- 为了更清楚的阐释本发明的技术方案, 体现发明优点。以下结合实施例对本发明进一步详细说明, 但并不将发明限制于所述实施例范围内。

- 以下实施例中: 实施例 1 以及实施例 4-9 说明该方法可以用于分析抗癌药物、代谢物、抗体、含有翻译后修饰的肽段、金属离子处理细胞裂解液后能量状态发生变化的蛋白及蛋白区域; 实施例 2 证明 PELSA 能产生更多对能量状态发生变化或结合配体有响应的肽段, 且这些肽段在 PELSA 中的变化幅度远大于在对比方法 LiP-MS 中鉴定到的变化幅度。实施例 3 证明 PELSA 是目前最灵敏的鉴定能量状态发生变化的蛋白及蛋白区域的方法。实施例 10 说明使用其他非特异性或特异性酶酶切, 通过检测含有两个能反映蛋白区域稳定性酶切位点的肽段也能用于能量状态发生变化的蛋白质的鉴定。实施例 11 说明 PELSA 的肽段定量也可以使用基于二甲基标记的定量方式。实施例 12 说明所述方法可用于确定配体与其结合区域的亲和力; 此外实施例 1-9 也说明该方法可以用于鉴定药物、广谱激酶抑制剂、代谢物、抗体、含有翻译后修饰的肽段、金属离子的结合蛋白及结合区域; 实施例 1-9 中配体与靶蛋白的结合会导致结合蛋白构象的变化, 因此说明该方法同样适用于蛋白构象变化的研究。

- 实施例 1 中, 100 nM lapatinib 处理 BT474 细胞裂解液后, 仅有 lapatinib 的已知靶蛋白 ERBB2 的能量状态发生了显著变化 ($-\log_{10}Pvalue > 5$), 且其能量状态发生变化的区域正好是 lapatinib 的作用区域即激酶结构域, 证明 PELSA 能用于鉴定能量状态发生变化的蛋白及区域、配体的结合蛋白及结合区域。根据 PELSA 的结果, 我们也发现当 lapatinib 浓度增加到 1 μ M 时, 则会出现较多的激酶脱靶蛋白; 实施例 2 通过比较 LiP-MS 和 PELSA 对 MTX 以及 SHP099 处理 HeLa 细胞裂解液后能量状态变化蛋白的鉴定证明 PELSA 能鉴定到更多对蛋白能量状态变化或配体结合有响应的肽段 (是 LiP-MS 鉴定到的 2-5 倍), 此外该部分肽段对结合配体或能量状态变化产生的响应是

LiP-MS 中的 4-6 倍。实施例 3 中, PELSA 分别在 K562 和 HeLa 细胞裂解液中鉴定到了 121 个和 111 个 staurosoprine 的激酶靶蛋白, 该数目分别是文献报道的 TPP 方法鉴定数目 (53) 的 2.1 倍以及文献报道的 LiP-Quant 方法在达到相同的激酶靶蛋白比例时所鉴定到的激酶靶蛋白数目 (9) 的 12.3 倍, 此结果说明该发明是目前最灵敏的鉴定能量状态发生变化的蛋白的方法。实施例 4-6 说明 PELSA 能够高灵敏的鉴定到弱的配体蛋白相互作用并定位配体作用区域; 实施例 7 中, PELSA 在 HeLa 裂解液中成功的鉴定到了抗体的抗原表位, 提示 PELSA 可以用于鉴定其他蛋白-蛋白作用界面; 实施例 8 中, PELSA 成功的在 BT474 细胞裂解液中鉴定到了酪氨酸磷酸化 (pYEEI) 的识别结构域, 为寻找其他翻译后修饰的识别结构域提供了一个强有力的工具。实施例 9 中, 加入锌离子后, PELSA 成功的在细胞裂解液中鉴定了 112 个锌离子结合蛋白并准确定位了锌离子结合的区域, 说明 PELSA 对配体与蛋白相互作用的鉴定与配体的大小无关。此外根据 PELSA 结果, 锌离子的加入稳定了钙离子结合 motif, EF-hand/EH motif, 而与 EF-hand motif 结合的 IQ motif 则被去稳定, 可能是由于锌离子与 EF-hand motif 结合, 导致其与 IQ motif 解离, 从而导致 IQ motif 去稳定, 该结果也提示 PELSA 可以用来研究复合物的结合界面。实施例 10 探究了在 PELSA 中是否可以用其他酶代替 trypsin, 结果显示尽管使用 chymotrypsin 及 proteinase K 肽段及蛋白的鉴定数目远小于 trypsin, 但也能实现对 MTX 靶蛋白的鉴定, 提示对高丰度靶标蛋白的鉴定不局限于使用 trypsin。实施例 11 探究了 PELSA 是否可以结合二甲基标记的定量方式实现对热休克蛋白抑制剂靶蛋白的鉴定, 结果显示在二甲基标记定量方式结合使用第二大变化的肽段的筛选策略, 能高特异性的鉴定热休克蛋白抑制剂的靶蛋白及脱靶蛋白。实施例 12 证明 PELSA 能准确测定配体与蛋白的结合亲和力。

实施例 1

PELSA 用于分析抗癌药 lapatinib 处理 BT474 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

- (1) 取一大皿 BT474 细胞样品 (细胞计数约为 5×10^7 个), 重悬在 1 mL 含有 1% (体积比) 的蛋白酶抑制剂 (Sigma, 货号 P8340-5 mL) 的 PBS 缓冲液中。液氮反复冻融三次 (在液氮中冷冻两分钟, 在 37°C 的水浴中融化两分钟, 循环三次), 得到细胞碎片及细胞裂解液的混合物, 在 500g, 4°C 离心 10 min, 取上清得到细胞裂解液, 用 PierceTM 660nm Protein Assay (Thermo, USA) 确定蛋白浓度。
- (2) 使用细胞裂解缓冲液 (含有 1% (体积比) 的蛋白酶抑制剂 (Sigma, 货号 P8340-5 mL) 的 PBS 缓冲液) 将所提取的细胞裂解液的蛋白浓度调整为 1 mg/mL, 取 4 份细胞裂解液至 EP 管中, 每份 50 μ L, 分别加入 0.5 μ L 不同母液浓度的 lapatinib (Selleck, 货号 S2111) 使其终浓度分别为 100 nM, 1 μ M, 10 μ M, 100 μ M (lapatinib 母液浓度分别对应为 10 μ M, 100 μ M, 1 mM, 10 mM; lapatinib 溶解在 DMSO 中) 作为实验组, 另取 50 μ L 细胞裂解液加入 0.5 μ L DMSO 作为对照组, 实验组和对照组均分别进行四次重复实验, 在室温下孵育 30 min。
- (3) 按照酶与蛋白 1:2 (质量比) 的比例于实验组和对照组中分别加入胰蛋白酶即 trypsin (Sigma, 货号 T1426), 在振荡器上 37°C, 1000 rpm 下酶切 1 min, 加入 165 μ L 含有 8 M 盐酸胍 (Sigma, 货号 G3272) 的 pH 8.2 的羟乙基哌嗪乙硫磺酸 (即 HEPES, 购自 Sigma, 货号 H3375) 缓冲液终止酶解, 得酶解液。

- (4) 分别向上述酶解液中加入终浓度为10mM的TCEP(Sigma, 货号C4706)和终浓度为40mM的CAA(Sigma, 货号22790)在95℃下加热5min, 冷却至室温后, 将样品进一步转移到10KDa超滤管(Sartorius, 货号VN01H02)中, 14000g离心50min, 收集透过膜的酶解下来的肽段, 再用200 μ L pH8.2 HEPES缓冲液洗超滤膜收集透过膜的溶液, 合并两次超滤下来的肽段。
- (5) 分别将超滤下来的肽段溶液通过装有2 mg HLB填料 (Waters, USA) tip柱 (规格为200 μ L)除盐, 冻干, 得肽段混合物。
- (6) 分别将上述肽段混合物用20 μ L 体积分数为0.1%的甲酸 (Sigma, 货号V900803) 复溶, 进行LC-MS/MS分析, 每个样品进行一次数据非依赖性(DIA)质谱采集。用Spectronaut软件(Biognosys, Switzerland)对质谱获得的谱图文件进行肽段的鉴定和定量, 得到每个肽段所属蛋白, 肽段在蛋白上的位置以及肽段在每个样品中的丰度信息。在本实施例中, 我们首先对该过程中PELSA酶切位点所处的蛋白二级结构进行了分析, 我们发现PELSA实验中59%的酶切位点来源于蛋白质的螺旋结构区 (图 2A), 此结果说明在PELSA中蛋白质的结构被破坏, 进而使得处于稳定的低能量状态的区域也被酶切下来了。
- (7) 对lapatinib处理组和未处理组的肽段的丰度进行Empirical Bayes t-test计算并得到每个肽段的P value和 fold change (变化倍数), 分别代表该肽段的丰度在加入lapatinib后变化的显著性以及变化的倍数。对于每一个蛋白, 使用其所有肽段中Pvalue最小 (显著性最强) 的肽段来代表该蛋白进行配体结合蛋白的筛选。
- 在该实施例中限定 $-\log_{10}Pvalue > 5$ 的蛋白为能量状态发生变化的蛋白, 图2B所示为以所有定量得到的蛋白的 $-\log_{10}P$ value为纵坐标, 以 $\log_2$ fold change (缩写为 $\log_2FC$)为横坐标作火山图。可以看到在100nM lapatinib的处理下在所有定量到的5774个蛋白中, 只有lapatinib的已知的靶蛋白ERBB2发生了显著变化。
- ERBB2 是结合在细胞膜表面的受体酪氨酸激酶, 其主要由胞外区, 跨膜区, 胞内区 (激酶区, 非激酶区) 组成, 其中 lapatinib 作用于 ERBB2 的激酶结构域。分析鉴定到的 ERBB2 的肽段, 如图 2C 所示, 横坐标表示蛋白序列从起始到结束 (蛋白 N 端到 C 端), 横坐标上的方块表示 lapatinib 作用的激酶结构域, 纵坐标表示 100 nM lapatinib 处理下 ERBB2 肽段丰度的变化倍数, 发现只有位于或非常毗邻该区域的肽段发生了显著变化 ($|\log_2FC| > 0.3$ & $-\log_{10}Pvalue > 2$), 说明该方法能确定药物的结合区域。
- 进一步发现较大的 lapatinib 浓度 (100 μ M, 10 μ M, 1 μ M)作用下, ERBB2 激酶结构域的肽段均更耐酶切 (图 2D)。此外我们注意到尽管在 lapatinib 的浓度为 100 nM 时, 仅有 ERBB2 被鉴定为能量状态发生了变化。而在增加 lapatinib 的浓度时有较多其他激酶例如 CHEK2, SLK, RIPK2 以及 YES1 也发生了能量状态的变化 (图 2E), 并且这些激酶中大多数变化的肽段也正好处于激酶结构域 (图 2F)。此结果说明 lapatinib 在高剂量时容易和其他激酶有混交反应, 指示 PELSA 可以用来评估药物的混交性, 进而用来指导药物的设计和合成。
- 此外我们也注意到非激酶蛋白 PTGES2 在高 lapatinib 浓度下也发生了能量状态的变化 (图 2E)。我们采用 Westernblotting 的方式来进一步验证。过程和条件同上, 与其不同之处在于: 一大皿 BT474 细胞 (细胞计数约为 5×10^7 个) 用 PBS 悬浮, 细胞裂解

和蛋白浓度测定同上，将获得的细胞裂解液均分为 8 份，加入等体积不同终浓度的 lapatinib(lapatinib 溶解在 DMSO 中) 或 DMSO，使 lapatinib 的终浓度依次为 100 μ M, 50 μ M, 10 μ M, 1 μ M, 0.1 μ M, 0.01 μ M, 0.001 μ M 及 0，在室温 25 $^{\circ}$ C 下孵育 30min。孵育结束之后按照酶与蛋白 1:40 的质量比例加入 trypsin，在 37 $^{\circ}$ C 下酶切 1min，酶切
5 结束之后 95 $^{\circ}$ C 加热终止酶解，并加入 1/4 蛋白液体积的 5*loading buffer 在 95 $^{\circ}$ C 下加热 5min。采用免疫印迹法来检测酶切后的蛋白剩余量，浓缩胶：80V，20min，分离胶 120V，60min。转膜：恒流 250mA，40min。封闭后，按照 1:800 的体积比例加入 PTGES2 一抗(Proteintech, USA) 室温孵育 1h，然后与山羊抗兔 HRP-IgG 二抗(Abcam, UK)室温孵育 1h，ECL 发光试剂 (Thermo Fisher Scientific, America)和 Fusion FX5 化学
10 发光系统 (Vilber Infnit, France)对酶切剩余蛋白成像，并根据条带的深浅对剩余蛋白进行定量。参考蛋白 GAPDH，于 4 $^{\circ}$ C 过夜孵育，兔抗鼠 HRP-IgG 二抗(Abcam, UK) 室温孵育 1h，发光检测。如图 2G 所示随着加入 lapatinib 的浓度的增加，条带越来越深，表示 PTGES2 被酶切剩余的蛋白量在逐渐增加，在 lapatinib 的浓度达到 50 μ M 时达到最大值，说明 PTGES2 抵抗酶切的能力依赖于 lapatinib 的浓度，进一步提示
15 PTGES2 可能为 lapatinib 的脱靶蛋白。

以上分析结果说明 PELSA 能够高特异性的鉴定配体的结合蛋白，并通过分析靶蛋白肽段的变化情况确定配体的结合区域；这种无歧视的组学层次上的靶蛋白筛选模式还能鉴定到其他脱靶蛋白，例如在本实施例中鉴定到的其他激酶以及 PTGES2 等，提示该方法能用来评价药物的混交性，进而对药物的设计和研发提供指导。此外
20 在肽段层次上分析发现只有配体结合区域的肽段发生了显著变化，提示该方法能揭示药物的作用区域。在用不同浓度的配体处理，靶蛋白的丰度变化值则有浓度依赖性变化，提示该方法可以确定配体与靶蛋白的亲合力。

实施例 2

比较 LiP-MS 和 PELSA 对甲氨蝶呤 (MTX)以及 SHP099 处理 HeLa 细胞裂解液
25 后能量状态发生变化的蛋白的鉴定。

为了体现本发明在能量状态发生变化的蛋白的鉴定的优势，本实施例平行比较了目前已有的基于酶切鉴定配体结合靶蛋白的方法 LiP-MS 以及本发明 (PELSA) 对 MTX 以及 SHP099 处理 HeLa 细胞裂解液后能量发生变化的蛋白的鉴定。其中 MTX 的作用靶蛋白为 DHFR, SHP099 的作用靶蛋白为 PTPN11。在本实施例中 PELSA 的过程和条件同实施例 1，与实施例 1 不同之处在于：所使用的细胞样品来源于 HeLa 细胞，所使用的配体分子分别为终浓度为 10 μ M 的 MTX (Selleck, 货号 S1210)或终浓度为 10 μ M 的 SHP099(Selleck, 货号 S8278)。LiP-MS 的实验流程主要参考文献 (Piazza I, et al, Nature Communication, 2020, 11(1):4200)，具体实施步骤如下（过程和条件同
30 实施例 1，不同之处在于）：

- 35 (1) 使用的细胞裂解液为 60 mM HEPES pH 7.5, 150 mM KCl 以及 1 mM $MgCl_2$ (Piazza I, et al, Nature Communication, 2020, 11(1):4200)。
- (2) 细胞裂解液与药物孵育结束后，按照酶与蛋白 1:100 (质量比)的比例加入 Proteinase K (Sigma, 货号 P2308)，在振荡器上 25 $^{\circ}$ C，1000 rpm 下酶切 4 min，在 98 $^{\circ}$ C 下加热 1 min,再加入等体积的体积分数为 10%的脱氧胆酸钠(Sigma, 货号 D6750)，
40 98 $^{\circ}$ C 下加热 4 min，使蛋白大片段变性，得蛋白溶液。

- (3) 在上述蛋白溶液中加入终浓度为 10 mM 的 TCEP(Sigma, 货号 C4706) 和终浓度为 40mM 的 CAA(Sigma, 货号 22790), 98 °C 下加热 5 min, 冷却至室温, 加入四倍体积 pH 8.2 的 60mM HEPES 缓冲液, 将脱氧胆酸钠的浓度稀释为 1%。
- (4) 按照酶与蛋白比例 1:100 (质量比) 加入 Lys-C (Wako chemicals) 酶切 4 h, 再按照酶与蛋白比例 1:50 (质量比) 加入 trypsin (Promega) 酶切 16 h。
- (5) 在酶切结束后的溶液中加入甲酸 (FA) 使其最终体积分数为 1.5%。静置 10 min, 待脱氧胆酸钠沉淀平衡后, 20000g, 室温离心 10 min (两次) 除去脱氧胆酸钠沉淀。
- (6) 除盐, 质谱定量, 软件分析, 数据检验过程同实施例 1。不同之处在于用 Spectronaut 进行鉴定时, 酶切参数设置为 Trypsin, Semi-specific。

如图 3A, 3B 所示, 如果将满足 $-\log_{10}Pvalue > 2$ (图中水平横虚线以上部分) 且 $|\log_2FC| > 0.3$ (图中竖直虚线外侧部分) 的肽段定义为对配体有响应的肽段, 在 MTX-DHFR 体系中, PELSA 鉴定到的对 MTX 有响应的 DHFR 肽段是 LiP-MS 的 2 倍; SHP099-PTPN11 体系中, PELSA 鉴定到的对 SHP099 有响应的 PTPN11 肽段是 LiP-MS 的 5.25 倍。此外 PELSA 中对 MTX 有响应的 DHFR 肽段的响应幅度是 LiP-MS 中的 4.3 倍(图 3C); PELSA 中对 SHP099 有响应的 PTPN11 肽段的响应幅度是 LiP-MS 中的 6.4 倍(图 3C)。例如, 在 MTX 和 DHFR 的作用体系中, 我们发现位于 MTX 的作用位点的一个肽段在 PELSA 实验中表现出了显著的变化, 而在 LiP-MS 结果中该肽段没有变化(图 3D)。该肽段处于低能量状态的螺旋结构, 在 LiP-MS 中轻微的酶切难以触及到该结构, 而 PELSA 的破坏性酶切使得蛋白能量状态较低的区域也能被酶切下来, 因此能检测到该肽段的变化。同样在 SHP099 和 PTPN11 的作用体系中, 一个处于 SHP099 结合位点且位于蛋白结构内部的肽段, 也仅在 PELSA 实验中检测到了变化(图 3E), 同样也说明 PELSA 能破坏蛋白的内部稳定结构, 从而检测到该处能量状态的变化。此结果说明与 LiP-MS 相比, PELSA 能鉴定到更多对配体有响应的靶蛋白的肽段, 且该部分肽段对配体的响应幅度远大于 LiP-MS 鉴定到的对配体有响应的靶蛋白肽段, 这也使得 PELSA 在对能量状态发生变化的蛋白的鉴定方面有更高的灵敏度。例如在本实施例中, 参与 MTX 的药代动力学过程的 GART (Mikkelsen T S et al, Pharmacogenet Genomics, 2011, 21(10): 679-686), 仅在 PELSA 中被鉴定为能量状态发生变化的蛋白, 而在 LiP-MS 中 GART 却没有任何变化(图 3A)。

该实施例说明本发明(PELSA)与现有方法 LiP-MS 相比, 能产生更多对配体结合有响应的肽段且响应肽段的响应幅度更大, 因此能鉴定到更多的配体结合蛋白, 这也部分解释了本发明在配体结合蛋白鉴定方面具有极高灵敏度的原因。

实施例 3

PELSA 用于分析广谱激酶抑制剂 staurosporine 处理 HeLa 及 K562 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

本实施例通过对 staurosporine 结合蛋白的鉴定以及和文献报道的结果对比, 显示出 PELSA 在鉴定能量状态发生变化的蛋白方面具有极高的灵敏度。

实验过程和条件同实施例 1, 与实施例 1 不同之处在于使用 K562 和 HeLa 细胞样品, 实验组使用的配体分子为终浓度为 20 μ M staurosporine (Selleck, 货号 S1421)。如图 4A 所示, 加入 staurosporine 后, PELSA 测得大量的激酶在 HeLa 及 K562 裂解液里被稳定(图 4A)。如果限定所鉴定到的靶蛋白中激酶所占的比例为 80%, PELSA 最

终在 K562 细胞裂解液中鉴定到 121 个被稳定的激酶（总共 143 个靶蛋白），在 HeLa 裂解液中鉴定到 111 个被稳定的激酶（总共 135 个靶蛋白）。在已发表的文献中，LiP-Quant 在达到相同的激酶比例条件下（即大于等于 80% 的靶蛋白为激酶），仅能鉴定到 9 个对 staurosporine 有响应的激酶（图 4B），TPP 鉴定到 53 个对 staurosporine 有响应的激酶（总共鉴定到 60 个靶蛋白）。此数据说明 PELSA 与现有方法相比具有极高的灵敏度（图 4B）。

我们进一步比较了 LiP-Quant, TPP 以及 PELSA 的蛋白和肽段的覆盖深度。我们发现尽管 LiP-Quant 比 PELSA 鉴定到更多的肽段（图 4C）且 LiP-Quant 所鉴定的蛋白质组的蛋白序列覆盖度整体上要大于 PELSA（图 4D），但是 PELSA 鉴定到对 staurosporine 有响应的激酶（即鉴定为靶蛋白的激酶）数目远远多于 LiP-Quant（111 vs. 9）。进一步分析发现 PELSA 鉴定为靶蛋白的激酶序列覆盖度要明显小于 LiP-Quant（图 4D），提示 PELSA 即使在序列覆盖度较低时也能鉴定结合蛋白为靶蛋白，而 LiP-Quant 由于含有很多无效的全 trypsin 酶切肽段，因此需要更高的序列覆盖度才能鉴定准确为靶蛋白。

即使在 TPP (K562) 比 PELSA-K562 有更多的蛋白鉴定的情况下（7673 vs. 6310）（图 4C），PELSA 鉴定到的激酶靶蛋白的数目依旧是 TPP 的 2.28 倍，进一步说明了 PELSA 在靶蛋白鉴定方面极高的灵敏度。分析 TPP 和 PELSA 鉴定到的激酶的熔点发现，TPP 对熔点很高或很低的激酶有歧视，导致不能被筛选出来，而 PELSA 对各个熔点范围的激酶均有较好的鉴定（图 4E）。

PELSA 在两个细胞系中共鉴定到 192 个 staurosporine 的靶蛋白，其中 154 个是激酶，激酶所占比例达到 80%（图 4F）。此外 PELSA 还能鉴定 staurosporine 的作用位点，图 4G 所示，PELSA 鉴定到的显著变化的肽段中在 K562 及 HeLa 裂解液中均有超过 92% 的肽段位于激酶结构域或者距离激酶结构域非常近（距离激酶结构域小于 10 个氨基酸残基）。该实验结果说明 PELSA 不仅在配体作用蛋白鉴定方面有极高的灵敏度且能准确鉴定配体在蛋白上的作用区域。

实施例 4

PELSA 用于分析代谢物叶酸（folate）处理 K562 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

不同于 lapatinib 与 ERBB2 之间强的相互作用（亲和力在 9 nM 左右），folate 对其结合蛋白的亲和力较弱，如文献报导 folate 与其靶蛋白 DHFR 之间的解离常数在 3-10 μM（Ozaki Y et al., Biochemistry, 1981, 20(11): 3219-3225），因此我们首先使用 folate 探究 PELSA 是否能用于分析亲和力较弱的配体与蛋白结合导致蛋白能量状态的变化。过程和条件同实施例 1，与实施例 1 不同之处在于：使用 K562 细胞样品，液氮反复冻融三次后，在 500g, 4℃ 离心 10 min 得到上清液后，用蛋白除盐柱（Zeba Spin desalting columns, Thermo Fisher Scientific）去除内源性的 folate，收集下来的蛋白用 Pierce™ 660nm Protein Assay（Thermo, USA）测定蛋白浓度，用细胞裂解缓冲液，即含有 1%（体积比）的蛋白酶抑制剂 cocktail（Sigma, 货号 P8340-5mL）的 PBS 缓冲液将蛋白浓度调整为 1mg/mL。所使用的配体分子为终浓度为 50 μM 的 folate（Sigma, 货号 F7879），其余步骤同实施例 1。

如图 5A 所示，folate 的已知靶蛋白 DHFR 的稳定性发生了最显著的变化，同时

将 DHFR 变化的肽段对应到其蛋白结构上(图 5B),发现 folate 的结合位点的蛋白区域稳定性变化最为显著。此外也发现三个能和 folate 类似物有相互作用的蛋白 ATIC, MTHFR, GART 的稳定性也发生了变化,并且其稳定的部位正好位于 folate 类似物的结合部位(图 5C-5E)。P3H1 是一个胶原蛋白脯氨酸羟化酶,文献报导 folate 能够在体内参与胶原蛋白脯氨酸羟化过程(Haustvast J, et al., British Journal of Nutrition, 1974, 32(2): 457-469)。我们的实验结果显示加入 folate 后 P3H1 的羟化酶结构域被稳定(图 5F),进而为 folate 参与胶原蛋白脯氨酸羟化过程提供了依据。

实施例 5

PELSA 用于分析代谢分子亮氨酸(leucine) 处理 K562 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

Leucine 对其靶蛋白 LARS1 和 SESN2 的解离常数分别为 95 μM (Kim S, et al, Cell Reports, 2021, 35(4): 109301)和 20 μM (Wolfson R L, et al, Science, 2016, 351(6268): 43-48),因此我们也使用 leucine 来评价 PELSA 是否能用于分析亲和力较弱的配体与蛋白结合导致蛋白质能量状态的变化。

与实施例 4 不同之处在于实验组所使用的配体分子为终浓度为 5 mM 的 leucine (Sigma, 货号 61819),其余步骤同实施例 4。如图 5G,我们发现 leucine 的已知靶蛋白 LARS1, LARS2, GLUD1 以及 SESN2 的能量状态发生了最显著的变化。LARS1 有两个亮氨酸结合位点,分别位于 CD 和 CP 结构域,分别对应于 synthetic site 和 editing site。我们发现 LARS1synthetic site (图中标注的 CP)的肽段变化幅度显著大于 editing site (图中标注的 CD)的变化幅度,而对于结合 leucine 没有贡献的 C-端结构域则没有任何的响应(图 5H),该结果提示 PELSA 具备同一蛋白上区分不同作用位点的能力。

我们也发现 leucine 的另一作用靶蛋白 SLC1A5 的稳定性也发生了显著变化(图 5G)。SLC1A5 是一个由胞内区,跨膜区以及胞外区组成的膜蛋白,而其发挥氨基酸运输功能的是位于残基 54-483 的 Na-dicarboxylate_symporter 结构域。我们鉴定到三个 SLC1A5 的肽段,最终只有位于该结构域的肽段(残基 190-212)发生了显著变化(图 5I),该结果说明我们的方法也能识别膜蛋白作用靶点并定位作用区域。有意思的是,我们发现 leucine 能稳定 PPIP5K1 以及 PPIP5K2 (图 5G),且这两个蛋白仅组氨酸磷酸酶结构域被稳定(图 5J)。

此实施例说明 PELSA 也能很好的鉴定代谢物和蛋白之间的弱的相互作用,并准确指出代谢物的结合位点。此外通过对 SLC1A5 的鉴定,我们发现 PELSA 对膜靶蛋白也有很好的鉴定效果。PELSA 也鉴定到了一些可能与 leucine 有相互作用的蛋白,为以后 leucine 的功能研究提供了基础。

实施例 6

PELSA 用于分析代谢分子二酮戊二酸 (αKG)处理 HeLa 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

过程和条件同实施例 4,与实施例 4 不同之处在于实验组所使用的配体分子为终浓度为 2mM 的 αKG (Sigma, 货号 75890-25g),所使用的细胞裂解液来源于 HeLa 细胞,将满足 $-\log_{10}\text{Pvalue} > 3.4$ 且 $\log_2\text{FC} < -0.5$ 的蛋白认为是 2mM αKG 稳定的蛋白。如图 5K,在 2 mM αKG 处理的 HeLa 细胞液中,PELSA 鉴定到 40 个蛋白被 αKG 稳

定, 其中 30 个蛋白是已知的 α KG 的作用蛋白;这个数目代表目前在一次实验中鉴定到的最多的已知 α KG 靶蛋白的数目。虽然有文献报道使用 LiP-MS 在大肠杆菌中鉴定 α KG 靶蛋白(Piazza I et al, Cell, 2018, 172 (1-15)), 但是在其鉴定到的 34 个靶蛋白中仅有 2 个是已知的 α KG 作用蛋白。此外 PELSA 测得蛋白二维稳定性变化图显示 (图 5L), 这些已知的 α KG 靶蛋白的能量变化区域正好位于 α KG 的结合区域。综上, 此实施例说明 PELSA 在分析亲和力较弱的代谢分子与蛋白相互作用时也具有极高的灵敏度, 并且能鉴定到代谢物的作用区域。

实施例 7

PELSA 用于分析抗体处理 HeLa 细胞裂解液后能量状态发生变化的蛋白及区域。

本实施例将通过对抗体处理 HeLa 细胞裂解液后能量状态发生变化的蛋白的分析说明 PELSA 能用于鉴定抗体的抗原结合表位。

过程和条件同实施例 1, 与实施例 1 不同之处在于: 采用 HeLa 细胞, 采用的配体分子为商品化的 DHFR 抗体(Wabways, China, RRID:AB_2877179)和商品化的 CDK9 抗体(Wabways, China, RRID: AB_2877178); 实验组为加入 2%(体积分数)的 DHFR 抗体或 1%(体积分数)的 CDK9 抗体, 对照组为同等体积的赋性剂;将满足 $-\log_{10}Pvalue > 5$ 的蛋白判定为能量状态发生变化的蛋白。如图 6A 所示为在细胞裂解液里鉴定抗体结合抗原表位的示意图。图 6B (左)为 DHFR 抗体处理 HeLa 细胞裂解液得到的蛋白质层次上的火山图, 图 6B(右)为 CDK9 抗体处理 HeLa 细胞裂解液得到的蛋白质层次上的火山图。DHFR 和 CDK9 的能量状态分别在加入 DHFR 抗体及 CDK9 抗体后发生了显著变化, 说明 PELSA 能够直接在细胞裂解液中筛选到抗体的抗原蛋白。其他标注出来能量状态发生变化的蛋白可能是由于抗体的非特异性吸附造成的。

进一步分析 PELSA 鉴定到所有 DHFR 肽段的变化情况, 可以看出能量状态变化的肽段正好坐落在抗体的已知识别表位(图 6C,图 6D)。观察到对于 CDK9, 有两个肽段发生了相反方向的变化。CDK9 抗体识别位于 CDK9 尾端的 PATTNQTEFERVF, 序列 NPATTNQTEFER 有一个位点正好处于该抗原表位内, 在加入抗体后产生的更少(图 6D), 而序列 NPATTNQTEFERVF (NPxxVF), 完全包含该抗原表位序列, 但是其 N 端的 trypsin 的酶切位点在该抗原表位外, 该肽段加入抗体后产生的反而更多(图 6D)。进一步分析发现两个肽段鉴定的强度和, 在加不加入抗体时差别很小(图 6E), 因此该肽段的反方向变化可以解释为: 由于抗体的保护作用, 导致序列 NPATTNQTEFER, C 端的 R 位点更难被酶切, 因此产生的更少, 进而导致本该没有响应的序列 NPATTNQTEFERVF(NPxxVF)剩下的更多。该结果说明 PELSA 在鉴定抗原表位中有很高的分辨率。

该实施例说明 PELSA 能很好的鉴定抗体与蛋白之间的相互作用, 并揭示抗体与蛋白的作用位点, 并提示 PELSA 也可用于其他蛋白-蛋白相互作用体系等。

实施例 8

PELSA 用于分析翻译后修饰的肽段处理 BT474 细胞裂解液后能量状态发生变化的蛋白及蛋白区域。

蛋白翻译后修饰在各项生命活动中发挥着重要作用, 其对生命活动的调节往往需要通过被下游蛋白识别并招募效应蛋白来实现。因此翻译后修饰的识别蛋白的鉴定对

翻译后修饰功能的理解以及疾病发生机制的研究都有重要意义。磷酸化酪氨酸-谷氨酸-谷氨酸-异亮氨酸(以下简称 pYEEI)已知被含有 SH2 结构域的蛋白识别。以下我们将通过对 pYEEI 处理 BT474 细胞裂解液后能量状态发生变化的蛋白及蛋白区域的分析来说明 PELSA 能用于翻译后修饰识别结构域的鉴定。

5 过程和条件同实施例 1, 与实施例 1 不同之处在于: 细胞裂解缓冲液里额外加入了 2mM 磷酸酶抑制剂以防止加入的磷酸化肽段被细胞裂解液里活性磷酸酶去磷酸化; 为了能使配体分子适用于 Pulldown 实验的分析以便于和 PELSA 的实验结果平行比较, 我们采用的配体分子为 100 μ MN 端偶联生物素的 pYEEI (以下缩写为 Biotin-pYEEI, 强耀生物科技代为合成); 为同时排除磷酸基团以及 YEEI 对整个蛋白
10 质组的影响, 在本实施例中采用两个对照组, 对照组所加入的配体分子分别为: 100 μ MN 端生物素取代的磷酸化丝氨酸-谷氨酸-谷氨酸-异亮氨酸 (缩写为 Biotin-pSEEI, 强耀生物科技代为合成) 及 100 μ MN 端生物素取代的酪氨酸-谷氨酸-谷氨酸-异亮氨酸 (缩写为 Biotin-YEEI, 强耀生物科技代为合成)。

Pulldown 实验方案如下: 采用 BT474 细胞, 细胞的裂解及细胞裂解液的获取,
15 浓度的测定及调整同实施例 1, 在将细胞裂解液的蛋白质浓度调整为 1mg/mL 后, 取 100 μ L 细胞裂解液加入终浓度为 100 μ M 的 Biotin-pYEEI 作为实验组, 取两份 100 μ L 分别加入终浓度为 100 μ M 的 Biotin-pSEEI 及终浓度为 100 μ M 的 Biotin-YEEI 作为对照组, 重复三次, 细胞裂解液与所加入的配体分子在室温孵育 30 min。孵育结束后每
20 组分别加入 200 μ L 的 avidin beads (Thermo, USA), 4 $^{\circ}$ C 过夜孵育。次日用洗涤液 (1% cocktail, 1%磷酸酶抑制剂及 0.5%NP40 的 PBS 溶液) 400 μ L 洗四次, 再用细胞裂解缓冲液 (1% cocktail, 1%磷酸酶抑制剂的 PBS 溶液) 400 μ L 洗四次。再用 100 μ L 含有 8M 盐酸胍的 HEPES 缓冲液 (pH 8.2) 洗脱蛋白, 重复两次。在洗脱得到的蛋白中加入终浓度为 10 mM 的 TCEP 及 40 mM 的 CAA, 95 $^{\circ}$ C 加热 5 min, 使蛋白充分烷基化。
25 将蛋白液转移到超滤管中, 14000 g 离心 30min, 去除膜透过液。再用 200 μ L 10 mM 碳酸氢铵 (NH_4HCO_3) 缓冲液洗超滤膜两次, 去除透过液。再加入 100 μ L 10mM NH_4HCO_3 将膜上保留的蛋白样品重悬, 并加入 2 μ g trypsin (Promega) 过夜酶解, 次日将超滤管在 14000 g 下离心 50 min 来收集酶解得到的肽段, 再用 100 μ L NH_4HCO_3 洗超滤管膜, 收集透过液来回收残留的肽段, 合并酶解液及第二次洗膜液并冻干。质谱定量, 软件分析, 以及数据检验过程同实施例 1。

30 图 7A 所示为在细胞裂解液中鉴定酪氨酸磷酸化识别结构域的示意图。以 YEEI 作为对照组, 经 Empirical Bayes t-test 得到 Pvalue(pYEEI/YEEI); 以 pSEEI 为对照组, 经 Empirical Bayes t-test 得到 Pvalue(pYEEI/pSEEI)。将这两个 Pvalue 取对数后做散点图, 如果将满足 $-\log_{10}\text{Pvalue}(\text{pYEEI}/\text{YEEI}) > 3.1$; $\log_2\text{FC}(\text{pYEEI}/\text{YEEI}) < 0$; $-\log_{10}\text{Pvalue}(\text{pYEEI}/\text{pSEEI}) > 3.1$; $\log_2\text{FC}(\text{pYEEI}/\text{pSEEI}) < 0$ 的蛋白定义为能量状态降低即被加入
35 的配体稳定的蛋白, 最终筛选到 28 个被 pYEEI 稳定的蛋白。在这 28 个蛋白中, 有 9 个含有 SH2 结构域的蛋白, 另外有 8 个和钙离子相关的蛋白 (图 7B)。而在 Pulldown 实验中, 由于洗涤的强度过大, 没有含有 SH2 结构域的蛋白被筛选出来 (图 7C)。PELSA 也实现了对 pYEEI 的识别结构域的鉴定: 如图 7D 及图 7E 所示, 只有位于 SH2 结构域的肽段在加入 pYEEI 后被稳定, 而其他部分的肽段的丰度保持不变。另外鉴定
40 到的钙离子相关的蛋白只在特定的 EF-hand 结构域被稳定 (图 7F)。

上述结果说明 PELSA 能通过对翻译后修饰的肽段处理细胞裂解液后能量状态发生变化的蛋白及蛋白区域的鉴定, 来实现翻译后修饰的识别蛋白及识别结构域的鉴定。

实施例 9

5 PELSA 分析金属离子处理 HeLa 细胞裂解液后能量状态发生变化的蛋白质及蛋白区域。

在此实施例中我们探究了 PELSA 是否适用于小的金属离子与蛋白结合导致蛋白质能量状态的变化。

过程和条件同实施例 1, 与实施例 1 不同之处在于: 采用 HeLa 细胞且采用的细胞裂解缓冲液为在实施例 1 的细胞裂解缓冲液基础上额外加入了 2 mM 乙二胺四乙酸钠 (EDTA, 购自 Sigma); 按照实施例 1 的方式得到细胞裂解液后, 用蛋白除盐柱 (Zeba Spin desalting columns, Thermo Fisher Scientific) 去除加入的 EDTA, 该蛋白除盐过程重复两次; 实验组加入的配体分子为终浓度为 30 μ M 的 ZnCl₂ (Sigma, 货号 450111-10G); 将满足 $-\log_{10}P\text{value} > 3$, $\log_2FC < -0.5$ 的蛋白认为是被锌离子稳定的蛋白。

15 图 8A 所示有大量的金属离子结合蛋白被锌离子稳定; 在所有鉴定到的被锌离子稳定的 280 个蛋白中, 有 66% 以上的蛋白即 185 个蛋白是已知的金属离子结合蛋白, 而在整个被鉴定到的蛋白质组中, 金属离子结合蛋白的比例只有 19%, 说明 PELSA 能通过对加入锌离子后对能量状态变化的蛋白的鉴定, 从而将金属离子结合蛋白筛选出来。文献报导锌离子能和钙离子结合蛋白的 EF-hand motif 结合 (Tsvetkov P O, et al, 20 Front Mol Neurosci, 2018, 11: 459), 同时能结合镁离子结合蛋白并占据镁离子的结合位点 (Dudev T, et al., Chemical Reviews, 2003, 103(3): 773-787), 提示这些二价金属离子对金属离子结合蛋白有混交性。与文献报导一致, 我们发现在 185 个被锌离子稳定的金属离子结合蛋白中 112 个是锌离子结合蛋白, 73 个其他金属离子结合蛋白(图 8B)。在 73 个其他金属离子结合的蛋白中 25 个是钙离子结合蛋白, 20 个是镁离子结合蛋白, 15 个铁离子结合蛋白, 9 个锰离子结合蛋白以及 6 个其他金属离子结合蛋白(图 8C)。在 112 个锌离子结合蛋白中, 60 个蛋白含有结合锌离子的锌指结构(图 8C)。分析这些含有锌指结构的蛋白, 我们发现位于锌指结构部分的肽段的 $\log_2FC$ 的分布中位数明显小于在锌指结构之外的肽段的 $\log_2FC$ 的分布中位数 (图 8D); 分析单个含有锌指结构的蛋白也可以发现只有位于锌指结构的肽段被显著稳定 (图 8E), 这些结果说明 PELSA 能够确定锌离子的结合位点。此外对于没有锌指结构的蛋白, 以 LIMA1 蛋白为例, 其含有一个 LIM 锌离子结合结构域, 在本次实验中该结构域有三条肽段被定量到(图 8F), 变化倍数最大的一条肽段是与锌离子直接结合的肽段, 而其他两条未直接参与锌离子结合的肽段则变化倍数非常小 ($|\log_2FC| < 0.3$, 图 8G), 该结果证明 PELSA 在结合位点定位方面具有极高的精确度。

35 分析鉴定到的钙离子结合蛋白, 27 个蛋白中 20 个包含 EF-hand/EH motif(钙离子结合 motif)。同含有锌指结构的蛋白一样, 位于 EF-hand/EH motif 的肽段的 $\log_2FC$ 的分布中位数明显小于 EF-hand/EH motif 之外的肽段的 $\log_2FC$ 的分布中位数, 说明锌离子作用在这些钙离子结合蛋白的 EF-hand/EH motif (图 8H)。另外对于 9 个含有铁离子结合结构域的蛋白, 位于铁离子结合结构域的肽段的 $\log_2FC$ 的分布中位数显著低于在铁离子结合结构域之外的肽段的分布中位数, 说明锌离子作用在这些铁离子结

合蛋白的铁离子结合区域(图 8I)。

如果配体和蛋白结合导致该蛋白从其原有的复合物中解离出来,则会导致与其形成复合物的伙伴蛋白去稳定。我们发现几个含有 IQ-motif 的蛋白发生了去稳定即 $-\log_{10}Pvalue > 6$; $\log_2FC > 0$ (图 8J),而且去稳定的肽段正好位于 IQ-motif 及其附近(图 8K)。IQ-motif 是已知的 EF-hand motif 的结合位点,而 EF-hand motif 能和锌离子结合(图 8H),因此 IQ-motif 的去稳定可能是由于 EF-hand motif 和锌离子结合导致了 IQ-motif 与 EF-hand motif 的解离。IQ-motif 和 EF-hand motif 正好是蛋白与蛋白相互作用的界面,该结果提示 PELSA 能用来分析细胞内蛋白复合物的组装和解离造成的能量状态的变化,并揭示复合物内的蛋白的结合界面。

我们也注意到一系列 26S 蛋白酶体调节亚基的组分蛋白 PSMC1-6 发生了去稳定(图 8J)。PSMC1-6 均含有一个 P-Loop-NTPase 结构域位于 PSMC1-6 所形成的复合物的结合界面上,PELSA 的结果显示只有位于该结构域的肽段才被显著的去稳定(图 8L),这可能是锌离子使 26S 蛋白酶体调节亚基解离,从而导致了 PSMC1-6 蛋白相互作用界面的去稳定。该结果进一步提示该方法可以用来研究蛋白复合物组装和解离的动态变化。

实施例 10

基于其他酶的 PELSA 分析甲氨蝶呤 (MTX) 处理 HeLa 细胞裂解液后能量状态发生变化的蛋白。

为了评估按照 PELSA 所述流程,其他酶是否能用来分析能量状态发生变化的蛋白,我们并行测试了 trypsin 以及另外两种具有不同酶切位点特异性的酶,糜蛋白酶(chymotrypsin,购自 Sigma,货号 C3142)和蛋白酶 K(proteinase K,也简称为 PK,购自 Sigma,货号 P2308),来分析 MTX 处理 HeLa 细胞裂解液后能量状态发生变化的蛋白。其中 chymotrypsin 主要的切割位点是在芳香族氨基酸的 N 端,proteinase K 则具有广泛的切割位点。

Trypsin-PELSA 的实施步骤与实施例 1 不同之处在于:使用 HeLa 细胞样品,实验组加入的配体分子为终浓度为 $10\mu\text{M}$ 的 MTX(溶解在 DMSO 中,母液浓度为 1mM ;购自 Selleck,货号 S1210),实验组和对照组均进行四次重复实验,在室温下孵育 30min,按照酶与蛋白 1:2(质量比)的比例加入 trypsin,在振荡器上 37°C , 1000 rpm 下酶切 1 min,加入 $165\mu\text{L}$ (即三倍样品体积)含有 8M 盐酸胍的 pH 8.2 HEPES 缓冲盐终止酶解。其余实验以及质谱定量,软件分析,数据检验过程同实施例 1。

Chymotrypsin-PELSA 的实施步骤与 Trypsin-PELSA 不同之处在于:将使用 trypsin 替换为 chymotrypsin;酶切条件为在振荡器上 25°C , 1000 rpm 下酶切 1 min;最终使用 Spectronaut 搜库时, digestion 参数设置酶切位点为: F,W,Y,L,M,其余步骤同 Trypsin-PELSA。

Proteinase K-PELSA 的实施步骤与 Trypsin-PELSA 不同之处在于:将使用的 trypsin 改为 proteinase K;酶切条件为在振荡器上 25°C , 1000 rpm 下酶切 1 min;最终使用 Spectronaut 搜库时, digestion 参数设置酶切位点为: trypsin,unspecific,其余步骤同 Trypsin-PELSA。

如图 9A 所示,最终 Trypsin-PELSA 鉴定到 69245 个肽段,而 Chymotrypsin-PELSA 及 Proteinase K-PELSA 鉴定到的肽段数目则远少于 Trypsin-PELSA(分别为 18027 及

28702), 与此相对应, Chymotrypsin-PELSA (2710)和 Proteinase K-PELSA (1937) 也远少于 Trypsin-PELSA 鉴定的蛋白数目 (5487) (图 9B)。且 Trypsin-PELSA 组鉴定到的蛋白能覆盖大多数的 Proteinase K-PELSA 组及 Chymotrypsin-PELSA 组鉴定到的蛋白 (图 9B)。这主要是由于 trypsin 产生的 tryptic 肽段相比其他酶产生的肽段更有利于质谱鉴定。在能量状态发生变化的蛋白鉴定方面, 由于 MTX 的已知结合蛋白 DHFR 是一个较高丰度的蛋白, 使用以上三种酶均能覆盖到该蛋白, 因而均能将其筛选出来 (图 9C)。该实施例说明 PELSA 不局限于使用 trypsin, 也可以使用其他特异性酶及非特异性酶来鉴定能量状态发生变化的蛋白, 但是可能面临蛋白及肽段鉴定数目少的问题。

10 实施例 11

基于二甲基标记定量的 PELSA 分析 3 种热休克蛋白抑制剂处理 HeLa 细胞裂解液后能量状态发生变化的蛋白及区域。

实施步骤如下:

- 15 (1) 采用 HeLa 细胞, 细胞破碎, 蛋白提取和定量步骤同实施例 1, 取六份 50 μ L 的细胞裂解液至 EP 管中, 其中三份分别加入 100 μ M geldanamycin, 100 μ M tanespimycin, 100 μ M ganetespib (母液浓度为 10mM; 均溶解在 DMSO 溶液中, 均购自 Selleck), 另外三份分别加入等量的 DMSO。在室温 25 $^{\circ}$ C 下孵育 30min, 孵育结束后采用 trypsin 与蛋白质量比例 1:2, 37 $^{\circ}$ C 下酶切 1 min, 100 $^{\circ}$ C 加热 5min 终止酶解。
- 20 (2) 向上述蛋白溶液中分别加入蛋白溶液三倍体积的含有 8M 盐酸胍的磷酸二氢钠缓冲液 (pH 6.5), 并加入终浓度为 10mM TCEP 和终浓度为 40mM CAA 在 95 $^{\circ}$ C 下加热 5min, 冷却至室温后, 将样品进一步转移到 10KDa 超滤管中, 14000 g 离心 50min, 收集酶解下来的肽段, 再用 200 μ L 磷酸二氢钠缓冲液洗超滤膜 (pH 6.5), 合并两次超滤下来的肽段。
- 25 (3) 二甲基标记: 加药组收集到的肽段加入 16 μ L 体积分数为 4% 的氘代甲醛 (Sigma, 货号 596388) 以及 16 μ L 浓度为 0.6 M 的氰基硼氢化钠 (Sigma, 货号 156159), 作为中标标记; 对照组加入 16 μ L 体积分数为 4% 的甲醛 (Sigma, 货号 252549) 以及 16 μ L 浓度为 0.6M 的氰基硼氢化钠作为轻标记, 在 30 $^{\circ}$ C 下反应 1h, 反应结束后再加入 10 μ L 体积分数为 10% 的氨水 (Sigma, 货号 338818) 反应 30min, 标记结束后将加药组和对照组的肽段等质量混合。
- 30 (4) 将二甲基标记后溶液加入 4.5 μ L 三氟乙酸 (Sigma, 货号 T6508) 酸化, 用装有 2 mg HLB 填料 (Waters, USA) 的 tip 柱 (规格为 200 μ L) 除盐, 冻干, 得肽段混合物。
- (5) 将上述肽段混合物用 30 μ L 体积分数为 0.1% 的甲酸复溶, 上样 1 μ g, 在数据依赖型数据采集模式下 (DDA), 每个样品进行两次 LC-MS/MS 分析, 因此每个肽段有
- 35 两针的定量结果。
- (6) 用 Maxquant 软件 (Cox, Germany) 分析 DDA 采集模式下所得到的谱图文件, 从而得到小肽段所属蛋白, 小肽段在蛋白上的位置以及小肽段在实验组和对照组丰度的比值 (fold change, 即 FC)。

40 在本实施例中选用了三个具有代表性的热休克蛋白抑制剂, 分别为含有苯醌基团的安沙霉素类热休克蛋白抑制剂 geldanamycin, tanespimycin, 以及全新结构的第二代

热休克蛋白抑制剂ganetespib (图10A), 其中geldanamycin和tanespimycin仅圆圈圈住的结构部分不同。采用第二大变化的肽段(即某一给定蛋白所有定量到的肽段中 $|\log_2FC|$ 第二大的肽段)代表蛋白, 质谱两针定量结果如图10B。若考虑两针定量结果 $|\log_2FC| > 1.4$ 为能量状态发生变化的蛋白, 热休克蛋白HSP90AB1, HSP90AA1, HSP90A1, HSP90ABP均被很好的筛选出来, 证明了基于二甲基标记定量的PELSA也能够很好的鉴定能量状态发生变化的蛋白。另外发现位于线粒体的热休克蛋白TRAP1仅在ganetespib中被鉴定为能量状态发生变化的蛋白。

HSP90 蛋白主要由三个 domain 组成, N 端 ATP 结合 domain, 中间 middle domain, 以及 C 端 domain, 其中 geldanamycin, tanespimycin 以及 ganetespib 均作用于 N 端 ATP 结合 domain。分析不同 domain 的肽段的丰度的变化倍数 (fold change, 即 FC), 如图 10C 所示, 以加入的药为 geldanamycin 为例, 只有 N 端 domain 即该药物结合区域的肽段经过 PELSA 分析后有明显的丰度变化, 进一步证明 PELSA 能确定能量状态发生变化的蛋白区域。

除了已知的靶蛋白, PELSA 也鉴定到了很多未知的脱靶蛋白。对 geldanamycin, PELSA 鉴定到了 PRDX 家族的几个蛋白发生了去稳定作用即 $\log_2FC > 1.4$ (图 10B)。文献报道 geldanamycin 能产生活性氧(ROS)导致严重的肝毒性(Clark et al, Free Radical Biology & Medicine, 2009, 1440-1449), 而 PRDX 家族蛋白能在细胞遭受氧化应激压力时清除 ROS 进而保护细胞。PRDX 家族蛋白的去稳定说明在高的 geldanamycin 浓度下 PRDX 家族蛋白结构被破坏, 这个结果也为 geldanamycin 的肝毒性机制提供了一种可能的解释。对与 geldanamycin 结构类似的小分子 tanespimycin, PELSA 鉴定到了与 geldanamycin 共有的脱靶蛋白 PRDX5 和 NQO1, 其中 NQO1 被报道能代谢安沙霉素类热休克蛋白抑制剂 (Reigan P D, et al, Molecular Pharmacology, 2011, 79(5): 823-832)。另外 PELSA 也鉴定到两个此前未被报道过的 ganetespib 的脱靶蛋白, MAT2A 和 AKR1C2 (图 10B)。为了进一步验证这两个蛋白能和 ganetespib 结合, 我们将这两个蛋白纯化出来, 然后用热迁移的方法看这两个蛋白是否与 ganetespib 有相互作用。图 10D 所示, 加入 ganetespib 后, MAT2A 和 AKR1C2 的热熔化温度均发生了明显的上调, 证实了 AKR1C2 和 MAT2A 与 ganetespib 之间的结合。另外热迁移的结果显示 AKR1C2 和 geldanamycin 之间也有结合, 且 geldanamycin 对 AKR1C2 的稳定作用要小于 ganetespib (图 10D), 同样在 PELSA 的结果中 geldanamycin 对 AKR1C2 的稳定作用也小于 ganetespib (图 10B)。

以上结果说明, 使用二甲基标记方式定量的 PELSA, 采用两条变化肽段的筛选策略能够准确确定能量状态发生变化的蛋白, 并很好的区分结构相似和结构不同的抑制剂的作用靶蛋白。体外蛋白纯化的热迁移实验证明 PELSA 鉴定的未知靶蛋白具有很高的可信度。该靶蛋白空间的构建为理解安沙霉素热休克蛋白抑制剂的肝毒性以及 ganetespib 的老药新用有重要的指导意义。

实施例 12

PELSA 评估热休克蛋白抑制剂与其靶蛋白的局部亲和力。

过程和条件同实施例 11, 与实施例 11 不同之处在于: 取 14 份 50 μ L 细胞裂解液分为实验组和对照组, 每组 7 份细胞裂解液, 在实验组中加入不同母液浓度的 geldanamycin 使其终浓度分别为 100 μ M, 10 μ M, 1 μ M, 100nM, 10nM, 1nM, 0.1nM

(geldanamycin 溶解在 DMSO 中), 对照组则均加入等体积的 DMSO。后续实验步骤, 质谱定量及软件分析过程同实施例 11。Tanespimycin 和 ganetespib 处理方式参照 geldanamycin。

图 11A 所示, 以 ganetespib 为例, 仅有位于 HSP90 蛋白 N 端结构域的肽段随着
5 药物浓度的增加其丰度的变化倍数也增加而后趋于稳定, 进一步根据这些肽段的变化
倍数随药物浓度的变化计算热休克蛋白和三个热休克蛋白抑制剂的亲和力。将至少有
12 针定量值的热休克蛋白的肽段 (总共 14 个定量值), 用 Prism 软件(GraphPad)将这
些 肽 段 拟 合 为 四 参 数 对 数 方 程
$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{-(\text{LogEC}_{50} - X) * \text{HillSlope}})$$

10 估原始数据和该方程的拟合的好坏, 将拟合程度好即 R^2 大于 0.9 的肽段提取出来作
为候选肽段, 同时要求候选肽段在最高配体浓度处理下丰度的变化幅度要在 30%及以
上。将所有满足条件的同一蛋白的肽段的变化倍数的中位数作为蛋白的在某一配体浓
度下的变化倍数即四参数对数方程 Y 对应的值, X 依旧代表配体浓度, 再次用 Prism
软件拟合四参数对数方程, 得到 EC_{50} (图 11B)。图 11C 所示, 则是将蛋白纯化出来,
15 然后用微量热泳动仪(MST)测三个抑制剂和 HSP90AA1 之间的亲和力 (K_d)。MST 测
得的亲和力 K_d 与 PELSA 计算出来的 EC_{50} 非常接近。

以上结果说明, 利用 PELSA 计算得到的亲和力与经典的测定亲和力的方法 (如本发
明所用的 MST) 测得的亲和力具有很好的一致性。此外该方法还可以得到肽段层次
上的亲和力数据, 这将有助于理解配体与蛋白之间的作用方式。

20

权 利 要 求 书

1. 一种探测能量状态发生变化的蛋白的方法，其特征在于：该方法包含以下步骤：
 - A、将待比较的两组或三组及以上处于不同能量状态的蛋白溶液(其中一组为基准对照组，其为能量状态未曾发生变化的蛋白溶液；其余为实验组，其为能量状态发生变化的蛋白溶液)与蛋白酶混合孵育 0.5-60 min，酶与蛋白样品的质量比例在 1:1-1:50 之间，使得蛋白质处于低能量状态的结构也被酶切破坏(即使得蛋白质发生破坏性酶切)，并直接产生含有两个能反映蛋白质区域稳定性酶切位点并适合自下而上(由肽段鉴定到蛋白鉴定)质谱分析的小肽段(分子量< 5KDa)；
 - B、将小肽段(分子量< 5-10 KDa)和蛋白质大片段(分子量 $\geq$ 10 KDa)分离；
 - C、使用定量蛋白质组学技术和软件分析(即质谱定量和分析)得到小肽段(分子量< 5KDa)的丰度，小肽段所属蛋白及小肽段在蛋白上的位置信息；
 - D、通过分析小肽段(分子量< 5KDa)在实验组和对照组样品中丰度的差异，判断能量状态发生变化的蛋白质；
 - E、通过分析丰度发生变化的小肽段(分子量< 5KDa)在所属蛋白中的位置来确定能量状态发生变化的蛋白区域。
2. 根据权利要求1所述方法，其特征在于：所述蛋白的不同能量状态包含但不限于通过下述一种或两种以上方式引起的能量状态的变化：
 - A、配体(包含药物分子、人或动物体内代谢物、植物提取物、核酸分子、金属离子、肽段、抗体、蛋白中的一种或两种以上)与蛋白相互作用导致的蛋白能量状态的变化；
 - 或 B、蛋白翻译后修饰导致的蛋白能量状态的变化；
 - 或 C、热刺激、渗透压力改变、变性剂刺激、氧化刺激、疾病导致的蛋白质能量状态变化中的一种或两种以上。
3. 一种探测配体与蛋白亲和力的方法，其特征在于：该方法包含以下步骤：
 - A、将一系列(或多个)不同终浓度(其中一个浓度点可为不添加配体的空白点作为对照组，其余为实验组)的配体与蛋白溶液分别进行混合孵育；
 - B、孵育结束后，将上述样品与蛋白酶在非变性的条件下孵育 0.5-60 min，酶与蛋白的质量比例为 1:1-1:50，使得蛋白质处于低能量状态的结构也被酶切破坏(即使得蛋白质发生破坏性酶切)，并直接产生大量含两个能反映蛋白质区域稳定性酶切位点并适合自下而上质谱分析的小肽段(分子量< 5KDa)；
 - C、将样品中适合质谱检测的小肽段(分子量< 5-10 KDa)与蛋白质大片段(分子量 $\geq$ 10 KDa)分离；
 - D、使用定量蛋白质组学技术和软件分析(即质谱定量和分析)得到小肽段(分子量< 5KDa)的丰度，小肽段所属蛋白以及小肽段在蛋白上的位置；
 - E、根据小肽段在不同配体浓度处理的样品中的丰度以及配体的浓度计算蛋白和配体之间的局部亲和力即半数有效浓度(EC_{50})，其中计算方程为四参数对数方程： $Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{-(\text{LogEC}_{50} - X) * \text{HillSlope}})$ ，Y 为纵坐标代表肽段丰度在实验组和对照组的变化倍数，X 为横坐标为不同的配体浓度，

Bottom 和 Top 分别为四参数对数方程对应曲线下端和上端的稳定值，单位与 Y 相同，Hill Slope 指的是曲线最大斜率的绝对值即曲线中点， EC_{50} 为所需计算的半数有效浓度， EC_{50} 越小代表配体与该肽段所属蛋白区域的亲和力越强；

5 所述一系列为至少 3 个以上，优选 5 个以上。

4. 根据权利要求 3 所述方法，其特征在于：配体可以是药物分子、动物或植物体内代谢物、植物提取物、核酸分子、金属离子、肽段、抗体、蛋白质或其他可以和蛋白发生相互作用的各类物质中的一种或两种以上。

5. 根据权利要求 1 或 3 所述方法，其特征在于：蛋白溶液包括含有一种蛋白质的单一蛋白溶液或含有两种以上蛋白质的混合蛋白溶液；混合蛋白溶液包含来源于
10 人、动物、植物或细菌的细胞或组织提取液；

蛋白提取应采用温和的提取方式包含但不局限于液氮冻融提取、液氮研磨提取或匀浆法提取；若提取过程采用提取缓冲液应当避免于其中使用破坏蛋白构象剂量的表面活性剂或变性剂，以使得蛋白能保持其待测构象；待测蛋白的存在形式可以为游离的蛋白和/或固载的蛋白。
15

6. 根据权利要求 1 所述方法，其特征在于：将处于不同能量状态的蛋白溶液在其待测构象下，采用相同的酶切条件，按照酶和蛋白的质量比例加入酶，使得蛋白质暴露的处于高能量状态的结构首先被酶切破坏展开，进而诱导蛋白质暴露出处于低能量状态的酶切位点，在较大量酶的存在下，使处于低能量状态的结构也被酶切破坏，直接产生用于自下而上质谱鉴定（由肽段鉴定到蛋白鉴定）的小肽段（分子量 < 5 KDa），并使处于不同能量状态的蛋白有不同的酶切程度；可选择的酶包括胰蛋白酶、谷氨酰蛋白内切酶、嗜热菌蛋白酶、蛋白酶 K 以及其他具有位点特异性或者非位点特异性酶的一种或两种以上；所用酶与蛋白的质量比例应在 1:1-1:50 之间，其中存在体积分数为 1%-2% 的蛋白酶抑制剂的情况下，优选 1:1-1:10；同时应调整酶切时间在 0.5-60 min；根据不同酶保持活性的温度范围选择
20 合适的酶切温度。

7. 根据权利要求 3 所述方法，其特征在于：将多个配体浓度处理的配体组和空白溶剂处理的对照组，分别按照酶和蛋白的质量比例加入酶，使得蛋白质暴露的处于高能量状态的结构首先被酶切破坏展开，进而诱导蛋白质暴露出处于低能量状态的酶切位点，在较大量酶的存在下，使处于低能量状态的结构也被酶切破坏，直接产生用于自下而上质谱鉴定（由肽段鉴定到蛋白鉴定）的小肽段（分子量 < 5 KDa），并使结合配体的蛋白和未结合配体的该蛋白有不同的酶切程度；可选择的酶包括胰蛋白酶、谷氨酰蛋白内切酶、嗜热菌蛋白酶、蛋白酶 K 以及其他具有位点特异性或者非位点特异性酶中的一种或两种以上；所用酶与蛋白的质量比例应在 1:1-1:50 之间，其中存在体积分数为 1%-2% 的蛋白酶抑制剂的情况下，优选 1:1-1:10；同时应调整酶切时间在 0.5-60 min；根据不同酶保持活性的温度范围选择
25 合适的酶切温度。

8. 根据权利要求 1 或 3 所述方法，其特征在于：分离小肽段和大的蛋白片段的方式包含但不局限于利用分子量的差异、亲疏水的差异、热稳定性的差异中的一种或
30 两种以上进行分离。
40

9. 根据权利要求1或3所述方法,其特征在于:所述定量蛋白质组技术是指酶切下来的肽段在质谱自下而上的分析模式(由肽段鉴定到蛋白鉴定)下定量,质谱数据采集模式包含数据依赖型采集(DDA)以及数据非依赖型采集(DIA);定量方式包含无标定量和/或标记定量(例如:二甲基标记定量、iTRAQ 标记定量、TMT 标记定量、TMTpro 标记定量中的一种或两种以上);
- 5 软件分析是指将质谱产生的原始谱图文件导入到搜库软件进行分析获得小肽段的定量信息,小肽段所属蛋白信息以及小肽段在蛋白上所处的位置信息;软件包括 Maxquant (Cox, Germany), MSFragger(Alexey, USA), Spectronaut (Biognosys, Switzerland)中的一种或两种以上。
- 10 10. 根据权利要求1或3所述方法,其特征在于:所述能量状态发生变化的蛋白的判断方法为根据样品中蛋白所含差异肽段的差异显著性 $-\log_{10}P$ value 的大小、肽段差异倍数 $|\log_2 \text{fold change}|$ 的大小、蛋白所含差异肽段的数目 N、肽段差异倍数是否随着加入配体的浓度呈剂量依赖性的变化中的一种或两种及以上因素的任意组合来判断能量状态发生变化的蛋白;
- 15 差异肽段差异的显著性的获得方式包括:将肽段在实验组和对照组得到的定量值进行显著性检验,可以是 Empirical Bayes t-test 或者 Student's t-test,得到的 P value 作为肽段在实验组和对照组差异显著性的指标,包含 $-\log_{10}P$ value $> 2\sim 5$ 的肽段的蛋白可认为是能量状态发生变化的蛋白;
- 20 差异肽段差异倍数的获得方式包括:搜库软件直接得到肽段丰度在实验组和对照组的差异倍数,数据检验如 Empirical Bayes t-test 或者 Student's t-test 得到差异倍数,包含 $|\log_2 \text{fold change}| > 0.3\sim 3$ 的肽段的蛋白可认为是蛋白能量状态发生变化的蛋白;
- 25 差异肽段变化数目处理方式包括:将每个蛋白中所有肽段按照差异显著性或者差异倍数的大小排序,差异显著性或者差异倍数最大的肽段定义为该蛋白第一大变化的肽段,第二大的肽段为该蛋白第二大变化的肽段,顺次类推第三大,第四大变化的肽段;将第一大变化的肽段或第二大变化的肽段或第三大或第四大变化的肽段的 $-\log_{10}P$ value 和 $\log_2 \text{fold change}$ 作为该蛋白的定量值,差异显著性 $-\log_{10}P$ value $> 2\sim 5$ 或者差异倍数 $|\log_2 \text{fold change}| > 0.3\sim 3$ 的蛋白认为是能量状态发生变化的蛋白,通过该方式确定的能量状态发生变化的蛋白包含 1 条或 2 条或 3 条或
- 30 4 条满足 $-\log_{10}P$ value $> 2\sim 5$ 或 $|\log_2 \text{fold change}| > 0.3\sim 3$ 的肽段;
- 肽段丰度的变化倍数是否随着加入配体的浓度呈剂量依赖性的变化的衡量标准包括:将通过上述方式得到的肽段在不同配体浓度下的变化倍数以及配体的浓度作为拟合四参数对数方程的变量,通过 Pearson 相关系数 R^2 评估原始数据与该方程拟合程度的好坏,若 $R^2 > 0.75\sim 0.95$ 则认为该肽段随加入配体的浓度有剂量
- 35 依赖性变化,则可以认为该肽段所属蛋白为配体的作用蛋白;
- 两种及以上因素的任意组合包含:同时满足上述任意两个或任意两个以上的条件,或将上述指标 $-\log_{10}P$ value, $|\log_2 \text{fold change}|$, N (差异肽段的数目), R^2 以任意比例组合得到综合的打分值来判断能量状态发生变化的蛋白。