



Concurso Público Fiocruz 2023

Pesquisador em Saúde Pública

Prova Discursiva

PE 45

Espectrometria de Massas Computacional

Espelho de Resposta

Pontuação de cada Questão Discursiva conforme Anexo II do Edital nº 3, de acordo com a Unidade detentora da vaga.

Espera-se que o candidato, no desenvolvimento do tema, tenha feito considerações técnicas adequadas sobre os seguintes pontos:

Questão 01

I-. Ao descrever DDA mencionar:

A varredura completa dos íons precursores em uma determinada faixa de massa, medindo sua relação massa-carga (m/z) e intensidade (MS_1).

Em seguida, um número pré-determinado de íons mais abundantes ou intensos (chamados de precursores) são selecionados para fragmentação e análise do m/z e intensidades dos íons-fragmento (íons-produto), assim gerando o mesmo número de espectros de massa sequenciais (MS_2 , MS/MS) por ciclo. Além do número de íons precursores mais intensos a serem selecionados, parâmetros como limiar de intensidade e tempo de exclusão são aplicados para selecionar apenas íons precursores acima de um determinado limiar e impedir que eles sejam selecionados mais de uma vez dentro de um determinado intervalo de tempo.

Modos de fragmentação mais utilizados incluem CID e HCD, para modificações ExD e UVPD.

As janelas de isolamento mais comuns para DDA estão entre 0.8 e 1.6 Th, garantindo principalmente isolamento de íons precursores únicos.

O uso de DDA em amostras complexas podem gerar espectros MS/MS quiméricos que contém fragmentos de múltiplos peptídeos.

Normalmente, os 10-20 íons precursores mais intensos são selecionados para fragmentação. No entanto, espectrômetros de massa recentes são capazes de operar em velocidades de sequenciamento superiores a 100 Hz.

Ao falar sobre DIA mencionar:

Os métodos DIA são baseados na fragmentação de um subconjunto ou de todos os íons precursores possivelmente gerados. Isso pode ser conseguido através de ciclos com janelas m/z predefinidas ao longo de toda a faixa de varredura ou fragmentando todos os íons precursores que são eluídos em um

determinado momento. Todos os precursores presentes em cada janela MS1 selecionada são co-isolados e, subsequentemente, co-fragmentados, independentemente de sua abundância.

Descrever os métodos DIA, SWATH MS, MSE, entre outros focando nos diferentes modos de aquisição dos dados. Acoplamento de métodos DIA com tecnologias de mobilidade iônica tem ajudado a reduzir a complexidade espectral e aumentar ainda mais a especificidade e sensibilidade da análise DIA.

II-. A abordagem DIA difere da DDA por não ter informações precisas sobre o precursor m/z.

A abordagem DDA foi amplamente adotada pela comunidade científica anteriormente e isso permitiu o desenvolvimento de configurações de aquisição de dados que são atualmente estabelecidas e incorporadas aos espectrômetros de massa como métodos pré-definidos. Além disso, existem vários pipelines computacionais que foram desenvolvidos para analisar espectros de fragmentação de peptídeos obtidos a partir de métodos DDA. Outra vantagem para a abordagem DDA é a identificação de peptídeos modificados. No entanto, deve-se notar que abordagens DIA têm sido propostas para fosfopeptídeos, peptídeos acetilados e glicosilados.

Os contras do DDA são que é um processo semi-estocástico, que resulta em uma baixa reprodutibilidade entre aquisições. Dependendo da velocidade de varredura, um número fixo de íons precursores pode ser selecionado para MS/MS durante cada ciclo, o que pode introduzir valores ausentes (missing values) tanto na identificação quanto na quantificação.

Os prós do DIA são que ele tem uma alta reprodutibilidade e cobertura do proteoma, e que ele pode quantificar todos os íons em cada janela, independentemente de sua abundância. A abordagem DIA tem mais robustez para quantificar proteínas. A reprodutibilidade da DIA está bem estabelecida em alguns estudos multi-laboratoriais. Isso permite a aquisição de grandes cortes amostrais com robustez para construir modelos de classificação dos pacientes e interpretar as alterações fisiopatológicas nas amostras analisadas. A digitalização de mapas de proteínas utilizando DIA permite a coleta dos espectros de fragmentos de todos os peptídeos presentes em uma amostra. Esses mapas podem ser reinterrogados posteriormente com base em novos conhecimentos biológicos ou clínicos. Os contras do DIA são que ele tem sensibilidade e especificidade menores do que o DDA, e que requer um fluxo de trabalho de processamento de dados mais complexo. Nos métodos DIA os espectros MS/MS são altamente multiplexados, o que implica a necessidade de diferentes algoritmos de busca.

Recentemente, novas estratégias de aquisição de dados foram propostas como o WWA, BoxCar e MDA.

III_ Algoritmos computacionais dedicados foram desenvolvidos para atribuir uma sequência peptídica baseada no espectro de fragmentação. As etapas computacionais desde a conversão de dados brutos até a apresentação de dados. Os dados brutos do LC-MS/MS são armazenados em formatos proprietários de fornecedores de instrumentos que precisam ser convertidos para um formato padrão, como mzdata, mzML ou mzXML que permitem a aplicação de diferentes algoritmos de pesquisa. Os dados brutos são convertidos usando o ProteoWizard que inclui msconvert que é capaz de converter dados brutos da maioria dos instrumentos em mzML, ou em um dos muitos outros formatos. Além da conversão de formato, o msconvert também pode executar uma ampla variedade de funções de filtragem de ruído e de coleta de picos para preparar dados para análise.

Para os dados de DDA, quatro estratégias principais foram estabelecidas para a identificação de espectros de massa em tandem derivados de peptídeos: 1. Busca em banco de dados de sequências (*database search*), onde os espectros são pesquisados contra um banco de dados de sequências de proteínas (ou peptídeos) de referência; 2. sequenciamento do espectro ou sequenciamento de novo, onde as sequências de aminoácidos são diretamente inferidas a partir dos espectros; 3. *Sequence-tag*, uma mistura de busca em banco de dados e sequenciamento *de novo* 4. Busca em bibliotecas espectrais, *spectral library search* (experimentais ou preditas), onde os espectros são pesquisados contra uma biblioteca de espectros de compostos conhecidos.

Descrever, pelo menos, uma dessas estratégias em detalhes. Focando no *database search*. O princípio da busca em bancos de dados é comparar espectros experimentais de MS/MS com um banco de dados de sequências de proteínas. Em particular, as proteínas presentes no banco de dados são digeridas e os peptídeos resultantes fragmentados *in silico*. Os espectros MS/MS derivados experimentalmente são pesquisados contra os espectros teóricos usando parâmetros definidos pelo usuário.

Descrever os princípios de três algoritmos de busca de banco de dados como Sequest, Mascot e Andromeda e ressaltar os pontos pros e contras de cada um.

Uma vez que o software de busca de banco de dados retorna a lista de peptídeos candidatos a explicar um espectro MS/MS juntamente com o seu *score*, é importante definir um ponto de corte para identificar os peptídeos corretos. Descrever por exemplo como é feito o cálculo do *E-value* e do False Discovery Rate (FDR).

Sequências peptídicas são utilizadas para inferir quais proteínas estão presentes na amostra analisada. Alguns peptídeos são exclusivos de uma proteína específica, enquanto outros são compartilhados entre proteínas, considerando o espaço de busca utilizado. Descrever alguns algoritmos de inferência como o ProteinProphet.

Para DIA, os espectros MS/MS são complicados, pois contêm os fragmentos de vários íons precursores presentes em cada janela. Os arquivos de espectros DIA contêm mais informações do que DDA porque eles incluem todos os íons em cada janela e não podem ser processados usando os algoritmos clássicos usados para dados DDA. O software para interpretar dados DIA pode ser dividido em baseados em biblioteca espectral (*spectral library-based*) e livre de biblioteca espectral (*spectral library-free*). Descrever os princípios dos dois métodos de pelo menos três software utilizados.

Ferramentas de geração de bibliotecas espectrais, como SpectraST e Skyline, podem usar os resultados de dados DDA para construir uma biblioteca espectral. Com os avanços da inteligência artificial, diversas soluções de software como Prosit, DeepMass e pDeep podem prever computacionalmente os espectros de fragmentos de peptídeos atingindo níveis de precisão comparáveis aos espectros derivados experimentalmente.

Descrever os princípios de software abertos como OpenSWATH, Skyline, DIA-Umpire, DeepNovo-DIA e mencionar software comercialmente disponíveis como Spectronaut e PEAKS para análise de dados DIA.

Questão 02

- A espectrometria de massas (EM) é uma ferramenta poderosa para estudar interações proteína-proteína. Algumas das técnicas baseadas em MS para caracterizar interações proteína-proteína são *Affinity Purification Mass Spectrometry* (AP-MS), marcação por proximidade (*proximity labeling*), *Cross-Linking Mass Spectrometry* (XL-MS), cromatografia de exclusão de tamanho acoplada à espectrometria de massas e co-agregação térmica de proximidade. Cada técnica descrita acima tem seu próprio fluxo de trabalho, incluindo análise experimental e computacional, mas algumas etapas comuns são compartilhadas entre as diferentes técnicas.

Descrever três das técnicas expostas acima em termos de estratégia experimental e computacional.

AP-MS. Esta técnica envolve o isolamento de uma proteína de interesse (isca) e seus parceiros interagentes (presas) de uma amostra biológica complexa usando um ligante por afinidade (STREP, c-MYC ou FLAG) ligada à proteína de interesse. O ligante pode ser inserido exogenamente usando um plasmídeo ou endogenamente usando a técnica CRISPR-Cas9. Os componentes proteicos individuais do complexo são identificados pela EM usando uma estratégia proteômica *bottom-up*. Usando AP-MS, não é possível determinar quais proteínas interagem diretamente com a isca e assim por diante na topologia complexa. Além disso, esta técnica não permite monitorar interações transitórias que são perdidas durante o preparo da amostra. Outra limitação do AP-MS é que a introdução de um ligante específico no terminal N ou C pode induzir mudanças na localização e estrutura da proteína afetando a formação do complexo. Devido à necessidade de lisar células em condições nativas, a identificação de proteínas de membrana é dificultada nesta abordagem. Uma vez que essas abordagens dependem da marcação de proteínas específicas, a investigação do interatoma global consome tempo e recursos. Para AP-MS, os dados MS1 e MS2 obtidos podem ser pesquisados usando mecanismos de busca em banco de dados para determinar a identidade e a quantidade das proteínas interagentes. Uma vez que as proteínas são identificadas, uma análise computacional adicional, como com a Análise de

Significância do INTERactome (SAINT), é necessária para atribuir a confiança ou probabilidade de uma IPP verdadeira. Descrever os princípios da etapa adicional de análise computacional.

- XL-MS é uma técnica que permite determinar quais membros de um complexo estão interagindo diretamente. De fato, esta técnica permite identificar a estrutura do complexo, identificando quais resíduos de cada proteína estão em contato direto. Além disso, XL-MS permite determinar também interações transitórias. A identificação de *crosslinked peptides* revela a proximidade espacial e a orientação das proteínas ou domínios que interagem. Esta técnica envolve a ligação covalente de duas ou mais proteínas interagindo ou domínios proteicos usando um *crosslinker* que reage com os grupos funcionais de resíduos específicos de aminoácidos. Vários resíduos de aminoácidos podem ser ligados e os *crosslinkers* podem abrigar outras propriedades químicas, como grupos de enriquecimento, sítios cliváveis e fragmentos isotópicos. XL-MS tem sido aplicado a complexos proteicos purificados, lisado celular *in vitro* e células *in vivo*. Após a adição de um *crosslinker*, as proteínas são então digeridas com proteases e os peptídeos resultantes são analisados por LC-MS/MS. Descrever os tipos de peptídeos gerados por XL-MS. Descrever os princípios de pelo menos três algoritmos de busca de dados XL-MS como XComb, CXMS, Protein Prospector, Kojak, pLink, SIM-XL, XQuest e XLinkX. Descrever como são selecionadas as identificações corretas.

- A co-agregação térmica por proximidade (TPCA) permite um perfil de todo o sistema da dinâmica de complexos proteicos em células. Este método é baseado nas abordagens de *thermal proteome profile* (TPP) ou *cellular thermal shift assay* (CETSA-MS) inicialmente desenvolvidas para identificar alvos de drogas. A TPP baseia-se no princípio de que o aquecimento de uma proteína causa a sua desnaturação e subsequente precipitação. Ao medir a fração solúvel de uma proteína em diferentes temperaturas pode resultar em um perfil de fusão. O perfil de fusão de uma proteína é característico e pode ser comparado em diferentes condições para elucidar o efeito de modificações pós-traducionais, conformações proteicas, interações fármaco-proteína e interações proteína-proteína. De fato, a TPCA é baseada na hipótese de que proteínas interagindo co-agregam após a desnaturação térmica, levando a solubilidade semelhante em diferentes temperaturas. O TPCA pode ser aplicado em lisados celulares ou em células intactas. Um fluxo de trabalho clássico para esta abordagem envolve o aquecimento a diferentes temperaturas, seguido de ultracentrifugação. A fração solúvel é digerida em peptídeos e os peptídeos resultantes são marcados com etiquetas isobáricas como TMT antes da análise de LC-MS/MS. A marcação auxilia na condensação do grande número de amostras reduzindo a variabilidade da quantificação. Os dados brutos de espectrometria de massa são processados para identificar e quantificar as proteínas medidas. Um software utilizado é o isobarQuant combinado com busca em banco de dados, como Mascot, Andromeda e Sequest. Uma abordagem de normalização é necessária entre as réplicas das diferentes temperaturas ou entre as condições. Curvas de fusão para cada proteína são derivadas. A fim de identificar complexos proteicos, a distância média entre as curvas de fusão é calculada entre todos os pares de subunidades de um complexo proteico. Esta abordagem pode ser aplicada a proteínas purificadas, lisados proteicos e células intactas, tecidos e biofluidos.