

Bioinformatik von Proteinsequenzen

von Dr. Horst Lohrer



© Getty Images Plus/iStock/non-exclusive

Die Methoden der Molekularen Biologie, insbesondere die Sequenzierung von DNA und Proteinen, haben gigantische Mengen digitalisierter Daten generiert. Für die Analyse und Visualisierung dieser Daten bietet die Informatik die notwendigen Methoden. Die Aufgabe der Bioinformatik ist es, sicherzustellen, dass die biologischen Daten geordnet abgelegt werden, abrufbar sind und durch kreative Kombination zu neuen Erkenntnissen führen. Die vorliegende Unterrichtseinheit für den Biologieunterricht der gymnasialen Oberstufe bietet eine Einführung in die Thematik. Dabei liegt der Schwerpunkt auf dem Aspekt der Biologie in der Bioinformatik. Es wird dabei ausschließlich mit Proteinsequenzen gearbeitet. Die einzelnen Kapitel werden themenbezogen eingeleitet und erläutern notwendige Begriffe für die Benutzung der Algorithmen. Die praktische Arbeit mit öffentlich zugänglichen Datenbanken und Analyseprogrammen steht im Vordergrund der Arbeit. Den Beteiligten dieser Unterrichtseinheit wird ein erster Blick in die Methoden der Bioinformatik eröffnet, was hoffentlich zu eigenen Recherchen motiviert.

Impressum

RAABE UNTERRICHTS-MATERIALIEN Biologie Sek. II

Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt. Es ist gemäß § 60b UrhG hergestellt und ausschließlich zur Veranschaulichung des Unterrichts und der Lehre an Bildungseinrichtungen bestimmt. Die Dr. Josef Raabe Verlags-GmbH erteilt Ihnen für das Werk das einfache, nicht übertragbare Recht zur Nutzung für den persönlichen Gebrauch gemäß vorgenannter Zweckbestimmung. Unter Einhaltung der Nutzungsbedingungen sind Sie berechtigt, das Werk zum persönlichen Gebrauch gemäß vorgenannter Zweckbestimmung in Klassenstärke zu vervielfältigen. Jede darüber hinausgehende Verwertung ist ohne Zustimmung des Verlages unzulässig und strafbar. Hinweis zu §§ 60a, 60b UrhG: Das Werk oder Teile hiervon dürfen nicht ohne eine solche Einwilligung an Schulen oder in Unterrichts- und Lehrmedien (§ 60b Abs. 3 UrhG) vervielfältigt, insbesondere kopiert oder eingescannt, verbreitet oder in ein Netzwerk eingestellt oder sonst öffentlich zugänglich gemacht oder wiedergegeben werden. Dies gilt auch für Intranets von Schulen und sonstigen Bildungseinrichtungen. Die Aufführung abgedruckter musikalischer Werke ist ggf. GEMA-meldepflichtig.

Für jedes Material wurden Fremdrechte recherchiert und ggf. angefragt.

Dr. Josef Raabe Verlags-GmbH
Ein Unternehmen der Klett-Gruppe
Rotebühlstraße 77
70178 Stuttgart
Telefon +49 711 62900-0
Fax +49 711 62900-60
meinRAABE@raabe.de
www.raabe.de

Redaktion: Anne Zörlein
Satz: Röser MEDIA GmbH & Co. KG, Karlsruhe
Bildnachweis Titel: Getty Images Plus/iStock/non-exclusive
Korrektorat: Josef Mayer

M 2d Austausch von Aminosäuren in Proteinen

Ketten von Aminosäuren erfüllen als Proteine bestimmte Funktionen, die von einer korrekten Tertiär- und Quartärstruktur abhängig sind. Führt eine Mutation in der codierenden Sequenz der DNA zum Austausch einer Aminosäure, dann ist dieser Austausch:

1. **folgenlos**, weil die neue Aminosäure die Aufgabe der substituierten Aminosäure vollständig erfüllt, oder weil der Austausch in eine neutrale Zone des Proteins fällt (= neutrale Mutation; kein Effekt in der natürlichen Selektion).
2. **katastrophal**, weil die neue Aminosäure die korrekte Tertiärstruktur zerstört oder an einer kritischen Stelle der Aminosäurekette die Aufgabe der substituierten Aminosäure nicht erfüllen kann (*loss of function*, negative Mutation).
3. **vorteilhaft**, weil mit der neuen Aminosäure die Funktion des Proteins besser an die Umweltbedingungen des Organismus angepasst ist und dessen *Fitness* erhöht wird (*gain of function*, positive Mutation).

Aminosäuren kommen verschieden häufig in Proteinen vor. Glycin, Alanin, Leucin und Lysin sind die häufigsten in Proteinen verwendeten Aminosäuren. Die Wahrscheinlichkeit eines Austausches mit einer anderen Aminosäure entspricht nicht der Häufigkeit des Vorkommens: Asparagin, Serin, Asparaginsäure und Glutamin sind die am häufigsten substituierten Aminosäuren (siehe Tabelle 3).

Tab. 3: Relative Häufigkeit von Aminosäuren in Proteinen in % und deren relative Mutabilität.⁵

Relative Häufigkeit des Vorkommens der Aminosäuren [%]			
Gly	8,9	Arg	4,1
Ala	8,7	Asn	4,0
Leu	8,5	Phe	4,0
Lys	8,1	Gln	3,8
Ser	7,0	Ile	3,7
Val	6,5	His	3,4
Thr	5,8	Cys	3,3
Pro	5,1	Tyr	3,0
Glu	5,0	Met	1,5
Asp	4,7	Trp	1,0

Relative Mutabilität von Aminosäuren (Ala = 100)			
Asn	134	His	66
Ser	120	Arg	65
Asp	106	Lys	56
Glu	102	Pro	56
Ala	100	Gly	49
Thr	97	Tyr	41
Ile	96	Phe	41
Met	94	Leu	40
Gln	93	Cys	20
Val	74	Trp	18

nach DAYHOFF (1978)

M 2e Fragen zu M 2

1. Nehmen Sie die Buchstabenfolge ihres Namens als Aminosäure-Sequenz eines Oligopeptides im 1-Buchstaben-Code. **Bilden** Sie nun einen möglichen genetischen Code ihres Namens für menschliche DNA mithilfe einer Code-Sonne. Falls Ihr Name mit diesen Buchstaben nicht gebildet werden kann, nehmen Sie den Namen DARWIN. Beachten Sie, dass das menschliche Genom zu 65 % aus A/T besteht.
2. **Informieren** Sie sich in Wikipedia (engl.) über den Strukturbegriff *protein domain*. **Beurteilen** Sie, ob es sich bei einem „Zink-Finger“ um eine Person handelt, die das Metall Zink besonders schnell erkennen kann.
3. Die vollständige Hydrolyse eines Proteins und die folgende qualitative und quantitative Analyse der Aminosäuren ergab folgende Summenformel:
(Gly)₂₅-(Ala)₂₄-(Pro)₄₅-(Val)₁₈-(Leu)₃₂-(Ile)₈-(Met)₄₂-(Phe)₁₁-(Tyr)₉-(Trp)₄-(Ser)₃₈-(Thr)₂₂-(Cys)₁₀-(Asn)₁₄-(Gln)₁₅-(Lys)₂₀-(His)₁₂-(Arg)₂₀-(Asp)₂₀-(Glu)₃₀.
Ermitteln Sie das Molekulargewicht des Proteins in [u] und [kDa].
4. **Informieren** Sie sich über den Begriff „Zentrales Dogma der Molekularbiologie“. **Stellen** Sie ein einfaches Schema dieses Dogmas **auf** und **erklären** Sie seine Bedeutung. **Erläutern** Sie, warum die Entdeckung des Enzyms Reverse Transkriptase der Retroviren das *zentrale Dogma* einschränkt.
5. Die PAM1-Matrix (Tabelle 4) gibt für jede Aminosäure die Wahrscheinlichkeit für ihren Austausch in Proteinen an. So wird z. B. Lysin (Lys, K) am häufigsten durch Arginin (Arg, R) ersetzt, jedoch nicht gegen Phenylalanin (Phe, F). **Erläutern** Sie diesen Unterschied anhand der jeweiligen genetischen Codes und der Eigenschaft des Restes der Aminosäuren.

M 3 Datenbanken

Datenbanken sind für die moderne biologische Forschung unerlässlich. Sie müssen sowohl dem Aspekt der Bereitstellung von Information genügen als auch durch Kombination von Information die Entdeckung neuer Information möglich machen. Es gibt mehr als 400 biologische Datenbanken, die sich auf verschiedene Datensammlungen spezialisiert haben. Alle Datenbanken setzen Anfragen in englischer Sprache voraus. Folgende Datenbanken können unterschieden werden:

- **Primäre Datenbanken:** enthalten Nukleotid- und Protein-Sequenzen, die wenigstens über Stichworte recherchierbar sind.
- **Sekundäre Datenbanken:** ermöglichen darüber hinaus weitere Analysen z. B. bezüglich der Struktur der Moleküle.
- **Abgeleitete Datenbanken:** enthalten gefilterte Informationen und bedienen oft spezielle Forschungsinteressen. Abgeleitete Datenbanken sind untereinander vernetzt und erlauben so die Heranziehung weiterer spezieller Information.

M 3a Zwei wichtige Datenbanken

Die **UniProt**-Datenbank ist eine Kollaboration des *European Bioinformatics Institute* (EMBL-EBI), des *Swiss Institute of Bioinformatics* (SIB) und der *Protein Information Resource* (PIR) der Georgetown University, Washington, und ist eine der wichtigsten Sammlungen von Protein-Sequenzen. Es handelt sich um eine abgeleitete Datenbank, d. h. jeder Eintrag wird von Spezialisten überprüft und relevante Links zu anderen Datenbanken hergestellt. UniProt ist in verschiedene Unter-Datenbanken unterteilt, welche spezifische Sammlungen von Proteinen beinhalten (z. B. Proteome von vollständig sequenzierten Genomen). Die Datenbank „UniProtKB (UniProt**knowledgebase**)“ ist für die hier vorgestellten Probleme am besten geeignet. Sie kann kostenfrei erreicht werden unter: www.uniprot.org.

M 3c Fragen zu M 3

1. **Informieren** Sie sich mithilfe von UniProt (www.uniprot.org/) über das Enzym ABF2_BACSU. **Recherchieren** Sie wie der vollständige Name des Enzyms lautet, um welchen Organismus es sich handelt und welche Aufgabe das Enzym erfüllt. **Vergleichen** Sie dies mit der gleichen Anfrage an PubMed. **Suchen** Sie in PubMed nach den Veröffentlichungen eines Autors mit Ihrem Namen.
2. **Extrahieren** Sie die Sequenz des Enzyms Katalase (*catalase*) des Menschen aus der UniProt-Datenbank. **Legen** Sie eine Word-Datei an, in der Sie die Funktion des Enzyms, seinen *Identifier* und seine FASTA-Sequenz **eintragen**.
3. Katalase ist ein wichtiges Enzym des zellulären Stoffwechsels des Menschen. Änderungen der Proteinsequenz (*polymorphism*) werden mit Stoffwechselstörungen in Verbindung gebracht. **Recherchieren** Sie Information zu dem Enzym Katalase in PubMed. Beschränken Sie Ihre Suche auf Übersichtsartikel (*review*). **Fassen** Sie die Informationen in einem einseitigen Dokument (plus Quellenangaben) mit dem Titel *Polymorphismen der Katalase und ihre Assoziation mit Stoffwechselstörungen* **zusammen**.
4. Geben Sie folgende Proteinsequenz in der Suchzelle ein: VNOHLCGSHLVEAL. **Nennen** Sie den Namen des Proteins und Arten, bei denen diese Sequenz vorkommt.
5. Die α -Helix ist eine häufige Sekundärstruktur in Proteinen. Die Berechnungen von RAMACHANDRAN legen nahe, dass bestimmte Aminosäuren eher in der Lage sind, diese Struktur aufzubauen, als andere. Als Arbeitshypothese formuliert: „In α -Helices von globulären Proteinen werden bestimmte Aminosäuren häufiger angetroffen, als es ihrem allgemeinen Vorkommen in Proteinen entspricht.“ **Beurteilen** Sie diese Aussage anhand folgender Proteine des Menschen: Cytochrom c (P99999), DNA-binding protein inhibitor (Q02535), Ribonuklease (P07998), Hsp40 (P25685), Myoglobin (P02144) und HLA class II histocompatibility antigen (P06340). Öffnen Sie jeweils das Datenblatt in UniProt, **informieren** Sie sich über die Funktion des Proteins, gehen Sie auf die 3-dimensionale Darstellung und identifizieren Sie die Aminosäuren in einer oder mehreren Helices durch Aktivierung der Position mittels Pointer. **Notieren** Sie die Sequenz der Helices in einem Word-Dokument. **Berechnen** Sie die relative Häufigkeit jeder Aminosäure und vergleichen Sie Ihre Daten mit der allgemeinen Häufigkeit von Aminosäuren (siehe Abbildung von UniProt).