Das Proteom. Die Proteomik. Bestimmung der Aminosäure-Sequenz C2BL

CC BY NC SA
© laborberufe.de

Erst das gesamte Arbeitsblatt oder einzelne Abschnitte eigenständig durcharbeiten. Dann Lernvideo mit der Nachbesprechung der Aufgaben anschauen: https://youtu.be/OwGIrrOQmJs :

1. Definition zweier wichtiger Begriffe

Die Gesamtheit aller Proteine in einem Lebewesen, einem Gewebe, einer Zelle oder einem Zellkompartiment wird als **Proteom** bezeichnet. Beispiele währen das Proteom des Menschen, das Proteom der Kartoffelknolle, der Bakterienzelle, oder des Zellkerns. Man kann die Proteine dabei in verschiedene funktionelle Gruppen einteilen.

Lernvideo (6 min) über die Vielfalt der Proteine und was Proteomik bedeutet: *„Proteomik – Vielfalt nach Plan“* https://youtu.be/ej77tOmU2oo

***1.1** Füllen Sie die im Lernvideos „Proteomik – Vielfalt nach Plan“ angesprochenen Proteingruppen.*

Protein-Hauptgruppe	**Zwei Beispiele (aus dem Film oder durch Eigenrecherche)**
Strukturproteine	
	Insulin

Das Proteom steht in einem Gleichgewicht ständiger Neusynthese von Proteinen bei gleichzeitigem Abbau nicht mehr benötigter Proteine. Damit ist das Proteom im Gegensatz zum relativ statischen Genom ständig Änderungen in seiner Zusammensetzung unterworfen. Diese Änderungen werden im Zuge der orts- und zeitabhängigen *Genexpression* über komplexe Regulationsprozesse gesteuert und werden maßgeblich durch Umweltstimuli, Krankheiten, Wirkstoffe und Medikamente beeinflusst. Das Proteom ist somit Spiegel seiner Umwelt und hoch dynamisch. So bilden Zellen, die thermischem Stress oder UV-Strahlung ausgesetzt werden, beispielsweise vermehrt *Hitzeschockproteine*, die zellinterne Strukturen vor diesen negativen Umwelteinflüssen schützen.

Nach der mittlerweile abgeschlossenen Sequenzierung des menschlichen Genoms (man kennt die genaue Basenabfolge aller Gene) verspricht man sich nun als nächstes von der Erforschung des Proteoms, mit der sich die Wissenschaft der **Proteomik** beschäftigt, große Durchbrüche für viele Bereiche. Durch die Proteomik am menschlichen Organismus, erhofft sich beispielsweise eine tiefe Einsicht in die Entstehung vieler Krankheiten und in weiterer Folge, die Entwicklung kausal wirkender Medikamente. **Die Hauptaufgaben der Proteomik ist die Separation, Identifikation und Charakterisierung von Proteinen.** In Bakterienzellen umfasst das Proteom je nach Art um die 1000 bis 10.000 verschiedene Arten von Proteinmolekülen, beim Menschen rechnet man mit 500.000 bis 1.000.000 Proteinspezies.

***1.2** Benennen und beschreiben Sie stichwortartig die drei im Lernvideo „Proteomik – Vielfalt nach Plan“ angesprochenen Mechanismen der Erhöhung der Proteinvielfalt (mehrere Proteine aus einem einzigen Gen). Je ein Satz genügt.*

Unter den bekannten Proteinen gehören einerseits *Massenproteine*, andererseits aber auch nur selten exprimierte Proteine, die nur in bestimmten Zelltypen und dort nur in wenigen (10 – 1000) Kopien vorliegen. Inzwischen ist bekannt, dass der Mensch etwa 20.344 ± 10 vermutete proteincodierende Gene besitzt, allerdings sind nur etwas mehr als 18.000 Proteine zu diesen Genen bekannt. Laut Genkarten muss es also noch weitere Proteine geben, die noch nicht bekannt/entdeckt sind und beispielsweise nur während der embryonalen oder fetalen Entwicklung exprimiert werden.

Werkzeuge zur Bestimmung der Aminosäuresequenz

2. Ansequenzierung statt vollständiger Sequenzierung

Die Notwendigkeit einer **vollständigen** Aminosäure-Sequenzanalyse eines Proteins im Labor ist sehr selten geworden, denn mittlerweile existieren Datenbanken, die bei Kenntnis einiger weniger kurzer charakteristischer Aminosäure-Abfolgen von Proteinfragmenten, einen Vergleich mit bisher bekannten Proteinen erlauben. Hat man einige wenige charakteristische Puzzlestücke eines Proteins, so kann anhand derer, die vollständige Aminosäuresequenz nachgeschlagen werden. Hierbei können auch weitere experimentelle Ergebnisse, wie der ungefähren molaren Masse, die Kenntnis der N- oder C-terminalen Aminosäure etc. die möglichen Kandidaten deutlich einschränken. Eie große Reduktion der Daten ergibt sich auch, wenn man den Wirtsorganismus kennt, was ja meist der Fall ist. Handelt es sich z.B. um ein humanes Protein, braucht die Datenbank nur die bekannten menschlichen Proteine zu durchsuchen. Es bleibt als praktische Laborarbeit also nur die Isolation des Proteins (z.B. Elektrophorese) und danach die sogenannte **Ansequenzierung**, also die Ermittlung einiger weniger Aminosäureabfolgen und evtl. Bestimmung weiterer Charakteristika (z.B. Molare Masse über die Elektrophorese).
Kleine mathematische Überschlagsrechnung: Geht man von ungefähr 20 proteinogenen Aminosäuren aus, so gibt es für die letzte Aminosäure 20 Möglichkeiten. Für die Abfolge der letzten beiden Aminosäuren zusammen gibt es 20·20 Möglichkeiten, also 400. Für die Abfolge der letzten 3 Aminosäuren sind es schon $20 \cdot 20 \cdot 20 = 20^3 = 8000$ Möglichkeiten. Ermittelt man beispielsweise die letzten 15 Aminosäuren, so gibt es rechnerisch schon mehrere Billiarden Möglichkeiten. Das sind viel mehr Möglichkeiten, als es bekannte Proteine gibt. Will heißen: So lange müssen die Aminosäureabfolgen also gar nicht sein, vor allem, wenn es mehrere bekannte Puzzlestücke gibt oder weitere Fakten bekannt sind, z.B. die Molekülmasse oder der Wirtsorganismus.

Man könnte meinen, dass dieses Verfahren voraussetzt, das das zu identifizierende Protein schon bekannt ist und seine Aminosäuresequenz in der Datenbank hinterlegt wurde. Aber selbst wenn das nicht der Fall ist, kommt man weiter! Man kann die Aminosäureabfolge in eine DNA-Basenabfolge übersetzen lassen und dann in Genom-Datenbanken weiter suchen! Hat man das Gen gefunden, so kann man mit einigen weiteren Kenntnissen zu diesem Gen (z.B. Intron-Regionen o.ä.) in eine hypothetische Aminosäureabfolge zurück übersetzen lassen, diesmal aber das gesamte Gen. So hat man dann die gesamte Aminosäuresequenz, oder zumindest große Teile davon (wenn man posttranslationale Modifikationen etc. ausblendet).
Eine *vollständige* Neusequenzierung **(*De-novo*-Sequenzierung)** ist deshalb sehr selten geworden.

3. Ein wichtiges Werkzeug zur Identifikation: Enzymatische Zerlegung in Fragmente

Statt die AS-Sequenz ein längeres Stück Protein (z.B. 20 AS) zu analysieren, kann man es einem **Verdau** aussetzen, also der Spaltung durch spezielle Enzyme: **Proteasen** und **Peptidasen** spalten Aminosäureabfolgen hydrolytisch an den Peptidbindungen. Hierbei gibt es **Exopeptidasen** (exo: gr: „von außen“, „außen“), die endständig je nach Enzym die N- oder C-terminale Aminosäure abspalten. Weiterhin gibt es **Endopeptidasen**, die innerhalb des Aminosäurestrangs spalten und zwar vor oder nach bestimmten Aminosäuren. So spaltet **Trypsin** hinter (d.h. Richtung C-Terminus) den basischen AS *Lysin (Lys)* und *Arginin (Arg)*. **Chymotrypsin B** spaltet bevorzugt an Peptidbindungen, deren Carbonylgruppe von einer aromatischen Aminosäure stammen, also hinter *Tyrosin (Tyr)*, *Tryptophan (Trp)* und *Phenylalanin (Phe)*.

3.1 Enzymatische Zerlegung: *Ein Polypeptid wurde einmal mit Trypsin und einmal mit Chymotrypsin behandelt und die Fragmente chemisch analysiert. Es finden sich folgendes Ergebnis:*

Hydrolyse mit dem Enzym Trypsin	***Hydrolyse mit dem Enzym Chymotrypsin B***
Trp-Phe-Arg	*Phe*
Leu-Gly-Leu-Leu-Phe	*Lys-Leu-Gly-Leu-Leu-Phe*
Ala-Leu-Gly-Met-Lys	*Ala-Leu-Gly-Met-Lys-Trp*
Ala-Ala-Ser-Met-Ala-Phe-Lys	*Arg-Ala-Ala-Ser-Met-Ala-Phe*

a) Widerspricht Fragment Nr. 2 beim Trypsinverdau (Leu-….-Phe) dem erklärenden Text? Begründen Sie!

b) Ermitteln Sie die Aminosäuresequenz

Die einzelnen Fragmente können bei Bedarf beispielsweise elektrophoretisch voneinander getrennt werden. Die anschließende Identifizierung, also die Bestimmung der Aminosäuren innerhalb der Fragmente, erfolgt heutzutage über die **Massenspektrometrie.** Mittlerweile sind die Computer so leistungsfähig geworden, dass man zunehmend auf die Trennung der Fragmente/Proteine verzichten kann. Mithilfe der Computer und Datenbanken können sozusagen nicht nur die Puzzlestücke (AS-Fragmente) zu einem ganzen Puzzle (Protein) kombiniert werden, sondern aus ein Gemisch aus Puzzlestücken verschiedener Puzzles gleichzeitig auf alle Puzzles identifiziert werden.

4. Crashkurs Massenspektroskopie

Überblick über alle massenspektrometrischen Verfahren/Zusammenfassung: Die zu untersuchenden Fragmente werden in die Gasphase überführt und ionisiert. Die Ionen werden anschließend durch ein elektrisches Feld beschleunigt und dem Analysator zugeführt, der sie nach ihrem Masse-zu-Ladung-Verhältnis *m/z* detektiert.

Zur Erzeugung geladener gasförmiger Probemoleküle gibt es verschiedene Verfahren. Am häufigsten verbreitet zur Peptiduntersuchung ist die MALDI/TOF-Technik. Die **M**atrix–**A**ssistierte **L**aser–**D**esorption–**I**onisierung (MALDI) ist ein dreistufiger Prozess:

- Die Probe wird in einer Matrix einbgebettet und auf einem Träger fixiert.
- Ein gepulster Laser-Strahl löst Moleküle als heißes Gas aus der Probe; dieser Vorgang wird Ablation bzw. Desorption genannt.
- Die Abgabe oder Annahme von Protonen wandelt die befreiten Moleküle zu Ionen um.

Schließlich werden die Ionen beschleunigt und ihre jeweilige Flugzeit (time of flight, TOF) registriert, bis sie auf einen Detektor treffen. Die Flugzeit korreliert nämlich mit ihrer Masse (etwas genauer, aber hier nicht weiter relevant: Masse-zu-Ladung-Verhältnis). Hierbei wird ausgenutzt, dass die Ionen beim Eintritt in den Analysator alle die gleiche Energie haben. Leichte Ionen müssen deshalb schneller sein als schwere. Daher erreichen beim Flug durch den Raum leichte Ionen den Detektor früher als schwere Ionen. Der Flugzeitanalysator besteht somit nur aus einem Rohr unter Vakuum mit einem Detektor am Ende. Die Massen-Spektren sind für bestimmte Aminosäureabfolgen, für ganze Proteine, für Gemische und ja, sogar für ganze Bakterien charakteristisch! So kann man mittlerweile kultivierte Bakterien anhand ihres Massenspektrums identifizieren, das großteils auf dem Proteom dieses Organismus basiert.

Krieg der Sterne im Spektrometer
https://youtu.be/0jeFpXHZ8W0

sehenswertes Lernvideo der Max-Planck-Gesellschaft
https://youtu.be/XcoZ8AdgNvY

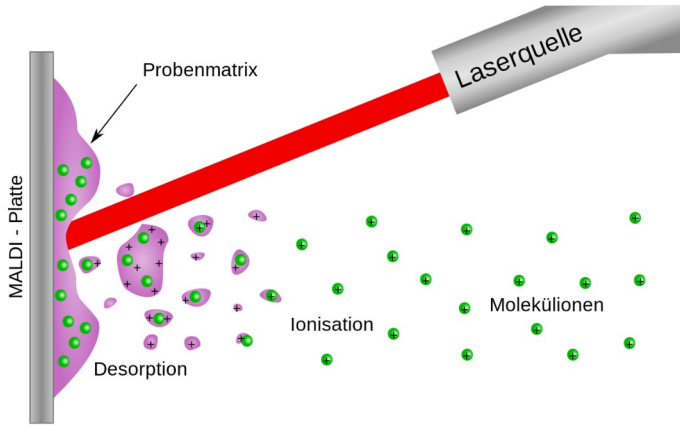


Abb. 4.1 Matrix-unterstützte Laser-Desorption/Ionisation (MALDI). Quelle: wikipedia.de Stichwort: MALDI-TOF

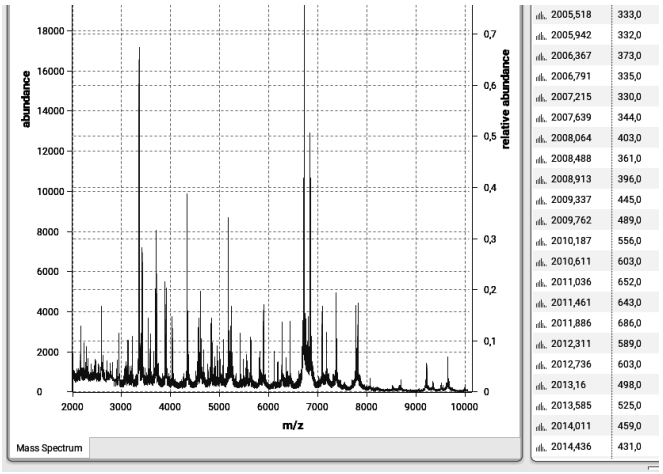
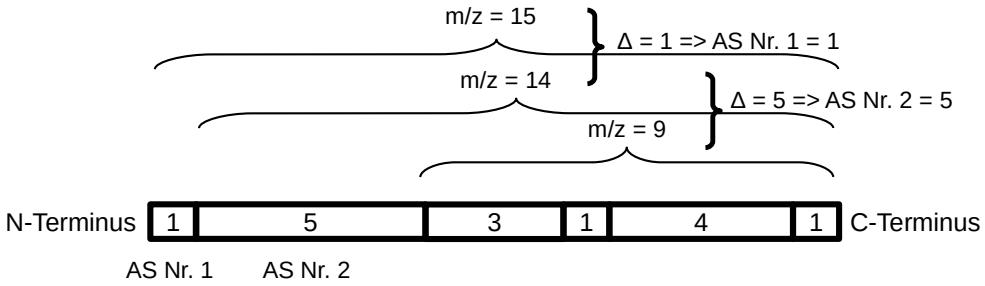


Abb. 4.2 Massenspektrum von *Bacillus cereus*
https://commons.wikimedia.org/w/index.php?curid=64607316 (verändert)

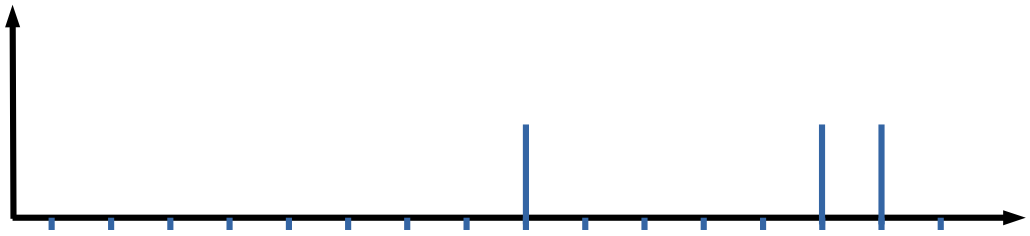
***4.1** Benennen Sie ein anderes Verfahren zur Erzeugung geladener Proteinfragmenten und beschreiben Sie kurz den wichtigsten Unterschied zu MALDI-TOF (vgl. Lernvideo: „Proteomik – eine Bibliothek von Proteinen“)*

Das Massenspektrum ist dann wie ein Fingerabdruck, und das Vorhandensein von Peaks, Peakhöhe und Ähnlichkeit von Spektren lässt sich die Materie mithilfe von Datenbanken identifizieren. Die Bestimmung der Aminosäuresequenz beruht darauf, dass die Aminosäurekette bevorzugt an den Peptidbindungen auseinander bricht. So entstehen charakteristische Restfragmente, die jeweils um eine Aminosäure kürzer sind. Die Massendifferenzen (Δ) benachbarter Peaks entsprechen daher der Masse der nächsten anhängenden Aminosäure.

Schematisches Beispiel-Peptid: *Hexapeptid mit 4 AS-Sorten (charakteristische „Massen" (m/z) = 1, 3, 4 und 5):*



4.2 *Hier das zum Beispiel-Peptid gehörende schematische Spektrum. Beschriften Sie die Achsen und vervollständigen Sie es. Die auch oben aufgeführten drei Peaks mit der höchsten „Masse (m/z), sind schon eingezeichnet. Beachten Sie nur das rechtsseitige Fragment (Richtung C-Terminus).*



Die Realität ist wegen unterschiedlicher Isotope die berücksichtigt werden müssen, etwas mehr Varianz in den Spaltungsstellen, Berücksichtigung der linksseitigen Fragmente etc. komplexer

5. Denovo-Sequenzierung über eine Tandem-Massenspektroskopie

Im Gegensatz zu der üblicheren Ansequenzierung mittels Massenspektrometrie und anschließender Datenbank-Recherche erfordert die **de-*Novo*-Peptidsequenzierung** keine Informationen (Datenbanken o.ä.), weshalb man auch von ***de novo*** spricht. Hier nutzt man sogenannte **Tandem-Massenspektrometer.** Wie der Name schon andeutet, handelt es sich um zwei hintereinander geschaltete Massenspektrometer: Die nach der ersten massenspektrometrischen Fragmentierung entstehenden Bruchstücke (aus jeweils wenigen AS bestehend) werden gleich massenspektrometrisch weiter analysiert (nach dem im vorangegangenen Abschnitt erwähnten Schema). Durch Aneinanderreihung der Fragmente, kann man wieder auf das gesamte Protein schließen, indem man nach Überlappungsregionen sucht (wie bei dem oben beschrieben enzymatischen Verfahren).

6. Wechsel auf genomische Methoden

Statt wenige Mikrogramm eines wertvollen Proteins direkt zu Sequenzierung zu opfern, kann man, wenn man die entsprechenden DNA-Abschnitte kennt, lieber deren Basenabfolge bestimmen und so auf die Aminosäuresequenz schließen. Das kann strategische Vorteile bieten, den DNA-Abschnitte lassen sich mittels PCR vervielfältigen (amplifizieren). So hat man ausreichend Ausgangsmaterial zur Sequenzierung. Die Sequenzierung erfolgt weiterhin heute vollautomatisch und ist wenig arbeitsintensiv.

Allerdings ist der rohe DNA-Abschnitt nicht immer passend zum fertigen Protein, denn häufig kommt es zu posttranskriptionalen (⇒ Spleißen) und posttranslationalen Modifikationen.

Sie sollten jetzt alle Aufgaben aus den Abschnitten 1 und 2 auf dem Aufgabenblatt zu Aminosäuren und Proteinen lösen können. Link: http://www.laborberufe.de/Biolaboranten%20-%202.%20Lehrjahr/Aufgaben_Aminosaeuren_Proteine_+_Lsg_BioLab.pdf