



Biophysikalisches Praktikum
Institut für Biophysik
Johann Wolfgang Goethe-Universität
Frankfurt am Main



Proteindatenbanken und Strukturanalyse von Proteinen am Computer

Prof. Dr. Achilleas Frangakis, Dr. Margot Frangakis, Katharina Eberl, Andreas Rau, Christian Weidemüller, Andreas Messmer, Dr. Georg Wille und Karin Hauser

1 Biologische Datenbanken

Biologische Datenbanken sind reich an Informationen aus den Biowissenschaften, die aus wissenschaftlichen Experimenten, veröffentlichter Literatur, Hochdurchsatz-Experimentiertechnologien und Computeranalysen stammen. Sie enthalten Informationen aus Forschungsgebieten wie Genomik, Proteomik, Metabolomik, Microarray-Genexpression und Phyloenetik. In biologischen Datenbanken enthaltene Informationen umfassen Genfunktion, Struktur, Lokalisation (sowohl zellulär als auch chromosomal), klinische Auswirkungen von Mutationen, und was uns höchstens betrifft, die Ähnlichkeiten von biologischen Sequenzen und Strukturen.

Das biologische Wissen wird auf viele verschiedene allgemeine und spezialisierte Datenbanken verteilt. Dies macht es manchmal schwierig, die Konsistenz der Informationen sicherzustellen. Integrative Bioinformatik ist ein Bereich, der versucht, dieses Problem durch Bereitstellung eines einheitlichen Zugangs zu lösen. Eine Lösung besteht darin, wie biologische Datenbanken auf andere Datenbanken mit Zugangsnummern verweisen, um das zugehörige Wissen miteinander zu verknüpfen.

Durch die Entwicklung neuer Methoden zur Sequenzanalyse und zur Strukturbestimmung steigt die vorhandene Datenanzahl von bekannten Sequenzen und Strukturen biologischer Moleküle nahezu

exponentiell an. Dieses stetige Wachstum von Sequenz- und Strukturdaten (vgl. Abb. 1 und Abb. 2) macht eine Speicherung und Verwaltung über zentrale Datenbanken unverzichtbar. Zudem garantieren Datenbanken eine nicht redundante, sichere Datenintegration verschiedener Datentypen, auf die effizient zugegriffen werden kann.

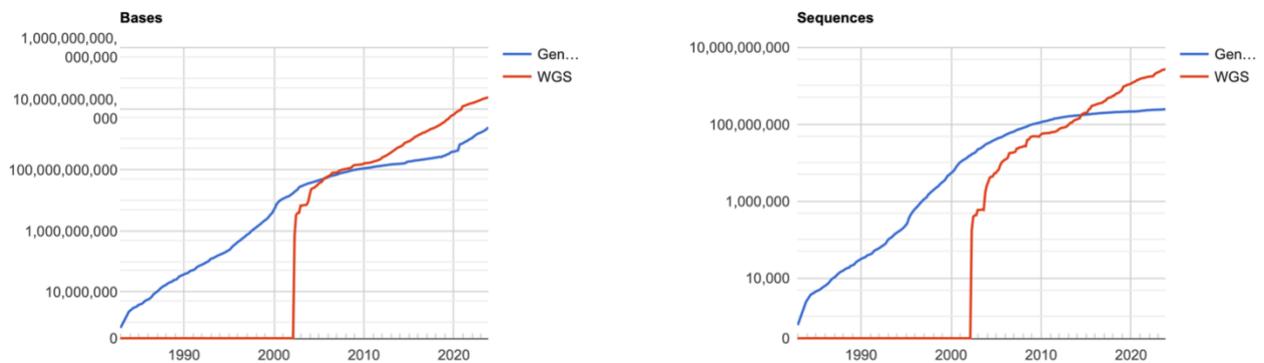


Abbildung 1a: Die Anzahl der Basen und die Anzahl der Sequenzdatensätze in der GenBank und WGS (Whole Genome Shotgun) pro Jahr seit 1982 (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>).

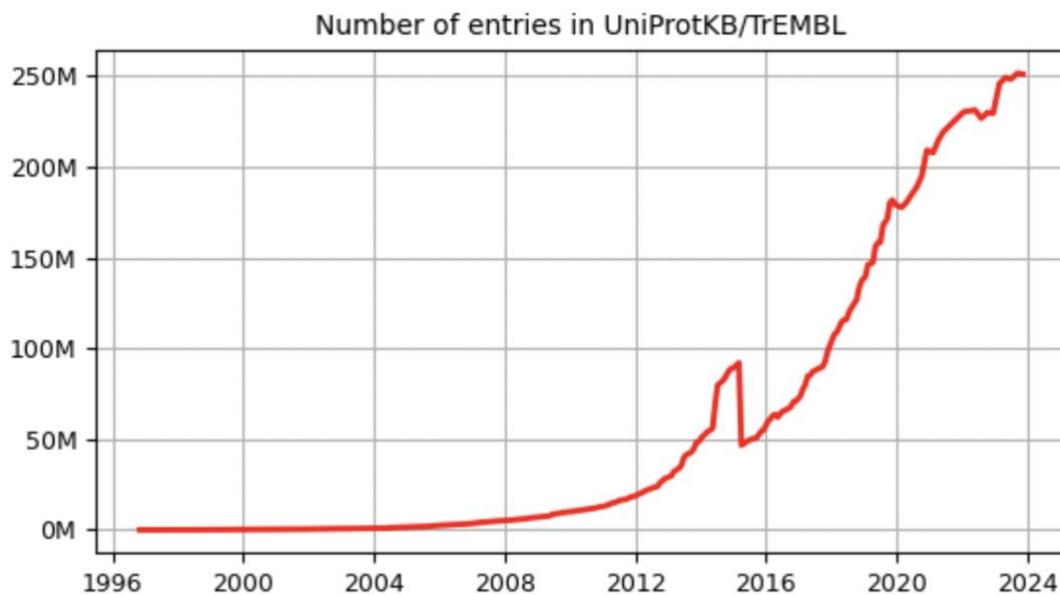


Abbildung 1b: Wachstum von Sequenzdaten (<https://www.ebi.ac.uk/uniprot/TrEMBLstats>)

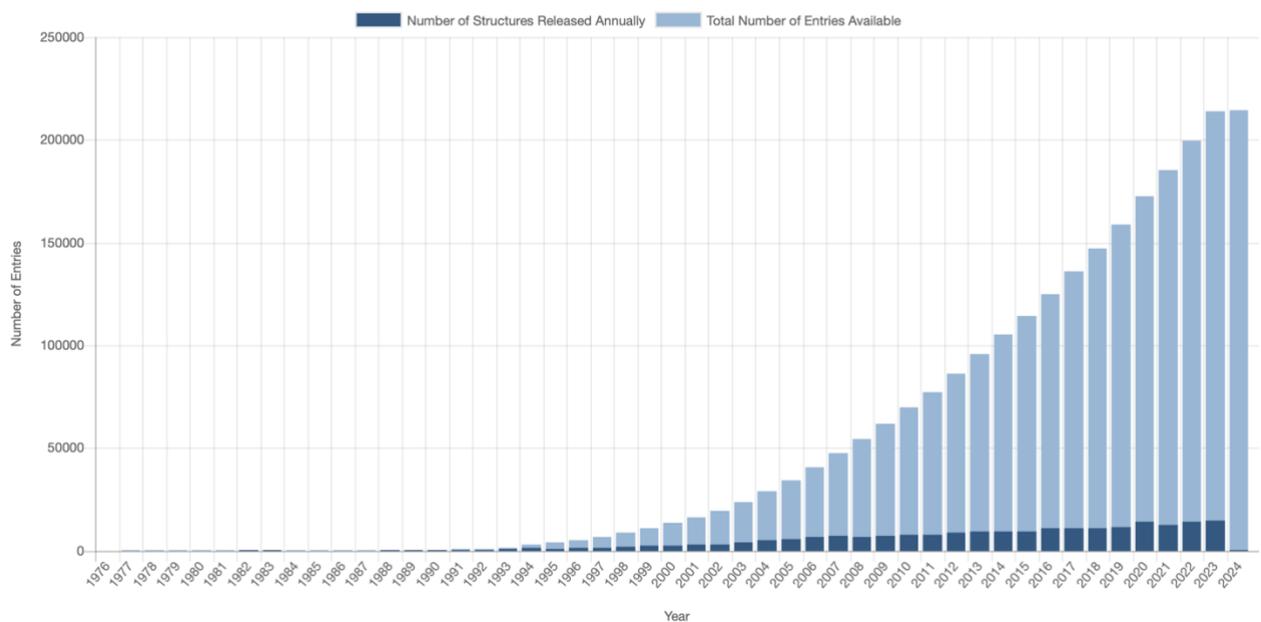


Abbildung 2: Wachstum von Strukturdaten (<https://www.rcsb.org/stats/growth/growth-released-structures>)

In diesen Datenbanken werden unterschiedlichen Typen von Daten (z.B. Sequenzen, 3D-Strukturen, Sequenzmotive) gespeichert. Abbildung 1a zeigt die Wachstum von Sequenzdatensätze in der GenBank und WGS (Whole Genome Shotgun) und Abbildung 1b zeigt die Wachstum von Sequenzdatensätze in der Uniprot Datenbank. Abbildung 2b zeigt die Wachstum von der 3D-Struktur-Datenbanken (PDB).

Jede Datenbank verwendet Basisfunktionalitäten, welche die einheitliche Verwaltung aller Daten, eine Integritätssicherung und eine Datensicherung garantieren. Auch der freie Zugriff auf die Daten durch mehrere Personen zum selben Zeitpunkt wird von einem Datenbankmanagementsystem ermöglicht. Außerdem können durch Import- oder Export-Funktionen Daten problemlos integriert oder exportiert werden. Die freie Verfügbarkeit und die Vielzahl der Daten bietet ein breites Anwendungsgebiet in der Wissenschaft, besonders in der Bioinformatik und Biophysik. Mit Hilfe der Daten können z.B. Simulationsrechnungen an Proteinen durchgeführt werden, die Aufschluss über die Funktion komplexer biologischer Systeme geben können.

2 Grundlagen

2.1 Strukturanalysen

Zur Bestimmung der dreidimensionalen Struktur eines Proteins mit hoher Auflösung werden folgende Methoden angewendet: die Röntgenstrukturanalyse, die Freie-Elektronen-Röntgenlasern (XFEL), die NMR-Spektroskopie (nuclear magnetic resonance) und die Elektronenmikroskopie. Die auf diese Weise gewonnenen Informationen über den Aufbau und die Atomkoordinaten einer bestimmten Struktur werden in Proteindatenbanken bereitgestellt (siehe Kap. 4.2).

2.1.1 Die Röntgenstrukturanalyse

Röntgenstrukturanalyse kann die Struktur von Proteinen bis zur atomaren Ebene aufzulösen. An den Atomen werden die Röntgenstrahlen gebeugt und die Spektren (Power Spektrum/ Beugungsmuster) können gemessen werden. Aus den Beugungsmustern können unter bestimmten Voraussetzungen die Elektronendichten und die Positionen der einzelnen Atome errechnet werden. Voraussetzung für diese leistungsfähige Methode ist, dass die Proteine als Kristall vorliegen und damit die einzelnen Atome regelmäßig angeordnet sind. Allerdings ist die Kristallisation von Membranproteinen und Proteinkomplexen technisch sehr aufwendig und stellt meist eine große Herausforderung dar.

2.1.2 Freie-Elektronen-Röntgenlasern (XFEL)

Eine neue Technologie, die als serielle Femtosekunden-Kristallographie bezeichnet wird, revolutioniert die Methoden der Röntgenkristallographie. Ein Freie-Elektronen-Röntgenlaser (XFEL) wird verwendet, um extrem kurze (nur Femtosekunden dauernde) und extrem helle Strahlungspulse zu erzeugen. Ein Strom winziger Kristalle (Nanometer bis Mikrometer groß) wird durch den Strahl geleitet, und jeder Röntgenpuls erzeugt ein Beugungsmuster aus einem Kristall, der dabei oft verbrennt. Aus bis zu Zehntausenden dieser einzelnen Beugungsmuster wird ein vollständiger Datensatz zusammengestellt. Die Methode ist sehr leistungsfähig, weil sie es Wissenschaftlern ermöglicht, molekulare Prozesse zu untersuchen, die über sehr kurze Zeiträume ablaufen, wie beispielsweise die Absorption von Licht durch biologische Chromophore.

2.1.3 NMR-Spektroskopie

Die NMR-Spektroskopie kann Strukturen von Proteinen in nativer Umgebung auflösen. Die Atomkoordinaten der Moleküle können mit einer ähnlichen Genauigkeit ermittelt werden wie es auch mit der Röntgenstrukturanalyse möglich ist. Bei der NMR-Methode wird mit Hilfe extrem starker Magnetfelder der Zustand des Kernspins beeinflusst. Beim Relaxieren aus dem angeregten Zustand wird Energie frei, die in Form von Strahlung im Radiofrequenzbereich abgegeben wird.

Die Frequenz ist spezifisch für die Atomart und von der unmittelbaren Umgebung des Atoms abhängig. Somit kann die Molekülstruktur daraus errechnet werden. Da das Molekül nicht in einer festen Kristallstruktur vorliegt und die eigentliche Messzeit sehr kurz ist ($< 1s$), können auch Konformationsänderungen in Proteinen mit dieser Methode detektiert werden. Eine Schwierigkeit besteht darin, dass die Kerne der wichtigsten biologischen Atome wie z.B. die des Kohlenstoffs keine kernmagnetische Resonanz aufweisen. Sie müssen durch entsprechend kernmagnetresonanzfähige Isotope des entsprechenden Elements ersetzt werden. Auch ist die Größe des zu untersuchenden Proteins stark begrenzt. Moderne NMR-Spektrometer, die mit einer Betriebsfrequenz bis zu 1.2 GHz, können Proteinstrukturen bis maximal 50 kDa (ca. 250 Aminosäuren) analysieren.

2.1.4 Elektronenmikroskopie

Die Elektronenmikroskopie kann auch die Struktur von Proteinen bis zur atomaren Ebene auflösen. Im Gegensatz zu der Röntgenstrukturanalyse haben Elektronenmikroskope hochauflösende Linsen und deshalb können die Proteine direkt in ihrer nativen Umgebung abgebildet werden. Dabei ist nicht notwendig (es ist sogar nicht erwünscht), dass die Proteine kristallisieren. Ähnlich wie bei der NMR-Spektroskopie können auch Konformationsänderungen in Proteinen mit dieser Methode detektiert werden. Dabei braucht die Elektronenmikroskopie weder den Einsatz von kernmagnetresonanzfähigen Isotopen noch ist sie an die Größe der Proteine gebunden. Insofern ist die Elektronenmikroskopie die am weitesten einsetzbare Methode und hat die wenigsten Nachteile.

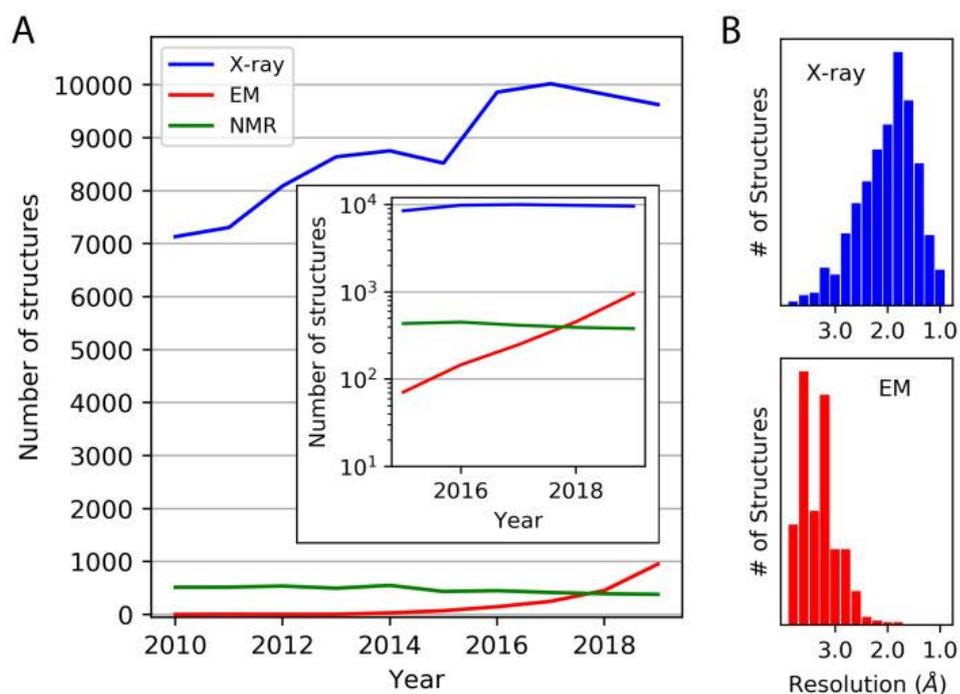


Abbildung 3: Wachstum der Anzahl der durch EM gelösten Strukturen im Vergleich zu NMR und Röntgenkristallographie.

2.2 Datenbanken

2.2.1 Sequenzdatenbanken

Sequenzierungstechnologien der nächsten Generation (Next-Generation Sequencing (NGS)) bezeichnet eine neuartige Technologie der Nukleinsäureanalytik. Im Gegensatz zu bisher bekannten Sequenzierungsverfahren können simultan mehrere hundert Millionen Fragmente in einer Probe sequenziert werden. Diese Hochdurchsatz- oder Parallelsequenzierungstechnologien haben gänzlich neue Dimensionen in der Nukleinsäureanalytik eröffnet. Die NGS-Parallelisierung der Sequenzierungsreaktionen erzeugt Hunderte von Megabasen bis Gigabasen an Nukleotidsequenz-Reads in einem einzigen Gerätelauf. Dies hat eine drastische Zunahme der verfügbaren Sequenzdaten ermöglicht und die Ansätze zur Genomsequenzierung in den biomedizinischen Wissenschaften grundlegend verändert [3,4].

Die DNA-Sequenzierung kann verwendet werden, um die Sequenz einzelner Gene, größerer genetischer Regionen (d. h. Cluster von Genen oder Operons), vollständiger Chromosomen oder ganzer Genome eines beliebigen Organismus zu bestimmen. Die DNA-Sequenzierung ist auch der effizienteste Weg, um RNA oder Proteine indirekt zu sequenzieren (zB Transkriptomsequenzierungen).

Der erste Schritt besteht in einer an das Ausgangsmaterial und an die Fragestellung angepassten Probenaufbereitung. So wird für De-novo-Genomsequenzierungen unbekannter Organismen genomische DNA präpariert, während bei Transkriptomanalysen unterschiedliche RNA-Populationen (mRNA, smallRNA, ncRNA) untersucht werden [5].

Eine Datenbank kann Sequenzen eines einzelnen Organismus, z. B. alle Proteine einer Spezies oder DNA-Sequenzen aller Organismen enthalten, deren Genom sequenziert wurde. Um Sequenzdaten von Proteinen zu erhalten, wird im Praktikumsversuch auf die UniProtKB-Datenbank zugegriffen. UniProtKB ist die übergreifende Datenbank, in der alle bekannten Proteinsequenzen gespeichert sind. Diese Datenbank besteht aus zwei Teilen, dem Swiss-Prot, der manuell annotiert wird, und dem TrEMBL, wo die Protein-Daten automatisch annotiert werden. Diese beiden Teile können aufgrund ihrer Synergistik nicht getrennt werden, da das manuelle Annotieren abhängig ist vom automatischen Annotieren und umgekehrt.

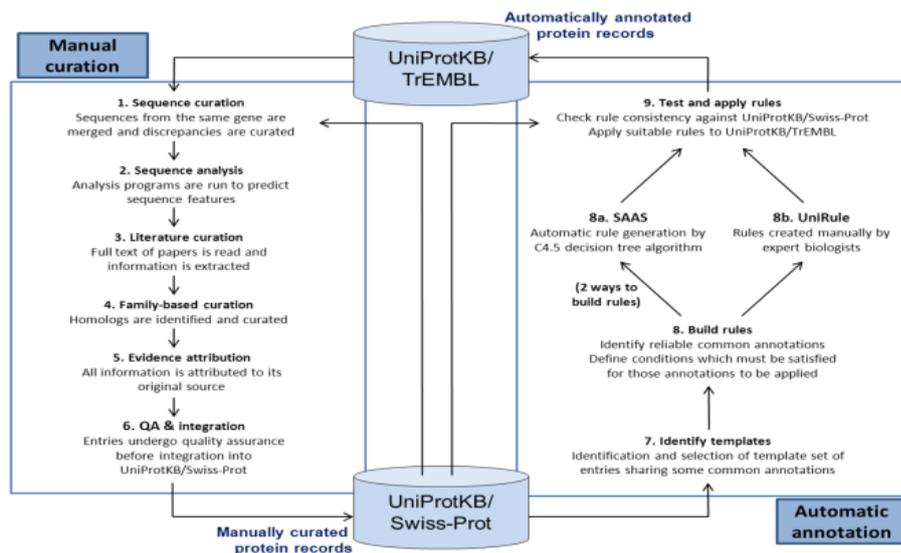


Abbildung 4: Unterschiede zwischen der annotierten Datenbank Swiss-Prot und der automatisch annotierten Datenbank TrEMBL (<https://www.uniprot.org/help/biocuration>).

UniProtKB/Swiss-Prot

Swiss-Prot ist der annotierte Teil der UniProt Proteinsequenz-Datenbank. Die SWISS-PROT Proteindatenbank enthält Sequenzeinträge von Proteinen, unterscheidet sich aber von UniProtKB/TrEMBL durch drei Kriterien:

- 1. Annotation:** In UniProtKB/Swiss-Prot können, wie in vielen Sequenzdatenbanken, zwei Klassen von Daten unterschieden werden: Die Kerndaten und die Annotationen. Die Kerndaten umfassen die Sequenzdaten, die Zitatinformationen und die taxonomischen Daten. Die Annotationen beschreiben die Funktionen eines Proteins, die Post-Translation-Modifikationen, die Domänen, die Sekundärstruktur, die Tertiärstruktur, die Quartärstruktur, aber auch Krankheiten, die mit einer Anzahl von Mängeln in einem Protein assoziiert sind, sowie Sequenzkonflikte, Varianten u.a.
- 2. Minimale Redundanz:** Viele Sequenzdatenbanken enthalten für eine gegebene Proteinsequenz separate Einträge, die unterschiedlichen Beiträgen entsprechen. In SWISS-PROT werden alle diese Daten zusammengefasst, um die Redundanz von Datenbanken zu minimieren. Kleinere Sequenzunterschiede werden in der Annotation vermerkt. Falls Konflikte zwischen verschiedenen Sequenzberichten existieren, werden diese in der Tabelle des entsprechenden Eintrags angezeigt.
- 3. Integration mit anderen Datenbanken:** Es ist wichtig, die Benutzer von molekularbiologischen Datenbanken mit einem Integritätsgrad zwischen den drei Typen von Datenbanken (Nukleinsäuresequenzen, Proteinsequenzen und Proteintertiärstrukturen) ebenso wie mit speziellen Datenbanken bekannt zu machen. SWISS-PROT hat Querverweise zu etwa 30 verschiedenen Datenbanken. Das intensive Netzwerk von Querverweisen erlaubt SWISS-PROT, einen Schwerpunkt auf die Zusammenarbeit der Biomolekulardatenbanken zu legen. Die Einträge in der SWISS-PROT-Datenbank sind so strukturiert, dass sie nicht nur für Benutzer, sondern auch für

Computerprogramme lesbar sind. Die Suche in SWISS-PROT-Datenbanken kann mit „BLAST“ (**B**asic **L**ocal **A**lignment and **S**earch **T**ool, <http://www.ncbi.nlm.nih.gov/blast/>) und „FASTA“ (**F**ast-**A**ll) erfolgen, welche nach Ähnlichkeiten zwischen Sequenzen suchen oder eine gegebene Sequenz mit einer Sequenz-Datenbank vergleichen [2, 3].

2.2.2 Proteinstrukturdatenbanken

Die Proteindatenbank (PDB: „Protein Data Bank“ <http://www.rcsb.org/>) ist eine zentrale Datenbank für Proteinstrukturen und wird vom *Research Collaboratory for Structural Bioinformatics* (RCSB) unterhalten. In 3D-Struktur-Datenbanken werden die Atomkoordinaten und einige strukturelle Parameter (z.B. B-Faktor) gespeichert und in sogenannten pdb-Dateien zusammengefasst. Abb. 5 zeigt einen Auszug aus einer pdb-Datei. In dieser Datei sind zu jedem einzelnen Atom des Proteins die Atomnummer (1), der Name des Atoms (2), sowie der Name der Aminosäure (3), zu dem das Atom gehört, die Polypeptidkette, zu der die Aminosäure gehört (4), die Residuennummer (5), die kartesischen Ortskoordinaten (6), die Besetzung (7) und der B-Faktor (8) gespeichert.

Der B-Faktor jedes Atoms ist eine Größe, die beschreibt, wie gut definiert die Position des Atoms im Modell ist. Je größer die Zahl, umso schlechter definiert ist das Atom in der experimentell bestimmten Elektronendichtekarte. Die Unordnung kann auf zwei Arten entstehen: statische Unordnung dadurch, dass das Protein leicht verschiedene Konformationen annehmen kann, und dynamische Unordnung dadurch, dass manche Teile des Proteins so flexibel sind, dass die Elektronendichtekarte verschmiert.

Neben die Atomkoordinaten des Proteins beinhaltet eine pdb-Datei auch die Koordinaten von möglichen Cofaktoren und Wassermolekülen. Im Kopfteil der Datei, dem „Header“, stehen Informationen über Name, Organismus, Auflösung, Methode und Literaturangaben.

Proteindatenbanken bieten zudem eine Fülle von Informationen über die biologische Funktion, über Mechanismen, die mit dieser Funktion verbunden sind und über Relationen zwischen Makromolekülen. Mit Hilfe der in den Datenbanken bereitgestellten Atomkoordinaten und einer Visualisierungssoftware, wie z.B. ChimeraX, kann die Proteinstruktur veranschaulicht werden (siehe Kap. 5).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
ATOM	2388	N	GLU	A	309	63.857	13.511	11.766	1.00	46.48
ATOM	2389	CA	GLU	A	309	64.927	13.901	12.669	1.00	47.48
ATOM	2390	C	GLU	A	309	64.709	15.221	13.413	1.00	46.13
ATOM	2391	O	GLU	A	309	63.777	15.965	13.113	1.00	43.62
ATOM	2392	CB	GLU	A	309	66.286	13.895	11.931	1.00	49.32
ATOM	2393	CG	GLU	A	309	66.375	14.558	10.552	1.00	49.95
ATOM	2394	CD	GLU	A	309	65.686	13.782	9.429	1.00	47.99
ATOM	2395	OE1	GLU	A	309	64.507	14.075	9.139	1.00	48.52
ATOM	2396	OE2	GLU	A	309	66.317	12.884	8.832	1.00	45.36
ATOM	2397	N	GLY	A	310	65.547	15.466	14.426	1.00	47.35
ATOM	2398	CA	GLY	A	310	65.476	16.694	15.211	1.00	42.61
ATOM	2399	C	GLY	A	310	64.432	16.768	16.311	1.00	41.75
ATOM	2400	O	GLY	A	310	64.224	17.833	16.887	1.00	41.59

Abbildung 5: Auszug aus einer pdb-Datei

2.3 Alignments – Ähnlichkeiten zwischen Sequenzen

Sequence Alignment ist die grundlegendste und wichtigste Operation in der Bioinformatik. Die Informationen über Funktion, Struktur und Evolution in einer biologischen Sequenz können durch Sequenzvergleiche gefunden werden. Sequenz-Alignment wird sehr häufig verwendet, um konservierte Regionen in einem Protein zu zeigen und kann auch verwendet werden, um Homologie zwischen einer Proteinfamilie und einem entfernt verwandten Mitglied zu zeigen. Auf diese Weise ist es auch möglich, die Evolution eines Organismus durch die Rekonstruktion eines Stammbaums anhand von Sequenzanalysen darzustellen.

Bei einem gegebenen Satz biologischer Sequenzen (RNA, Proteine, DNA) besteht der Zweck eines Sequenz-Alignment-Verfahrens darin, die Sequenzen so auszurichten, dass entweder ihre evolutionäre, funktionelle oder ihre strukturelle Verwandtschaft widerspiegelt wird.

Die auf der Sequenz basierende strukturelle Beziehung wird immer wichtiger, da ein Sequenz-Alignment für die Vorhersage der Proteinstruktur erforderlich ist, was in Abschnitt 2.3.4 hervorgehoben wird.

2.3.1 Sequenz-Alignment

Das sogenannte Sequenz-Alignment ist eine Standardmethode, um Aminosäure- oder Nukleotidsequenzen zu vergleichen. Aus Alignments lassen sich Erkenntnisse über Abstammung und Verwandtschaften ableiten. Alignments werden nach zwei Kriterien unterteilt. Zum einen wird je nach Anzahl der Sequenzen zwischen paarweisen und multiplen Alignments unterschieden, zum anderen zwischen lokalen und globalen Alignments. Im Gegensatz zum globalen Alignment werden beim lokalen Alignment nur Ausschnitte der jeweiligen Sequenzen miteinander verglichen. Im Folgenden wird näher erläutert, wie Alignments berechnet werden.

Beim globalen paarweisen Alignment werden zwei gegebene Sequenzen x und y miteinander verglichen (siehe auch Übungsaufgabe 1.3). Um Ähnlichkeiten dieser Sequenzen zu berechnen, werden die Sequenzen in einer Matrix so übereinander geschrieben, dass ähnliche Teilbereiche übereinander stehen. Zu diesem Zweck werden Lücken, sogenannte „Gaps“, eingefügt.

X: QWR**T**AHNN X: QWRT--AHNN
 Y: QWRTGYAHNN Y: QWRTGYAHNN

Um den Längenunterschied in der Sequenz X auszugleichen, wurde eine Lücke eingefügt. Nach Löschen der Gaps müssen wieder die ursprünglichen Sequenzen X bzw. Y vorliegen. Ein Gap kann eine oder mehrere leere Positionen umfassen, die in der Sequenz durch Minuszeichen (-) gekennzeichnet werden. Eine weitere Bedingung ist, dass es keine Spalte geben darf, in der in allen Sequenzen nur Lückensymbole übereinander stehen, da sonst die Sequenzen unendlich verlängert werden könnten.

2.3.2 „Scores“

Um die Qualität eines Alignments zu bewerten, vergibt man „Scores“ (Punktzahlen für Übereinstimmungen), die sich in der Regel aus dem „Substitutions-Score“ und dem „Gap-Score“ zusammensetzen. Der Substitutions-Score bewertet Spalten der Alignments, in denen zwei Zeichen übereinander stehen. Dabei werden hohe Punktzahlen für ähnliche Aminosäuren und niedrige, gegebenenfalls negative Punktzahlen für sehr unterschiedliche Aminosäuren vergeben. Das hängt damit zusammen, dass z.B. die Wahrscheinlichkeit für einen Austausch einer polaren Aminosäure mit einer anderen polaren Aminosäure größer ist als der Austausch mit einer unpolaren Aminosäure. Bekannte Substitutionsmatrizen sind die PAM- und die BLOSUM-Matrix, z.B. die BLOSUM 62 (vgl. Abb. 5). Bei dieser Substitutionsmatrix handelt es sich um eine symmetrische

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9	-1	-1	-3	0	-3	-3	-3	-4	-3	-3	-3	-3	-1	-1	-1	-1	-2	-2	-2
S	-1	4	1	-1	1	0	1	0	0	0	-1	-1	0	-1	-2	-2	-2	-2	-2	-3
T	-1	1	4	1	-1	1	0	1	0	0	0	-1	0	-1	-2	-2	-2	-2	-2	-3
P	-3	-1	1	7	-1	-2	-1	-1	-1	-1	0	-2	-1	-2	-3	-3	-2	-4	-3	-4
A	0	1	-1	-1	4	0	-1	-2	-1	-1	-2	-1	-1	-1	-1	-1	-2	-2	-2	-3
G	-3	0	1	-2	0	6	-2	-1	-2	-2	-2	-2	-2	-3	-4	-4	0	-3	-3	-2
N	-3	1	0	-2	-2	0	6	1	0	0	-1	0	0	-2	-3	-3	-3	-3	-2	-4
D	-3	0	1	-1	-2	-1	1	6	2	0	-1	-2	-1	-3	-3	-4	-3	-3	-3	-4
E	-4	0	0	-1	-1	-2	0	2	5	2	0	0	1	-2	-3	-3	-3	-3	-2	-3
Q	-3	0	0	-1	-1	-2	0	0	2	5	0	1	1	0	-3	-2	-2	-3	-1	-2
H	-3	-1	0	-2	-2	-2	1	1	0	0	8	0	-1	-2	-3	-3	-2	-1	2	-2
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5	2	-1	-3	-2	-3	-3	-2	-3
K	-3	0	0	-1	-1	-2	0	-1	1	1	-1	2	5	-1	-3	-2	-3	-3	-2	-3
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5	1	2	-2	0	-1	-1
I	-1	-2	-2	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4	2	1	0	-1	-3
L	-1	-2	-2	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4	3	0	-1	-2
V	-1	-2	-2	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4	-1	-1	-3
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6	3	1
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	2
W	-2	-3	-3	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

Abbildung 6: BLOSUM 62-Substitutionsmatrix

Matrix, welche als Einträge die Score-Punkte für alle möglichen Konfigurationen von Aminosäuren enthält, die hier im Einbuchstaben-Code aufgelistet sind. Der kleinste Score-Wert beträgt -4 (geringe Wahrscheinlichkeit für ein Austauschen; quasi "Strafpunkte", wenn diese beiden Aminosäuren im Alignment an der gleichen Position stehen) und der höchste Score-Wert beträgt 11 (hohe Wahrscheinlichkeit für ein Austauschen, also "Bonuspunkte", wenn diese beiden Aminosäuren im Alignment an der gleichen Position stehen) [4].

2.3.3 *Multiple Sequence Alignment (MSA)*

Ein Multiple Sequence Alignment (MSA) steht für das gleichzeitige Analysieren mehrerer Sequenzen. Im Vergleich zum paarweisen Alignment liefert es genauere Informationen über die Aminosäureverteilung an einzelnen Positionen. Solche Verteilungen können nicht nur Aufschluss über konservierte Bereiche geben, sondern sind auch die Grundlage für profilbasierte Datenbanksuche und phylogenetische Analysen.

Die Berechnung eines multiplen Alignments von Sequenzen ist zeitaufwendiger als bei einem paarweisen Alignment. Deswegen verwendet man heuristische Ansätze, um diesem Problem entgegen zu wirken. Multiple Alignments werden in globale multiple Alignments und lokale multiple Alignments unterteilt. Das globale multiple Alignment fasst die Sequenzen in einem Block unter Einführung von Gaps zusammen. Das lokale multiple Alignment sucht in den Sequenzen nach Blöcken großer Ähnlichkeit und erstellt dann blockweise multiple Alignments ohne Gaps. Eine Anwendung von lokalen multiplen Alignments ist die Analyse von Proteindomänen.

2.3.4 *Präzise Vorhersage der Proteinstruktur mit 3-Track Neural Networks* *AlphaFold2 und RoseTTAFold*

Die Vorhersage der Proteinstruktur kann verwendet werden, um die dreidimensionale Form eines Proteins aus seiner Aminosäuresequenz zu bestimmen. Dieses Problem ist von grundlegender Bedeutung, da die Struktur eines Proteins seine Funktion maßgeblich bestimmt [5]. Nach 50 Jahren Entwicklung der Proteinstrukturvorhersage hat ein Team namens DeepMind von Google ein System namens AlphaFold2 entwickelt mit einer Deep-Learning-Architektur um genaue Vorhersagen der Abstände zwischen Aminosäureresten zu treffen.

AlphaFold2 wurde mit Hunderttausenden von experimentell ermittelten Proteinstrukturen und -sequenzen in der PDB und anderen Datenbanken trainiert. Angesichts einer neuen Sequenz sucht es

zunächst nach verwandten Sequenzen in Datenbanken, die Aminosäuren identifizieren können, die sich tendenziell gemeinsam entwickelt haben, was darauf hindeutet, dass sie im 3D-Raum nahe beieinander liegen. Die Struktur vorhandener verwandter Proteine bietet eine weitere Möglichkeit, Abstände zwischen Aminosäurepaaren in der neuen Sequenz abzuschätzen [6].

AlphaFold iteriert Hinweise aus diesen parallelen Spuren hin und her, während es versucht, die 3D-Positionen von Aminosäuren zu modellieren, und aktualisiert seine Schätzung kontinuierlich. Insbesondere die Verwendung eines AI-Mechanismus namens „Aufmerksamkeit“, um zu bestimmen, welche Aminosäureverbindungen für seine Aufgabe in jedem Moment am wichtigsten sind. AlphaFold2 erreicht eine hohe Genauigkeit, auch für Proteine mit weniger homologen Sequenzen [7,8].

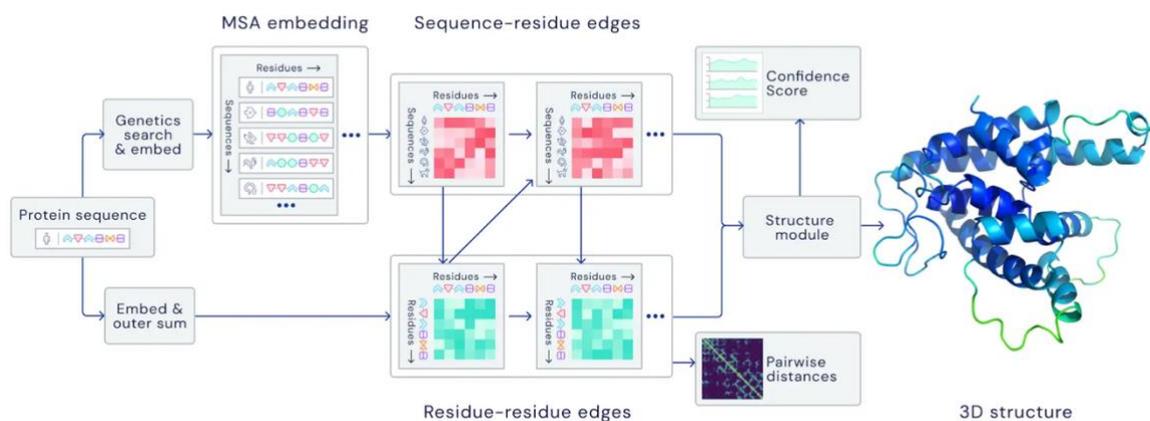


Abbildung 7: Ein Überblick über die wichtigsten Neuronales Netzwerk Modellarchitekturen. Das Modell arbeitet sowohl über evolutionär verwandte Proteinsequenzen als auch über Aminosäurerestpaare, wobei iterativ Informationen zwischen beiden Darstellungen weitergegeben werden, um eine Struktur zu erzeugen [7].

Die Abhängigkeit des Netzwerks von Informationen über verwandte Proteinsequenzen bedeutet, dass AlphaFold einigen Einschränkungen unterliegt. Es ist nicht darauf ausgelegt, die Auswirkung von Mutationen, wie sie beispielsweise Krankheiten verursachen, auf die Form eines Proteins vorherzusagen. Es wurde auch nicht darauf trainiert, zu bestimmen, wie Proteine ihre Form in Gegenwart anderer interagierender Proteine oder Moleküle wie Medikamente ändern. AlphaFold kann auch nicht Proteinen vorhersagen, die in unterschiedlichen Konformationen unterschiedliche Strukturen annehmen können. Vorhersagen gelten für isolierte Strukturen, während viele Proteine neben Liganden wie DNA und RNA, Fettmolekülen und Mineralien wie Eisen funktionieren und diese vom aktuellen Algorithmus nicht berücksichtigt werden können. Die Modelle verfügen daher über Werte die als Genauigkeit bezeichnet werden (Abbildung 8), die das Vertrauen des Netzwerks in seine Vorhersage für jede Aminosäureeinheit eines Proteins messen [8].

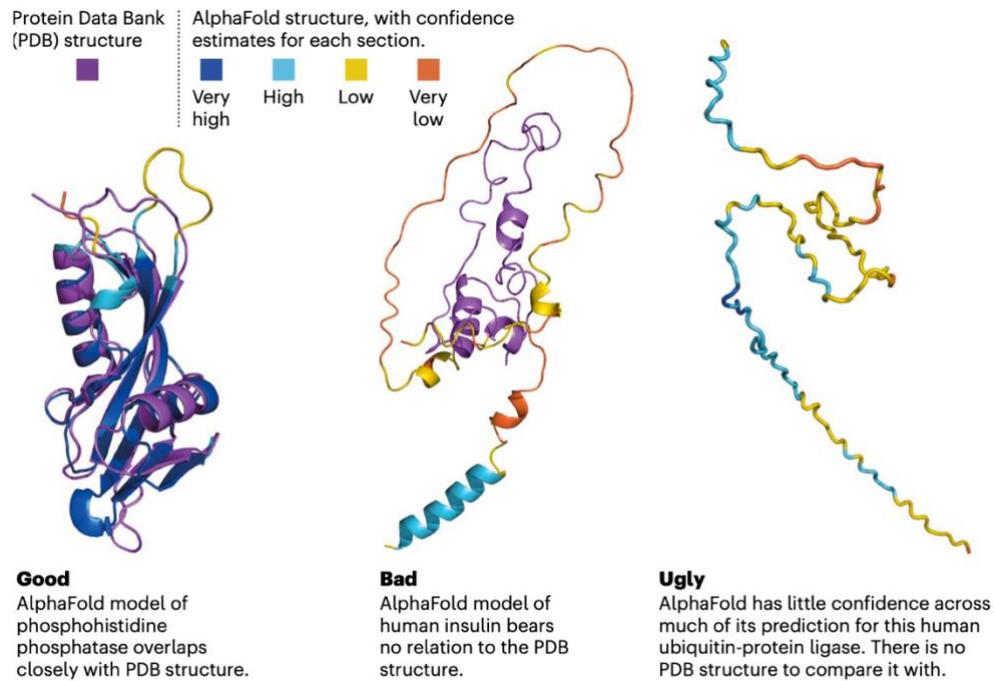


Abbildung 8: Die Vorhersagen von AlphaFold für ein gefaltetes Protein sind mit Vertrauensschätzungen versehen. Die Überlagerung jedes Modells mit der experimentell bestimmten Struktur (falls verfügbar) zeigt die Genauigkeit der Vorhersage [8].

2.3.5 AlphaFold Protein Structure Database, kuratiert von EMBL-EBI

DeepMind und das European Bioinformatics Institute (EMBL-EBI) des EMBL haben sich zusammengetan, um die AlphaFold-Datenbank zu erstellen, um der wissenschaftlichen Gemeinschaft Proteinstrukturvorhersagen frei zugänglich zu machen. Die neueste Datenbankversion enthält über 200 Millionen Einträge und bietet eine breite Abdeckung von UniProt (dem Standard-Repository für Proteinsequenzen und Annotationen). Es werden einzelne Downloads für das menschliche Proteom und für die Proteome von 47 weiteren Schlüsselorganismen bereitgestellt, die für Forschung und globale Gesundheit wichtig sind. Es werden auch Downloads für die gesamte manuell kuratierte Teilmenge von UniProt (Swiss-Prot) bereitgestellt (<https://alphafold.ebi.ac.uk>).

3 Quellen und Literatur

- [1] Growth of GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>
- [2] Apweiler, R: Introduction to Molecular Biology Databases: <http://www.ebi.ac.uk/swissprot/Publications/mbd1.html>
- [3] Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol.* 2018 Apr;122(1):e59. doi: 10.1002/cpmb.59. PMID: 29851291; PMCID: PMC6020069.

- [4] <https://www.igb.fraunhofer.de/de/forschung/in-vitro-dagnostik/next-generation-sequencing.html>
- [5] https://en.wikipedia.org/wiki/DNA_sequencing
- [6] Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, 28, 45–48
- [7] <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure>
- [8] E. Callaway. (2022) What's next for AlphaFold and the AI protein-folding revolution. *Nature*, Vol 604.
- [9] Thompson et al (2020) Advances in methods for atomic resolution macromolecular structure determination *F1000Research* 9(Faculty Rev):667
- [10] S. Henikoff and J.G. Henikoff (1992) Amino acid substitution matrices from protein blocks. *P.N.A.S. USA*, 89, 10915-10919.
- [11] Baek et al. (20 August 2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557):871-876. doi: 10.1126
- [12] <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>
- [13] Senior et.al. (30 January 2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577. doi: 10.1038
- [14] <https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery>
- Jumper, J et al. Highly accurate protein structure prediction with AlphaFold (*Nature* 2021).

4 Versuchsdurchführung

Benutzung der Rechner im Pool-Raum Physik

Für den Versuch verwenden wir die Linux-Rechner des Hochschulrechenzentrums im Pool-Raum Physik. Um sich dort einzuloggen, benötigen Sie Ihren HRZ-Login und das Passwort dazu. Weitere Informationen finden Sie unter: www.rz.uni-frankfurt.de/services/linux

Führen Sie die folgenden Strukturanalysen durch und antworten Sie auf die Fragen in Ihrem Protokoll. Bitte beantworten Sie die Vorbereitungsfragen im Zusammenfassungs- und Einführungsteil Ihres Protokolls sorgfältig.

4.1 Vorbereitung

1. Woher kommen die Sequenzdaten in den Datenbanken?
2. Was ist der Unterschied zwischen den Swissprot- und TrEMBL-Teilen der UniprotKB?
3. Woher kommen die Strukturdaten in den Datenbanken?
4. Wie können Sie die Ähnlichkeit zweier Proteinsequenzen bzw. zweier Proteinstrukturen quantifizieren?
5. Welche Vor- und Nachteile haben die verschiedenen Methoden zur Proteinstrukturbestimmung?

6. Was ist mit dem Begriff Apoprotein gemeint?
7. Welche Vorberechnung ist für die Proteinstrukturvorhersage erforderlich und auf welche Basis wird das Neurale Netzwerk trainiert?
8. Warum sind Strukturmethoden immer noch notwendig, wenn die Struktur eines Proteins vorhergesagt werden kann?

4.2 Sequenz-Alignments

1. Suchen Sie die Aminosäuresequenz des Cytochrom b (*cytochrome b*) vom Afrikanischen Elefanten (*Loxodonta africana*) in der UNIPROT-Datenbank (<http://www.uniprot.org>). Benutzen Sie im Abschnitt "Sequences" des Datenbankeintrags das BLAST-Tool für eine Suche nach ähnlichen Sequenzen. Tabellieren Sie nur die Score-Punkte der besten fünf unter den von „UniProtKB/Swiss-Prot“ überprüften Einträgen (die mit dem gelben statt grauen Sternen). Warum ist es wichtig, manuell kuratierte Datensätze für eine Sequenzdatenbanksuche zu verwenden?
2. Suchen Sie aus der Trefferliste die Sequenzen des Cytochrom b vom Indischen Elefanten (*Elephas maximus*) und des Mammuts (*Mammuthus primigenius*) und notieren Sie deren UniProt ID's. Lassen Sie ein Alignment der Cytochrom-b-Sequenzen der drei Elefantenarten berechnen. Selektieren Sie hierzu die Kästchen von die drei verschiedenen Cytochrome und verwenden Sie den Link „Align“. Wie viele Unterschiede zwischen den einzelnen Sequenzen gibt es (genaue Anzahl angeben)? Ist das Mammut eher mit dem Indischen oder dem Afrikanischen Elefanten verwandt? Kann man allein auf Grund der Sequenzen behaupten, dass es sich um eng verwandte Arten handelt? Wie könnten Sie Ihre Hypothese testen?
3. Mit welcher Methode wurde die Struktur des Cytochrom B des Mammuts bestimmt und welche Genauigkeit hat die Struktur?

4.3 Proteinstrukturdatenbank

1. Suchen Sie nach 3D-Molekülstrukturen des Proteins Cytochrom b der Ratte in der Proteindatenbank (<http://www.rcsb.org/pdb/>). Suchen Sie 5 verschiedene Strukturen aus, darunter einen möglichen Wildtyp, zwei Mutanten (davon eine NMR- und eine X-Ray-Struktur) sowie eine Apoproteinstruktur. Nutzen Sie für die Suche die 'Advanced Search' rechts oben auf der PDB-Webseite und suchen Sie dann z.B. nach der Substanz, dem Organismus, ob es Mutationen gibt, ob es Liganden gibt und so weiter. Fassen Sie folgende Daten in einer Tabelle zusammen: pdb-ID, NMR- oder X-Ray-Struktur, Wildtyp/Mutante, Auflösung & R-Faktor (für X-Ray) bzw. Anzahl der Modelle (NMR).
2. Welchen wichtigen Anhaltspunkt über die Qualität einer Struktur liefert der R-Faktor?

3. Worin unterscheiden sich pdb-Dateien aus Elektronenmikroskopie, NMR- und Röntgenstrukturanalysen? Nennen Sie mindestens drei Unterschiede.
4. Öffnen Sie die pdb-Datei der Cytochrom b Mutante (pdb-ID: 1EUE) entweder direkt auf der PDB-Webseite oder mit einem Texteditor nach dem Herunterladen und lesen Sie sich den sogenannten „Header“ durch. Hier erhalten Sie u.a. Informationen über die Sequenz, Aminosäuren, Sekundärstrukturelemente und weitere Moleküle, die in der Struktur vorkommen, z.B. über Cofaktoren, Inhibitoren und Wassermoleküle (falls vorhanden geben Sie Art und Anzahl der Moleküle an). Geben Sie die Anzahl der α -Helices und β -Faltblätter an.

4.4 Proteinstrukturvorhersage

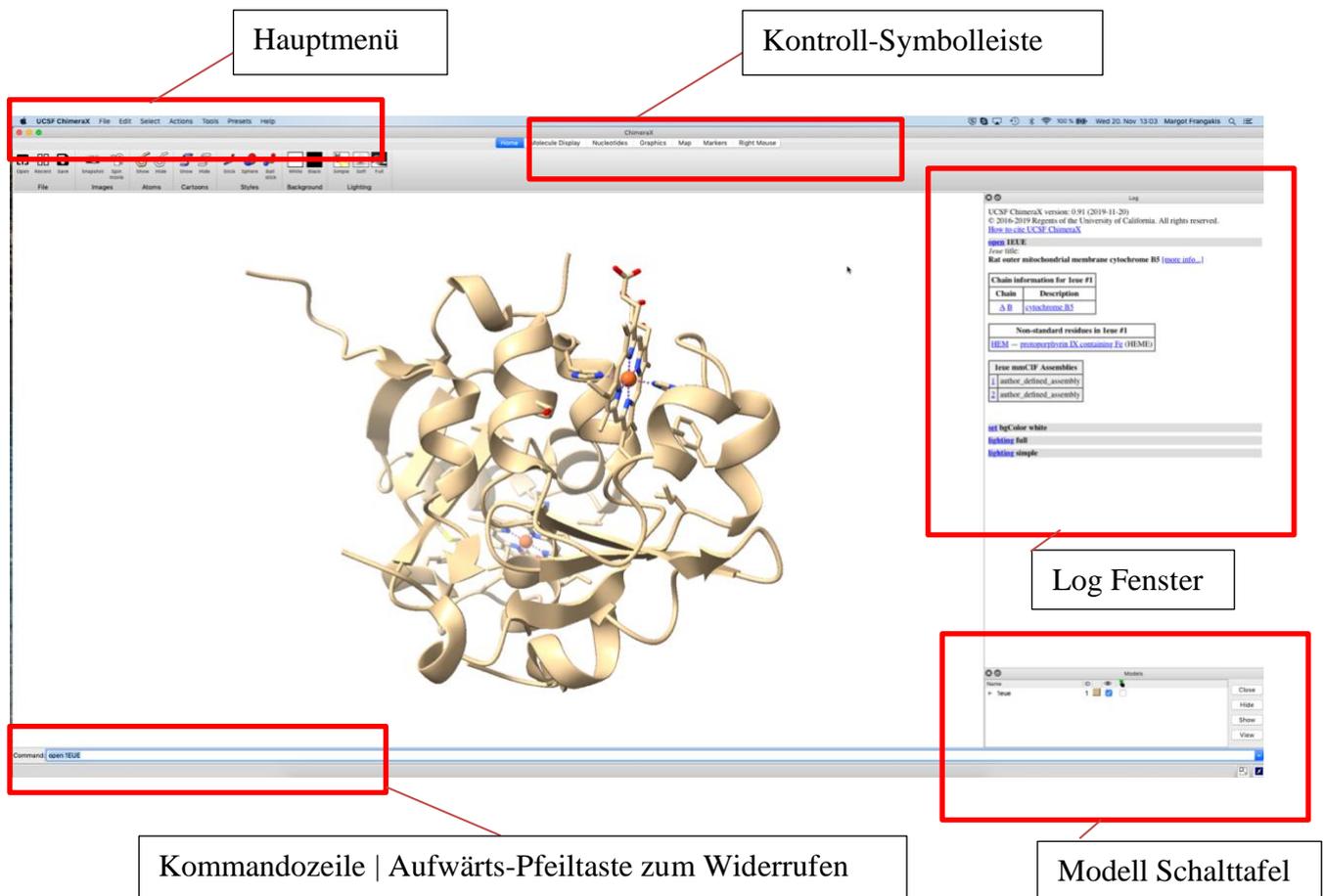
ColabFold (<https://colab.research.google.com>) ist eine benutzerfreundliche Notebook-basierte Umgebung für schnelle und bequeme Vorhersagen der Proteinstruktur. Seine Strukturvorhersage wird von AlphaFold2 und RoseTTAFold in Kombination mit einer schnellen Generierungsphase für mehrere Sequenzen unter Verwendung von MMseqs2 unterstützt. Der Vorteil des Colab-Notebooks besteht darin, dass jeder Benutzer die Eingaben der GPU-Maschinen „ausleihen“ und unbekannte Sequenzen eingeben kann, um die Struktur vorherzusagen.

1. Suchen Sie in der Datenbank <https://alphafold.ebi.ac.uk> eine Struktur eines Cytochrom-b-Proteins aus einem ausgestorbenen Organismus (außer dem Mammot) und geben Sie die Uniprot-ID und Genauigkeit an.
2. Ein Teil der Struktur von *Mycoplasma pneumonia* P116 wird durch Alphafold (Uniprot-ID E1QBX4) vorhergesagt, jedoch hat sich dieses Protein als Dimer erwiesen und die Datenbank zeigt nur das Monomer. Verwenden Sie den Colab-Server, um eine vorhergesagte Struktur für das Dimer dieser Sequenz mit AlphaFold2 zu berechnen. Dazu muss man ein Google-Konto haben oder schnell eines mit Ihrer aktuellen E-Mail-Adresse erstellen. Die Berechnung dieser Strukturen nimmt Zeit in Anspruch. Bitte fahren Sie mit Abschnitt 4.5 fort und verwenden Sie anschließend ChimeraX, um den Unterschied zwischen Dimer und Monomer zu beschreiben.

4.5 Darstellung von 3D-Strukturen mit ChimeraX

UCSF ChimeraX ist das molekulare Visualisierungsprogramm der nächsten Generation aus der Ressource für Bio-Computing, Visualisierung und Informatik (RBVI), einer Forschungsquelle des NIH für biomedizinische Technologie (Goddard et al. 2018).

Die Proteindaten können beispielsweise von einer Datenbank heruntergeladen und in ChimeraX geöffnet werden. ChimeraX bietet eine Vielzahl von Möglichkeiten, die biomolekularen Strukturen zu betrachten und spezifische Teile eines Proteins hervorzuheben. Mit ChimeraX lassen sich in begrenztem Maße Veränderungen an Proteinen und deren Auswirkung auf die Struktur beobachten. Die Anwendung ist daher ein nützliches Werkzeug in der Biologie, Bioinformatik und Pharmazie. Sie finden das Programm ChimeraX unter Main MATE menu → Education → UCSF ChimeraX.



Kontroll-Symboleiste

Hierüber lassen sich die allgemeinen Einstellungen wie z.B. die Hintergrundfarbe oder auch Darstellungseinstellungen der zu betrachtenden Struktur vornehmen. Proteindateien (z.B. im .pdb-Format) können über "*Home* → *File* → *Open...*" geöffnet werden.

ChimeraX Hauptgrafikfenster

In diesem Fenster sieht man eine dreidimensionale Ansicht des geöffneten Proteins, welches mittels der Maustasten frei vergrößert (mit dem Mausrad), verschoben (rechte Taste) und gedreht werden (linke Taste) kann. Das Programm öffnet die Struktur als vorgegeben in „cartoon“-Darstellung. Linien, die andere Interaktionen als kovalente Interaktionen anzeigen, werden als Pseudobindungen bezeichnet.

Log Fenster

Das Log Fenster gibt ausgeführte Befehle mit erweiterten Kürzungen wieder. Es werden auch Nachrichten und andere Ausgabeinformationen angezeigt, einschließlich möglicherweise Hyperlinks und Befehlsausführungslinks.

Modell Schalttafel

Es listet die aktuellen Modelle mit Spalten auf:

Name – Modellname

ID – Modellnummer

Ein Farbfeld, auf das geklickt werden kann, um die Farbe interaktiv zu ändern

Augensymbol – Kontrollkästchen ein- / ausblenden

Handsymbol – Auswahl-Kontrollkästchen (ein Kontrollkästchen mit "-" bedeutet, dass das Modell teilweise ausgewählt ist).

Kommandozeile

Geben Sie hier die Befehle ein, wenn Sie nicht mit der Kontroll-Symboleiste arbeiten. Die Befehle werden immer gespeichert, auch wenn das Programm schließt. **Wichtig:** Um die Befehle wieder anzuzeigen, benutzen Sie die Aufwärts-Pfeiltasten. Mit dem Befehl "undo" oder Ctrl Z können Sie Ihre Befehle immer schrittweise zurücksetzen. Mit dem Pfeil rechts neben der Kommandozeile können Sie die Befehle speichern mit "Command History → save".

Für die Nutzung des Kommandozeile sind die Hierarchical specifiers wichtig (siehe Tabelle unten). Eine detaillierte Übersicht über ChimeraX-Befehle finden Sie im ChimeraX-Log Fenster. Klicken Sie dafür auf einen Hyperlink für ein Befehl oder mit "help Befehl" in der Kommandozeile.

Hierarchical Specifiers

Symbol	Reference Level	Definition	Examples
#	model	model number assigned to the data in ChimeraX (hierarchical, with positive integers separated by dots: N, N.N, N.N.N, <i>etc.</i>)	#1 #1.3
/	chain	chain identifier (case-insensitive unless both upper- and lowercase chain IDs are present)	/A
:	residue	residue number <i>OR</i> residue name (case-insensitive)	:51 :glu
@	atom	atom name (case-insensitive)	@ca

- Geben Sie den Befehl **'open 1EUE'** ein. Klicken Sie im Tab "Home" in die Symbolleiste, klicken Sie auf Background "white", um einen weißen Hintergrund dazustellen. Um eine atomische Vorstellung vom Protein zu bekommen, färben Sie zuerst das Protein in blau mit dem Befehl **color #1 slate blue**. Danach geben Sie folgende Befehle ein: **show #1 atoms; hide #1 ribbons; color #1 byhetero; select :HOH; hide sel**. Selektiere jetzt die HEM Gruppe und färbe es in einer beliebigen Farbe. Speichern Sie dann ein Bild davon in Ihrem Ordner mit dem Befehl: **save /path_to_your_directory/image1.png supersample 4**. Stellen Sie sicher, dass beim Erstellen Ihres Bildes keine Atome oder Cartoons ausgewählt sind (grüner Umriss). Benutzen Sie jetzt entweder das GUI auf den Navigationstasten "Molecule Display" oder die Kommandozeile, um ein cartoon Darstellung des Proteins anzuzeigen mit atomischer Darstellung der HEM Gruppe. Lassen Sie sich nur die A-Kette anzeigen (in ChimeraX werden die Ketten mit **/X** bezeichnet). Benutzen Sie dafür den Befehl **hide** oder **del**.
- Laden Sie nun eine weitere Mutante (pdb-ID:1AQA) desselben Proteins dazu und überlagern Sie die beiden Strukturen. (ChimeraX-Befehl **match**, Hilfe mit **help match**). Im Log Fenster finden Sie die Parameter des Überlagerns und der Ergebnisse. Lassen sich räumliche Unterschiede feststellen? Wenn ja, beschreiben Sie diese. Geben Sie den RMSD-Wert (**root mean square deviation**) für das Alignment der beiden Strukturen an.
- Identifizieren Sie die Aminosäuren des Cytochrom b (pdb-ID: 1EUE), die die Hämgruppen binden. Stellen Sie das Häm und die Aminosäuren, die an der Bindung beteiligt sind, in der "sticks"-Darstellung (**select zone ligand 3 protein res t**) oder im Hauptmenü (Select → zone → 3A from the currently selected atoms) und das Protein selbst in der „cartoon“-Darstellung dar. Bestimmen Sie die Abstände dieser Aminosäuren zum zentralen Eisenatom des Häms (Symbolleiste: Right Mouse → Distance). Vermerken Sie, welcher Abstand zu welcher Aminosäure gehört. Zur besseren Veranschaulichung empfiehlt es sich, die ausgewählten Bereiche des Proteins farbig unterschiedlich zu markieren. Zum Beispiel, um nur das zentrale Eisenatom zu färben (**select #1/A:201@FE**) und (**color sel Farbe**) (Bild speichern, s.o.). Die Schriftgröße können Sie mit dem Befehl **label height Größe** ändern.

4. Laden Sie mit ChimeraX die Struktur des Proteins Calbindin (pdb-ID: 3ICB). Geben Sie auch hier wichtige Strukturdaten an (Organismus, Auflösung etc., siehe Aufg. 4.3.1). Identifizieren Sie die Calciumliganden (Aminosäurereste, die das Calcium binden) und bestimmen Sie die Abstände der sauren Liganden zu den Calciumionen, benutzen Sie (**label #X pseudobonds**) und (Symbolleiste: Right Mouse → Label) oder (**label @@display**), um die beteiligten Aminosäuren zu beschriften. Verwenden Sie (**cartoon suppress false**), um unzusammenhängende Aminosäuren mit dem Rückgrat zu verbinden. Benennen Sie die Abstände eindeutig oder zeigen Sie ein Bild mit den eingezeichneten Abständen und den beschrifteten Aminosäuren. Laden Sie die Apoproteinstruktur *ICLB* und überlagern Sie das Proteinrückgrat der beiden Strukturen. Worin liegen die Unterschiede zwischen der Apo- und der Holostruktur im Bereich der Bindungsstelle (Bild speichern und beschreiben)?

5. Laden Sie mit ChimeraX die Struktur des Membranproteins Porin (pdb-ID: 1BH3). Stellen Sie die Farbe des Proteins mit „wheat“ Farbe dar. Beschreiben Sie die Struktur. Wie ist das Porin in die Membran eingebettet? Stellen Sie die aromatischen Aminosäuren (Tyrosin und Tryptophan) in atomistische Darstellung und farbig dar (Bild). Wo sind sie im Protein angeordnet? Welche Funktion könnten sie haben? Verstecken Sie nun das Tyrosin und Tryptophan atomistische Darstellung und Verfärbung. Stellen Sie die geladenen Aminosäuren Glutamat, Aspartat, Lysin und Arginin in atomistische Darstellung und farbig dar (Bild). Wo befinden sich diese fast ausschließlich (Tipp: Drehen Sie die Darstellung, um einen Überblick die Orientierung der Seitenketten zu bekommen)? Lassen Sie sich außerdem die „Beweglichkeit“ des Proteins mittels des B-Faktors anzeigen (Symbolleiste: Molecule Display → b-factor) oder (**color bfactor #X range 10,60**) in der Kommandozeile. Welche Regionen sind besonders beweglich?

6. Laden Sie mit ChimeraX die Strukturen der Bakterielle ATP-Synthase (pdb-ID: 6n2y, 6n2z, 6n30). Die drei Strukturen sind durch Elektronenmikroskopie aufgelöst worden und zeigen die Konformationsänderung der Synthase. Geben Sie den Befehl **open emdb:9333** ein, um ein Elektronendichtekarte von der Synthase zu öffnen. Welche der drei pdb-Dateien lassen sich aus dieser Elektronendichtekarte erstellen? Gehen Sie auf der Website der Elektronenmikroskopie-Datenbank (<https://www.ebi.ac.uk/pdbe/emdb>) und geben Sie die Datei Nummer bei 'search' ein. Welche Auflösung hat die Elektronendichtekarte? Analysieren Sie die Bewegung der Synthase durch den Befehl **morph #x,y,z**. Geben Sie bei x,y,z die

Modell Nummer ein. Beschreiben Sie in einem Absatz die Bewegung der Synthase, die durch Konformationsänderung entsteht und dadurch die Funktion des Komplexes.